Psychometrika (2026), **91**, 761–788
doi:10.1017/psy.2026.10083

PSYCHOMETRIC SOCIETY

APPLICATION AND CASE STUDIES - ORIGINAL

From Digital Data to Psychological Insights: Making Sense of Mobile-Sensing Data through Integrative Preprocessing Pipelines

Ramona Schoedel[1,2], Larissa Sust[2], Philipp Sterner[2,3] and David Goretzko[4]

[1]Charlotte Fresenius Hochschule, Germany; [2]Department of Psychology, LMU Munich, Germany; [3]Technical University of Munich, Germany; [4]Goethe University Frankfurt, Germany

Corresponding author: Ramona Schoedel; Email: ramona.schoedel@charlotte-fresenius-uni.de

(Received 31 January 2025; revised 5 January 2026; accepted 5 January 2026)

This manuscript is part of the special section on Integrating and Analyzing Complex High-Dimensional Data in Social and Behavioral Sciences Research. We thank Drs. Eric F. Lock and Katrijn Van Deun for serving as Co-Guest Editors.

Abstract

Psychological research has long centered around questionnaire assessments, but now digital devices, especially smartphones, enable the collection of real-world behavioral data through mobile sensing. While this data collection method offers unique opportunities, it also introduces new methodological challenges, as mobile-sensing data are highly complex and high in dimensionality (i.e., timestamped events with millisecond resolution), requiring advanced preprocessing to derive psychologically meaningful variables. This article highlights these challenges by reviewing the current state of data preprocessing based on app usage logs from smartphones. Afterward, it presents three preprocessing cases that vary in complexity across the dimensions of *data enrichment*—which involves adding context to raw data by integrating information from external and internal sources (including ecological momentary assessments)—and *data aggregation*—which entails summarizing data in different ways, from basic descriptive statistics to sophisticated machine-learning models. For each case, potential pitfalls are identified, and extensions are discussed to refine our preprocessing pipelines and accommodate different data types and research questions. By outlining these preprocessing strategies, this manuscript demonstrates the rich potential of mobile-sensing data for extracting nuanced behavioral variables beyond simple person-level summaries and aims to inspire the development of more advanced research questions based on sensing data.

Keywords: data preprocessing; digital behavioral data; digital phenotyping; mobile sensing; variable extraction

1. Introduction

In the past, psychological research has traditionally emphasized questionnaire assessments, often overlooking the study of actual behavior (Funder, 2009). While investigating behavior in the field was previously challenging due to factors like high costs, time constraints, and intrusiveness (Baumeister et al., 2007), it is now facilitated by the rise of digital devices. Using off-the-shelf electronics such as smartphones and smartwatches, researchers can automatically gather behavioral (and situational)

data from people's everyday lives through mobile sensing. More specifically, mobile sensing enables the passive collection of data from system logs and native sensors present in these devices via designated research apps (for an introduction to the method, see Mehl et al., 2024). These sensing apps can be installed on participants' own devices, making data collection less intrusive and more financially economical and environmentally friendly compared to predecessors like portable cameras or audio recorders (Miller, 2012; Schoedel & Mehl, 2024). As a result, sensing data can be collected over extended periods in longitudinal study designs.

For a long time, mobile sensing has primarily been implemented on smartphones, as these devices are widespread in the Western world (GSMA, 2025) and are with their users most of the time (Dey et al., 2011). While smartwatches and other wearables are now becoming increasingly important—especially for collecting physiological and movement-related data (Fuller et al., 2020; Wac, 2018)—smartphones still play a central role in mobile-sensing research. Their ability to combine passive sensing with active data collection through ecological momentary assessments (EMAs; Wrzus & Neubauer, 2023) makes them especially useful for capturing behavioral data alongside participants' momentary subjective experiences (Conner & Mehl, 2015).

Drawing on the unprecedented accessibility and variety of behavioral data collected from smartphones or other sensing devices, researchers have started to explore a range of psychological phenomena in everyday life, for example, personality traits (Stachl, Au, et al., 2020), sociability (Harari, Müller, et al., 2020), mood states (Spathis et al., 2019), chronotype (Schoedel et al., 2020), and symptoms of depression and anxiety (Moshe et al., 2021). Thereby, new terms have emerged to describe this line of research, including Psychoinformatics in psychometrics (Markowetz et al., 2014), Personality Sensing in the differential context (Harari, Vaid, et al., 2020), and Digital Phenotyping in mental health research (Insel, 2017).

While mobile sensing holds great promise for providing valuable insights from everyday life into psychological phenomena, it also poses considerable methodological challenges that researchers must overcome. The data generated by sensing apps are highly complex (i.e., timestamped event data), arrive in large volumes (i.e., dozens of events per second), and encompass various modalities (e.g., usage logs, text data, and GPS coordinates). Furthermore, they often contain inconsistencies due to logging errors or discrepancies between devices. Consequently, unlike item responses from self-report questionnaires, these high-dimensional data require extensive preprocessing efforts to derive meaningful behavioral variables that are suitable for studying the phenomena of interest. However, most psychologists lack training in handling mobile-sensing data and are often left without methodological guidance (Wrzus & Schoedel, 2023). As a result, the field often relies on conventional analysis strategies. These approaches could be complemented by emerging methods that allow researchers to more fully leverage the rich potential of these data and thereby enhance both theoretical and empirical insights.

To advance mobile-sensing research, we begin by summarizing the current state of data processing and then present three use cases that go beyond current practices to demonstrate more advanced methods. For this purpose, we use app usage logs as an exemplary starting point for extracting behavioral variables. We systematically report our preprocessing efforts, focusing on two key dimensions: *data enrichment*, which reflects the extend to which raw sensing data are combined with contextual information, and *data aggregation*, which captures how variables are summarized across individual data points.

2. Data collection

To illustrate our preprocessing pipelines and use cases, we use an exemplary dataset collected in the Smartphone Sensing Panel Study (SSPS; Schoedel & Oldemeier, 2020). This study was part of the interdisciplinary PhoneStudy research project at LMU Munich and was conducted in collaboration with the Leibniz Institute for Psychology (ZPID). The aim of the SSPS was to create a benchmark dataset for the research community, comprising longitudinal and high-dimensional sensing data, along with self-report data about a wide range of psychological phenomena. The data collection was comprised of

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

mobile sensing, EMAs, and online surveys, but here, we mostly concentrate our report on the unique sensing data. Additional procedures can be found in our preregistered study protocol (Schoedel & Oldemeier, 2020) and initial publications by große Deters and Schoedel (2024), Reiter and Schoedel (2024), and Schoedel et al. (2023).

2.1. Transparency and openness statement

All procedures of the SSPS received approval from the responsible ethics committee at LMU Munich and complied with the General Data Protection Regulation (GDPR). Before data collection, all participants provided their informed consent, which they could withdraw at any point during the study without giving a reason.

The analyses presented in this manuscript are purely exploratory and only serve illustrative purposes for the preprocessing pipelines proposed here. We also provide an OSF repository with our online supplemental materials (OSMs) and the code for data preprocessing and analysis. All analyses were conducted in the statistical software R (version 4.2.1 for the basic preprocessing and data enrichment steps; version 4.4.1 for data aggregation steps; R Core Team, 2024). For reproducibility purposes, we utilized the package management tool groundhog (Simonsohn & Gruson, 2024). While the data privacy prevents us from sharing the raw sensing data, we provide the set of aggregated variables extracted here in our repository.

2.2. Study procedures

The SSPS took place between May and November 2020 for either three or six months, depending on random group assignment. All data were collected using our custom research app, *PhoneStudy*, which participants installed on their personal smartphones at study onset. Subsequently, the app began to continuously log various mobile-sensing data (see the next section). In addition, the app administered—depending on the group assignment—three to six monthly online surveys (approx. 30 minutes) and one or two 14-day EMA waves, collecting self-reports on various psychological constructs (see Schoedel & Oldemeier, 2020 for an overview of instruments).

The total sample of the SSPS comprised data from 850 participants, collected according to quotas that represented the German population in terms of age, gender, education, income, religion, and relationship status in 2020. Only persons between 18 and 65 years could participate. Furthermore, participants had been required to be the sole users of a smartphone running on the Android operating system (version 5 or higher) due to technical reasons.

In this manuscript, we used only a fraction of the complete data set, specifically the demographics from the first survey (May 2020), as well as the sensing data collected during the first and second EMA waves (07/27/2020–08/09/2020; 09/21/2020–10/04/2020) and the corresponding EMAs (situation perception and sleep diary). We only included participants with at least three sensing days and who answered at least 10 EMAs in wave 1. These components of the SSPS exhibited a sample size of $N = 538$, from which 473 participants provided their demographic information: ages ranged from 18 to 65, with an average of 41 years ($SD = 12.6$). Additionally, 45.7% ($n = 216$) of participants identified as female, while 54.3% identified as male ($n = 257$).

2.3. Data structure

The PhoneStudy app continuously collected data from participants' personal smartphones and spanned 13 distinct sensing modalities, including screen status, notifications, usage logs (for phone, apps, keyboard, and music player), connectivity reports (power and headphone plug, WiFi, Bluetooth, and flight mode), and sensor data (GPS and physical activity). Data were logged as timestamped entries and, depending on the sensing modality, stored with varying specifications (see Table 1). Beyond sensing, the

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

app administered EMAs and stored the corresponding self-reports and their timestamps in a separate table.

To illustrate different cases of sensing data analysis, we will concentrate on the app usage logs, which provide rich behavioral information and allow for extracting variables at different levels of granularity (Sust, Talaifar, et al., 2023). In addition, to further contextualize app usage, we will incorporate data from GPS sensors as well as self-report data from EMAs. We describe these sensing modalities in more detail below, but refrain from reporting on any other modalities and refer interested readers to the methods section of Schoedel et al. (2023).

2.3.1. App usage logs

App usage was logged in an event-based manner, meaning the PhoneStudy app captured data points whenever they occurred, specifically, whenever an app was used. The resulting logs[1] contain timestamped information regarding the package names of the used apps and event types (see Table 1). Package names serve as unique identifiers for each app (e.g., *com.facebook.katana* for Facebook or *bbc.mobile.news.ww* for BBC News) on distribution platforms such as the Google Play Store or on the devices where they are installed. Event types are defined by Android's accessibility services (see Parry & Toth, 2025 for a complete documentation) and encompass a variety of actions, including launch (type 1: activity resumed), transition to background (type 2: activity paused), or closure (type 23: activity stopped). As illustrated in Table 1, using a single app—for instance, opening, navigating within, and then closing the app Facebook—produces a sequence of recorded app events. Below, we discuss how these raw logs can be aggregated to behavioral variables on app usage.

2.3.2. GPS sensor data

GPS sensor data were logged via three different logging modes to provide an accurate representation of the user's environment while conserving battery life. In particular, they were logged (a) at fixed intervals (e.g., every 10–60 minutes, depending on the smartphone model), (b) based on changes (i.e., whenever the coordinates altered significantly) using the Google Fence application programming interface (API), and (c) at the precise moment when EMAs were opened using the Google Snapshot API. The resulting sensing logs contain timestamped GPS coordinates, specifying (among other parameters) the latitude and longitude through the Fused Location Provider API (see Table 1). In Section 4.1, we use these data to provide context for app usage and present one (out of many possible) approaches to preprocess mobile-sensed GPS coordinates.

2.3.3. EMA data

EMAs were scheduled in a pseudo-randomized manner, with two to four questionnaires presented in one of four equally sized sections of the day (from 7 a.m. to 10 p.m. on weekdays and from 9 a.m. to 11 p.m. on weekends), while ensuring a minimum of 60 minutes between samplings. Questionnaires were prompted via a notification as soon as participants first used their smartphones actively after the scheduled time, to avoid provoking artificial smartphone usage (van Berkel et al., 2019). The EMAs contained a varying number of short questions about participants' current mood and situation, as well as sleep-related questions (only in the first EMA per day). While analyzing stand-alone EMA data falls outside the scope of our manuscript, these data are often combined with mobile-sensing data to connect subjective experiences with objective behavioral or situational information at the momentary level (e.g., Elmer et al., 2025; Schoedel et al., 2023). Hence, we will incorporate EMA data into our preprocessing pipelines in Sections 4.2 and 4.3.

[1]The data structure shown in Table 1 is not specific to the research app we used, but represents the data structure output by Android (see also Parry & Toth, 2025).

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

Table 1. Exemplary logs of screen status, app usage, and GPS sensors

Timestamp	Activity	Event	Name	Package name	Latitude	Longitude
2020-07-27 21:40:02	GPS				48.150509	11.580299
2020-07-27 21:50:07	SCREEN	ON_LOCKED				
2020-07-27 21:50:09	GPS				48.155722	11.582963
2020-07-27 21:50:11	SCREEN	ON_UNLOCKED				
2020-07-27 21:50:13	APPS	1 (activity resumed)	Facebook	com.facebook.katana		
2020-07-27 21:50:58	APPS	2 (activity paused)	Facebook	com.facebook.katana		
2020-07-27 21:50:59	APPS	1 (activity resumed)	Facebook	com.facebook.katana		
2020-07-27 21:52:32	APPS	2 (activity paused)	Facebook	com.facebook.katana		
2020-07-27 21:52:33	APPS	1 (activity resumed)	Facebook	com.facebook.katana		
2020-07-27 21:53:50	APPS	2 (activity paused)	Facebook	com.facebook.katana		
2020-07-27 21:53:50	APPS	1 (activity resumed)	unknown	com.android.pacprocessor		
2020-07-27 21:53:51	APPS	1 (activity resumed)	BBCnews	bbc.mobile.news.ww		
2020-07-27 21:53:51	APPS	2 (activity paused)	unknown	com.android.pacprocessor		
2020-07-27 21:53:55	APPS	23 (activity stopped)	Facebook	com.facebook.katana		
2020-07-27 21:57:40	APPS	2 (activity paused)	BBCnews	bbc.mobile.news.ww		
2020-07-27 21:57:41	SCREEN	OFF_LOCKED				
2020-07-27 22:20:02	GPS				48.155722	11.582963
2020-07-27 23:45:06	SCREEN	ON_LOCKED				
2020-07-27 23:45:09	SCREEN	ON_UNLOCKED				
2020-07-27 23:45:12	APPS	1 (activity resumed)	WhatsApp	com.whatsapp		

Note: Timestamp sorted logs of screen status, app usage, and GPS location from an Android smartphone, collected with the PhoneStudy mobile-sensing app. This is artificially generated data that has been simplified for illustrative purposes (e.g., removal of other recording modalities and adjustment of variable values).

3. State-of-the-art preprocessing

Building on the app usage logs, we now explore current approaches to extracting variables from mobile-sensing data. We begin by introducing basic preprocessing steps and then review the current state of the art—that is, the methods most commonly used by researchers to analyze such data to date (see Parry & Toth, 2025).

3.1. Data cleaning and preparation

The event sequences in Table 1 reveal several issues related to the sensing of app usage events, which we needed to consider before getting into further preprocessing. Specifically, app usage events usually exhibit ambiguous patterns, particularly due to multiple launch events (type 1) being recorded within one sequence when users navigate within the components (e.g., menu levels) of an app (Parry & Toth, 2025). When counting the Facebook app launches in Table 1, it may appear as if the app was used three times instead of only once. Furthermore, there are often delays when logging closure events (type 23) after the app has not been actively used for a while (Parry & Toth, 2025). At first glance, the raw sensing data in Table 1 may imply that the app Facebook is being closed at 21:53:55, even though the user started engaging with another app a few seconds prior. The complexity of these logging sequences is further exacerbated by system apps (e.g., *com.android.pacprocessor*), which generate logs without active user interaction through background processes like battery optimization tasks or network management (see also Parry & Toth, 2025; Schoedel et al., 2022). Because of these ambiguous patterns, it is not possible to extract app usage variables by simply counting certain event types without some initial data cleaning.

Instead, it is essential first to define consecutive app usage sessions and to label the data patterns corresponding to one session. Only after this basic cleaning procedure can we derive meaningful variables on app usage behaviors. Here, as a first step, we removed app events created by system apps that are not produced through active user engagement and may have occurred during the usage session of a proper app. Then, consistent with standard practices in mobile-sensing research (Parry & Toth, 2025), we defined the start time of an usage session as the first event when an app A was launched (type 1) and the end time as the app's last logging event, before the launch of another app B (type 1) or the screen turning off. To implement this rationale, we developed a rule-based function to detect and label smartphone usage patterns within the dataset. This function first labeled all app events that occurred within the same smartphone usage session, that is, between the onset and offset of the phone's screen, and assigned them a unique identifier. Within each smartphone session, the function then labeled all app events that occurred within the same app usage session, that is, between the launch (type 1) of an app A and either the launch of another app B (type 1) or the end of the smartphone usage session. Again, all events belonging to one app usage session obtained a unique identifier. Finally, based on these labeled data, we grouped the raw logs by unique app usage session identifier. We generated a new summary table with one entry per app usage session, stating its start and end times and the corresponding package name. This new table of app usage events then served as the starting point for all further preprocessing steps outlined below (see Table 2).

It should be noted that, for the sake of brevity, we have presented a simplified version of our app usage session labeling approach in this article. We have not delved into the specifics of how we handled logging anomalies, which can vary depending on the smartphone or operating system versions. For a more comprehensive understanding of our custom labeling function, we direct readers to the annotated preprocessing R scripts in our OSM. Moreover, for additional details on the preprocessing procedures used to extract screen and app usage sessions, we encourage readers to consult the thorough introduction provided by Parry and Toth (2025).

3.2. Current preprocessing approaches

After labeling usage sessions, we can now identify which app was used at each point in time. By calculating the duration of each session as the interval between its start and end times, we can

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

Table 2. Summary table of individual app usage sessions after basic preprocessing of app usage logs

ID	Package name	Start	End	Duration	App category	Location category	EMA ID	Bedtime ID
34	com.google.android.apps.maps	2020-07-27 19:00:42	2020-07-27 19:05:14	4.53	Orientation	Not at home		
35	com.whatsapp	2020-07-27 19:05:15	2020-07-27 19:08:13	2.97	Communication	Not at home		
36	com.whatsapp	2020-07-27 20:11:23	2020-07-27 20:14:44	3.35	Communication	Not at home		12
37	com.facebook.katana	2020-07-27 20:14:45	2020-07-27 20:16:36	1.85	Social Media	Not at home	25	12
38	com.whatsapp	2020-07-27 20:16:37	2020-07-27 20:22:00	5.38	Communication	Not at home	25	12
39	bbc.mobile.news.ww	2020-07-27 20:52:08	2020-07-27 20:58:24	6.27	News	Not at home	25	12
40	com.facebook.katana	2020-07-27 21:50:13	2020-07-27 21:53:50	3.62	Social Media	At home		12
41	bbc.mobile.news.ww	2020-07-27 21:53:51	2020-07-27 21:57:40	3.82	News	At home		12
42	com.whatsapp	2020-07-27 23:45:12	2020-07-27 23:46:05	0.88	Communication	At home		
43	com.google.android.deskclock	2020-07-28 06:52:05	2020-07-28 06:52:35	0.50	Time	At home		

Note: Each individual app usage session includes the app's package name, start and end times, duration, and various labels introduced in our preprocessing pipelines. In Case 1, app sessions were linked to location information. Sessions were labeled with an EMA ID if they occurred within 60 minutes of an EMA questionnaire and with a Bedtime ID if they fell within the 3-hour window before the participants' self-reported sleep time on a given day. The gray-shaded area is a summary of the raw app logs presented in Table 1.

Table 3. Summary statistics for different usage quantities

	Number of app users *n*	Average daily number of app usage *M* (SD)	Average daily duration of app usage *M* (SD)
Facebook	319	7.25 (10.22)	25.91 (33.24)
Instagram	273	6.02 (8.46)	16.29 (22.49)
TikTok	55	2.37 (4.54)	33.61 (41.71)
Social media	418	11.68 (16.40)	37.01 (46.28)
At home	362	7.08 (12.48)	29.92 (38.63)
Not at home	362	3.45 (8.00)	10.89 (15.48)
Communication	538	29.47 (26.07)	30.26 (28.87)
At home	466	15.22 (16.91)	20.24 (24.05)
Not at home	466	11.39 (16.41)	12.02 (15.02)

Note: Overall sample size $N = 538$.

also determine how long each app was used (Table 2). However, to move from this still relatively raw sensing data to psychologically meaningful variables, additional preprocessing is necessary. The strategies commonly used for this purpose vary along the key dimensions of data enrichment and data aggregation. In the following sections, we explore both dimensions in more detail using our example of app usage logs.

3.2.1. Data enrichment

The app sessions in Table 2 allow us to focus either on individual apps, such as Facebook, Instagram, or TikTok (see the first three rows in Table 3), or to group apps based on their functional similarities to reduce dimensionality. To create such groups, we need external information on the apps' similarities, which we can obtain through data enrichment.

In our example, we used the open-source categorization proposed by Schoedel et al. (2022) and assigned each package name one category (see the column *App Category* in Table 2). This category system was specifically designed for psychological research. The authors developed and validated a taxonomy of 26 behaviorally grounded, unambiguous categories (e.g., Audio Entertainment, Career, and Food) and manually classified over 3,000 commonly used Android smartphone apps through an iterative process. In the examples of this manuscript (see Table 3), we first focus on Schoedel et al., 2022 two categories *social media* and *communication*, which showed inter-rater agreements of 0.63 and 0.71 and summarize 21 and 66 apps, respectively.

Another common source for enriching app usage logs is the default categorization provided by commercial app distribution platforms, such as the Google Play Store (see Böhmer et al., 2011) or the App Store (see Gordon et al., 2019). Both manual and default category systems have advantages and drawbacks. Manually created taxonomies often only cover the most common apps used in a given sample (as in Schoedel et al., 2022), so the aggregated variables tend to systematically underrepresent less common apps, leading to measurement error (see Sust, Talaifar, et al., 2023). While default categorizations avoid this issue, they are designed with marketing in mind and may not offer optimal groupings based on app functionality (see Sust, Talaifar, et al., 2023).

Although no studies, to our knowledge, have systematically investigated and compared different app categorization approaches based on their psychometric properties, enriching app usage logs through categories offers some advantages over analyzing individual apps. First, examining the use of single apps is only meaningful if most participants in the sample use that specific app (Sust, Talaifar, et al., 2023). However, with over two million apps available in the Google Play Store (44matters, 2025), it is unlikely that any single app is universally used across samples. Our example in Table 3 illustrates this point.

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

Only about half of the participants used popular social media apps like Facebook (59.3%) and Instagram (50.7%), resulting in sparse data. Here, the enrichment through external data helped summarize detailed technical events into broader behavioral units, reducing data sparsity and allowing us to capture more behavioral occurrences (Sust, Talaifar, et al., 2023). In Table 3, 78% of the sample used at least one social media app and, compared to Facebook users, the most popular single social media app in our example, category-wise aggregation produced nearly 100 additional observations. A second advantage of using categories is that category labels improve psychological interpretability by summarizing app functionalities, making it easier to connect app use to specific behavioral tendencies like socializing (Sust, Talaifar, et al., 2023). Conversely, analyzing individual apps has limited psychological relevance since researchers usually focus on the behaviors these apps facilitate—essentially their functionalities—which often overlap (Sust, Talaifar, et al., 2023). Third, app categories are less susceptible to shifts in meaning than individual apps because they include various apps and can grow as new apps enter the market. The popularity and functionality of each app can change over time, leading to shifts in usage patterns that can affect the reliability of research results based on a single app. For example, using TikTok might have been seen as innovative in early 2020 in Germany (as shown by the low adoption numbers in Table 3), but it may now be considered quite mainstream and could become outdated in a few years.

To put this into context, it is important to note that whether to apply data enrichment depends on the type of sensing data. As mentioned earlier, in the case of app usage logs, there are several benefits to enriching logs with categories. However, variables based on individual apps remain psychologically interpretable and are still valuable for research. Similarly, some types of sensing data do not require enrichment because their data points are inherently meaningful and sparsity is not an issue. For example, screen status (such as the number of smartphone usage sessions) or call logs (like the duration of outgoing calls) fall into this category. Conversely, other types of mobile-sensing data, such as those from GPS sensors listed in Table 1, cannot be easily interpreted on their own and are not useful without further enrichment. Like app categorizations above, data from external sources can be used for enrichment in these cases. For instance, the timestamped GPS data points in Table 1 can be supplemented with information from weather databases, external map providers that detail types of places (e.g., restaurants, cafés, or shops), or census databases containing population statistics for specific regions or countries (Müller et al., 2022).

Regardless of sensing modality, the process of enriching sensing data with external information can vary greatly in complexity depending on whether the external data are provided in a ready-to-use format (like default app categorizations or types of places from map providers) or need additional preprocessing before enrichment. As an example, consider music player logs that record the titles of played songs and can be enriched with song-level information to develop music preference variables. Song-level data, such as genres or audio features, can be directly obtained from third-party providers like Spotify (Anderson et al., 2021). In contrast, song titles can also be enriched using textual features of their lyrics, as shown by Sust, Stachl, et al. (2023). They first obtained the lyrics for the songs in their smartphone-sensed music logs and then used natural language processing techniques, such as latent Dirichlet allocation (Blei et al., 2003), to identify thematic labels. Assigning these labels to the corresponding music logs allowed for the calculation of lyrics-based preference variables. Pipelines that include preprocessing of external data for enrichment can also be applied to app usage logs (such as text descriptions of the apps) and other modalities.

3.2.2. Data aggregation

While it is theoretically possible to work directly with the (enriched) session-wise app usage data (e.g., relating them to time of day), psychologists usually aggregate them to derive more interpretable variables. For this purpose, the usage sessions in Table 2 can be combined either at the individual level or by app category. This process of quantifying app usage behavior involves two considerations, which we outline below.

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

On the one hand, we need to determine a time frame for data aggregation. This time frame can range from hourly to daily, weekly, or over the entire study period. For our example in Table 3, we used a common approach by first aggregating session-wise data per day. We defined the boundaries of a day based on waking hours (from 6:00 a.m. to 5:59 a.m.) instead of calendar days to better reflect human behavior. Then, we averaged these daily metrics across each participant's study days to obtain person-level variables. We decided to first aggregate data at the daily level before aggregating at the person-level to allow for additional data quality checks (e.g., verifying valid study days and technical completeness of app logging). Overall, the choice of an appropriate time frame depends on the research question and should be carefully considered, as it can influence the results and conclusions of the study (Langener, Stulp, et al., 2024; Schoedel et al., 2020).

On the other hand, we have to choose an aggregation method. Typically, researchers usually quantify the frequency or duration of app usage within a predefined time frame using basic summary metrics. In our example, we calculated person-level metrics by averaging the number and duration of daily app usage across individual apps and app categories over the available study days (see Table 3). We used the median as a measure of central tendency because it is more robust to outliers, which can occur due to logging and labeling errors in the sensing data, than the arithmetic mean. Of course, many alternative approaches exist for aggregating app usage sessions. Besides frequency and duration, more advanced quantifiers like the ratio of a specific app's usage to total app usage can also be used (e.g., Schoedel et al., 2018). As shown in our example, variables defined differently (such as frequency versus duration) can display different patterns and may reflect different aspects of app usage behavior. Additionally, summary metrics can extend beyond central tendency to include measures of dispersion (e.g., variability and range), density, and robust alternatives such as the Huber mean (e.g., Stachl, Au, et al., 2020).

When selecting their aggregation procedure, researchers also need to decide how to handle missing values in the sensing data. Importantly, missingness must be considered at every aggregation step, such as when summarizing daily frequencies or durations and when combining them to person averages across study days. At the lower level, when summarizing app usage sessions, it is crucial to carefully determine whether data were unavailable (for example, due to technical logging failures) or whether participants simply did not exhibit the target behavior (such as not using apps of a certain category). Once logging errors are ruled out through plausibility checks, missing data can be recorded as zeros in the respective behavioral aggregates. Consequently, in our example, days without any sessions from the apps or app categories in Table 3 were assigned a zero. At the higher level, when aggregating daily values for each person, researchers actively control the number of missing values by setting a minimum threshold for study days to be included. In our example, we calculated person-level medians across all study days with available sensing data. In doing so, the median could have been derived from just one study day or from multiple study days, depending on participants' data availability. Alternatively, we could have set a minimum number of study days required to compute the average, and if that threshold was not met, recorded a missing value. In this context, it is crucial to balance the amount of missing data in a variable with the number of data points (or study days) needed to accurately assess average behavioral tendencies.

All of these considerations regarding the selection of an appropriate aggregation method apply not only to app usage logs but also to other sensing modalities. Apart from the basic aggregation described here, more complex methods can be applied, as seen in our use cases below.

3.2.3. Data enrichment and aggregation in practice

When selecting preprocessing strategies, there is no universally correct solution for data enrichment or aggregation—appropriate choices depend on the specific research question and data characteristics. As a result, researchers face considerable degrees of freedom when preprocessing raw sensing data.

In this context, it should also be recognized that these two dimensions introduced above cannot always be viewed as separate, sequential, and independent steps in the data preprocessing workflow. Instead, they are sometimes performed in a reverse order (see our Case 1) and can even overlap often

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

(see our Case 2 and Case 3). For example, suppose researchers want to find each participant's most visited place—such as a restaurant, café, or outdoor area—over the study period. In that case, they would first identify the most frequently visited location using aggregation methods and then enhance this GPS position with an external map database.

4. Preprocessing use cases

Building on the state-of-the-art approaches outlined above, we now present three preprocessing use cases that extend the concepts of data enrichment and aggregation to more complex solutions for extracting variables from mobile-sensing data. Each case demonstrates how contextualizing variables through the integration of multiple data modalities—specifically, GPS and EMA—can yield more insightful variables. The aggregation applied in these cases expands to new temporal scopes (i.e., hourly windows around EMA prompts) and progresses from basic summarization to statistical and predictive modeling, thereby enabling more nuanced, within-person analyses. As the complexity grows, the boundaries between enrichment and aggregation become less distinct. For each case, we outline an exemplary research question, the detailed preprocessing pipeline, and ideas for additional research questions and preprocessing approaches. Importantly, although some analyses may be interesting on their own, they mainly serve preprocessing purposes here. Their primary goal is to generate variables that can be used later for formal modeling in psychology.

4.1. Case 1: Data integration

In the state-of-the-art preprocessing described above, we enriched raw sensing data using external information (on app categories). However, another powerful approach to enrichment involves leveraging internal data—specifically, by integrating different sensing modalities to derive more contextualized and nuanced variables. As outlined in Section 2.3, our mobile-sensing app collected various types of data concurrently, including usage logs and sensor data. By combining these modalities, we can enrich behavioral variables with additional contextual information—such as temporal, physical, spatial, social, or digital context (Harari & Gosling, 2023). In Case 1, we demonstrate this internal enrichment by integrating data from two sensing modalities: we enhance category-wise app usage metrics with location labels derived from GPS data.

4.1.1. Exemplary research question

To illustrate how app usage can be further contextualized through data from additional sensing modalities, we extract variables that capture social media and communication app usage across different locations—specifically, comparing behavior when participants are at home versus away. These context-aware variables not only reveal location-based differences in app use but can also be analyzed in relation to psychological outcomes, such as symptoms of mental illness, well-being, or personality traits. For instance, higher social media use at home may be linked to lower well-being, since both frequent social media activity (e.g., Valkenburg, 2022) and more time at home have been connected to reduced psychological well-being (e.g., Müller et al., 2020).

4.1.2. Preprocessing approach

To enrich app usage sessions with location information, we first needed to preprocess the raw GPS sensor data, which often contains noise. In our experience, the accuracy of GPS logs depends on technical differences between smartphone manufacturers (hardware) and Android versions (software). Specifically, GPS data points are frequently scattered around participants' exact locations (e.g., at home) even if they did not move (Müller et al., 2022). To address this scattering, researchers should first identify key locations and then label the raw GPS data points based on their distance to these locations.

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

(We recommend Müller et al. (2022), for a comprehensive introduction to working with GPS data in psychological research.)

To identify key locations, we began by calculating a distance metric between GPS data points (longitude and latitude) using the Haversine metric ($(d_{Haversine} = 2r \cdot arcsin(\sqrt{sin^2(\frac{\phi_2-\phi_1}{2})+cos(\phi_1)\cdot cos(\phi_2)\cdot sin^2(\frac{\lambda_2-\lambda_1}{2})}))$, with ϕ_i representing the latitude of a location i, λ_i being the longitude of that location, and r = 6, 378.137km serving as the radius of a spherical approximation of Earth) using the geosphere package (Hijmans, 2024). We then applied the density-based spatial clustering of applications with noise (DBSCAN) algorithm, as implemented in the R package of the same name by Hahsler et al. (2019). Originally proposed by Ester et al. (1996), this unsupervised machine-learning algorithm identifies clusters in spatial data (here: two-dimensional data) without the need to specify the number of clusters in advance. In more detail, DBSCAN identifies clusters by grouping nearby data points while marking points in low-density regions as noise. The algorithm requires two key parameters: the radius of the neighborhood (ε) and the minimum number of points required to form a dense region (*MinPts*). In other words, it has to be determined how far apart spatial data points can be and how many spatial data points are necessary to form a cluster. A point is classified as a core point if its neighborhood contains at least the specified number of *MinPts* points and otherwise as a border or noise point. DBSCAN iteratively expands clusters of the core points by merging the original clusters of core points when they are within the specified distance ε of each other. The resulting output includes various clusters (i.e., key locations) and any points outside these clusters are considered noise points (see Hahsler et al., 2019).

Our analyses were based on one of the two 14-day EMA waves within the SSPS (compare Section 2.2). However, to identify key locations, we used all GPS data points collected over the entire three- to six-month study period to ensure a larger dataset and facilitate the identification of more robust clusters. To reduce computational costs, we randomly sampled up to 5,000 GPS data points per participant if more data were available.

The key location of interest in our analysis was *home*. Similar to previous studies (Müller et al., 2022; Saeb et al., 2015), we defined home as the place where participants spent most of their time between 1:00 a.m. and 5:00 a.m. For each participant, we filtered all GPS data points recorded during this nighttime window and then applied the distance metric and the DBSCAN clustering algorithm (ε = 30 [meters]; *MinPts* = 3 [data points]) as described above. We selected the hyperparameters (ε = 30 meters, MinPts = 3) based on prior experience with similar GPS datasets (Schoedel et al., 2023). While these parameters produced meaningful clusters in our data, alternative settings may also be appropriate depending on GPS sampling rate, accuracy, and point density (Müller et al., 2022). Researchers may adjust these values to optimize cluster detection for their own data and ensure robustness of results. While manually optimizing the hyperparameters of DBSCAN arguably makes the most sense in typical psychological research settings, researchers can also perform systematic hyperparameter tuning or even rely on automated procedures (e.g., utilizing reinforcement learning, Zhang et al., 2022) if the computational resources and sample size allow it. In this case, based on the hyperparameters specified above, we performed DBSCAN without extensive hyperparameter tuning, selected the cluster visited most frequently during that time window, and calculated its center to determine the home location for each participant.

Afterward, we revisited the raw sensing data (see Table 1) and calculated the distance between each GPS point and the identified home cluster using the Haversine metric. Data points within a radius of ε = 30 meters were labeled as *at home*, while all others were labeled as *not at home*. Since GPS data in this study were not just collected at regular intervals but also in response to location changes (see Section 2.3.2), we extended the respective label to all sensing data points recorded between two consecutive GPS logs.

To extract the usage quantities of interest (i.e., social media and communication app usage at home and not the home), we used the summary table (see Table 2) as a starting point. As described in Section 3.2, we enriched the app usage sessions by their respective app category. Additionally, we incorporated the previously assigned GPS labels from the raw sensing data and added them as a new column in Table 2

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

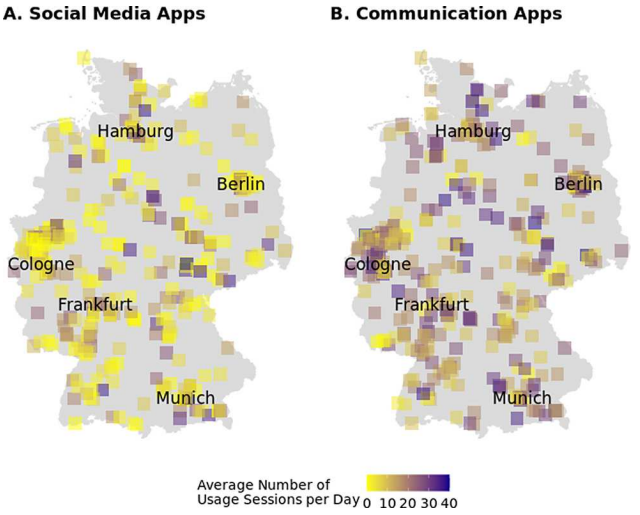

Figure 1. Average daily number of uses of social media and communication apps by GPS-based home location, across all participants recruited in Germany.

(see column *Location Category*). When more than one GPS label was assigned to a single app usage session (e.g., when participants moved locations while app use), we used the GPS label corresponding to the location where the app was initially opened. However, such cases were very rare due to the typically short duration of app usage sessions.

Next, we filtered the data to include only app usage sessions from the social media and communication categories and grouped them by location. In line with state-of-the-art procedures, we calculated the average (i.e., median) daily total number and duration of app usage sessions per category and location across study days (see Table 3).

4.1.3. Final variables

The lower part in Table 3 provides an overview of summary statistics by app category and location. Due to GPS data availability, location-dependent app usage could only be extracted for a part of the sample, with social media app usage recorded for 67.3% of participants and communication app usage for 86.6%. Both app categories were used more frequently and for longer periods when participants were at home, with these discrepancies being especially pronounced for social media apps, which were, on average, used twice as often and three times as long at home compared to other locations. Social media apps were used less frequently than communication apps, regardless of location. This descriptive finding is illustrated in Figure 1, which shows the average daily number of social media and communication app sessions, grouped by participants' home locations across Germany. In subsequent formal analyses, these location-dependent app usage variables could be related to various person-level variables, as suggested in Section 4.1.1.

4.1.4. Outlook

In Case 1, we demonstrated how raw sensing data can be enriched internally by integrating different sensing modalities. Specifically, we used spatial context derived from GPS data to transform general app usage metrics—commonly applied in psychological research—into more nuanced, contextualized variables. While the aggregation approach remained basic (see Section 3.2), the enrichment process was more sophisticated: rather than relying on external sources, we drew on internal data, which first required preprocessing steps such as identifying participants' home locations. This integration of

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

data streams enables researchers to capture everyday behaviors within meaningful situational contexts, thereby offering deeper insights for addressing specific psychological research questions.

There are, of course, various ways to combine app usage with GPS data beyond the pipeline from our example. For instance, GPS data points recorded outside the home could be further categorized by location types (e.g., restaurants and shops) if they are first enriched with external data (see Müller et al., 2022 and Section 3.2). Additionally, app usage logs could be integrated with other sensing modalities as long as the contextual data are logged simultaneously during app use (e.g., sensor data and connectivity reports). For example, Do et al. (2011) utilized logs of nearby Bluetooth devices to distinguish between app usage when users were alone versus when others were nearby, and Böhmer et al. (2011) used accelerometer data (from physical smartphone sensors) to examine app usage during different physical activities. Similarly, different types of contextual sensing data can also be combined for a more detailed picture. For example, Rüegger et al. (2020) merged Bluetooth and GPS logs to infer participants' social contexts. They classified anonymized Bluetooth signals based on location (e.g., home or workplace) to label them concerning social interaction partners. Devices detected at home were assumed to represent close contacts, such as family or partners, while those at work were linked to less emotionally supportive contacts, like colleagues. Beyond different sensing modalities, sensing data may also be integrated with information from other sources such as self-reports (see Section 4.2). Finally, another promising direction is the integration of sensing data across devices, as wearables like fitness trackers and smartwatches become more accessible for research (Schoedel & Mehl, 2024). Combining physiological data from wearables with smartphone screen time, app usage, or music logs could offer rich, contextual behavioral insights for psychological research. Despite its great potential, data integration remains underexplored in current mobile-sensing research, offering promising opportunities for future studies.

It should, however, be noted that increasing the information density in variables generally comes with costs. On the one side, different sensing modalities require unique handling during preprocessing—whether due to different logging modes (e.g., event-based, change-based, or interval-based) or the modality-specific information involved (e.g., app package names versus GPS coordinates). As a result, preprocessing pipelines become more complicated, with different steps for each sensing modality. This makes the pipelines longer and requires more analytical decisions, ultimately introducing additional degrees of freedom for researchers. On the other side, combining multiple sensing modalities involves a balance between the richness of the variables and data sparsity. While multiple data modalities enable the observation of more specific behaviors, they also depend on data being available across modalities, reducing the number of observations for certain analyses. Compared to the broader app usage variables in our advanced preprocessing example, the sample size shrank when focusing on more location-specific variables because some participants had few GPS data points, which were not enough to identify home clusters. In summary, researchers need to decide whether adding sensing modalities will provide enough data points for their analysis.

4.2. Case 2: Data integration via statistical modeling

In our state-of-the-art example and Case 1 presented in Section 4.1, sensing data were first enriched and then aggregated at the person level using simple metrics such as the median. Alternatively, researchers aiming to integrate data across modalities may develop more sophisticated variables that directly reflect the relationships between the different data types. In doing so, enrichment and aggregation become intertwined, which is most effective when applied to data at a detailed, moment-to-moment level. While establishing relationships through (bivariate) covariances or correlations is possible, more complex statistical models can offer advantages for several reasons, including (a) condensing information, (b) correcting for data dependencies (e.g., in longitudinal data), and (c) identifying outliers or high-leverage cases. In Case 2, we present a relatively simple example of how data from different modalities can be combined through statistical modeling to create variables for subsequent formal analysis. We again use categorized app usage sessions but incorporate them with self-reports from our EMA data. Specifically,

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

we model multivariate relationships between app usage and self-reports using (semi-)parametric models and extract model information at the observation level (i.e., individual model parameters) as final variables. Along with a traditional nomothetic approach, we also employ an idiographic approach, which is gaining popularity in psychology (e.g., Beck & Jackson, 2022; Bringmann et al., 2017; Wright et al., 2019).

4.2.1. Exemplary research question

To demonstrate how relationships modeled among different sensing modalities can generate input variables for further formal modeling, we draw on an example from situation research. In this field, scholars often examine how individuals subjectively perceive situations and how individual differences in these perceptions relate to person-level characteristics such as personality traits (e.g., Kritzler et al., 2020). Staying within the scope of our app usage example, we focus on perceptions of sociality in the context of social media and communication app use—specifically, how participants interpret the social nature of situations in which they use apps from these categories. To investigate this, we extract variables that capture both the direction and strength of the association between app usage (by category) and participants' momentary perceptions of sociality as reported in EMA responses. These variables reflect individual differences in social reactivity—that is, how a person's perception of sociality varies in relation to their app usage. Of course, this is an interesting analysis on its own. However, in the context of this manuscript, such person-level indicators of (social) reactivity can be used in subsequent analyses to explore broader psychological questions, such as whether stronger associations between social app use and perceived sociality are linked to loneliness. Beyond such nomothetic considerations, we also extract variables from autoregressive effects, representing the stability of perceived sociality across situations. As situation perception is linked to affective states (Horstmann & Ziegler, 2019), stronger stability could, for example, indicate persistent negative affect patterns or even mental health issues. In sum, such associations within app usage and self-reported sociality could be interesting variables for research in social, personality, and clinical psychology.

4.2.2. Preprocessing approach

To model the relationship between app usage and concurrent perceptions of sociality, we needed to extract sensing variables at a momentary level (rather than the person-level variables considered so far) and in a timely relation to EMA instances. Since app usage data result from interactions with the respective app, they cannot be generated while completing the EMA questionnaire. Therefore, we could not extract app sessions at the exact moment of the EMA; instead, we had to aggregate them over a time window surrounding the EMA instance (see Schoedel et al., 2023). The choice of the time frame length is somewhat arbitrary and lacks clear guidance in current mobile-sensing research. Hence, we selected 60 minutes (30 minutes before and after the EMA) based on practical considerations: this duration was enough to capture multiple app usage sessions while remaining short enough to prevent overlap between consecutive EMA instances (which could occur 60 minutes apart). EMA data were stored in a separate data table. We developed a rule-based function that extracted the EMA start times from this table and then labeled all app usage sessions in Table 2 occurring within a 60-minute time frame around these start times with a unique identifier. This approach allowed us to match all app usage sessions linked to the same EMA with the respective self-reports. Within each EMA time frame, we calculated the total duration of app usage sessions for the social media and communication categories.

Regarding the EMA data, we selected a dichotomous item adapted from the S8-I scale that assesses situation perception in terms of the situational eight DIAMONDS (Rauthmann & Sherman, 2015). This item reflected participants' self-reported perception of sociality, meaning they indicated whether they believed their current situation allowed for or required social interactions.

To integrate these two sources of information, we modeled the relationship between perceived sociality and app usage durations from social media and communication categories across various

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

situations using both a nomothetic and an idiographic approach. In both models, we used sociality self-reports as the criterion variable and aimed to estimate model parameters as new person-level variables.

As a nomothetic approach, we employed a classical binomial generalized linear mixed model (GLMM; e.g., Bolker et al., 2009) with a logit-link due to our dichotomous outcome variable. We ran this model based on a sample including all participants with at least 10 EMA instances. We utilized the *lme4* package (Bates et al., 2015) and accounted for between-participant heterogeneity in the GLMM by incorporating a random intercept and random slopes (see Equation (1)). This mixed model quantifies, inter alia, inter-individual differences in the strength of association between communication and social media app use, and sociality:

$$g(\mathbb{E}[Y_{ij}]) = \beta_0 + \beta_{SM} SM_{ij} + \beta_{Com} Com_{ij} + b_{0j} + b_{SM,j} SM_{ij} + b_{Com,j} Com_{ij} \quad (1)$$

with $g()$ being the logit link to account for the binary response variable, Y_{ij} being the perceived sociality of person j in instance i, SM and Com being the duration of social media and communication app usage, respectively, and β describing the fixed and b the random effects.

This model considers the effects of individual participants (i.e., the random intercept and slopes), but does not explicitly account for temporal correlations. We decided not to include autoregressive effects per default in our nomothetic modeling procedure for Case 2 for several reasons. First, the distances between measurement occasions varied both within and between participants. Second, observations per participant were relatively few, and, third, high stability of perceived sociality over time was unlikely (see also the discussion below). However, we (a) performed a sensitivity analysis with an autoregressive effect (see Equation (2)) to check the latter assumption and to address potential temporal effects and (b) conducted individual regression analyses for each participant to illustrate how autoregressive effects could serve as person-specific variables in an idiographic modeling approach:

$$g(\mathbb{E}[Y_{ij}]) = \beta_0 + \beta_{SM} SM_{ij} + \beta_{Com} Com_{ij} + \beta_{AR} Y_{i-1,j} + b_{0j} + b_{SM,j} SM_{ij} + b_{Com,j} Com_{ij} \quad (2)$$

with β_{AR} quantifying the autoregressive effect using a lag1-variable $Y_{i-1,j}$ that describes the perceived sociality of the previous instance $i-1$.

Instead of explicitly modeling temporal correlations through autoregressive effects as in Equation (2), researchers could also incorporate these effects using temporal correlation models (e.g., Ver Hoef et al., 2010). These pseudo GLMMs enable to account for temporal correlations across residuals and heteroscedastic data structures, rather than assuming residuals are independent when controlling for clustering with random effects. For example, the R-package *nlme* (Pinheiro & Bates, 2000) allows researchers to easily include such temporal correlations. To address issues with non-equidistant measurement occasions that impact discrete-time modeling of autoregressive effects, continuous-time models can also be employed (e.g., Driver et al., 2017).

As an idiographic approach, we fit person-specific models with autoregressive effects to examine intra-individual trajectories while incorporating sensing and EMA data (see Equation (3) for such a model with social media app usage as a predictor and a lag1-variable to quantify an autoregressive effect):

$$\text{logit}(\mathbb{P}(Y_i = 1)) = \beta_0 + \beta_{SM} SM_i + \beta_{AR} Y_{i-1} \quad (3)$$

with Y_i being a participant's perceived sociality at measurement occasion i, Y_{i-1} being the lag1-variable indicating the perceived sociality of the previous time point, and β_{SM} and β_{AR} being the effect of social media app usage and the autoregressive effect, respectively.

4.2.3. Final variables

At the nomothetic level, our mixed logistic regression model was fit to a sample of 421 participants, with a total of 8,887 situations, of which 57.2% were perceived as social. The model identified a significant[2]

[2] Assuming a conventional α-level of 0.05, 0.01, or 0.001.

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

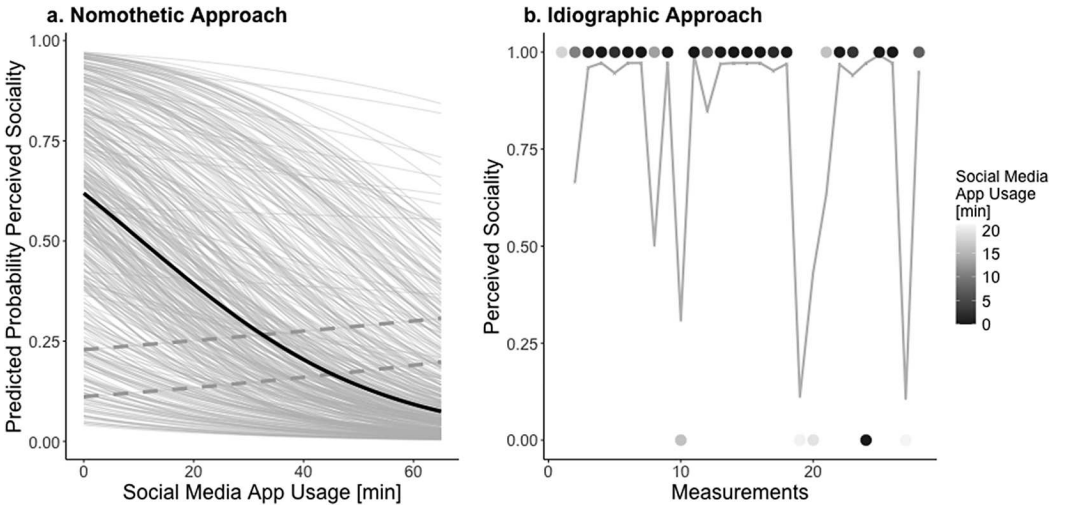

Figure 2. (a) GLMM predictions for all participants given average communication app usage. The solid dark line represents the fixed effect, while the dashed lines illustrate participants with inverse relationships. (b) Time series of Participant 377 and the predicted sociality perception based on an individual logistic regression model with an autoregressive effect.

fixed effect for social media app usage ($\hat{\beta}_{SM} = -0.046, p < 0.001$), whereas the duration of communication app usage was not related to the perceived sociality of situations ($p = 0.608$). Since the random effect variance was also higher for social media app usage ($\hat{\tau}_{SM} = 0.00173$ vs. $\hat{\tau}_{Com} = 0.00120$),[3] we focus on inter-individual differences regarding how social media app usage relates to perceived sociality across situations. Panel (a) in Figure 2 illustrates a notable level of inter-individual variation among participants, with two individuals exhibiting an association in the opposite direction (see dotted lines).[4] The individual slopes from this model could now be used as variables in subsequent formal analyses, for example, to explore how individual differences in perceiving sociality during social media usage connect to person-level outcomes (see Section 4.2.1).

To model these associations in an idiographic manner, we fitted the individual logistic regression with autoregressive effects to the sensing and EMA data of one exemplary participant with many observations (user ID: 377). However, this individual still exhibited only 28 non-equidistant measurements, which may reduce the statistical power of our significance tests and the accuracy of the parameter estimates. Therefore, the following results should be interpreted with caution. The model showed that the duration of social media app use was negatively associated with the perceived sociality of a situation ($\hat{\beta}_{SM-377} = -0.269, p = 0.011$), and the lag variable was not significantly related ($p = 0.422$), indicating low stability of perceived sociality across situations for this participant. Panel (b) in Figure 2 depicts the model predictions for this example individual. Similar to the nomothetic approach, a participant's slope parameter (e.g., $\hat{\beta}_{SM-377}$) could serve as a predictor for a specific outcome in subsequent formal analyses. Depending on the research question, the autoregressive effect (i.e., the estimated coefficient from the logistic regression) may also be a useful variable. While perceptions of sociality varied widely across different daily situations for our exemplary participant, the autoregressive effect may look different in other contexts.

[3]Note that due to the logit link, these numbers are relatively small.

[4]Note that including an autoregressive effect using a lag1-variable as described in Equation (2) did not alter these findings. Social media app usage—unlike communication app usage—remained a significant predictor, and the same two individuals showed positive associations between social media use and perceived sociality, while the overall fixed effect indicated a negative relationship.

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

4.2.4. Outlook

Case 2 expands on our previous example from Section 4.1 on both preprocessing dimensions. While also enriching data internally, this time we combined different data sources instead of just two sensing modalities, specifically mobile-sensing and EMA data. To achieve this, we modified our aggregation methods in two ways. First, we introduced a new time frame for data aggregation so that sensing data were aggregated not at the daily or person level, but at the level of (i.e., in timely correspondence to) EMA instances. Second, to identify variables that represent the relationship between these two data sources, we used statistical models at the intra-individual observation level. This more advanced approach combined data enrichment and aggregation into a single step, allowing researchers to capture highly contextualized variables that merge behavior and perception. By leveraging statistical model parameters for data aggregation, this use case illustrates how researchers can move beyond simple behavioral metrics to derive complex variables suited for addressing more nuanced psychological research questions.

Depending on the research question, hypotheses, and available dataset, the preprocessing pipeline in Case 2 can vary considerably in complexity, as different statistical modeling approaches may be employed. In our example above, we used relatively simple linear models and compared a nomothetic approach with an idiographic one. When choosing between the two, it should be noted that person-specific models are most effective when extensive longitudinal data with multiple measurements per individual are available. Conversely, when fewer measurement occasions are present, mixed models help distinguish within- and between-person variability, enabling the creation of person-specific variables (e.g., by considering random effect terms) while reducing outliers and sampling errors caused by small sample sizes at the within-person level through regularization. Expanding beyond linearity, researchers might also explore non-linear effects using a GAMLSS (Stasinopoulos et al., 2018) or other semi-parametric methods to derive both person-specific means and individual variances as variables. Since the purpose of this statistical modeling step is to generate new variables for further analysis, interpretability may take priority when selecting a model. Therefore, researchers must carefully choose a modeling strategy that adequately captures key data features while providing interpretable values at the desired level (e.g., person-specific scores that can be used to examine psychological outcomes).

Variables capturing relationships within the collected data can be derived not only between EMA self-reports and various sensing modalities but also across different sensing modalities or even within a single modality. For example, this approach may be useful when studying sequences of app usage, such as by applying Markov models (e.g., Zhang et al., 2016). While, technically, all types of self-report and sensing modalities can be integrated, the chosen data sources might limit the model options if certain assumptions are violated. For instance, EMAs are frequently collected only a few times per day and at non-fixed intervals (Wrzus & Neubauer, 2023) and often have systematically missing data (Reiter & Schoedel, 2024), making them unsuitable for modeling person-specific effects in discrete-time models and requiring continuous-time modeling (see, e.g., Driver et al., 2017; Koch et al., 2023; Voelkle et al., 2012). Likewise, most sensing modalities, such as app usage, are not recorded in an interval-based way but depend on active user engagement. Therefore, the data structure must be taken into account when choosing a statistical model for variable extraction.

4.3. Case 3: Data substitution via predictive modeling

Building on Case 2, preprocessing can extend beyond modeling relationships between data sources to substituting data instances based on these associations. This approach mainly targets the substitution of EMA responses, which impose a significant burden on participants in large-scale longitudinal studies (Wrzus & Neubauer, 2023). EMA responses can be predicted from passively collected sensing data, which would be helpful if not all participants answer the EMAs or if EMAs are only collected during certain periods of the study. In Case 3, we demonstrate how EMA and mobile-sensing data can be integrated to address this challenge. Similar to Case 2, enrichment and aggregation become intertwined—only here, we use machine learning instead of statistical modeling to establish

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

relationships within the data. Additionally, instead of estimating parameters for each participant, our goal is to train a supervised machine-learning model capable of predicting single instances of behaviors or experiences. These predictions can then serve as stand-ins for variables of interest in subsequent analyses, especially when data points are missing by design or at random. In this way, researchers could develop machine-learning-based prediction models in one study and apply these models in follow-up studies to drop EMAs and substitute future instances with model predictions based on the mobile-sensing data (Wiernik et al., 2020). In our example, we train a basic random forest model to predict EMA self-reports based on various app usage quantities and evaluate its performance to see how well our model can estimate unseen data points for future formal analyses.

4.3.1. Exemplary research question

To illustrate our substitution approach and predict self-reports in EMAs from mobile-sensing data, we focus on an example from sleep research. Besides wearables like fitness trackers, smartphones can also serve as "sleep sensors." Accordingly, there are early empirical hints that evening app use may be a relevant predictor for sleep outcomes (Pillion et al., 2022). Two key outcomes here are sleep duration and morning relaxation, both of which are typically measured via self-reports in field studies (Carney et al., 2012). In this study, we use app usage data before going to bed to infer these outcomes from passively sensed app usage. Our goal is to see if these predicted values can reliably replace EMAs. If they can, we could use the prediction model in a new study to reduce the need for repeated sleep self-reports. This approach would facilitate long-term studies of sleep patterns and their variations in real-world settings by passively tracking sensing data, removing the burden of daily EMAs on participants.

4.3.2. Preprocessing approach

To predict self-reported sleep outcomes based on app usage before bedtime, we needed to extract our sensing variables during the period before participants went to sleep each day. For this purpose, we applied a similar timely aggregation procedure as in Case 2 and as described in große Deters et al. (under review). That is, we developed a rule-based function to extract self-reported times of falling asleep from EMAs and labeled all app usage sessions in Table 2 occurring within a 3-hour window before this sleep time with a unique identifier. Next, we aggregated the total number and duration of app usage sessions across these 3 hours for different app categories. This time, we not only included the social media and communication categories but extracted app usage for all 25 categories (excluding system apps) from Schoedel et al. (2022). These categories included, for example, audio entertainment (e.g., music apps), finance (e.g., banking apps), gaming, internet, and shopping.

To determine sleep outcomes, we used four items from the Consensus Sleep Diary (Carney et al., 2012), which were collected during the first EMA instance each day. First, to assess sleep duration, we selected three items that asked participants to report their time of attempting to fall asleep at night, the time it took to fall asleep, and the time of waking up in the morning. We only included days where the reported sleep time was after 7 p.m. and before 3 a.m, and we calculated the time offset between going to sleep and waking up. As a second sleep outcome, we used an item on morning relaxation, where participants indicated how rested or refreshed they felt upon waking (6-point Likert scale, from 1 ["not rested at all"] to 6 ["very rested"]). Observations with missing values in either of the outcome variables were removed in a list-wise manner.

In the final enrichment and aggregation step, we predicted the two sleep outcomes of sleep duration and morning relaxation based on app usage numbers and durations across various categories before going to sleep using two random forest models (Breiman, 2001). Random forests were selected as a representative state-of-the-art machine-learning approach that performs well "off the shelf" and typically requires no extensive hyperparameter tuning, thus reducing computational costs for our small use case (e.g., Sterner et al., 2023). We employed 10-fold cross-validation to detect overfitting. Since our goal was to develop a model that could be applied to sensing data in a new study (i.e., to replace

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

EMA-based sleep outcome measures), we split data instances by participant to avoid information leakage between training and testing sets. This means that all instances belonging to the same participant were either in the training set or the test set together. For performance evaluation, we used the proportion of explained variance (R^2) and the root mean squared error (RMSE).

4.3.3. Final variables

We trained and evaluated our random forest models on a sample of 534 participants, with a total of 8,429 days[5] with complete sleep outcome data, averaging 16 days per participant (range: 1–24 days).

For both outcomes, the random forest models showed a poor prediction performance. For sleep duration, the average performance across 10 cross-validation test sets was $R^2 = -0.05$ and $RMSE = 2.52$. For morning relaxation, the cross-validated performance was $R^2 = -0.04$ and $RMSE = 1.16$. The negative coefficients of determination indicate that the models performed worse than the baseline constant model, which, in this case, simply predicted the mean of the respective outcome variable.

Had we been able to predict sleep outcomes from app usage data with satisfactory performance, we could have used the random forests to generate predictions for new samples, replacing the time-consuming EMA with sensing-based proxy variables. For this purpose, we would have re-trained the random forest models on the full available dataset and used these final models for making predictions. In our OSM, we provide the R code for the complete use case. However, as both random forest models failed to learn meaningful relationships between sleep outcomes and evening app usage, we refrain from executing these steps here. A more detailed description of a typical machine-learning modeling pipeline is provided by Pargent et al. (2023) and Sterner et al. (2023).

4.3.4. Outlook

Our final case adopted a predictive modeling approach to combine two data sources, extending Case 2 by further enhancing both enrichment and aggregation at the same time. Like before, this case merged mobile sensing with EMA data, but this time not to create a new person-level variable. Instead, we meant to substitute (unseen) self-reported data points with model predictions from passive sensing data, which would be especially useful for obtaining proxies for EMAs, which are burdensome to collect (see also Reiter & Schoedel, 2024). However, for this method to produce predictions that can be used as data points in subsequent analyses, the data involved must show a substantial relationship that the chosen model can detect. Unfortunately, in our example above, that was not the case, as the two self-reported sleep outcomes could not be successfully predicted from mobile-sensed app usage. Therefore, Case 3 can be viewed more as a vision for the future and a source of inspiration.

We identified several common challenges in predicting EMA self-reports from sensing data that likely contributed to the poor predictive performance of the random forest models. First, self-reports often contain substantial measurement error (in the sense of reduced reliability), which can diminish the predictive performance of machine-learning models, especially when capturing non-linear relationships (Jacobucci & Grimm, 2020). This issue is especially problematic for EMA self-reports, which often consist of single-item measures that make it difficult to assess and model measurement error (for example, by estimating a latent variable model). Therefore, if researchers aim to develop prediction models using EMA data, they should gather psychometrically sound self-report measures to obtain more reliable results. Second, sensing data exhibit a sparsity problem (see Section 3.2) that becomes more noticeable when considering event-based modalities (such as app usage) in relation to randomly timed EMA instances. For example, the 3-hour window before sleep often lacked any app usage sessions from categories like finance or dating. This sparsity limits the ability of random forests to learn systematic relationships between the predictor variables and outcomes. Additionally, it is important to note that the strength of random forest models is in their capacity to learn from many predictor

[5]In contrast to Case 2, we here included data from the study's first and second EMA waves, allowing us to increase the number of observations for the machine-learning model, as sleep was measured only once per day.

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

variables at once. In Case 3, we only used app usage variables to stay within this outlet's scope, but it might be more effective to combine different sensing modalities, with thousands of potential sensing variables to consider (e.g., Stachl, Au, et al., 2020). Third, it is difficult to identify the best algorithm and hyperparameter settings to effectively capture relationships in the given data. In our example, we chose a simple random forest because it can handle high-dimensional feature spaces (in our case, 50 features, but often more with sensing data, see Stachl, Au, et al., 2020), while keeping computational costs low by being not very sensitive to hyperparameter tuning (Goretzko & Ruscio, 2024; Probst, Boulesteix, et al., 2019; Probst, Wright, et al., 2019). Alternatively, we could have used other types of algorithms capable of handling high-dimensional feature spaces, such as boosting algorithms, penalized regression models, or neural networks. Combining more advanced models, like gradient boosting machines (e.g., the XGBoost; Chen & Guestrin, 2016), with more complex tuning, could produce better prediction results.

Beyond these general considerations of predictive modeling, the data substitution approach raises questions of generalizability—particularly in relation to evaluation strategies and model selection. In our example, we adopted a conservative evaluation strategy by applying a cross-validation scheme to the entire dataset and splitting the dataset by participants to prevent data leakage from repeated measures. We did this to achieve realistic performance estimates for applying the model to new participants in a different study, which aligns with the practical goal of using such predictions as proxies for EMA responses in samples with sensing data but no EMA. Alternatively, we could have trained our models on data from the first few study days and tested them on the remaining days. This approach would likely have yielded better predictions but would have resulted in a model specifically tailored to the participants in our sample. However, this method would be sufficient when planning to collect EMA data during the initial days of a longer sensing study and to build a study-specific prediction model to replace EMA instances for the rest of the study. Such a model could also be useful if participants miss individual EMAs during the study or for planned missing data designs. The predictive performance of this latter approach could be further improved by using more advanced machine-learning models explicitly designed to consider the nested structure of longitudinal data (i.e., repeated measures of participants). Since our goal above was to predict new observations from new participants, only fixed effects mattered in our example, and new observations would have received the random effects' expected value of zero. However, when aiming to predict new observations for participants included during model development—meaning when the model is not intended to generalize to new participants—adding random effects to handle correlated data could potentially enhance predictive performance. For this purpose, the advanced implementations of random forest models specifically for longitudinal data described in Hu and Szymczak (2023) could be used. Alternatively, researchers might apply the general mixed effects approach to machine learning explained in Kilian et al. (2023), which combines random effects modeling with various machine-learning models in a model-agnostic way. However, it should be noted that mixed effects models can be computationally expensive, and researchers should carefully consider whether this type of model is appropriate for their preprocessing needs based on their specific research goals and the availability of time and computational resources.

Even if researchers take these considerations seriously and achieve stronger predictive performance, several challenges remain. First, they must decide when their predictions are accurate enough to replace self-reported data in future formal analyses. This threshold likely depends on the specific research question and its practical implications, and it may not align with the established standards for evaluating prediction performance. Additionally, researchers need to define the conditions under which their predictions remain useful, as relationships in the data may change between study contexts or over time (e.g., due to seasonal effects; Wiernik et al., 2020). Second, researchers should recognize that the predictions inherit the reliability issues associated with the self-reports used to train the model (Wiernik et al., 2020). For example, the ground truth in our example, self-reported sleep duration, has been shown to be biased by systematic over-reporting (Lauderdale et al., 2008), so the same will likely apply to our predictions. Third, when replacing self-reports with machine-learning predictions, the question of whether these predictions represent psychological measurements arises (Stachl, Pargent, et al., 2020).

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

One initial obstacle in this context may be that the discriminant validity of sensor-based predictions has often been found to be problematic in the past (Wiernik et al., 2020). Finally, researchers should be aware of circularity issues when using sensing-based predictions (e.g., sleep duration predicted from app usage) for subsequent analyses if those also include (parts of) the same data (e.g., social media app usage in the evening) used to generate predictions. Due to these open questions, the preprocessing pipeline from Case 3 requires more conceptual development before it becomes a feasible approach in research practice.

5. Discussion

This manuscript presents three exemplary use cases for preprocessing mobile-sensing data that go beyond the current state of the art and expand on data enrichment and aggregation to develop more advanced solutions for variable extraction. Importantly, while some analyses in these use cases may serve as interesting research questions on their own, they primarily aim to generate new variables for formal modeling in psychological research here. Each case provides a detailed discussion of methodological decisions, opportunities, challenges, and potential extensions beyond the specific example. Therefore, we conclude with some overall considerations for choosing preprocessing approaches and offer an outlook on future directions for preprocessing mobile-sensing data.

5.1. Practical considerations

As illustrated across the case-specific outlooks above, researchers must consider several factors when selecting an appropriate preprocessing pipeline for their sensing data.

First, all of our preprocessing cases focus on mobile-sensing data collected from smartphones, specifically through Android-based logging. Hence, the data structure and preprocessing steps described throughout this manuscript are tailored to the exemplary dataset used here. However, the methodological strategies and considerations discussed above and below also apply to sensing data from other platforms, including iOS devices, smartwatches, fitness trackers, and, to some extent, digital footprints from online platforms (e.g., Facebook likes and Tweets; Stier et al., 2020). Of course, researchers will need to adapt specific preprocessing steps based on the technical characteristics and constraints of their specific sensing platform or data source.

Second, our three cases illustrate several *trade-offs* associated with different levels of complexity in data enrichment and aggregation. These include balancing factors such as sparsity, interpretability, and information density of the resulting variables on one side, against the required methodological skills, time, and computational costs for implementation on the other. For example, the straightforward state-of-the-art preprocessing to extract variables from single apps is easy to implement, yet it produces sparse variables that lack clear psychological meaning. In contrast, external enrichment through app categories involves additional considerations but results in variables that are less sparse and more interpretable. Going further, Case 1 produces contextualized variables by combining sensing modalities but requires extra steps to preprocess different modalities, which can, in turn, increase sparsity again. Case 2 demands significantly higher computational effort and advanced methodological expertise to model relationships within the data and extract the model parameters as new variables, providing a high level of detailed, temporally specific information. Lastly, Case 3 could even substitute data points entirely, offering practical utility in replacing burdensome EMAs, but it remains vulnerable to failure despite considerable methodological and computational efforts. Comparing all approaches highlights the trade-offs inherent in selecting more or less complex preprocessing strategies. The optimal level of complexity will mainly depend on the specific research question and study design.

Third, besides these trade-offs, increased complexity in preprocessing always comes with more *researcher degrees of freedom* as more preprocessing steps are added. However, the number of preprocessing decisions can also vary greatly within each case, depending on factors like how app categories

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

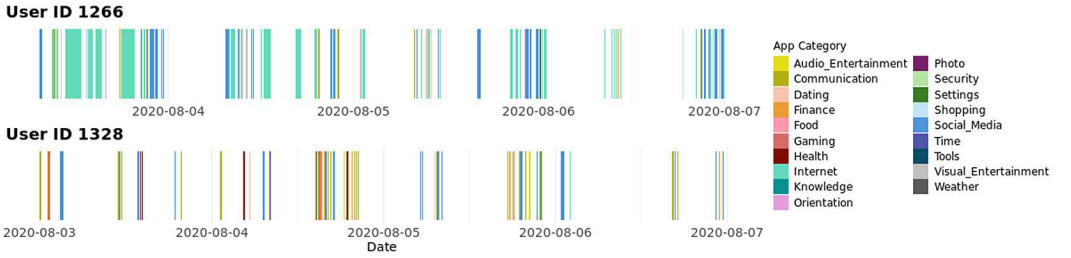

Figure 3. App usage patterns by category over several study days for two randomly selected participants.

are created (e.g., manually versus default categories, see Section 3.2) or which model is chosen for integrating data (see Case 2). Previous research has shown that variations in preprocessing pipeline design can significantly affect study results, revealing researcher degrees of freedom as a threat to reproducibility in mobile-sensing research (Langener, Stulp, et al., 2024; Schoedel et al., 2020). To address this and avoid questionable research practices, preprocessing decisions should ideally be preregistered in as much detail as possible (Langener, Siepe, et al., 2024). However, due to the complexity of preprocessing mobile-sensing data, not all decisions can be predicted, especially when unexpected problems, such as missing or mislogged data, occur despite careful piloting. This underscores the need for transparent and thorough reporting of preprocessing pipelines to ensure reproducibility (Wrzus & Schoedel, 2023). In this context, mobile-sensing researchers should begin to incorporate existing data quality frameworks (e.g., ISO 8000 and FAIR) as structured guides for transparent reporting (for an overview, see Miller et al., 2025).

5.2. General outlook

Beyond the case-specific extensions discussed above, two important characteristics of mobile-sensing data remain underexplored in our cases but may offer valuable directions for future preprocessing approaches.

One critical characteristic of mobile-sensing data that our preprocessing cases only superficially addressed is their *temporal resolution*. In Figure 3, we illustrate category-wise app usage sessions, with temporal resolution measured in seconds across several study days for two randomly selected participants. Our preprocessing examples approached these temporal dynamics by aggregating sensing data over different time frames. In the state-of-the-art preprocessing and Case 1, we first generated daily measures and then combined them across study days to produce person-level variables, which may be used in subsequent formal analyses at the inter-individual level. Such an approach was, for example, employed to explore the relationship between smartphone usage and personality traits (Stachl, Au, et al., 2020). In contrast, the same variables can be obtained as state-level observations (e.g., daily or hourly) to explore intra-individual questions (Rüegger et al., 2020). Additionally, sensing data can be aggregated over more specific time frames related to certain events. Conversely, in Case 2, we extracted variables within 60 minutes surrounding EMA instances, and in Case 3, we concentrated on the 3 hours before sleep. Overall, in all our cases, single events (such as app usage sessions) were aggregated—albeit over different time frames—simplifying the complex temporal structure of the mobile-sensing data. To better exploit the temporal resolution shown in Figure 3, researchers could consider preprocessing methods that enable a more detailed mapping of individual behavioral trajectories, such time-varying autoregressive models (e.g., Bringmann et al., 2017), non-linear latent growth curve models for individually spaced time intervals (e.g., Sterba, 2014), or continuous-time models (e.g., de Haan-Rietdijk et al., 2017; Driver et al., 2017).

Another important extension involves the *high dimensionality* of mobile-sensing data. While raw sensing data are inherently high-dimensional (i.e., comprising large volumes of event-level data), this characteristic was not fully addressed in our preprocessing cases. That is, because our approach to

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

variable extraction was guided by theory, prioritizing interpretable variables that align with content-related psychological research questions—for example, investigating whether the social reactivity to app usage from Case 2 is linked to loneliness. This theory-driven strategy limited our use of fully bottom-up, data-driven preprocessing techniques, which often generate less interpretable features. Nonetheless, we briefly touched on more data-driven methods in Case 1, where we used unsupervised clustering of GPS data as a form of dimensionality reduction. In principle, the structure of mobile-sensing data allows for more data-driven approaches—for example, applying reduction techniques such as principal component analysis or clustering to identify (latent) patterns in app usage behavior. While these methods treat the data as static multivariate vectors, functional data analysis—recently reviewed in the context of wearable sensor data (Acar-Denizli & Delicado, 2024)—offers a complementary perspective by modeling data as continuous functions over time, thus capturing temporal dynamics more naturally. Following this idea, functional principal component analysis (e.g., Shang, 2014) can perform dimensionality reduction on time-dependent processes and therefore could potentially be used with (intensive) sensing data. Addressing the sequentiality of (mobile) sensing data, Peters et al. (2024) recently applied neural network architectures to capture patterns in raw app usage events and estimated intra-individual predictability of social media app use, which could also be adapted for variable extraction. In our three use cases, we intentionally adopted a theory-driven preprocessing strategy, which we believe is well-suited for addressing traditional psychological research questions. However, more data-driven approaches may become increasingly important when the goal is to maximize predictive accuracy—for example, in clinical applications such as depression detection (Squires et al., 2023) or suicide prediction (Linthicum et al., 2019). Ultimately, the choice of preprocessing strategy depends on whether the resulting variables should reflect theoretically-derived behaviors and whether the format and dimensionality of the resulting variables are suitable for subsequent formal modeling (e.g., inferential vs. predictive modeling).

To conclude, this manuscript first summarized the current state of data preprocessing in mobile-sensing research and then presented three use cases that go beyond common practices. These preprocessing cases demonstrate the rich potential of mobile-sensing data for extracting nuanced behavioral variables and aim to inspire the development of more sophisticated, theory-driven research questions. While the presented preprocessing pipelines represent only a small subset of the many possible approaches, they highlight the variation in complexity along the dimensions of data enrichment and aggregation. As methodological standards for mobile sensing in psychological research are still emerging (Schoedel & Mehl, 2024), future work will need to address unresolved conceptual and analytical challenges and may develop even more advanced preprocessing strategies. Our use cases offer a starting point for researchers seeking guidance on selecting an appropriate level of complexity for their own preprocessing pipelines.

Data availability statement. We cannot share the raw smartphone sensing data to preserve individuals' privacy under the European General Data Protection Regulation. However, we provide the datasets with the aggregated variables that were generated during preprocessing for this article and the corresponding preprocessing R code. All the analyses reported in this article are based on these datasets. The aggregated data and all materials are available in the project's OSF-repository: https://osf.io/tmuhe.

Acknowledgements. We thank the PhoneStudy team for their diligent work on the PhoneStudy App and the panel study. For data collection in the context of the Smartphone Sensing Panel Study, our special thanks go to our cooperation partners, the Leibniz Institute of Psychology (ZPID).

Author contributions. R.S. and L.S. shared first authorship and contributed equally to this work.

Funding statement. The Smartphone Sensing Panel Study, which produced the dataset used in this article, is a joint project of LMU Munich and the Leibniz Institute for Psychology (ZPID). ZPID provided most of the funding for the implementation of the described panel study. This work was supported by the German Research Foundation under Grant No. 516600480.

Competing interests. The authors declare none.

Statement on the use of AI tools. During the preparation of this manuscript, the authors used ChatGPT 4o to improve the readability of the text and to correct language errors (January 2025). After using this tool, the authors reviewed and edited the content of the publication for which they take full responsibility.

References

44matters. (2025). Google play vs the apple app store: App stats and trends. https://42matters.com/stats

Acar-Denizli, N., & Delicado, P. (2024). *Functional data analysis on wearable sensor data: A systematic review*. Preprint, arXiv:2410.11562.

Anderson, I., Gil, S., Gibson, C., Wolf, S., Shapiro, W., Semerci, O., & Greenberg, D. M. (2021). Just the way you are: Linking music listening on Spotify and personality. *Social Psychological and Personality Science*, *12*(4), 561–572. https://doi.org/10.1177/1948550620923228

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beck, E. D., & Jackson, J. J. (2022). Personalized prediction of behaviors and experiences: An idiographic person–situation test. *Psychological Science*, *33*(10), 1767–1782. https://doi.org/10.1177/09567976221093307

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.7551/mitpress/1120.003.0082

Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). Falling asleep with angry birds, Facebook and kindle: A large scale study on mobile application usage. In *Proceedings of the 13th international conference on human computer interaction with mobile devices and services*, ACM, New York, 47–56. https://doi.org/10.1145/2037373.2037383

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*(3), 409–425. https://doi.org/10.1037/met0000085

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The consensus sleep diary: Standardizing prospective sleep self-monitoring. *Sleep*, *35*(2), 287–302. https://doi.org/10.5665/sleep.1642

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM. https://doi.org/10.1145/2939672.2939785

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life in N. P. R. In S. S. Kosslyn (Ed.), *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource* (pp. 1–15). Wiley Hoboken. https://doi.org/10.1002/9781118900772.etrds0010

de Haan-Rietdijk, S., Voelkle, M. C., Keijsers, L., & Hamaker, E. L. (2017). Discrete-vs. continuous-time modeling of unequally spaced experience sampling method data. *Frontiers in Psychology*, *8*, 1849. https://doi.org/10.3389/fpsyg.2017.01849

Dey, A. K., Wac, K., Ferreira, D., Tassini, K., Hong, J.-H., & Ramos, J. (2011). Getting closer: An empirical investigation of the proximity of user to their smart phones. In *Proceedings of the 13th international conference on ubiquitous computing*, ACM, New York, 163–172. https://doi.org/10.1145/2030112.2030135

Do, T. M. T., Blom, J., & Gatica-Perez, D. (2011). Smartphone usage in the wild: A large-scale analysis of applications and context. In *Proceedings of the 13th international conference on multimodal interfaces*. ACM, New York, 353–360. https://doi.org/10.1145/2070481.2070550

Driver, C. C., Oud, J. H., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*, 1–35. https://doi.org/10.18637/jss.v077.i05

Elmer, T., Fernández, A., Stadel, M., Kas, M., & Langener, A. (2025). Bidirectional associations between smartphone usage and momentary well-being in young adults: Tackling methodological challenges by combining experience sampling methods with passive smartphone data. *Emotion*, *25*(5), 1065–1078. https://doi.org/10.1037/emo0001485

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X., (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, *96*(34), 226–231.

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Logan Slade, L., & Taylor, N. G. A. (2020). Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: Systematic review. *JMIR mHealth and uHealth*, *8*(9), e18694.

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality*, *43*(2), 120–126. https://doi.org/10.1016/j.jrp.2008.12.041

Gordon, M. L., Gatys, L., Guestrin, C.,Bigham, J. P., Trister, A., & Patel, K. (2019). App usage predicts cognitive ability in older adults. In *Proceedings of the 2019 chi conference on human factors in computing systems*. ACM, New York, 1–12. https://doi.org/10.1145/3290605.3300398

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

Goretzko, D., & Ruscio, J. (2024). The comparison data forest: A new comparison data approach to determine the number of factors in exploratory factor analysis. *Behavior Research Methods*, *56*(3), 1838–1851. https://doi.org/10.3758/s13428-023-02122-4

große Deters, F., Reiter, T., & Schoedel, R. (under review). From swipe to sleep: A daily diary study using smartphone sensing to examine smartphone usage before bedtime and sleep outcomes [under review].

große Deters, F., & Schoedel, R. (2024). Keep on scrolling? Using intensive longitudinal smartphone sensing data to assess how everyday smartphone usage behaviors are related to well-being. *Computers in Human Behavior*, *150*, 107977. https://doi.org/10.1016/j.chb.2023.107977

GSMA. (2025). The mobile economy 2025. (Tech. rep.) [PDF report]. GSMA. PDF report. https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-economy/wp-content/uploads/2025/04/030325-The-Mobile-Economy-2025.pdf

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). Dbscan: Fast density-based clustering with R. *Journal of Statistical Software*, *91*, 1–30. https://doi.org/10.32614/cran.package.dbscan

Harari, G. M., & Gosling, S. D. (2023). Understanding behaviours in context using mobile sensing. *Nature Reviews Psychology*, *2*(12), 767–779. https://doi.org/10.1038/s44159-023-00235-3

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., Rentfrow, P. J., Campbell, A. T., & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality and Social Psychology*, *119*(1), 204. https://doi.org/10.1037/pspp0000245

Harari, G. M., Vaid, S. S., Müller, S. R., Stachl, C., Marrero, Z., Schoedel, R., Bühner, M., & Gosling, S. D. (2020). Personality sensing for theory development and assessment in the digital age. *European Journal of Personality*, *34*(5), 649–669. https://doi.org/10.1002/per.2273

Hijmans, R. J. (2024). Geosphere: Spherical trigonometry. [R package version 1.5-20]. https://CRAN.R-project.org/package=geosphere

Horstmann, K. T., & Ziegler, M. (2019). Situational perception and affect: Barking up the wrong tree? *Personality and Individual Differences*, *136*, 132–139. https://doi.org/10.1016/j.paid.2018.01.020

Hu, J., & Szymczak, S. (2023). A review on longitudinal data analysis with random forest. *Briefings in Bioinformatics*, *24*(2), bbad002.

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA*, *318*(13), 1215–1216. https://doi.org/10.1001/jama.2017.11295

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science*, *15*(3), 809–816. https://doi.org/10.1177/1745691620902467

Kilian, P., Ye, S., & Kelava, A. (2023). Mixed effects in machine learning—A flexible mixedML framework to add random effects to supervised machine learning regression. *Transactions on Machine Learning Research*.

Koch, T., Voelkle, M. C., & Driver, C. C. (2023). Analyzing longitudinal multirater data with individually varying time intervals. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(1), 86–104. https://doi.org/10.1080/10705511.2022.2096612

Kritzler, S., Krasko, J., & Luhmann, M. (2020). Inside the happy personality: Personality states, situation experience, and state affect mediate the relation between personality and affect. *Journal of Research in Personality*, *85*, 103929. https://doi.org/10.1016/j.jrp.2020.103929

Langener, A. M., Siepe, B. S., Elsherif, M., Niemeijer, K., Andresen, P. K., Akre, S., Bringmann, L. F., Cohen, Z. D., Choukas, N. R., Drexl, K., Fassi, L., Green, J., Hoffmann, T., Jagesar, R. R., Kas, M. J. H., Kurten, S., Schoedel, R., Stulp, G., Turner, G. & Jacobson, N. C. (2024). A template and tutorial for preregistering studies using passive smartphone measures. *Behavior Research Methods*, *56*, 8289–8307. https://doi.org/10.3758/s13428-024-02474-5

Langener, A. M., Stulp, G., Jacobson, N. C., Costanzo, A., Jagesar, R. R., Kas, M. J., & Bringmann, L. F. (2024). It's all about timing: Exploring different temporal resolutions for analyzing digital-phenotyping data. *Advances in Methods and Practices in Psychological Science*, *7*(1), 25152459231202677. https://doi.org/10.1177/25152459231202677

Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K., & Rathouz, P. J. (2008). Self-reported and measured sleep duration: How similar are they? *Epidemiology*, *19*(6), 838–845. https://doi.org/10.1097/ede.0b013e318187a7b0

Linthicum, K. P., Schafer, K. M., & Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral Sciences & the Law*, *37*(3), 214–222. https://doi.org/10.1002/bsl.2392

Markowetz, A., Błaszkiewicz, K., Montag, C., Switala, C., & Schlaepfer, T. E. (2014). Psycho-informatics: Big data shaping modern psychometrics. *Medical Hypotheses*, *82*(4), 405–411. https://doi.org/10.1016/j.mehy.2013.11.030

Mehl, M. R., Eid, M., Wrzus, C., Harari, G. M., Ebner-Priemer, U. W., & Insel, T. R. (2024). *Mobile sensing in psychology: Methods and applications*. Guilford Press.

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Miller, R., Chan, S. H. M., Whelan, H., & Gregório, J. (2025). A comparison of data quality frameworks: A review. *Big Data and Cognitive Computing*, *9*(4), 93. https://doi.org/10.3390/bdcc9040093

Moshe, I., Terhorst, Y., Opoku Asare, K., Sander, L. B., Ferreira, D., Baumeister, H., Mohr, D. C., & Pulkki-Råback, L. (2021). Predicting symptoms of depression and anxiety using smartphone and wearable data. *Frontiers in Psychiatry*, *12*, 625247. https://doi.org/10.3389/fpsyt.2021.625247

https://doi.org/10.1017/psy.2026.10083 Published online by Cambridge University Press

Müller, S. R., Bayer, J. B., Ross, M. Q., Mount, J., Stachl, C., Harari, G. M., Chang, Y.-J., & Le, H. T. (2022). Analyzing GPS data for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *5*(2), 1–11. https://doi.org/10.1177/25152459221082680

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the relationships between mobility behaviours and indicators of subjective well–being using smartphone–based experience sampling and GPS tracking. *European Journal of Personality*, *34*(5), 714–732. https://doi.org/10.1002/per.2262

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science*, *6*(3). https://doi.org/10.1177/25152459231162559

Parry, D., & Toth, R. (2025). Extracting meaningful measures of smartphone usage from android event log data: A methodological primer. *Computational Communication Research*, *7*(1), 1–32. https://doi.org/10.5117/CCR2025.1.8.PARR

Peters, H., Bayer, J. B., Matz, S. C., Chi, Y., Vaid, S. S., & Harari, G. M. (2024). Social media use is predictable from app sequences: Using LSTM and transformer neural networks to model habitual behavior. *Computers in Human Behavior*, *161*, 108381. https://doi.org/10.1016/j.chb.2024.108381

Pillion, M., Gradisar, M., Bartel, K., Whittall, H., & Kahn, M. (2022). What's "app"-ning to adolescent sleep? Links between device, app use, and sleep outcomes. *Sleep Medicine*, *100*, 174–182. https://doi.org/10.1016/j.sleep.2022.08.004

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in s and s-plus*. Springer. https://doi.org/10.1007/b98882

Probst, P., Boulesteix, A.-L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*, *20*(53), 1–32.

Probst, P., Wright, M. N., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/widm.1301

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rauthmann, J. F., & Sherman, R. A. (2015). Ultra-brief measures for the situational eight diamonds domains. *European Journal of Psychological Assessment*, *32*(2), 165–174. https://doi.org/10.1027/1015-5759/a000245

Reiter, T., & Schoedel, R. (2024). Never miss a beep: Using mobile sensing to investigate (non-) compliance in experience sampling studies. *Behavior Research Methods*, *56*(4), 4038–4060. https://doi.org/10.3758/s13428-023-02252-9

Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How are personality states associated with smartphone data? *European Journal of Personality*, *34*(5), 687–713. https://doi.org/10.1002/per.2309

Saeb, S., Zhang, M., Kwasny, M., Karr, C. J., Kording, K., & Mohr, D. C. (2015). The relationship between clinical, momentary, and sensor-based assessment of depression. In *2015 9th international conference on pervasive computing technologies for healthcare (pervasivehealth)*. IEEE. https://doi.org/10.4108/icst.pervasivehealth.2015.259034

Schoedel, R., & Mehl, M. (2024). Mobile sensing methods. In H. T. Reis, T. West, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (3rd ed., pp. 297–321). Cambridge University Press. https://doi.org/10.1017/9781009170123.014

Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., Bischl, B., Hussmann, H., & Stachl, C. (2018). Digital footprints of sensation seeking. *Zeitschrift für Psychologie*, (4), 232–245. https://doi.org/10.1027/2151-2604/a000342

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (2023). Snapshots of daily life: Situations investigated through the lens of smartphone sensing. *Journal of Personality and Social Psychology*, *125*(6), 1442–1471. https://doi.org/10.1037/pspp0000469

Schoedel, R., & Oldemeier, M. (2020). Basic protocol: Smartphone sensing panel study. Leibniz Institut für Psychologische Information und Dokumentation (ZPID). https://doi.org/10.23668/psycharchives.15508

Schoedel, R., Oldemeier, M.,Bonauer, L., & Sust, L. (2022). Systematic categorisation of 3, 091 smartphone applications from a large-scale smartphone sensing dataset. *Journal of Open Psychology Data*, *10*(7), 1–10. https://doi.org/10.5334/jopd.59

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day–night behaviour patterns. *European Journal of Personality*, *34*(5), 733–752. https://doi.org/10.1002/per.2258

Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, *98*(2), 121–142. https://doi.org/10.1007/S10182-013-0213-1

Simonsohn, U., & Gruson, H. (2024). Groundhog: Version-control for CRAN, Github, and Gitlab packages. [R package version 3.2.0]. https://CRAN.R-project.org/package=groundhog

Spathis, D., Servia-Rodriguez, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019). Passive mobile sensing and psychological traits for large scale mood prediction. In *Proceedings of the 13th EAI international conference on pervasive computing technologies for healthcare*. https://doi.org/10.1145/3329189.3329213

Squires, M., Tao, X., Elangovan, S., Gururajan, R., Zhou, X., Acharya, U. R., & Li, Y. (2023). Deep learning and machine learning in psychiatry: A survey of current progress in depression detection, diagnosis and treatment. *Brain Informatics*, *10*(1), 10. https://doi.org/10.1186/s40708-023-00188-6

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences*, *117*(30), 17680–17687. https://doi.org/10.1073/pnas.1920484117

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality*, *34*(5), 613–631. https://doi.org/10.1002/per.2257

Stasinopoulos, M. D., Rigby, R. A., & Bastiani, F. D. (2018). GAMLSS: A distributional regression approach. *Statistical Modelling*, *18*(3–4), 248–273. https://doi.org/10.1177/1471082x18759144

Sterba, S. K. (2014). Fitting nonlinear latent growth curve models with individually varying time points. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 630–647. https://doi.org/10.1080/10705511.2014.919828

Sterner, P., Goretzko, D., & Pargent, F. (2023). Everything has its price: Foundations of cost-sensitive machine learning and its application in psychology. *Psychological Methods*. https://doi.org/10.1037/met0000586

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). *Integrating survey data and digital trace data: Key issues in developing an emerging field*. Sage Publications. https://doi.org/10.1177/0894439319843669

Sust, L., Stachl, C., Kudchadker, G., Bühner, M., & Schoedel, R. (2023). Personality computing with naturalistic music listening behavior: Comparing audio and lyrics preferences. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.75214

Sust, L., Talaifar, S., & Stachl, C. (2023). Mobile application usage in psychological research. In M. R. Mehl, M. Eid, C. Wrzus, G. M. Harari, & U. W. Ebner-Priemer (Eds.), *Mobile sensing in psychology: Methods and applications* (pp. 184–214). Guilford Press.

Valkenburg, P. M. (2022). Social media use and well-being: What we know and what we need to know. *Current Opinion in Psychology*, *45*, 101294. https://doi.org/10.1016/j.copsyc.2021.12.006

van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2019). Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports. *International Journal of Human-Computer Studies*, *125*, 118–128. https://doi.org/10.1016/j.ijhcs.2018.12.002

Ver Hoef, J. M., London, J. M., & Boveng, P. L. (2010). Fast computing of some generalized linear mixed pseudo-models with temporal autocorrelation. *Computational Statistics*, *25*, 39–55. https://doi.org/10.1007/s00180-009-0160-1

Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: Relating authoritarianism and anomia. *Psychological Methods*, *17*(2), 176. https://doi.org/10.1037/a0027543

Wac, K. (2018). From quantified self to quality of life. In Rivas, H., & Wac, K. (Eds), *Digital health*. Healthcare informatics-(pp. 83–108). Springer.

Wiernik, B. M., Ones, D. S., Marlin, B. M., Giordano, C., Dilchert, S., Mercado, B. K., Stanek, K. C., Birkland, A., Wang, Y., Ellis, B., Yazar, Y., Kostal, J. W., Kumar, S., Hnat, T., Ertin, E., Sano, A., Ganesan, D. K., Choudhoury, T., & Al'absi, M. (2020). Using mobile sensors to study personality dynamics. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000576

Wright, A. G., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing personality assessment on the person: Modeling general, shared, and person specific processes in personality and psychopathology. *Psychological Assessment*, *31*(4), 502–515. https://doi.org/10.1037/pas0000617

Wrzus, C., & Neubauer, A. B. (2023). Ecological momentary assessment: A meta-analysis on designs, samples, and compliance across research fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Wrzus, C., & Schoedel, R. (2023). Transparency and reproducibility in mobile sensing research. In M. R. Mehl, M. Eid, C. Wrzus, G. M. Harari, & U. W. Ebner-Priemer (Eds.), *Mobile sensing in psychology: Methods and applications* (pp. 53–81). Guilford Press.

Zhang, R., Peng, H., Dou, Y., Wu, J., Sun, Q., Li, Y., Zhang, J., & Yu, P. S. (2022). Automating DBSCAN via deep reinforcement learning. In *Proceedings of the 31st ACM international conference on information & knowledge management*, ACM, New York, 2620–2630. https://doi.org/10.1145/3511808.3557245

Zhang, X., Wang, C., Li, Z., Zhu, J., Shi, W., & Wang, Q. (2016). Exploring the sequential usage patterns of mobile internet services based on Markov models. *Electronic Commerce Research and Applications*, *17*, 1–11. https://doi.org/10.1016/j.elerap.2016.02.002

Cite this article: Schoedel, R., Sust, L., Sterner, P. and Goretzko, D. (2026). From Digital Data to Psychological Insights: Making Sense of Mobile-Sensing Data through Integrative Preprocessing Pipelines. *Psychometrika*, 761–788. https://doi.org/10.1017/psy.2026.10083