

Original Article

*Both authors contributed equally.

Cite this article: De Calheiros Velozo J *et al* (2024). Designing daily-life research combining experience sampling method with parallel data. *Psychological Medicine* **54**, 98–107. <https://doi.org/10.1017/S0033291722002367>

Received: 5 July 2021

Revised: 26 June 2022

Accepted: 11 July 2022

First published online: 30 August 2022

Key words:

Daily life; ecological momentary assessments; experience sampling; passive monitoring; passive sensing

Author for correspondence:

Jeroen Habets,

E-mail: j.habets@maastrichtuniversity.nl

Designing daily-life research combining experience sampling method with parallel data

Joana De Calheiros Velozo^{1,*}, Jeroen Habets^{2,*} , Sandip V. George³, Koen Niemeijer⁴, Olga Minaeva³, Noëmi Hagemann¹, Christian Herff², Peter Kuppens⁴, Aki Rintala^{1,5}, Thomas Vaessen^{1,6}, Harriëtte Riese³ and Philippe Delespaul⁷

¹Department of Neurosciences, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium; ²Department of Neurosurgery, School of Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands;

³Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; ⁴Department of Psychology and Educational Sciences, Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium; ⁵Faculty of Social and Health Care, LAB University of Applied Sciences, Lahti, Finland; ⁶Department of Neurosciences, Mind Body Research, KU Leuven, Leuven, Belgium and ⁷Department of Psychiatry and Neuropsychology, School of Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

Abstract

Background. Ambulatory monitoring is gaining popularity in mental and somatic health care to capture an individual's wellbeing or treatment course in daily-life. Experience sampling method collects subjective time-series data of patients' experiences, behavior, and context. At the same time, digital devices allow for less intrusive collection of more objective time-series data with higher sampling frequencies and for prolonged sampling periods. We refer to these data as parallel data. Combining these two data types holds the promise to revolutionize health care. However, existing ambulatory monitoring guidelines are too specific to each data type, and lack overall directions on how to effectively combine them.

Methods. Literature and expert opinions were integrated to formulate relevant guiding principles.

Results. Experience sampling and parallel data must be approached as one holistic time series right from the start, at the study design stage. The fluctuation pattern and volatility of the different variables of interest must be well understood to ensure that these data are compatible. Data have to be collected and operationalized in a manner that the minimal common denominator is able to answer the research question with regard to temporal and disease severity resolution. Furthermore, recommendations are provided for device selection, data management, and analysis. Open science practices are also highlighted throughout. Finally, we provide a practical checklist with the delineated considerations and an open-source example demonstrating how to apply it.

Conclusions. The provided considerations aim to structure and support researchers as they undertake the new challenges presented by this exciting multidisciplinary research field.

Background

The experience sampling method (ESM) is a powerful diary-based tool to assess subjective daily-life data (Christensen, Barrett, Bliss-Moreau, Lebo, & Christensen, 2003; Eisele *et al.*, 2020; Myin-Germeys *et al.*, 2018; Palmier-Claus *et al.*, 2011). Typically, users complete an identical (or quasi-identical) questionnaire repeatedly throughout the day over the course of several days, weeks, or months. These questionnaires are often scheduled using semi-randomized cues and responding is time limited to avoid biased- or back-filling. ESM has been extensively used in the health care sector to describe individuals' wellbeing and symptom course (Myin-Germeys *et al.*, 2018) as well as to evaluate therapeutic effects in mental and physical health [Corrigan-Curay, Sacks, & Woodcock, 2018; FDA (Food and Drug Administration), 2022; Oyinlola, Campbell, & Kousoulis, 2016].

Technological advances in wearable devices and passive sensing tools are expected to revolutionize health care. These tools are increasingly used in combination with ESM to enhance daily-life monitoring and to explore a much wider and comprehensive array of questions (Rehg, Murphy, & Kumar, 2017). Throughout this paper we will refer to wearables and passive sensing data as 'parallel data'. Broadly, parallel data are any data that are collected in parallel to, and with the purpose of supplementing ESM. Examples of parallel data include but are not restricted to physiological (heart rate, blood pressure, movement), environmental (geolocation), or behavioral (smartphone usage) data. These data have been combined with ESM to study addiction (Bertz, Epstein, & Preston, 2018), affective disorders (Cousins *et al.*, 2011,

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



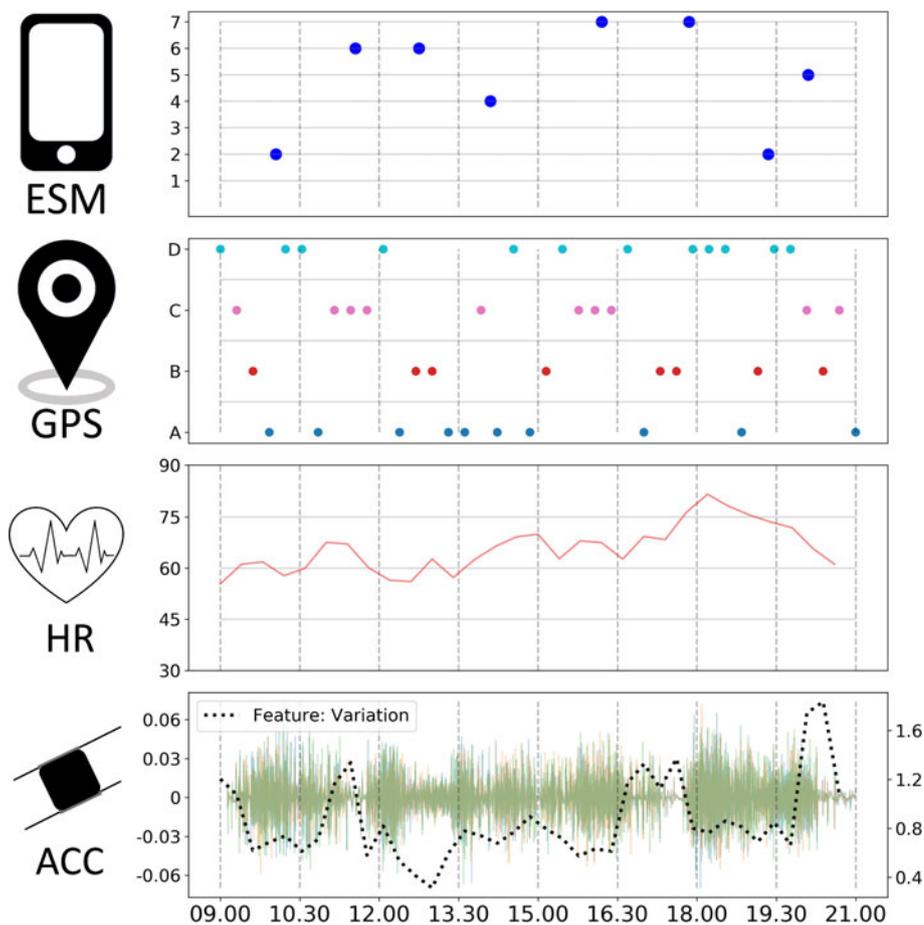


Fig. 1. Visualization of ESM and parallel data collected over 1 day. The ESM panel represents one ESM variable answered on a 7-point Likert scale. ESM assessments are collected semi-randomly in eight 90-min blocks, each containing one random beep. Geolocation (GPS) data-points are collected every 15 min and are classified, for demonstration purpose, in four different categories (A to D). Heart rate (HR) can be unobtrusively recorded by wrist-worn devices over periods of circa 20 min (Graham et al., 2019). For the accelerometer (ACC) signal, the raw tri-axial signal is shown. The summarizing feature is the variation (SciPy.Stats.Variation – SciPy v1.6.2 Reference Guide, n.d.) of the resulting signal vector magnitude (black dotted line, right y-axis).

p. 11; Kim et al., 2019; Minaeva et al., 2020), schizophrenia (Kimhy et al., 2017), and movement disorders (Heijmans et al., 2019a), among others.

Experience sampling and parallel data can be complementary to each other, which explains the growing interest in combining them. ESM can capture the variability within the day of variables that are not, or hardly, measurable with parallel data, such as affect, perceptions, and contextual cues and events (Myin-Germeys et al., 2018). Parallel data, on the other hand, are better suited to capture processes that are hard to measure subjectively such as physiological parameters like heart rate and skin conductance (van Halem, van Roekel, Kroencke, Kuper, & Denissen, 2020), and behaviors that are notoriously difficult to report such as internet usage (Yuan et al., 2019). In addition, parallel data can be collected passively and non-intrusively allowing for higher sampling frequencies (Fig. 1) and longer sampling periods with lower burden on participants (Barnett, Torous, Reeder, Baker, & Onnela, 2020). By combining ESM and parallel data one can overcome each method's limitations and explore their full potential. The combined methodology is promising for both ambulatory research and clinical practice focusing on real-life or real-time symptom tracking (Nahum-Shani et al., 2016). Especially its role in just in time adaptive interventions (JITAs) is highly anticipated and expected to change the health care landscape. JITAs are personalized interventions that are provided directly in daily-life, at the right time, and adapted to the patient's needs (Nahum-Shani et al., 2016; Sharmin et al., 2015).

However, as is often the case, technological advances outpace their scientific evidence and the common lack of parallel

ecologically valid, contextual information complicates parallel data analysis, validation, and reproducibility (Shortliffe, 1993; Stuppel, Singerman, & Celi, 2019; Tackett, Brandes, King, & Markon, 2019). Currently available ESM and parallel data monitoring guidelines focus too often on their respective data sources or their specific use case, and lack dedicated general directions to guide researchers in designing studies that combine them (Baumeister & Montag, 2019; Janssens, Bos, Rosmalen, Wichers, & Riese, 2018; Mehl & Conner, 2013; Myin-Germeys et al., 2018; Palmier-Claus et al., 2011; Rehg et al., 2017). Reproducibility is further threatened by the lack of standardization and the large heterogeneity in measures, methods, and approaches used to combine these two data types (Vaessen et al., 2021).

Consequently, as part of the Belgian-Dutch Network for ESM Research in Mental Health, an expert group focused on combining ESM and parallel data came together to formulate clear points to consider in the various stages of designing such a study. Rather than a rigid guideline, these are general considerations aimed at providing researchers with the necessary structure and support to design and conduct meaningful and reproducible research combining ESM and parallel data.

How to design a study that combines ESM with parallel data?

Research question and hypotheses

The ESM and parallel data are typically combined to (1) enrich subjective self-assessments of ESM with an 'objective' proxy or

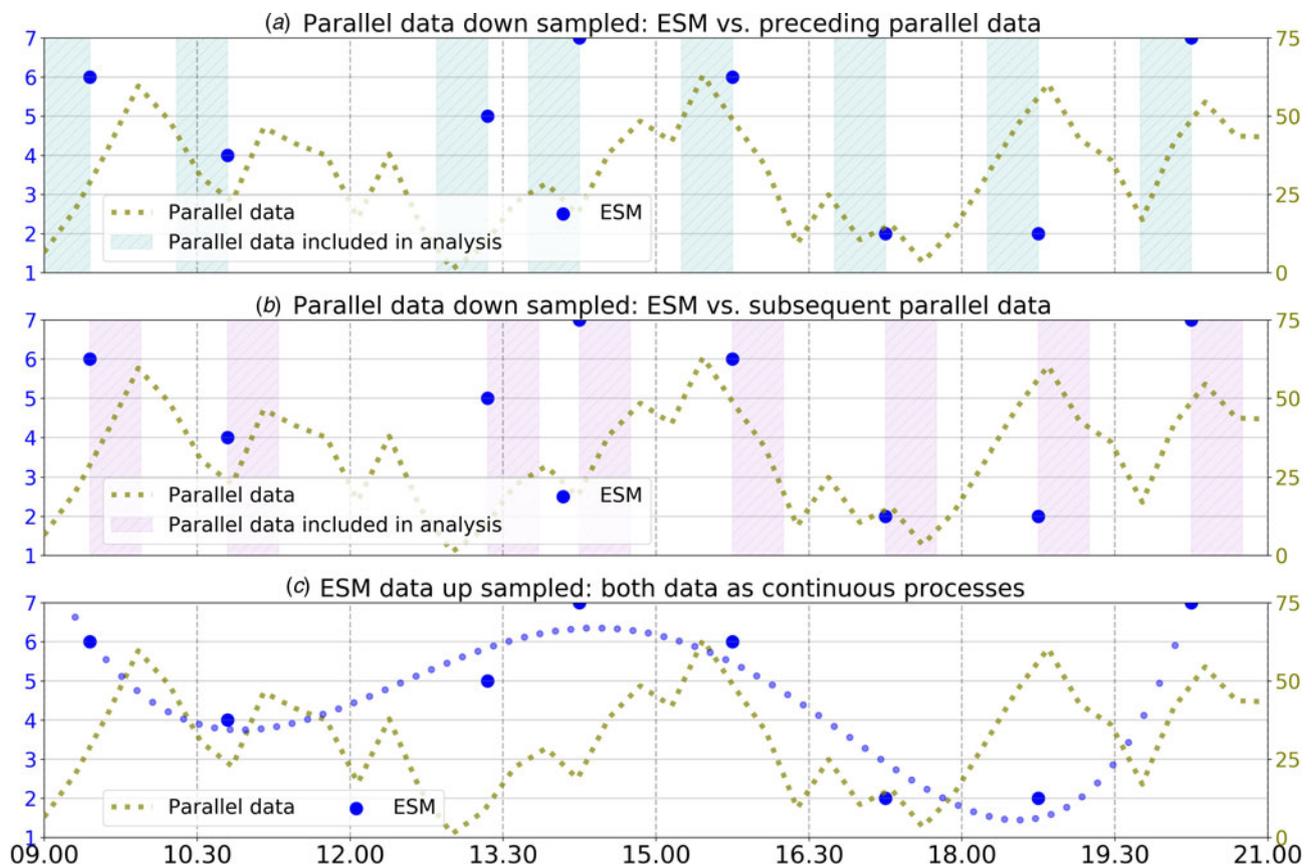


Fig. 2. Visualization of three different pre-processing approaches to combine ESM and parallel data. (a and b) Two examples where parallel data are down-sampled for analysis (i.e. a summary statistic will be computed over a specific period), which leads to eight data-points per day for analysis. The parallel data used can be taken from the window preceding (i.e. a) or proceeding (i.e. b) each ESM observation (blue dot), dependent on the research question and hypothesis. (c) An example where ESM data are up-sampled (i.e. in-between observations are generated), which leads to two continuous time series for analysis. Panel (c) uses a least squares polynomial fit (Numpy.Polyfit – NumPy v1.20 Manual, 2022) for demonstration purposes. (a) Parallel data down-sampled: ESM v. preceding parallel data; (b) parallel data down-sampled: ESM v. subsequent parallel data; (c) ESM data up-sampled: both data as continuous processes.

complementary data source, or (2) provide interpretable contextual or ‘ground truth’ labels next to volatile high-frequency time series. During the definition of the research question and its hypotheses, the limitations of both data types have to be carefully considered. A graphical representation of the anticipated results of the two variables of interest can be useful to grasp the fluctuation patterns within our variables of interest (Fig. 1). During this hypothetical exploration of the combined dataset, the question ‘What information does ESM data add to parallel data (or vice versa)?’ has to be leading. An understanding of the data collection, operationalization, and analytical techniques is therefore required. Answering this question helps to consider whether the combined use of ESM and parallel data is justified, meaning whether each data type has its own unique contribution while also maintaining synergy.

Since the variables of interest will result from two different data types and cover different timelines, it is essential to understand the expected temporal relationship between those variables and to specify the assumptions about the direction of the association of interest. For instance, are we interested in the parallel data features such as for example GPS location, preceding, following, or simultaneous to the ESM assessment of for example momentary anxiety (Fig. 2)? Or are we interested in relationships over time, the effect of moment n on moment $n + 1$, and if so, how long should the time lag be? In addition, the duration of parallel data corresponding with the ESM measure has to be determined. All these considerations should lead to a schematic draft of the

temporal relationship between the variables of interest as seen in Fig. 2. Preregistration of the hypotheses and the study design will help solidify the expected associations and increase the robustness of the results (Nosek & Lakens, 2014).

It is important to note that these directional patterns can help ‘unpack’ mutually occurring temporal relationships but they cannot prove causal relations due to the observational nature of data collected in the flow of daily-life (Holleman, Hooge, Kemner, & Hessels, 2020). Causality depends on several co-occurring factors and therefore cannot be claimed in such study designs (Rohrer, 2018; Rubin, 2007). Researchers who are interested in a causal relationship in an ecologically valid context should consider an experimental design in which a daily-life variable, for example stress, is actively manipulated in order to test its effect on another variable (Smets, De Raedt, & Van Hoof, 2019). Although, no absolute claims about causality can be made, ecological data collection allows more control over a certain variable in a more natural environment. This benefit has to be carefully weighed against the threat that it poses to the ecological validity (Holleman *et al.*, 2020). Findings should thus be interpreted with this in mind.

Variable fluctuation and volatility

Once the variables of interest and the nature of their relationship have been identified, we need to understand their variability (i.e. how much do they fluctuate?) and volatility (i.e. how fast do they

fluctuate?). These factors will determine data collection technicalities such as sensitivity (what is the smallest detectable symptom difference?), frequency (e.g. continuous high frequent sampling, multiple times per day or week), timing (e.g. morning, evening, event-triggered), and duration (days, weeks, months). Since the aim is to combine two separate data time series, it is essential at this time to define the temporal resolution of the anticipated 'outcome variables' or 'fluctuation scores' of both data types. If one data type results in outcomes of a higher frequency, a valid and meaningful aggregation method has to be designed to enable matching information from both data types.

ESM variables may have different fluctuations and volatilities, for instance depressive feelings may fluctuate slowly compared to anxiety, which may be more volatile. While it is tempting to select a higher ESM sampling frequency, this may cause increased participant burden, lower compliance, and lower data quality (Eisele et al., 2020; Fuller-Tyszkiewicz et al., 2013; Trull & Ebner-Priemer, 2020). Akin to parallel data variables, sampled data-points should not miss relevant fluctuations, but large amounts of redundant data should also be avoided.

Specific temporal associations between variables can drive the timing of the data collection. An association that includes a time-lag will determine the interval-period between consecutive assessments. On the other hand, when the focus is on a specific event-dependent time window, a data collection strategy based on a specific occurrence will be required; for example, triggering an ESM measurement following changes in physiology (van Halem et al., 2020).

It is of note, that variable fluctuations can vary depending on the population's health, social, economic, or cultural characteristics (Okun, 2019). It is therefore advised to use explicitly validated variables or conduct a pilot study testing whether a protocol captures the expected fluctuations in the population of interest. Furthermore, studies including multiple parallel data types should address these questions for each data type separately.

The above-mentioned questions should be carefully addressed prior to data collection and ideally be pre-registered in one of the open science platforms most relevant to the specific field. There are many available platforms and guidelines to help explore different options (Kathawalla, Silverstein, & Syed, 2021). Due to the multiple types of data and the different steps necessary to track and report research, a platform that allows for greater flexibility, such as the Open Science Framework (OSF), is advised. The OSF provides a common place to enact all open science practices, such as pre-registration, data storage and sharing, code sharing, pre-prints, to name a few (Foster & Deardorff, 2017).

Data analysis

Thus far, defining detailed hypotheses and their variables of interest including their intended operationalization and expected fluctuation patterns have been discussed. It is time to determine the pre-processing and statistical analysis that best answers our research question. Although a detailed statistical discussion is beyond the scope of this paper, it is necessary to highlight the importance of choosing the right analysis prior to data collection. The chosen methods of data pre-processing and analysis will likely influence the required study design, but also potentially limit the ability to answer the intended research question.

Literature exists on data analysis for ESM specifically, which helps researchers to consider and perform for instance power size calculations (Lafit et al., 2020; Scherbaum & Pesner, 2019) and multilevel statistics which account for the data hierarchy as

well as disentangle between- and within-person differences (Bolger & Laurenceau, 2013; Mehl & Conner, 2013; Singer, Willett, Willett, & Willett, 2003). On the other hand, parallel data sources can require various analytic approaches depending on the type and format of the data. Some of these approaches are already showcased specifically in relation to ESM (Baumeister & Montag, 2019; Rehg et al., 2017). However, in the case of parallel data sources that are not yet referenced, it is advised to consult existing analyses in the specific field of interest and ideally with similar variables.

Broadly, ESM and parallel data studies are typically limited to either describing the association between variables, or developing and/or validating a model that predicts a variable prospectively based on the other variable(s) (Baumeister & Montag, 2019; Pencina, Goldstein, & D'Agostino, 2020; Yarkoni & Westfall, 2017). To avoid spurious findings in predictive modeling, cross-validation is advised by splitting the collected data in a separate 'training data set' and a 'test data set', both containing all types of collected data (Kubben, Dumontier, & Dekker, 2018).

Likewise, we recommend pre-registering the statistical analysis plan, including the code, prior to data collection. Once the data are collected, we advocate that it should be shared in an appropriate database so that other researchers may replicate the work (Turkylmaz-van der Velden, Dintzner, & Teperek, 2020).

How to collect parallel data that truly capture the variable of interest?

Device selection

Device selection is a vital component of studies measuring parallel data, and an important decision to ensure good compliance and data quality. There are many commercial wearable devices which collect data with minimal intervention, but all of these perform differently and have specific limitations that may change by (patient) populations (Fuller et al., 2020; Lai et al., 2020; Nelson & Allen, 2019). It is important to ensure the device collects accurate data that are reliable and valid. Moreover, the device must capture the required data with the right frequency and be validated in the appropriate population. Although they are not available for every application, systemic reviews or comprehensive guidelines exist to help researchers select wearables devices for specific scientific applications (Kunkels, van Roon, Wichers, & Riese, 2021; Nelson et al., 2020).

In addition to data quality, there are other topics deserving attention such as patient comfort and burden, privacy regulations, data security, storage and ownership, and battery life (Rehg et al., 2017). If real-time evaluation is desired, for example in case of event-dependent ESM assessment, connectivity and data sharing issues should be considered (Cornet & Holden, 2018; Kohrt et al., 2019; Trifan, Oliveira, & Oliveira, 2019).

Furthermore, it is essential that devices reliably log the timestamps in universal comparable time. Timestamps simply note when an assessment took place. Some devices need to be synchronized at the beginning of a recording session, and some are subject to drift, which means the timestamp accuracy decreases over time. Overall, inaccuracies within second ranges are negligible since ESM answers do not represent events at a (micro-)second level.

An important distinction in the available devices is whether they provide raw data or proxy data. Data provenance should be well-known prior to data collection, that is the various pre-

processing steps taken to transform raw data into meaningful information (Rehg et al., 2017). Proxy data, often provided by commercial devices, are already processed or summarized into assessment scores, such as activity rates per day or per hour. When proxy data are preferred, or it is not possible to obtain raw data, it is vital that the algorithm used to compute the proxy data is known, or at least well-understood, and most importantly validated (Feehan et al., 2018; Horton, Stergiou, Fung, & Katz, 2017). Not understanding the essence of the proxy data may heavily affect the validity and interpretation of the obtained findings.

At this stage it is also relevant to consider a data management plan (Wilkinson et al., 2016). That is where data will be stored (short and long term), how will it be preserved, and who will have access to it. A proper data management plan is essential for all studies but especially in this case where there are many data sources with different formats and sizes. Part of this plan should include comprehensive information on what the datasets contain, if the data are raw or have been through any preprocessing steps.

Sampling frequency

Sampling frequency should be defined carefully for the reasons stated above. For variables with a stable volatility, the *Nyquist theorem* can be used. The Nyquist theorem is commonly used in signal processing, and dictates the sampling frequency to be larger than twice the frequency of the smallest fluctuation in the variable of interest (Bogdan, 2009). Violating the Nyquist theorem by under-sampling can lead to aliasing; the incorrect extraction of peaks and frequencies from a raw signal. Aliasing is more applicable for high-frequency sampling of parallel data than data collected with ESM.

On the contrary, variables with unstable volatilities are more complicated, such as stress-reactivity, or geolocation. Methodologies from studies or reviews assessing the same variable can often provide evidence regarding specific sampling frequencies. Some examples are heart rate variability (Shaffer & Ginsberg, 2017), GPS-based out-of-home activity (Kondo et al., 2020; Liao, Song, Robertson, Cox-Martin, & Basen-Engquist, 2020; Zeng, Fraccaro, & Peek, 2019), and accelerometry-based activity monitoring (Kolar et al., 2020; Niazi et al., 2017). It is important to stress that under-sampling issues cannot be resolved by simply collecting more data over longer periods of time. Larger datasets which still do not capture the fluctuation of the variable(s) of interest will not lead to meaningful interpretations.

How to translate data into meaningful information?

Feature extraction

For a meaningful interpretation of parallel data, information must be extracted from the raw parallel data in a way that it represents the variable of interest. In signal processing terms, the values containing this information are called features. The period of data used to calculate one feature is called the feature window. In some specific cases, the raw data contain the desired information, and no feature extraction is required (e.g. body temperature at specific moments). However, in general, raw parallel data will need to be pre-processed via feature extraction. The type and timescale of features is dependent on the type of data, and the exact variable of interest. For example, proximities to outdoor

natural environments can be extracted from GPS-data per 10 min (Kondo et al., 2020), while physiological features like heart rate or movement need to be calculated over (milli)seconds. Choosing, or finding, the right feature window size is important since various window sizes may lead to different results (Heijmans, Habets, Kuijf, Kubben, & Herff, 2019b). For some parallel data or hypotheses, aggregation of high-frequency features over longer windows might be necessary.

For both data types, it is important to consider these (pre-) processing steps including how to store and annotate the raw data, the features, and preferably the code. For this it is highly recommended to pre-register the study prior to conducting data collection. Publishing detailed scripts of the performed pre-processing steps and analyses, including possible post hoc or additional analysis, will further improve the study's scientific quality and reproducibility. Several resources are available to help with these steps such as the ESM pre-registration template (Kirtley, Lafit, Achterhof, Hiekkaranta, & Myin-Germeys, 2020) and guidance for scientific data care (Goodman et al., 2014).

Temporal feature aggregation

Assuming the parallel data features and ESM answers differ in sampling frequency, we can regard these two time series of data-points both as snapshots of an ongoing, continuous fluctuating process (Fig. 2). To describe an association between them, we need to either up-sample, or down-sample one of them, or both. Up-sampling, via value imputation or extrapolation, tends to generate uncertainty, especially when it needs to be done repeatedly. Down-sampling however, can lead to important information loss. It is common practice to down-sample the parallel data to ESM data sampling frequency.

The most straightforward comparison between ESM and parallel data is to regard each completed ESM questionnaire as a single event, and to compare it with the parallel data collected in the corresponding time window (see Fig. 2). In this case, the higher-frequency parallel data are down-sampled via feature extraction. For this we need to define the duration (how many seconds or minutes), and timing (prior or after the ESM completion) of the window of parallel data corresponding with the ESM variable. The correct duration and timing will depend on the subjective experience that is assessed with the ESM item, the ESM instruction given to the participant, and the formulated hypothesis. The selected parallel data window will then be translated into the variable(s) of interest via the described feature extraction process (Habets et al., 2021).

The comparison of ESM and parallel data in which both data types are regarded as ongoing processes over time is based on a different theoretical principle and requires different statistical approaches. This could be especially relevant for hypotheses focusing on continuous processes. Instead of down-sampling the parallel data, up-sampling of the ESM data is required via for example extrapolation (Fig. 2). It is important to carefully consider each statistical method's limitations and potential bias.

Missing data

Similar to the statistical analyses, a detailed description of missing data management is beyond the scope of this work. However, ignoring it entirely would be a significant omission. In general, missing data can be handled in various ways (Little & Rubin, 2002b). Broadly, it is important to assess whether the missing

Table 1. Considerations for study design checklist

Level	Considerations for study design
Focus of the research	<input type="checkbox"/> What is the association of interest? <input type="checkbox"/> Is there support in the literature for a specific association or is exploratory research required?
Variables	<input type="checkbox"/> What are the specific variables of interest? <input type="checkbox"/> What is their unique contribution? <input type="checkbox"/> What is known about each variable's fluctuation and volatility? <input type="checkbox"/> Is there literature on the appropriate sampling frequency to capture meaningful changes in the dimension of interest? If not, based on its fluctuation and volatility, what would be the minimum required frequency to capture meaningful variability?
Association of interest	<input type="checkbox"/> Is there a temporal order between variables? If so, which one? Keeping in mind that causal associations can scarcely be established in such study designs. <input type="checkbox"/> What window size is necessary for parallel data to reflect an ESM assessment?
Device selection	<input type="checkbox"/> What devices are available to collect the measures of interest? <input type="checkbox"/> Has the device been validated in the population of interest? <input type="checkbox"/> How is data managed, saved, and distributed? <input type="checkbox"/> Is there sufficient storage and battery life? <input type="checkbox"/> Is raw data accessible? If not, is data provenance known??
Data analyses	<input type="checkbox"/> What analyses are anticipated? <input type="checkbox"/> How much data is necessary to perform these analyses? <input type="checkbox"/> How will missingness be handled??

data-points are missing at random or may represent completion bias. Especially in the case of ESM, missing data can be caused by (disease) specific reasons and can contain significant information (Cursio, Mermelstein, & Hedeker, 2019). In addition to asking participants why they missed entries, analyzing both the corresponding parallel data as well as the non-missing ESM data can be considered. Many sources already exist to help researchers handle missing data in a multilevel structured dataset (van Ginkel, Linting, Rippe, & van der Voort, 2020), as well as in time-series data (De Waal, Pannekoek, & Scholtus, 2011). There are also resources to consider regarding for example data imputation techniques (Beard et al., 2019; van Breda et al., 2016) and the possible bias it may introduce (Little & Rubin, 2002a).

How to bring these considerations into practice?

Practical checklist

To provide researchers with an easy overview of the key elements to consider when conducting a study that combines ESM and parallel data, we provide a detailed checklist to guide in the different stages of the study design (see Table 1). This is intended to be a useful advisory research tool, rather than a rigid guideline.

Open-source example

A comprehensive practical example has been drafted to demonstrate how to apply all the different considerations mentioned in this paper (see Table 2). Here, we present how the checklist can be used in practice. For easy access and replicability, we use a publicly available dataset containing ESM and parallel data (i.e. movement data) collected from 20 patients with Parkinson's disease over the course of 14 consecutive days, without restrictions, in daily-life (Habets et al., 2021; Habets & Kubben, 2020). The ESM assessment contains psychological items assessing affect and mood, as well as questions on motor symptoms, physical ability, and contextual questions (Habets et al., 2020). The movement data consist of raw acceleration (accelerometer) and rotation (gyroscope) time-series data derived from a wrist-worn

movement sensor. From these data Parkinson motor symptom variables, such as tremor, were calculated. The database and accompanying Python Notebooks (Perez & Granger, 2007) with example code on data pre-processing and data merging have been published (Habets, 2020; Habets et al., 2021) and are available at: <https://zenodo.org/record/4734199#.YJAOZRQza3J>. For computational details and background on the movement data derived tremor scores we refer to a previous publication (Heijmans et al., 2019b). A detailed description of this example can be found in the Supplementary Material.

For this example, we are interested in the effect that Parkinsonian tremor severity has on negative affect. Interesting within-subject relationships between mood and tremor severity are showed in a longitudinal $n=1$ study (van der Velden, Mulders, Drukker, Kuijff, & Leentjens, 2018). This study only used ESM data and did not include any parallel data. With complementary movement data, we want to reproduce and further explore this relationship. Since this data repository is already available, we assess whether it is suitable to answer our question.

Conclusion

The ESM is increasingly combined with parallel data, collected by passive sensing devices or wearable sensors in daily-life. These methods hold a great potential to contribute to ambulatory monitoring and personalized health care. However, combining these data types in research or clinical practice comes with new challenges for which specific guidance is lacking. We presented several important and useful considerations to support researchers in every stage of designing a research project. We stressed the importance of understanding the fluctuations and the temporal resolution within the two separate data timelines and their operationalization. We further described essential considerations on device selection and feature extraction and aggregation, as well as their statistical analysis. Finally, we underlined necessary methods to ensure transparency and reproducibility.

The provided recommendations aim to guide researchers in conducting meaningful research combining state-of-the-art

Table 2. Practical example of how to apply the considerations of study design checklist

Level	Considerations for study design
Focus of the research	<ul style="list-style-type: none"> ☑ What is the association of interest? <ul style="list-style-type: none"> – Tremor episodes and negative affect (NA) in Parkinson's disease. ☑ Is there support in the literature for a specific association or is exploratory research required? <ul style="list-style-type: none"> – There is evidence that more tremor episodes lead to more NA in Parkinson's disease.
Variables	<ul style="list-style-type: none"> ☑ What are the specific variables of interest? <ul style="list-style-type: none"> – NA is a composite measure computed as the average of the items: mood, depression, anxiety, negatively related to joy, happiness, positive excitement (ESM). – Tremor: measured with movement features of accelerometer and gyroscope time series (parallel data). ☑ What is their unique contribution? <ul style="list-style-type: none"> – Subjective affect in daily-life. – Passive, unobtrusive measure of tremor. ☑ What is known about each variable's fluctuation and volatility? <ul style="list-style-type: none"> – NA fluctuates heavily within and between days. – Tremor can fluctuate heavily in severity over minutes, but will also fluctuate over hours based on dopaminergic medication states. ☑ Is there literature on the appropriate sampling frequency to capture meaningful changes in the dimension of interest? If not, based on its fluctuation and volatility, what would be the minimum required frequency to capture meaningful variability? <ul style="list-style-type: none"> – 7 to 10 assessments a day have shown to reliably capture NA variability without over-burdening Parkinson patients. – Tremor is a relatively volatile Parkinson motor symptom, and episodes can be less than a minute. Tremor severity must be assessed every ten to thirty seconds. At least 25–50 Hz is required for tremor detection. Current 200 Hz is oversampled but can be reduced during the pre-processing stage.
Association of interest	<ul style="list-style-type: none"> ☑ Is there a temporal order between variables? If so, which one? Keeping in mind that causal associations can scarcely be established in such study designs. <ul style="list-style-type: none"> – We are interested in NA after tremor, so parallel data is down-sampled, and windows are computed prior to the timestamp of the ESM assessment. ☑ What window size is necessary for parallel data to reflect an ESM assessment? <ul style="list-style-type: none"> – Due to the high volatility of NA, we down-sample the parallel data to windows of 30 minutes prior to each ESM assessment.
Device selection	<ul style="list-style-type: none"> ☑ What devices are available to collect the measures of interest? <ul style="list-style-type: none"> – Several ESM platforms with user-interfaces suitable for Parkinson patients are available. – Wrist-worn wearable movement sensors. ☑ Has the device been validated in the population of interest? <ul style="list-style-type: none"> – Data collection with the applied methodology was reported to be feasible in this specific population. ☑ How is data managed, saved, and distributed? <ul style="list-style-type: none"> – ESM data is directly transferred to a server when connected to WiFi. – Parallel data is primarily stored on the device and is later downloaded during a research visit. ☑ Is there sufficient storage and battery life? <ul style="list-style-type: none"> – There is sufficient storing capacity and participants receive clear instructions about battery life and charging methods. ☑ Is raw data accessible? If not, is data provenance known? <ul style="list-style-type: none"> – Yes, all raw data is available.
Data analyses	<ul style="list-style-type: none"> ☑ What analyses are anticipated? <ul style="list-style-type: none"> – Nested multi-level analyses on the ESM data have to show significant fluctuations in NA within patients, within days, but also enable comparison between patients based on differences to inter-patient means. – Parallel data will be analyzed with signal processing techniques in which relevant movement features (e.g. power, variance, variability, spectral decomposition scores, smoothness) will be extracted over short time-windows of 1 second to several minutes. ☑ How much data is necessary to perform these analyses? <ul style="list-style-type: none"> – Assuming the occurrence of frequent fluctuations in tremor severity, fourteen days of data would be sufficient to make meaningful conclusions. ☑ How will missingness be handled? <ul style="list-style-type: none"> – If missing parallel data are caused by technical difficulties this can be accepted and excluded. If the wearable devices are not worn frequently, the researcher has to explore anamnestically or in the ESM data whether there is a pattern between those moments. – Similar to parallel data, the moments without ESM answers should be explored in the parallel data for potentially high tremor severity.

methods of daily-life monitoring: ESM and parallel data. Careful reflection prior to data collection is fundamental to conduct a valid study that is effective in capturing the investigated processes. Doing so will result in combined datasets of better quality, rather than quantity, and new insights into ambulatory health care.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291722002367>.

Financial support. No direct funding is received for this work.

Conflict of interest. None of the authors have a conflict of interest due to the reported financial disclosures.

References

- Barnett, I., Torous, J., Reeder, H. T., Baker, J., & Onnela, J.-P. (2020). Determining sample size and length of follow-up for smartphone-based digital phenotyping studies. *Journal of the American Medical Informatics Association: JAMIA*, 27(12), 1844–1849. doi:10.1093/jamia/ocaa201.
- Baumeister, H., & Montag, C. (2019). *Digital phenotyping and mobile sensing: New developments in psychoinformatics* (pp. 31–260). Switzerland: Springer Nature. doi:10.1007/978-3-030-31620-4.
- Beard, E., Marsden, J., Brown, J., Tombor, I., Stapleton, J., Michie, S., & West, R. (2019). Understanding and using time series analyses in addiction research. *Addiction (Abingdon, England)*, 114(10), 1866–1884. doi:10.1111/add.14643.
- Bertz, J. W., Epstein, D. H., & Preston, K. L. (2018). Combining ecological momentary assessment with objective, ambulatory measures of behavior

- and physiology in substance-use research. *Addictive Behaviors*, 83, 5–17. doi:10.1016/j.addbeh.2017.11.027.
- Bogdan, M. (2009). Sampling rate and aliasing on a virtual laboratory. *Journal of Electrical and Electronics Engineering*, 2, 121–124.
- Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York City, NY: Guilford Press.
- Christensen, T. C., Barrett, L. F., Bliss-Moreau, E., Lebo, K., & Christensen, T. C. (2003). A practical guide to experience-sampling procedures. *Journal of Happiness Studies*, 4(1), 53–78. doi:10.1023/a:1023609306024.
- Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of Biomedical Informatics*, 77, 120–132. doi:10.1016/j.jbi.2017.12.008.
- Corrigan-Curay, J., Sacks, L., & Woodcock, J. (2018). Real-world evidence and real-world data for evaluating drug safety and effectiveness. *JAMA: The Journal of the American Medical Association*, 320(9), 867–868. doi:10.1001/jama.2018.10136.
- Cousins, J. C., Whalen, D. J., Dahl, R. E., Forbes, E. E., Olino, T. M., Ryan, N. D., & Silk, J. S. (2011). The bidirectional association between daytime affect and nighttime sleep in youth with anxiety and depression. *Journal of Pediatric Psychology*, 36(9), 969–979. doi:10.1093/jpepsy/jsr036.
- Cursio, J. F., Mermelstein, R. J., & Hedeker, D. (2019). Latent trait shared-parameter mixed models for missing ecological momentary assessment data: Latent trait shared-parameter mixed models. *Statistics in Medicine*, 38(4), 660–673. doi:10.1002/sim.7989.
- De Waal, T., Pannekoek, J., & Scholtus, S. (2011). *Handbook of statistical data editing and imputation* (Vol. 563). New York City, NY: John Wiley & Sons.
- Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., ... Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, 29(2), 136–151. doi:10.1177/1073191120957102.
- Feehan, L. M., Geldman, J., Sayre, E. C., Park, C., Ezzat, A. M., Yoo, J. Y., ... Li, L. C. (2018). Accuracy of Fitbit devices: Systematic review and narrative syntheses of quantitative data. *JMIR MHealth and UHealth*, 6(8), e10527. doi:10.2196/10527.
- Food and Drug Administration. (2022). Retrieved June 6, 2022, from U.S. Food and Drug Administration website <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>.
- Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA*, 105(2), 203–206. doi: 10.5195/jmla.2017.88.
- Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., ... 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. doi:10.2196/18694.
- Fuller-Tyszkiewicz, M., Skouteris, H., Richardson, B., Blore, J., Holmes, M., & Mills, J. (2013). Does the burden of the experience sampling method undermine data quality in state body image research? *Body Image*, 10(4), 607–613. doi:10.1016/j.bodyim.2013.06.003.
- Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., ... Slavkovic, A. (2014). Ten simple rules for the care and feeding of scientific data. *PLoS Computational Biology*, 10(4), e1003542. doi:10.1371/journal.pcbi.1003542.
- Graham, S. A., Jeste, D. V., Lee, E. E., Wu, T.-C., Tu, X., Kim, H.-C., & Depp, C. A. (2019). Associations between heart rate variability measured with a wrist-worn sensor and older adults' physical function: Observational study. *JMIR MHealth and UHealth*, 7(10), e13757. doi:10.2196/13757.
- Habets, J., Heijmans, M., Herff, C., Simons, C., Leentjens, A. F., Temel, Y., ... Kubben, P. (2020). Mobile health daily life monitoring for Parkinson disease: Development and validation of ecological momentary assessments. *JMIR MHealth and UHealth*, 8(5), e15628. doi:10.2196/15628.
- Habets, J., & Kubben, P. (2020). *EMA and wearable sensor monitoring in PD* [Data set]. Amsterdam, The Netherlands: DataverseNL, DANS KNAW.
- Habets, J. G. V., Heijmans, M., Leentjens, A. F. G., Simons, C. J. P., Temel, Y., Kuijff, M. L., ... Herff, C. (2021). A long-term, real-life Parkinson monitoring database combining unscripted objective and subjective recordings. *Data*, 6(2), 22. doi:10.3390/data6020022.
- Heijmans, M., Habets, J., Kuijff, M., Kubben, P., & Herff, C. (2019b). Evaluation of Parkinson's disease at home: Predicting tremor from wearable sensors. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2019*, pp. 584–587. doi:10.1109/EMBC.2019.8857717.
- Heijmans, M., Habets, J. G. V., Herff, C., Aarts, J., Stevens, A., Kuijff, M. L., & Kubben, P. L. (2019a). Monitoring Parkinson's disease symptoms during daily life: A feasibility study. *NPJ Parkinson's Disease*, 5(1), 21. doi:10.1038/s41531-019-0093-5.
- Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The 'real-world approach' and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11, 721. doi:10.3389/fpsyg.2020.00721.
- Horton, J. F., Stergiou, P., Fung, T. S., & Katz, L. (2017). Comparison of Polar M600 optical heart rate and ECG heart rate during exercise. *Medicine and Science in Sports and Exercise*, 49(12), 2600–2607. doi:10.1249/MSS.0000000000001388.
- Janssens, K. A. M., Bos, E. H., Rosmalen, J. G. M., Wichers, M. C., & Riese, H. (2018). A qualitative approach to guide choices for designing a diary study. *BMC Medical Research Methodology*, 18(1), 140. doi:10.1186/s12874-018-0579-6.
- Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra Psychology*, 7, 1. doi:10.1525/collabra.18684.
- Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., & Kim, N. (2019). Depression prediction by using ecological momentary assessment, Actiwatch data, and machine learning: Observational study on older adults living alone. *JMIR MHealth and UHealth*, 7(10), e14149. doi:10.2196/14149.
- Kimhy, D., Wall, M. M., Hansen, M. C., Vakhrusheva, J., Choi, C. J., Delespaul, P., ... Malaspina, D. (2017). Autonomic regulation and auditory hallucinations in individuals with schizophrenia: An experience sampling study. *Schizophrenia Bulletin*, 43(4), 754–763. doi:10.1093/schbul/sbw219.
- Kirtley, O., Lafit, G., Achterhof, R., Hiekkaranta, A. P., & Myin-Germeys, I. (2020). *A template and tutorial for (pre-)registration of studies using experience sampling methods (ESM)*. Charlottesville, VA: Open Science Framework. doi:10.17605/OSF.IO/2CHMU.
- Kohrt, B. A., Rai, S., Vilakazi, K., Thapa, K., Bhardwaj, A., & van Heerden, A. (2019). Procedures to select digital sensing technologies for passive data collection with children and their caregivers: Qualitative cultural assessment in South Africa and Nepal. *JMIR Pediatrics and Parenting*, 2(1), e12366. doi:10.2196/12366.
- Kolar, D. R., Neumayr, C., Roth, M., Voderholzer, U., Perthes, K., & Schlegl, S. (2020). Testing an emotion regulation model of physical activity in adolescents with anorexia nervosa: A pilot ecological momentary assessment. *European Eating Disorders Review: The Journal of the Eating Disorders Association*, 28(2), 170–183. doi:10.1002/erv.2706.
- Kondo, M. C., Triguero-Mas, M., Donaire-Gonzalez, D., Seto, E., Valentin, A., Hurst, G., ... Nieuwenhuijsen, M. J. (2020). Momentary mood response to natural outdoor environments in four European cities. *Environment International*, 134(105237), 105237. doi:10.1016/j.envint.2019.105237.
- Kubben, P., Dumontier, M., & Dekker, A. (2018). *Fundamentals of clinical data science* (M. D. Pieter Kubben Andre Dekker, Ed.). Springer International Publishing.
- Kunkels, Y. K., van Roon, A. M., Wichers, M., & Riese, H. (2021). Cross-instrument feasibility, validity, and reproducibility of wireless heart rate monitors: Novel opportunities for extended daily life monitoring. *Psychophysiology*, 58(10), e13898. doi:10.1111/psyp.13898.
- Lafit, G., Adolf, J., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2020). Selection of the number of participants in intensive longitudinal studies: A user-friendly shiny app and tutorial to perform power analysis in multilevel regression models that account for temporal dependencies. *PsyArXiv preprint*: <https://psyarxiv.com/dq6ky/>.
- Lai, B., Sasaki, J. E., Jeng, B., Cederberg, K. L., Bamman, M. M., & Motl, R. W. (2020). Accuracy and precision of three consumer-grade motion sensors during overground and treadmill walking in people with Parkinson disease:

- Cross-sectional comparative study. *JMIR Rehabilitation and Assistive Technologies*, 7(1), e14059. doi:10.2196/14059.
- Liao, Y., Song, J., Robertson, M. C., Cox-Martin, E., & Basen-Engquist, K. (2020). An ecological momentary assessment study investigating self-efficacy and outcome expectancy as mediators of affective and physiological responses and exercise among endometrial cancer survivors. *Annals of Behavioral Medicine*, 54(5), 320–334. doi:10.1093/abm/kaz050.
- Little, R. J. A., & Rubin, D. B. (2002a). Estimation of imputation uncertainty. In R. J. Little & D. B. Rubin (Eds.), *Statistical analysis with missing data* (pp. 75–93). Hoboken, NJ: John Wiley & Sons, Inc.
- Little, R. J. A., & Rubin, D. B. (2002b). Missing data in experiments. In R. J. Little & D. B. Rubin (Eds.), *Statistical analysis with missing data* (pp. 24–40). Hoboken, NJ: John Wiley & Sons, Inc.
- Mehl, M. R., & Conner, T. S. (2013). *Handbook of research methods for studying daily life*. New York City, NY: Guilford Publications.
- Minaeva, O., Riese, H., Lamers, F., Antypa, N., Wichers, M., & Booi, S. H. (2020). Screening for depression in daily life: Development and external validation of a prediction model based on actigraphy and experience sampling method. *Journal of Medical Internet Research*, 22(12), e22634. doi:10.2196/22634.
- Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, 17(2), 123–132. doi:10.1002/wps.20513.
- Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., ... Murphy, S. A. (2016). Just-in-time adaptive interventions (JITAI) in mobile health: Key components and design principles for ongoing health behavior support. *Annals of Behavioral Medicine*, 52(6), 446–462. doi:10.1007/s12160-016-9830-8.
- Nelson, B. W., & Allen, N. B. (2019). Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. *JMIR MHealth and UHealth*, 7(3), e10828. doi:10.2196/10828.
- Nelson, B. W., Low, C. A., Jacobson, N., Areán, P., Torous, J., & Allen, N. B. (2020). Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *NPJ Digital Medicine*, 3(1), 90. doi:10.1038/s41746-020-0297-4.
- Niazi, A. H., Yazdanehpas, D., Gay, J. L., Maier, F. W., Ramaswamy, L., Rasheed, K., ... Buman, M. (2017). Statistical analysis of window sizes and sampling rates in human activity recognition. *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies*. Setubal, Portugal: SciTePress – Science and Technology Publications.
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45(3), 137–141. doi:10.1027/1864-9335/a000192.
- Numpy.polyfit – NumPy v1.22 Manual. (2022). Retrieved June 6, 2022, from Numpy.org website <https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html>.
- Okun, S. (2019). The missing reality of real life in real-world evidence. *Clinical Pharmacology and Therapeutics*, 106(1), 136–138. doi:10.1002/cpt.1465.
- Oyinlola, J. O., Campbell, J., & Kousoulis, A. A. (2016). Is real world evidence influencing practice? A systematic review of CPRD research in NICE guidances. *BMC Health Services Research*, 16, 1. doi:10.1186/s12913-016-1562-8.
- Palmier-Claus, J. E., Myin-Germeys, I., Barkus, E., Bentley, L., Udachina, A., Delespaul, P. A. E. G., ... Dunn, G. (2011). Experience sampling research in individuals with mental illness: Reflections and guidance: Experience sampling research in individuals with mental illness. *Acta Psychiatrica Scandinavica*, 123(1), 12–20. doi:10.1111/j.1600-0447.2010.01596.x.
- Pencina, M. J., Goldstein, B. A., & D'Agostino, R. B. (2020). Prediction models – Development, evaluation, and clinical application. *The New England Journal of Medicine*, 382(17), 1583–1586. doi:10.1056/NEJMp2000589.
- Perez, F., & Granger, B. E. (2007). IPython: A system for interactive scientific computing. *Computing in Science & Engineering*, 9(3), 21–29. doi:10.1109/mcse.2007.53.
- Rehg, J. M., Murphy, S. A., & Kumar, S. (2017). *Mobile health: Sensors, analytic methods, and applications*. Basel, Switzerland: Springer.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. doi:10.1177/2515245917745629.
- Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, 26(1), 20–36. doi:10.1002/sim.2739.
- Scherbaum, C. A., & Pesner, E. (2019). Power analysis for multilevel research. In S. E. Humphrey & J. M. LeBreton (Eds.), *The handbook of multilevel theory, measurement, and analysis* (pp. 329–352). Washington: American Psychological Association.
- Scipy.stats.variation – SciPy v1.8.1 Manual. (n.d.). Retrieved June 6, 2022, from Scipy.org website <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.variation.html>.
- Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5, 258. doi:10.3389/fpubh.2017.00258.
- Sharmin, M., Raij, A., Epstien, D., Nahum-Shani, I., Beck, J. G., Vhaduri, S., ... Kumar, S. (2015). Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. *Proceedings of the ACM International Conference on Ubiquitous Computing. UbiComp (Conference)*, 2015, pp. 505–516. Retrieved from doi:10.1145/2750858.2807537.
- Shortliffe, E. H. (1993). The adolescence of AI in medicine: Will the field come of age in the '90s? *Artificial Intelligence in Medicine*, 5(2), 93–106. doi:10.1016/0933-3657(93)90011-q.
- Singer, J. D., Willett, J. B., Willett, C. W. E. P. J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.
- Smets, E., De Raedt, W., & Van Hoof, C. (2019). Into the wild: The challenges of physiological stress detection in laboratory and ambulatory settings. *IEEE Journal of Biomedical and Health Informatics*, 23(2), 463–473. doi:10.1109/JBHI.2018.2883751.
- Stuppelle, A., Singerman, D., & Celi, L. A. (2019). The reproducibility crisis in the age of digital medicine. *NPJ Digital Medicine*, 2(1), 2. doi:10.1038/s41746-019-0079-z.
- Tackett, J. L., Brandes, C. M., King, K. M., & Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual Review of Clinical Psychology*, 15(1), 579–604. doi:10.1146/annurev-clinpsy-050718-095710.
- Trifan, A., Oliveira, M., & Oliveira, J. L. (2019). Passive sensing of health outcomes through smartphones: Systematic review of current solutions and possible limitations. *JMIR MHealth and UHealth*, 7(8), e12649. doi:10.2196/12649.
- Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology*, 129(1), 56–63. doi:10.1037/abn0000473.
- Turkylmaz-van der Velden, Y., Dintzner, N., & Teperek, M. (2020). Reproducibility starts from you today. *Patterns (New York, N.Y.)*, 1(6), 100099. doi:10.1016/j.patter.2020.100099.
- Vaessen, T., Rintala, A., Otsabryk, N., Viechtbauer, W., Wampers, M., Claes, S., & Myin-Germeys, I. (2021). The association between self-reported stress and cardiovascular measures in daily life: A systematic review. *PLoS One*, 16(11), e0259557. doi:10.1371/journal.pone.0259557.
- van Breda, W., Pastor, J., Hoogendoorn, M., Ruwaard, J., Asselbergs, J., & Riper, H. (2016). Exploring and comparing machine learning approaches for predicting mood over time. In *Innovation in Medicine and Healthcare Conference 2016* (pp. 37–47). Springer. doi:10.1007/978-3-319-39687-3_4.
- van der Velden, R. M. J., Mulders, A. E. P., Drukker, M., Kuijff, M. L., & Leentjens, A. F. G. (2018). Network analysis of symptoms in a Parkinson patient using experience sampling data: An $n=1$ study: Symptom network analysis in Parkinson's disease. *Movement Disorders*, 33(12), 1938–1944. doi:10.1002/mds.93.
- van Ginkel, J. R., Linting, M., Rippe, R. C. A., & van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of Personality Assessment*, 102(3), 297–308. doi:10.1080/00223891.2018.1530680.
- van Halem, S., van Roekel, E., Kroencke, L., Kuper, N., & Denissen, J. (2020). Moments that matter? On the complexity of using triggers based on skin conductance to sample arousing events within an experience sampling framework. *European Journal of Personality*, 34(5), 794–807. doi:10.1002/per.2252.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR guiding principles for scientific

- data management and stewardship. *Scientific Data*, 3(1), 160018. doi:10.1038/sdata.2016.18.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 12(6), 1100–1122. doi:10.1177/1745691617693393.
- Yuan, N., Weeks, H. M., Ball, R., Newman, M. W., Chang, Y. J., & Radesky, J. S. (2019). How much do parents actually use their smartphones? Pilot study comparing self-report to passive sensing. *Pediatric Research*, 86(4), 416–418. doi:10.1038/s41390-019-0452-2.
- Zeng, Y., Fraccaro, P., & Peek, N. (2019). The minimum sampling rate and sampling duration when applying geolocation data technology to human activity monitoring. In D. Riaño, S. Wilk, & A. ten Teije (Eds.), *Artificial Intelligence in Medicine. AIME 2019. Lecture Notes in Computer Science*, Vol. 11526. Cham: Springer. doi:10.1007/978-3-030-21642-9_29.