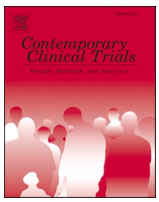
Contemporary Clinical Trials 113 (2022) 106661

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial



Considerations to address missing data when deriving clinical trial endpoints from digital health technologies

Junrui Di [a,*], Charmaine Demanuele [a], Anna Kettermann [b], F. Isik Karahanoglu [a], Joseph C. Cappelleri [a], Andrew Potter [b], Denise Bury [c], Jesse M. Cedarbaum [d,e], Bill Byrom [f]

[a] Pfizer Inc., United States of America
[b] FDA, United States of America
[c] Sanofi, United States of America
[d] Yale University School of Medicine, United States of America
[e] Coeruleus Clinical Sciences LLC, United States of America
[f] Signant Health, United Kingdom

ARTICLE INFO

Keywords:
Digital health technologies
Missing data
Statistical methods for missing data
DHT deployment

ABSTRACT

Digital health technologies (DHTs) enable us to measure human physiology and behavior remotely, objectively and continuously. With the accelerated adoption of DHTs in clinical trials, there is an unmet need to identify statistical approaches to address missing data to ensure that the derived endpoints are valid, accurate, and reliable. It is not obvious how commonly used statistical methods to handle missing data in clinical trials can be directly applied to the complex data collected by DHTs. Meanwhile, current approaches used to address missing data from DHTs are of limited sophistication and focus on the exclusion of data where the quantity of missing data exceeds a given threshold.

High-frequency time series data collected by DHTs are often summarized to derive epoch-level data, which are then processed to compute daily summary measures. In this article, we discuss characteristics of missing data collected by DHT, review emerging statistical approaches for addressing missingness in epoch-level data including within-patient imputations across common time periods, functional data analysis, and deep learning methods, as well as imputation approaches and robust modeling appropriate for handling missing data in daily summary measures. We discuss strategies for minimizing missing data by optimizing DHT deployment and by including the patients' perspectives in the study design. We believe that these approaches provide more insight into preventing missing data when deriving digital endpoints. We hope this article can serve as a starting point for further discussion among clinical trial stakeholders.

1. Introduction

The ongoing innovation and miniaturization of sensors have led to a wide demand for wearables to track personal health and wellness, which has become hugely popular in many application areas including drug development and clinical research. It is estimated that one in five adult Americans owns a smartwatch or a wearable activity tracker [1]. As suggested by the U.S. Food & Drug Administration (FDA), Digital Health includes "mobile health, health information technology, wearable devices, telehealth and telemedicine, and personalized medicine" [2]. The application of personal Digital Health Technologies (DHTs), defined as systems that use computing platforms, software, and sensors for healthcare and related uses [3], affords the opportunity to measure new endpoints from domains impractical or impossible to measure before, leading to a better understanding of more meaningful health outcome measures and disease characterization.

We are beyond the point where DHTs can only give us limited discrete data such as step counts. For instance, emerging approaches demonstrate the ability of one sensor located on the lumbar region to quantify the spatiotemporal characteristics of gait in Parkinson's Disease [4]. Other studies have demonstrated using wrist-worn accelerometers to quantify nocturnal scratching as an additional measure of pruritis alongside patient-reported itch ratings in atopic dermatology studies [5,6]. DHTs can provide measurements of human behaviors and health

* Corresponding author.
E-mail address: junrui.di@pfizer.com (J. Di).

https://doi.org/10.1016/j.cct.2021.106661
Received 18 August 2021; Received in revised form 23 November 2021; Accepted 18 December 2021
Available online 22 December 2021

status at high sampling frequency, and thus provide information with higher granularity. They can be used to measure episodic constructs and diurnal variability using continuous recordings over long periods of time in real-life settings. When used in free-living (day-to-day life, out of clinic) settings in clinical research, DHTs enable more frequent measurements of disease signs and symptoms, which can provide a more granular and holistic picture of health status in real-time relative to the more traditional snapshot of in-clinic assessments. DHTs provide a picture of real-world functioning which can be difficult to assess in the clinic and may lead to more pertinent and meaningful endpoints to understand disease and treatment impact and to provide more personalized care.

Many DHTs, such as accelerometers and continuous glucose monitors (CGM), collect data passively and frequently, leading to rich streams of time series data from which outcome measures are derived using algorithms. Careful consideration is needed when deriving clinical endpoints in drug development programs to ensure that they are robust and reliable to enable regulatory decision making [7]. While much work has been reported on the evidence to support device selection and reliability of measures [7,8], one less developed area is how to deal with missing data observed in these time series. Reaching a consensus on how to prevent and handle missing data, either due to participant compliance or technology related issues, will increase the integrity of DHT-based clinical study results and inform future studies. Therefore, in this article, we discuss characteristics of missing data from DHTs, review emerging methodologies to address missing data arising from DHTs, and provide strategies for minimizing missing data by optimizing DHT deployment in clinical studies.

2. Commonly used DHTs and their data

In this paper, we will focus on data derived from 2 types of DHTs, accelerometers and CGMs.

The accelerometer is the key component of modern wearable physical activity (PA) and sleep trackers (also known as actigraphy), which are often used to provide continuous and objective measures of 24-h activity and sedentary behavior. An accelerometer collects high resolution motion-induced acceleration signals (e.g., 30–100 Hz), which provide information about activity types and intensities. These raw acceleration signals are aggregated and summarized over epochs (e.g., 1 min) to quantify PA and sleep. Activity counts (or other similar metrics such as activity index [9] or Euclidean norm minus one [10]) per one-minute epoch are such a summary measure, and represent a proxy measure of activity level. These measures are produced either by the devices' proprietary algorithms, and therefore differ across devices and manufacturers [11], or by open source algorithms applied to the raw sensor data, which enable greater reproducibility and offer interpretational benefits [12]. Multiple features or daily summaries can then be extracted from the epoch-level counts data to describe different aspects and physiological characteristics of PA and sleep, such as total daily time spent in sedentary behavior or total sleep time [13]. PA monitors have wide applicability across a broad range of diseases including cardiovascular, respiratory, and metabolic indications.

CGMs are wearable devices that measure interstitial glucose levels continuously throughout the day, with some monitors taking measurements as often as every 5 min [14]. The time series nature of CGM data provides a characterization of the temporal (i.e., changes over time) glucose profile. Thus, the use of CGMs has become more popular in clinical practice, specifically in diabetes, since they provide instantaneous response to therapy decision, lifestyle modifications, and identification of patterns of hypoglycemia [15]. While glycosylated hemoglobin (HbA1c) is considered the key surrogate marker for the progression of long-term complications in diabetes, it does not provide information about acute glycemic changes (blood glucose variability) that reflect complications such as hypo- and hyperglycemia. It has been established that 14 days of CGM data provide a good estimate of glucose metrics for a 3-month period [16]. A list of core CGM metrics has now been recommended for use in clinical practice based on the consensus opinion of an international expert group. These include mean glucose level, predicted HbA1c, and the percentage of time glucose values are contained within a defined range (time in range, TIR) [17].

There are similarities in the time series nature of the data collected by CGM and actigraphy devices. Typically, for both CGM and actigraphy, a hierarchical structure of data types can be obtained and summarized by:

1. **Raw signal**. This is the high throughput, high-frequency, and minimally processed data directly from the sensor. Specifically, for tri-axial accelerometry signal, it is the acceleration (in the unit of g) at sub-second level (e.g. 30 Hz to 100 Hz) that can be measured along the x, y, and z axes.
2. **Epoch-level data**. For actigraphy data, raw acceleration can be summarized into epoch-level (e.g., 1 min) time series, such as activity counts. For CGM, glucose levels can be recorded every 5 min or 15 min. Epoch-level data can be used to derive the temporal trends of physiological behaviors such as diurnal patterns. One-minute epoch counts for PA data and 5-min epoch samples for CGM data represent 1440 and up to 288 daily data points for PA and CGM data respectively.
3. **Daily summaries**. From these epoch-level data, actigraphy summaries such as daily time spent in sedentary behavior, measures of activity quantity and intensity, and basic sleep parameters such as total sleep time, sleep efficiency and sleep disturbance can be derived [13]. CGM epoch-level data can be summarized to obtain mean glucose, diurnal glycemic variability, predicted HbA1c, and TIR measures [17].

In this manuscript, we will focus on considerations for handling missing data related to epoch-level data and daily summaries assuming multiple days/weeks of measurement. Though these DHTs are designed to minimize burden for the participant, missing data remains an important consideration. In each case, data may be missing for a variety of reasons including device malfunction, data transfer errors, data artifacts leading to inestimable outcomes measures, and compliance issues leading to periods of non-wear during the monitoring period.

3. Common practices for missing data in DHTs

In general, common approaches to deal with missing data in DHTs focus on the exclusion of data where the quantity of data collected is considered insufficient to reliably estimate the required outcome measures.

3.1. Common approaches to handle missing data in actigraphy-measured PA

Published studies on PA using actigraphy rarely report detailed methodology for handling missing data [18]. Commonly used approaches tend to focus on the concepts of "valid days" and "numbers of valid days." A "valid day" represents a 24-h period within which the number of minutes of data recording exceeds a given threshold. The "number of valid days" is a measurement interval that contains at least a defined number of valid days. For example, patients may be asked to wear a PA device from awakening to going to bed on each day within a 7-day interval; however, researchers may consider the data valid for estimation of PA if the device was worn for at least 10 h on at least 5 of these days (for example). Definitions of valid day and number of valid days needed for robust estimation of PA outcome measures vary in the literature.

Herrmann et al. [19] used datasets from the National Health and Nutrition Examination Survey (NHANES) [20] to conclude that using 12 h or less of wear data significantly underestimated time spent in activity

and sedentary behavior. Hart et al. [21] estimated that 3–4 days and 5 days of measurement in older adults is sufficient to enable representative measurement of PA and sedentary behavior respectively within a study period of 21 days. In their review of activity monitoring in chronic obstructive pulmonary disease (COPD), Byrom and Rowe recommended 5 or more days of at least 10 h recording each day to estimate PA and sedentary behavior outcome measures in COPD patients [18].

While robust estimation of PA outcome measures is important, discarding data that does not meet the valid days definitions is also undesirable. Catellier et al. [22] illustrated this dilemma, showing in their dataset that inclusion of all data produced lower estimates of PA due to bias from some days of limited wear time, whereas excluding data based on different valid day rules may introduce bias due to differences in activity between valid and invalid days.

In addition, while the importance of measuring across the majority of the wakeful day in estimation of parameters such as total PA (e.g., total counts per day) or total time spent in different levels of PA (e.g., sedentary, light, moderate to vigorous (MVPA) activity) is acknowledged, not all outcome measures reflect total daily measures. Cadence or real-world walking speed measures may only require estimation during periods of purposeful walking and so measurement across a small number of walking episodes per day may suffice for reliable estimation of these measures.

3.2. Common approaches to missing data in CGM

One of the most popular CGM-derived metrics is the time in range (TIR) [23]. TIR is a summary statistic that describes the amount of time that the patient spends with blood sugar levels between 70 and 180 mg/dL. It was proposed by Xing et al. [24] to require CGM data from at least 70% of monitoring days in order to have a valid TIR observation. Similarly, the International Consensus on Use of CGM recommends using a minimum of 14 consecutive days of data with approximately 70% of possible CGM readings over those 14 days for valid assessment of glycemic control using CGM data [25].

At this point, there are no perfect robust methods of dealing with missing CGM data. Often, in time intervals when the proportion of non-missing blood glucose measurements falls below 70%, the data for those days are discarded. Alternatively, missing blood glucose data are addressed using a linear interpolation process [26]. However, limitations of this approach are seen when the missing observations span a large time interval, or when the rate of blood glucose change is not linear. It is an open question as to whether the data on days with sufficient measurements can be considered representative of the data on the days with a large amount of missed measurements.

4. Common missing data methodologies and their limitations

4.1. Missing data mechanisms in data collected from clinical trials

Missing data are ubiquitous in almost all clinical trials which are used to evaluate a causal link between treatment and disease [27,28]. Conclusions drawn from clinical trials with missing data can vary depending on the assumptions made and the analytical method chosen, and there is no universal method to analyze data with missingness [29]. Therefore, regulatory authorities and industry groups have been developing standards. For example, the European Medicines Agency (EMA) provided their “Guideline on Missing Data in Confirmatory Clinical Trials” [30]. Funded by FDA, the US National Research Council (NRC) released guidelines on the “Handling of Missing Data in Clinical Trials” [31]. In addition, the International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) released the E9R1 Addendum on “Estimands and Sensitivity Analysis in Clinical Trials” [32].

Since this has been covered by many publications already, here, we will only briefly review the suggested treatment of missing data in clinical trials. The first step is always to evaluate the missing data mechanism to make a plausible assumption. As initially proposed by Little and Rubin, there are commonly three missing mechanisms, including: 1) Missing completely at random (MCAR), where the likelihood of missing data is unrelated to any observed or unobserved variables; 2) Missing at random (MAR), where the likelihood of missing data is related to observed variables but not to unobserved variables; and 3) Missing not at random (MNAR), where the likelihood of missing data depends on unobserved data [33]. For clinical trials, depending on the assumptions made, a wide range of approaches can be applied, such as complete case analyses, single imputation (mean or conditional mean), last observation carried forward (LOCF, now considered an outdated approach that in most cases does not adequately reflect reality), likelihood-based analysis, and multiple imputation, which has been widely studied and compared [27,34,35].

4.2. Missing data associated with DHT applications

The more complex data structure and more frequently acquired measurements bring additional technical difficulty in treating missing data from DHTs such as actigraphy and CGMs. However, the principles of missing data mechanisms described above should still be applied in general. Table 1 provides some simple examples of causes of missing data from DHTs for each of the three mechanisms (MCAR, MAR, and MNAR).

It is not immediately obvious how commonly used methods can be applied to treat missing data in these continuous measurements. For example, if there is an interval of missing epoch-level activity counts data due to non-wear, it is not straightforward to consider methods like multiple imputation due to the continuous nature of the data, and missing imputation for densely measured time series is not common practice.

Moreover, it is not entirely intuitive how these mechanisms can be applied to the more complicated scenarios where high-frequency data is collected in unsupervised free-living environments where more than one of the missing mechanisms may exist. Therefore, one must be extremely cautious when making assumptions on the reasons for missingness as

Table 1
Missing data mechanisms and examples in DHTs applications.

Mechanism	Definition	Examples
MCAR	Likelihood of missing data is unrelated to any observed or unobserved variables, i.e., the missingness is unrelated to the specific participants being studied.	Missing data due to device malfunction or data transfer error.
MAR	Likelihood of missing data is related to observed variables but not to unobserved variables, i.e. the missingness is related to the participant but can be predicted from other data known about that participant.	Data loss is more likely for women compared to men because of the form factor of the DHT (e.g., size, positioning).
MNAR	Likelihood of missing data depends on the unobserved data, i.e., there is a specific reason (related to the outcome measure) that is unobserved for the missingness.	Patient fails to wear the device during times of severe symptoms (e.g., during hospitalization) A Parkinson’s disease patient stays indoors and removes the DHT while practicing pastimes during times when motor symptoms, such as tremor and bradykinesia, are at their lowest. Participants are more likely to wear PA trackers on active days compared to inactive days

MCAR: Missing completely at random; MAR: missing at random; MNAR: missing not at random.

there are a multiplicity of factors and intercurrent events that can affect and confound the patterns of missing data, including but not limited to: participants' daily schedules, employment status, DHT usability/acceptability (e.g., attitude towards charging and wearing devices), disease stage and state, symptoms severity, health and device literacy of participants, and seasonality effects/weather patterns.

5. Emerging methodologies to deal with missing data in DHTs

In this section, we review and propose potential remedies to handle missingness in both continuous epoch-level data and daily summary data based on their characteristics.

5.1. Continuous epoch-level data

Study protocols with DHTs may require participants to wear the devices continuously (unless, for example, they must remove it to shower or to charge the device) to gain comprehensive knowledge about their physical behaviors or physiological patterns while considering the time-of-day effect. A good example is the physical activity monitoring part of the NHANES 2011–2012 [36]. Epoch-level data can be used to extract temporal patterns such as circadian rhythmicity [37,38] and is the foundation for deriving commonly used summaries of sleep and PA [13].

Missingness in epoch level data can occur when there is device malfunction, a participant forgets to wear the device, or intentionally removes the device (e.g., to charge it, shower, or swim). Imputing epoch-level data makes sense only when part of the day is missing: if there are no data collected at all or if the vast majority of data is missing for a specific day, it is not appropriate to impute missing epochs for that day. Therefore, a typical convention is to first decide the amount of data required for inclusion of that day and apply an imputation mechanism. This is similar to the concept of defining a valid day as described above (e.g., at least 10 h of wear time during the day is required for that day to be considered in the analysis of PA) [39,40]. Within such a valid day, missing data at specific times can be handled by the following approaches.

5.1.1. Median/mean imputation based on other days with available data

Protocols that deploy continuous wearable sensors in free living environments typically include multiple days of measurement (e.g., most commonly used is 7–14 days around certain clinic visits). If the interest of the study is to explore epoch-level time series data across multiple days, a median/mean imputation can be adopted to impute missing epochs based on other days with available data. In this case, for participants with enough valid wear days, if a certain time period of data is missing for some but not all days, we can use the median/mean from the same time period across the other valid days when there is no missingness to impute the data for the missing period, assuming a constant daily routine. The choice between mean and median value depends on the distribution of the measurement of interest.

For example, in the activity monitoring portion of the Baltimore Longitudinal Study of Aging (BLSA) [41], the adopted data processing procedure was as follows. The first step was to remove days with more than 5% of data missing; then for each of the remaining days within each participant, a particular missing minute was imputed as the average activity counts of the same minute over all the remaining days with available data for that minute interval [42–44]. This approach has been shown to be fairly robust to different modeling approaches by sensitivity analyses [13,44].

This approach is more applicable to the MCAR and MAR mechanisms. If data are MNAR, for example, a participant always takes off the device during the same time period across all recorded days due to disease-related symptoms or to take medical treatments, this approach will provide biased estimates. This method does not recognize that the imputed values are not known, hence it underestimates the standard errors.

5.1.2. Novel functional data analysis approaches

Most common and standard statistical methods are typically designed for multivariate data. For instance, regression analysis is used to assess the association between the response variable and independent variables; while principal component analysis (PCA) is used to reduce the dimension of multivariate data whilst retaining most of the information through the covariance of all the variables [45]. Since epoch-level data generated by DHTs are, by nature, functions of time as opposed to scalar variables, these methods cannot be directly employed. Therefore, statistical methodologies other than multivariate approaches should be considered for modeling epoch-level data and addressing their missingness.

Various techniques have been developed to address missing values in time series data such as smoothing, interpolation using splines or other nonparametric approaches, and kernel methods [46–48]. The core idea is to use the underlining functional characteristics such as smoothness. Functional data analysis is a set of statistical methodologies to study smooth functional (i.e., as a function of time or location) behaviors of curves over a continuum [49,50]. Epoch-level data measured by DHTs can be considered as a function of time and thus suitable for functional data analysis approaches.

A wide variety of functional principal component analyses (FPCAs) or functional regressions (scalar-on-function, function-on-scalar, or function-on-function) have been applied to data collected by CGM and actigraphy [38,51–53]. FPCA has extended PCA by exploring the covariance of functional observations at different time points, while functional regression has extended regular regression by replacing scalar response or independent variables as functions of time. For example, Gaynaova et al. introduced multilevel functional regression to quantify the blood glucose levels measured by CGM for multiple days [51]. Goldsmith et al. explored associations between covariates and diurnal profiles of actigraphy-measured using function-on-scalar regression [53] with functional activity profile as the response. Among others, FPCAs have been used extensively to extract nonparametric measures of circadian rhythmicity which have been linked to multiple health outcomes [38,52]. Due to the growing acceptance and utility of FPCA, we will focus on explaining how it can be used to address missing data.

Let us first review the basic concepts of FPCA. For a function $y_i(t)$ from observation i (e.g., the average epoch-level activity profile across all days for participant i) at time t (time of the day, $t = 1$ to 1440 min), FPCA can be formulated and represented in a lower dimensional space (with dimension K) as

$$y_i(t) = f_0(t) + f_i(t) + \epsilon_i(t) \approx f_0(t) + \sum_{k=1}^{K} \phi_k(t)\xi_{ik} + \epsilon_i(t)$$

In this model, $f_0(t)$ is the fixed mean effect (intercept), $\epsilon_i(t)$ is the random error, and $f_i(t)$ is the participant specific functional effect. $f_i(t)$ can be represented by the sum of product of eigenfunction $\phi_k(t)$ and principal component (PC) score ξ_{ik} over number of principal components (PCs) k. The eigenfunction $\phi_k(t)$ can be obtained by spectral decomposition of the covariance function $Cov(y_i(t), y_i(s))$, which is equivalent to eigenvectors in a multivariate PCA. It represents the kth functional principal component (FPC) capturing a specific dominant feature or mode of variation. The PC score, ξ_{ik}, can be obtained by projecting the function onto the corresponding eigenfunction which are uncorrelated random variables.

In the example of epoch-level actigraphy measurement, each of $\phi_k(t)$ represents a specific data-driven diurnal pattern and the corresponding score ξ_{ik} can be directly modelled [13,52]. Since all K PC scores are uncorrelated by design, one can include all of them as predictors in a statistical model. A natural extension to this method is to consider multiple days of measurement ($y_{ij}(t)$, where j represents the number of days). This generates a multilevel FPCA (MFPCA) based on a two-way

ANOVA model, which separates the main effect into inter-participants effect and intra-participant effect, and decomposes them separately based on Karhunen-Loeve (KL) expansion [54,55]. We will not expand on the details here but, from an application point of view, the obtained eigenfunctions and scores representing the inter-participant effects have similar interpretation to FPCA, and the obtained eigenfunctions and scores for intra-participant effect can represent patterns for within-participant day-to-day variation.

These frameworks were originally based on densely sampled functional data (e.g., for minute level actigraphy data, all participants have 1440 min of data observed in each day). There are various extensions of those frameworks to incorporate irregularly or sparsely measured functional data. In fact, sparse FPCA or sparse MFPCA has been widely used to analyze sparse functional or longitudinal data [56–60].

Therefore, epoch-level data from DHTs with missingness can be treated as sparse functional data, and sparse FPCA or sparse MFPCA can be directly applied to acquire nonparametric circadian rhythm measures as used for dense data without missingness.

On the other hand, if the goal is to fully view (or estimate) a specific daily activity trajectory profile for a participant with some missing epoch level data, sparse FPCA also provides a way to predict and impute those time periods with missing values. For example, Grigsby et al. used FPCA to predict individual trajectories of child growth from sparse data [61]. The intuition for such an approach is that, since the estimation of FPCA is based on the decomposing sample covariance function from all observations, to impute missingness at a specific time period for one observation, we can always "borrow" information from other observations.

Even though, to our knowledge, sparse FPCA approaches have not yet been adopted to address missingness in epoch-level data from DHTs, given their flexibility and ease of implementation, they have good potential and deserve further exploration. Fortunately, tools to deal with functional and sparse functional data with efficient computational capability are already available in well-established statistical analysis software. For example, popular R packages for functional data include "refund" [62], "fda" [63], and "mgcv" [64].

5.1.3. Advanced deep learning methods

Recent advancements in deep learning architectures have been applied to continuous data collected by DHTs (both epoch-level and raw data), particularly for prediction problems [65]. For instance, a wide variety of deep learning algorithms have been used for sleep detection from wrist-worn actigraphy [66–68], which is of significant clinical interest for numerous therapeutic areas. Most published works are based on recurrent neural networks (RNN) or convolutional neural networks (CNN) that directly take time series data instead of inputting hand-crafted features. RNN has been used to study sequential or temporal data, making continuous DHT measurements a suitable application area [69]. CNN, commonly used for computer vision and imaging processing, has also been shown to be able to design optimal time invariant local feature extractors from input data, which can be used in DHT measurements as well [70].

As suggested by Chollet, who developed one of the most popular deep learning libraries,"Keras" [71], deep neural networks trained on large samples can handle missing data [72]. The easiest approach is to input missing values as something that is meaningless in the data (e.g. -9999, if it is not already meaningful), and the network will learn from the exposure that this value represents missingness. Moreover, advanced extensions of deep learning architectures have been developed that allow missing values with different missing patterns into the models [46,73–75]. There is no doubt that, with the amount of available data growing rapidly, these approaches will become more streamlined and generate standard approaches to deal with real-world DHT measurements.

5.2. Daily summary data

From the 24-h epoch-level data, a number of daily summaries can be derived. For example, as previously mentioned, actigraphy provides summaries such as daily time spent in sedentary behavior and total night-time sleep minutes, and CGM provides summaries such as daily mean glucose, estimated HbA1C, TIR, and glycemic variability. Even though these daily summary data are not as frequently measured as epoch-level data (e.g., day vs. minute), they still need to be carefully dealt with and their missingness properly addressed. Traditional clinical endpoints are normally measured once at each clinic visit. For studies that incorporate DHTs, participants are typically asked to wear the devices/sensors for multiple days (most commonly 7 days or 14 days) around specific clinic visits, yielding repeated measurements around each visit. Summarizing DHT data on the daily level is natural as participants are typically instructed to wear the device over the course of a day [76].

Here, we focus on the situation where one or multiple days of daily summaries are missing. For example, in the week around a visit, only summaries of 3 days are available out of 7. This can be caused by a device malfunction in the middle of the week, participants not wearing the device long enough (e.g., less than 10 h) for that day to be considered valid, or participants not wearing the devices at all for those days (as opposed to temporally taking off the device due to charging or showering) due to vacation or hospitalization. Similar to deciding on the definition of a valid day from epoch-level data, for daily summaries data, the first step typically involves determining the minimum number of days required to consider that week as valid and include it in the analysis. For example, one can define 3 out of 7 days of valid or complete days of summaries as the cut-off for a valid week. If the number of valid days is below this threshold, the data for that participant for that week is considered insufficient, no imputation is carried out, and the whole 7-day-data for that participant will be removed due to low data quality. A review of these definitions can be seen in a recent publication [77]. For this type of situation, we believe the following procedures can be further considered to deal with missing data.

5.2.1. Robust data processing and feature engineering

Even for the daily summary data, we typically have more repeated measurements than traditional clinical measurements. Therefore, the richness of the data can help us address the missingness. In a clinical study, as we mentioned before, we have multiple days of DHT data for a clinic visit. As such, sometimes it is not necessary for us to investigate each day separately. One common approach is to take the median/mean or other statistics across all available days within the monitoring period. This simple but effective method reduces the day-to-day variability which is known to exist in DHT measurements [78]. Some extensions of this approach include but are not limited to: 1) using a multi-day mean/median for consecutive days (e.g., model a 3-day average across 14 days), 2) taking the mean for weekday and weekend separately [79], 3) taking the weighted mean based on the amount of missingness (e.g., days providing summary values based on less missing epoch data should be weighted more), and 4) while assuming exchangeability across adjacent weeks, substituting a missing day with the daily summary from the same day of the following week which is known as day-substitution (e.g. a missing Tuesday can be substituted by the following Tuesday if it is observed) [76]. One should keep in mind that such simple imputation typically results in an underestimation of the variance of the estimates.

Another recommendation goes back to the way that the summaries are derived, which is sometimes referred to as feature engineering. In this approach, it is considered more robust to work with normalized daily summaries, as these are less sensitive to the amount of wear time. For example, for actigraphy-measured time spent in different activity levels (e.g., sedentary, MVPA), instead of working with absolute values in minutes, one can consider the proportion or percentage of the total wear time spent in that activity level [80]. Similarly, other than daily

time spent in sedentary or active states, several studies suggested exploring active to sedentary transition probability (ASTP) as a normalized value (bounded in 0 and 1) regardless of daily wear time and revealed its relationship with various health outcomes and functional measurements [43,44,52,81,82]. By working with robust normalized summaries like these, it is possible to relax the definition of a valid day based on the amount of wear time, and thus potentially reduce the amount of missing data. Therefore, it is of great interest for researchers across multiple scientific domains such as statistics, engineering, computer science, and epidemiology to collaborate on developing more of these novel summary measures.

5.2.2. Robust modeling schemes

Even if the goal is to model multiple days of data, one can consider robust modeling schemes. The mixed models with repeated measures (MMRM) approach does not enforce any formal imputation and aims to estimate the mean treatment effect with all available data (some of them can be incomplete) while considering participant-specific effects and within-participant correlation. This approach assumes data are MAR (which is not likely to be always true in reality) and that dropouts would behave similarly to other patients within the same treatment group [83]. For this reason, this approach has gained more popularity than single imputation approaches and has been used to account for missing data in clinical trials [84].

Another maximum likelihood based missing data method is the Expectation-Maximization (EM) algorithm when the joint distribution of the missing data and the observed data is explicit. [85,86]. Under the MAR assumption, the EM algorithm, which entails an iterative procedure, does not complete the missing data but estimates the parameters directly by maximizing the complete data log likelihood [85]. Starting with an initial guess of the parameter value, in each iteration, the EM algorithm begins with an E step to compute the expectation of complete data log likelihood with respect to the conditional distribution of the missing data and ends with an M step that determines the value of the parameter to be estimated by maximizing the complete data log likelihood. This method has been used to impute missing data in daily activity metrics measured by accelerometers [22]. However, since the EM algorithm does not involve taking the derivative of the log likelihood, it cannot directly provide uncertainty measurements for the estimation procedure. Therefore, it is not as popular as MMRM in clinical studies, where statistical tests and inferences for the estimated parameters are the primary goals [85].

5.2.3. Missing data imputation

DHT-measured daily summaries share some of the similarities to the traditional repeated measurements from multiple visits. Data imputation, which is recommended for conventional clinical measurements can thus be applied here as well. Typically, single imputation methods such as mean or conditional mean imputation, and last observation carried forward (LOCF) which assumes that the measurement remains constant after dropout, are not appropriate because they tend to underestimate the variance and can sometimes produce biased estimations [34].

Multiple imputation (MI) is the most commonly used technique to deal with complex incomplete data [87]. Multivariate imputation with chained equation (MICE) is a particular MI approach under the assumption that, conditioning on the variables used in the imputation procedure [88], the missing data are MAR, and uses Markov Chain Monte Carlo (MCMC) methods to draw imputations over the conditional densities [89]. A typical MI has three steps: 1) impute the missing observations using their regression model-based predictive distribution based on observed values, 2) repeat the first step to create multiple imputed datasets, and analyze each of them using some statistical model to provide the estimation of interest, and 3) pool the results of interest from all the datasets together and derive their standard errors [33]. By doing so, MI can preserve the relationship between variables in the data while also simultaneously accounting for the uncertainty about these relations [89].

One advantage of using MI is to possibly incorporate auxiliary variables (variables that are predictive of missingness of the outcome and also of the value of the outcome) in the imputation model in addition to the variables in the primary analysis model, which may make the MAR assumption more plausible [76]. Examples of such auxiliary variables include Body Mass Index (BMI) and weather data (e.g. temperature, rainfall, sunshine, length of day) which have known association with human behaviors such as PA [76,90,91].

Even though the underlying theory is complex, MI has still been widely used in clinical trials as it has been shown to produce valid statistical inference while incorporating uncertainty due to missing data. The MI methodology can be implemented easily with popular statistical software such as the "mice" [89] and "Amelia" [92] R packages, and the SAS "proc mi" procedure [35]. It has been discussed that "Amelia" is more convenient in handling missingness in time series data (e.g. day-to-day summaries) since it contains elements such as polynomial time trends or lagged variables, as compared to "mice" where users need to construct such elements. "Amelia" has limited capacity for handling non-normal variables [93]. Ji et al. conducted a simulation study and demonstrated the advantages of MI in handling missing data in longitudinal studies for the following reasons: 1) it preserves the original observed time intervals which is more suitable for time series data and model; 2) it takes into account the uncertainty of the imputed values; and 3) most importantly, statistical packages such as "mice" and "Amelia" are highly flexible in incorporating various imputation models, accommodating different data types, and having no constraints on number of measurement time points [93].

5.2.4. Treat partially observed days as censored data

As we have discussed previously, one common practice of defining a day with missing summary relies on the definition of valid days, which is based on thresholding wearing time. Using actigraphy measured daily step counts as an example, Tackney et al. proposed a new framework for handling missing accelerometry data [76]. Instead of discarding days with wear time lower than the pre-defined thresholds and treating the day as missing, one can additionally consider partially observed days. This will leave us with three types of data, 1) when wear time is greater than the threshold, we have a day with fully observed value, 2) when wear time is between zero and the threshold, we have a day with partially observed value, and 3) when device is not worn at all, we have a day with missing value. The advantage of having partially observed values is to retain more information than discarding it completely. Intuitively, partially observed data can be regarded as right-censored data since the true daily summary value is higher than what is observed due to insufficient amount of wear time, and the partially observed value serves as the lower bound for that day. Similarly, the lower bound will be zero for those with completely missing daily summaries. Imputation can be done via Tobit regression which is a method for estimating linear relationship between variables when the outcome is censored [94].

All these considerations to handle missingness for both epoch-level (as described in Section 5.1) and daily summaries (as described in Section 5.2) are summarized in Fig. 1.

5.3. Other considerations

The analysis procedure to address missingness in DHT data involves multiple decision-making steps and parameters, such as: determining minimum amount of daily wear time and minimum number of valid days, deciding whether mean or median should be used to impute missing epoch-level data or combining daily summaries around a clinic visit, and controlling the level of smoothness and amount of variation retained for the FPCA approaches. It is wise to always pre-specify additional sensitivity analyses to test the reliability of the results, as suggested for clinical trials in general [28,29,83]. Since most of the

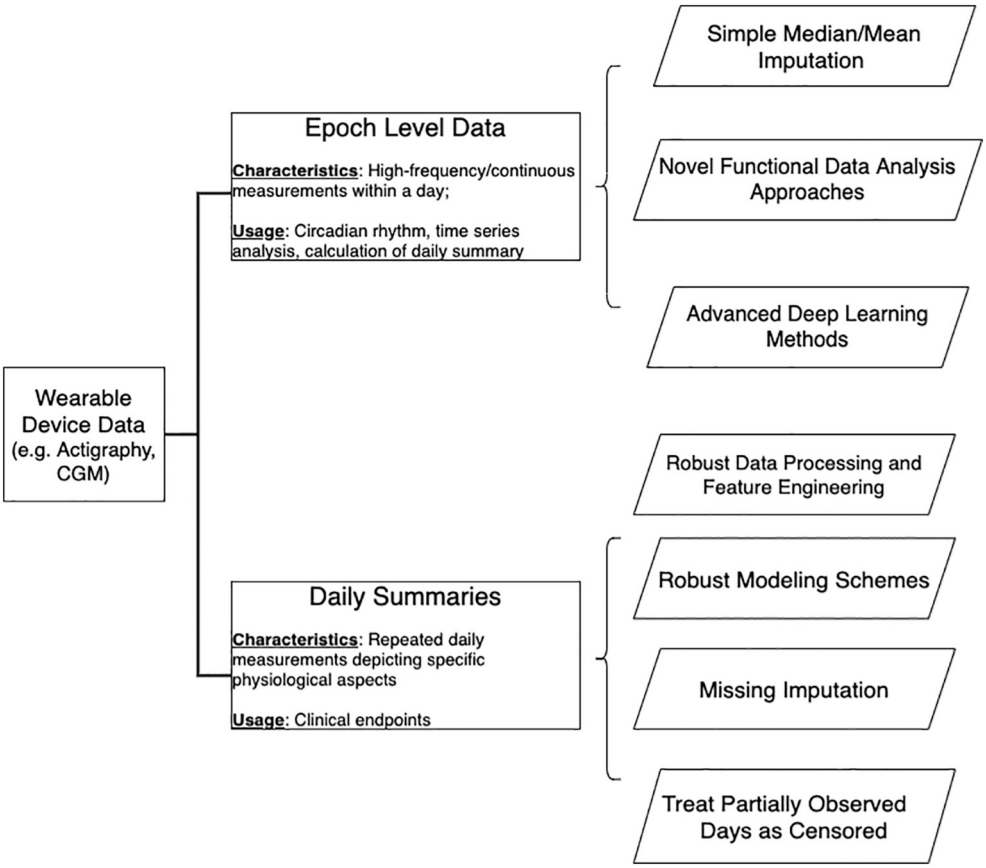


Fig. 1. Analytical framework to address missingness in DHT data.

robust modeling and imputation approaches depend on the MAR assumption, sensitivity analyses can always be performed to examine the effects of different assumptions [34,95]. It also aligns with the ICH's "estimands framework" to properly conduct sensitivity analysis and consider the intercurrent events [32].

Moreover, technological advancement has enabled data from multiple devices or from multiple sensors within the same device to be acquired simultaneously (e.g. actigraphy + CGM, actigraphy + heart rate) [96]. Although multi-modal sensor integration is not yet widely deployed and well-studied, given the co-dependency of different human physiological behaviors such as the association between activity and glucose levels, and activity and heart rate, it is possible to gain insight from one sensor to inform reasons for missing data from the other.

6. Prevention of missing data from DHTs

The best and perhaps the easiest treatment of missing data is always its prevention [97]. This means that researchers should always deploy DHTs thoughtfully to minimize missing data. For clinical studies with DHTs, we provide the following considerations.

6.1. Configuration of the device

As an example, consider the accelerometer where most research grade devices allow researchers to configure parameters that control the data collection mechanism, such as the sampling frequency (typical range: 30–100 Hz). Higher sampling rates can detect behaviors with more granularity, but also impose greater demands on the device's battery and storage capacity, which may introduce greater risk of data loss. It is important for researchers to determine the optimal sampling rate depending on the concept of interest to be measured.

6.2. Device placement location

The wear location of the device depends on the measurements to be taken. A lumbar-worn accelerometer is used to reliably measure gait, while wrist-worn devices are typically accurate enough for quantifying PA in general and number of steps per day [98], although they typically overestimate PA and underestimate sedentary time [99] . However, wrist worn devices offer better compliance compared with lumbar-worn devices due to ease and comfort of wear resulting in lower patient burden [100–103]. Therefore, whenever possible, it is always a good strategy to opt for the location that is less cumbersome to patients to increase compliance.

6.3. Collection of additional information for contextual information

Given that DHTs are more commonly used in the free-living and unsupervised environments, to safeguard against or at least mitigate data missingness, we can consider the use of qualitative data such as electronic patient diaries collected using a smart phone application on a regular basis to acquire contextual information such as intercurrent events. For example, a participant could indicate in the diary that she/he could not wear the device due to hospitalization or severe symptoms, or alternatively due to bad weather. Meanwhile, information such as weather can be collected or additionally acquired outside the scope of the study to serve as auxiliary variables for imputation model. This additional information can be extremely helpful to understand and classify the missing data mechanisms.

6.4. Incorporating the patient perspective and optimizing DHT deployment

At the design phase of the study, it is essential to capture the voice of the patients and their caregivers to help develop and apply strategies for

DHT deployment that reduce burden, encourage participation, and improve compliance. Understanding the preferred form factors, perceived usability and comfort of the DHTs in targeted forums such as patient panels can help reduce data loss and increase DHTs' benefit to patients. During deployment, data loss can be further reduced by: (a) having a comprehensive patient and site personnel training on the use of the DHTs; (b) offering patients clear audio-visual instructions related to the DHT use (such as charging); (c) having a support system (e.g., 24/7 support [104]) that patients and caregivers can reach out to in case of technical difficulties; and (d) developing strategies to efficiently ship back-up devices when device malfunction happens.

7. Discussion

In recent years, several studies have demonstrated the potential of DHTs to impact decision making in drug development [105]. In leveraging these technologies, a number of methodological gaps in the current knowledge have emerged. Industry consensus groups have sought to address these gaps including defining the process and supporting evidence to determine a "fit-for-purpose" DHT with appropriate measurement properties to support regulatory decision making [7,8] and the validation of new digital endpoints derived from their data [106,107]. However, a gap area that has not been extensively reviewed and discussed is addressing the practical challenges associated with missing data arising from the use of these technologies in free-living environments. Dealing with missing data is a vital consideration for clinical research, as different missing data approaches are typically based on various assumptions that, if violated, can lead to biased estimates of treatment effects, and impact the reliability of study results. Therefore, understanding the sources and nature of missing data (MAR, MNAR, MCAR) is important to the interpretation of study outcomes, and understanding thresholds above which missing data degrades the device/algorithms' ability to accurately capture the desired input is critical to evaluating the integrity of DHT-based study results.

In this article, we discussed how missing data from DHTs fit into the classic missing data mechanism (MCAR, MAR, and MNAR) and provided relevant examples. Since it is not directly intuitive how commonly used statistical methods for handling missing data are directly applicable to complex and continuously measured data collected by DHTs, we then presented a review of statistical and data processing methods that show promise in dealing with missingness in the analysis of both epoch-level and daily summary data. We specifically highlight the use of functional data analysis approaches for epoch-level data and robust feature engineering approaches to generate daily summary data that are robust to missingness as new ideas to be exercised in future studies. Finally, we provided considerations on preventing the collection of missing data from DHTs from study-design and device deployment points of view.

We believe the approaches discussed in this article are well suited to inspire further methodological innovation and applications to correctly handle DHT missing data, enabling the derivation of reliable and clinically meaningful digital endpoints for use in clinical trials. This paper presents a valuable reference for researchers in providing thoughtful and appropriate approaches to address missingness. Although in this article we focused mainly on technical considerations, the content can benefit other clinical trial stakeholders such as payers as well as journal editors/reviewers and commercial organizations.

Since clinical trials deploying DHTs are still not considered common practice, there is still plenty of room for improvement. First, from a study design and operation point of view, it is critical for clinical study teams, clinical sites, and healthcare providers to carefully summarize lessons learned from previous studies in deploying DHTs to optimize study design, increase data quality and reduce missingness. In this article, we provided the most important considerations to prevent missing data such as device location and device configuration. Future work should aim to develop systematic standard operating procedures for deploying DHTs by learning from previous studies and hearing the patients' perspectives. Secondly, from data analysis and data optimization point of view, it is up to the statisticians, data scientists, and other researchers to start adopting the emerging technical approaches discussed in this article (and others that are not). By doing so, more researchers understand the essence of these approaches, get more familiar with them, and think about their practical limitations, and gradually develop more suitable methodologies. What we have discussed, by no means, should be considered as the final technical solution as some of these solutions are still evolving. It is only when the field applies them, we can start to build a systematic and unified protocol to deal with missing data collected by continuous DHT monitoring. We hope this will inspire development of standards that could be applied in future clinical trials.

Disclaimer

This manuscript originated out of the Drug Information Association (DIA) digital endpoints working group. The manuscript's content is solely the responsibility of the authors and does not necessarily represent the official views of the Food and Drug Administration, or the United States Department of Health and Human Services.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Junrui Di, Charmaine Demanuele, F. Isik Karahanoglu , and Joseph C. Cappelleri are all employees and stockholders of Pfizer Inc. Denise P. Bury is an employee of Sanofi.

References

[1] E. Volgels, About one-in-five americans use a smart watch or fitness tracker, Pew Res. Cent. (2020). https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ (accessed August 16, 2021).
[2] Food and Drug Administration, What is Digital Health. https://www.fda.gov/medical-devices/digital-health-center-excellence/what-digital-health, 2020 (accessed August 16, 2021).
[3] Food and Drug Administration, National Instiution of Health, BEST (Biomarkers, EndpointS, and other Tools) Resource. https://www.ncbi.nlm.nih.gov/books/NBK338448/, 2016 (accessed August 16, 2021).
[4] N. Mahadevan, C. Demanuele, H. Zhang, D. Volfson, B. Ho, M.K. Erb, S. Patel, Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device, Npj Digit. Med. 3 (2020) 5, https://doi.org/10.1038/s41746-019-0217-7.
[5] M.P. Smith, K. Ly, Q. Thibodeaux, T. Weerasinghe, J.J. Wu, G. Yosipovitch, T. Bhutani, W. Liao, Emerging methods to objectively assess pruritus in atopic dermatitis, Dermatol. Ther. (Heidelb). 9 (2019) 407–420, https://doi.org/10.1007/s13555-019-0312-3.
[6] N. Mahadevan, Y. Christakis, J. Di, J. Bruno, Y. Zhang, E.R. Dorsey, W.R. Pigeon, L.A. Beck, K. Thomas, Y. Liu, M. Wicker, C. Brooks, N.S. Kabiri, J. Bhangu, C. Northcott, S. Patel, Development of digital measures for nighttime scratch and sleep using wrist-worn wearable devices, Npj Digit. Med. 4 (2021) 42, https://doi.org/10.1038/s41746-021-00402-x.
[7] M.K. Walton, J.C. Cappelleri, B. Byrom, J.C. Goldsack, S. Eremenco, D. Harris, E. Potero, N. Patel, E. Flood, M. Daumer, Considerations for development of an evidence dossier to support the use of mobile sensor technology for clinical outcome assessments in clinical trials, Contemp. Clin. Trials. 91 (2020), 105962, https://doi.org/10.1016/j.cct.2020.105962.
[8] Digital Medicine Society, The Playbook: Digital Clinical Measures. https://playbook.dimesociety.org/, 2020 (accessed August 16, 2021).
[9] J. Bai, C. Di, L. Xiao, K.R. Evenson, A.Z. LaCroix, C.M. Crainiceanu, D.M. Buchner, An activity index for raw accelerometry data and its comparison with other activity metrics, PLoS One 11 (2016), e0160644, https://doi.org/10.1371/journal.pone.0160644.
[10] V.T. van Hees, L. Gorzelniak, E.C. Dean León, M. Eder, M. Pias, S. Taherian, U. Ekelund, F. Renström, P.W. Franks, A. Horsch, S. Brage, Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity, PLoS One 8 (2013), e61691, https://doi.org/10.1371/journal.pone.0061691.
[11] D. John, P. Freedson, ActiGraph and actical physical activity monitors: a peek under the hood, Med. Sci. Sports Exerc. 44 (2012) S86–S89, https://doi.org/10.1249/MSS.0b013e3182399f5e.
[12] M. Karas, J. Bai, M. Strączkiewicz, J. Harezlak, N.W. Glynn, T. Harris, V. Zipunnikov, C. Crainiceanu, J.K. Urbanek, Accelerometry data in health

research: challenges and opportunities, Stat. Biosci. 11 (2019) 210–237, https://doi.org/10.1007/s12561-018-9227-2.
[13] J. Di, A. Spira, J. Bai, J. Urbanek, A. Leroux, M. Wu, S. Resnick, E. Simonsick, L. Ferrucci, J. Schrack, V. Zipunnikov, Joint and individual representation of domains of physical activity, sleep, and circadian rhythmicity, Stat. Biosci. 11 (2019) 371–402, https://doi.org/10.1007/s12561-019-09236-4.
[14] D. Rodbard, Continuous glucose monitoring: a review of successes, challenges, and opportunities, Diabetes Technol. Ther. 18 (2016), https://doi.org/10.1089/dia.2015.0417. S2-3-S2-13.
[15] A.L. Carlson, D.M. Mullen, R.M. Bergenstal, Clinical use of continuous glucose monitoring in adults with type 2 diabetes, Diabetes Technol. Ther. 19 (2017), https://doi.org/10.1089/dia.2017.0024. S-4-S-11.
[16] T.D. Riddlesworth, R.W. Beck, R.L. Gal, C.G. Connor, R.M. Bergenstal, S. Lee, S. M. Willi, Optimal sampling duration for continuous glucose monitoring to determine long-term glycemic control, Diabetes Technol. Ther. 20 (2018) 314–316, https://doi.org/10.1089/dia.2017.0455.
[17] T. Battelino, T. Danne, R.M. Bergenstal, S.A. Amiel, R. Beck, T. Biester, E. Bosi, B. A. Buckingham, W.T. Cefalu, K.L. Close, C. Cobelli, E. Dassau, J.H. DeVries, K. C. Donaghue, K. Dovc, F.J. Doyle, S. Garg, G. Grunberger, S. Heller, L. Heinemann, I.B. Hirsch, R. Hovorka, W. Jia, O. Kordonouri, B. Kovatchev, A. Kowalski, L. Laffel, B. Levine, A. Mayorov, C. Mathieu, H.R. Murphy, R. Nimri, K. Nørgaard, C.G. Parkin, E. Renard, D. Rodbard, B. Saboo, D. Schatz, K. Stoner, T. Urakami, S.A. Weinzimer, M. Phillip, Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range, Diabetes Care 42 (2019) 1593–1603, https://doi.org/10.2337/dci19-0028.
[18] B. Byrom, D.A. Rowe, Measuring free-living physical activity in COPD patients: deriving methodology standards for clinical trials through a review of research studies, Contemp. Clin. Trials. 47 (2016) 172–184, https://doi.org/10.1016/j.cct.2016.01.006.
[19] S.D. Herrmann, T.V. Barreira, M. Kang, B.E. Ainsworth, Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach, Br. J. Sports Med. 48 (2014) 278–282, https://doi.org/10.1136/bjsports-2012-091410.
[20] Centers for Disease Control and Prevetion, National Health and Nutrition Examination Survey, (n.d.). https://www.cdc.gov/nchs/nhanes/ (accessed August 16, 2021).
[21] T.L. Hart, A.M. Swartz, S.E. Cashin, S.J. Strath, How many days of monitoring predict physical activity and sedentary behaviour in older adults? Int. J. Behav. Nutr. Phys. Act. 8 (2011) 62, https://doi.org/10.1186/1479-5868-8-62.
[22] D.J. Catellier, P.J. Hannan, D.M. Murray, C.L. Addy, T.L. Conway, S. Yang, J. C. Rice, Imputation of missing data when measuring physical activity by accelerometry, Med. Sci. Sports Exerc. 37 (2005) S555–S562, https://doi.org/10.1249/01.mss.0000185651.59486.4e.
[23] R.W. Beck, R.M. Bergenstal, T.D. Riddlesworth, C. Kollman, Z. Li, A.S. Brown, K. L. Close, Validation of time in range as an outcome measure for diabetes clinical trials, Diabetes Care 42 (2019) 400–405, https://doi.org/10.2337/dc18-1444.
[24] D. Xing, C. Kollman, R.W. Beck, W.V. Tamborlane, L. Laffel, B.A. Buckingham, D. M. Wilson, S. Weinzimer, R. Fiallo-Scharer, K.J., Ruedy for the juvenile diabetes res, optimal sampling intervals to assess long-term glycemic control using continuous glucose monitoring, Diabetes Technol. Ther. 13 (2011) 351–358, https://doi.org/10.1089/dia.2010.0156.
[25] T. Danne, R. Nimri, T. Battelino, R.M. Bergenstal, K.L. Close, J.H. DeVries, S. Garg, L. Heinemann, I. Hirsch, S.A. Amiel, R. Beck, E. Bosi, B. Buckingham, C. Cobelli, E. Dassau, F.J. Doyle, S. Heller, R. Hovorka, W. Jia, T. Jones, O. Kordonouri, B. Kovatchev, A. Kowalski, L. Laffel, D. Maahs, H.R. Murphy, K. Nørgaard, C.G. Parkin, E. Renard, B. Saboo, M. Scharf, W.V. Tamborlane, S. A. Weinzimer, M. Phillip, International consensus on use of continuous glucose monitoring, Diabetes Care 40 (2017) 1631–1640, https://doi.org/10.2337/dc17-1600.
[26] S.J. Fonda, D.G. Lewis, R.A. Vigersky, Minding the gaps in continuous glucose monitoring: a method to repair gaps to achieve more accurate glucometrics, J. Diabetes Sci. Technol. 7 (2013) 88–92, https://doi.org/10.1177/193229681300700110.
[27] J.D. Dziura, L.A. Post, Q. Zhao, Z. Fu, P. Peduzzi, Strategies for dealing with missing data in clinical trials: from design to analysis, Yale J. Biol. Med. 86 (2013) 343–358. http://www.ncbi.nlm.nih.gov/pubmed/24058309.
[28] C.M. DeSouza, A.T.R. Legedza, A.J. Sankoh, An overview of practical approaches for handling missing data in clinical trials, J. Biopharm. Stat. 19 (2009) 1055–1073, https://doi.org/10.1080/10543400903242795.
[29] R.J. Little, R. D'Agostino, M.L. Cohen, K. Dickersin, S.S. Emerson, J.T. Farrar, C. Frangakis, J.W. Hogan, G. Molenberghs, S.A. Murphy, J.D. Neaton, A. Rotnitzky, D. Scharfstein, W.J. Shih, J.P. Siegel, H. Stern, The prevention and treatment of missing data in clinical trials, N. Engl. J. Med. 367 (2012) 1355–1360, https://doi.org/10.1056/NEJMsr1203730.
[30] European Medicines Agency, Guideline on Missing Data in Confirmatory Clinical Trials. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-missing-data-confirmatory-clinical-trials_en.pdf, 2010 (accessed August 16, 2021).
[31] National Research Council, The Prevention and Treatment of Missing Data in Clinical Trials, National Academies Press, Washington, D.C., 2010, https://doi.org/10.17226/12955.
[32] ICH, Addendum on Estimand and Sensitivity Analysis in Clinical Trials, (n.d.). https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf (accessed August 16, 2021).
[33] R.J.A. Little, D.B. Rubin, Stat. Anal. Missing Data, 2002 doi: 10.2307/1533221.
[34] R. Faria, M. Gomes, D. Epstein, I.R. White, A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials, Pharmacoeconomics. 32 (2014) 1157–1170, https://doi.org/10.1007/s40273-014-0193-3.
[35] H. Kang, The prevention and handling of the missing data, Korean J. Anesthesiol. 64 (2013) 402, https://doi.org/10.4097/kjae.2013.64.5.402.
[36] Centers for Disease Control and Prevetion, National Health and Nutrition Examination Survey (NHANES), Physical Activity Monitor (PAM) Procedures Manual. https://www.cdc.gov/nchs/data/nhanes/nhanes_11_12/Physical_Activity_Monitor_Manual.pdf, 2012 (accessed August 16, 2021).
[37] M.R. Marler, P. Gehrman, J.L. Martin, S. Ancoli-Israel, The sigmoidally transformed cosine curve: a mathematical model for circadian rhythms with symmetric non-sinusoidal shapes, Stat. Med. 25 (2006) 3893–3904, https://doi.org/10.1002/sim.2466.
[38] S.Y. Xu, S. Nelson, J. Kerr, S. Godbole, E. Johnson, R.E. Patterson, C.L. Rock, D. D. Sears, I. Abramson, L. Natarajan, Modeling temporal variation in physical activity using functional principal components analysis, Stat. Biosci. 11 (2019) 403–421, https://doi.org/10.1007/s12561-019-09237-3.
[39] R.P. Troiano, Large-scale applications of accelerometers: new frontiers and new questions, Med. Sci. Sports Exerc. 39 (2007) 1501.
[40] C. Tudor-Locke, S. Camhi, R. Troiano, A catalog of rules, variables, and definitions applied to accelerometer data in the national health and nutrition examination survey, 2003–2006, Prev. Chronic Dis. (2012), https://doi.org/10.5888/pcd9.110332.
[41] L. Ferrucci, The baltimore longitudinal study of aging (BLSA): a 50-year-long journey and plans for the future, journals Gerontol, Ser. A Biol. Sci. Med. Sci. 63 (2008) 1416–1419, https://doi.org/10.1093/gerona/63.12.1416.
[42] J.A. Schrack, V. Zipunnikov, J. Goldsmith, J. Bai, E.M. Simonsick, C. Crainiceanu, L. Ferrucci, Assessing the "physical cliff": detailed quantification of age-related differences in daily patterns of physical activity, J. Gerontol. Ser. A Biol. Sci. Med. Sci. 69 (2014) 973–979, https://doi.org/10.1093/gerona/glt199.
[43] J.A. Schrack, P.-L. Kuo, A.A. Wanigatunga, J. Di, E.M. Simonsick, A.P. Spira, L. Ferrucci, V. Zipunnikov, Active-to-sedentary behavior transitions, fatigability, and physical functioning in older adults, J. Gerontol. Ser. A. (2018), https://doi.org/10.1093/gerona/gly243.
[44] A.A. Wanigatunga, J. Di, V. Zipunnikov, J.K. Urbanek, P.-L. Kuo, E.M. Simonsick, L. Ferrucci, J.A. Schrack, Association of total daily physical activity and fragmented physical activity with mortality in older adults, JAMA Netw. Open 2 (2019), e1912352, https://doi.org/10.1001/jamanetworkopen.2019.12352.
[45] S. Karamizadeh, S.M. Abdullah, A.A. Manaf, M. Zamani, A. Hooman, An overview of principal component analysis, J. Signal Inf. Process. 04 (2013) 173–175, https://doi.org/10.4236/jsip.2013.43B031.
[46] Z. Che, S. Purushotham, K. Cho, D. Sontag, Y. Liu, Recurrent neural networks for multivariate time series with missing values, Sci. Rep. 8 (2018) 6085, https://doi.org/10.1038/s41598-018-24271-9.
[47] D.M. Kreindler, C.J. Lumsden, The effects of the irregular sample and missing data in time series analysis, Nonlinear Dynamics Psychol. Life Sci. 10 (2006) 187–214. http://www.ncbi.nlm.nih.gov/pubmed/16519865.
[48] K. Rehfeld, N. Marwan, J. Heitzig, J. Kurths, Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Process. Geophys. 18 (2011) 389–404, https://doi.org/10.5194/npg-18-389-2011.
[49] A.A. Georgiev, J. Ramsay, B. Silverman, Functional data analysis, Technometrics. 40 (1998) 260, https://doi.org/10.2307/1271190.
[50] J.O. Ramsay, Functional Data Analysis, Wiley Online Library, 2006, https://doi.org/10.1007/b98888.
[51] I. Gaynanova, N. Punjabi, C. Crainiceanu, Modeling continuous glucose monitoring (CGM) data during sleep, Biostatistics. (2020), https://doi.org/10.1093/biostatistics/kxaa023.
[52] A. Leroux, J. Di, E. Smirnova, E.J. Mcguffey, Q. Cao, E. Bayatmokhtari, L. Tabacu, V. Zipunnikov, J.K. Urbanek, C. Crainiceanu, Organizing and analyzing the activity data in NHANES, Stat. Biosci. 11 (2019) 262–287, https://doi.org/10.1007/s12561-018-09229-9.
[53] J. Goldsmith, X. Liu, J.S. Jacobson, A. Rundle, New insights into activity patterns in children, found using functional data analyses, Med. Sci. Sports Exerc. 48 (2016) 1723–1729, https://doi.org/10.1249/MSS.0000000000000968.
[54] C.-Z. Di, C.M. Crainiceanu, B.S. Caffo, N.M. Punjabi, Multilevel functional principal component analysis, Ann. Appl. Stat. 3 (2009) 458–488, https://doi.org/10.1214/08-AOAS206.
[55] H. Shou, V. Zipunnikov, C.M. Crainiceanu, S. Greven, Structured functional principal component analysis, Biometrics. 71 (2015) 247–257, https://doi.org/10.1111/biom.12236.
[56] C. Di, C.M. Crainiceanu, W.S. Jank, Multilevel sparse functional principal component analysis, Stat. 3 (2014) 126–143, https://doi.org/10.1002/sta4.50.
[57] F. Yao, H.-G. Müller, A.J. Clifford, S.R. Dueker, J. Follett, Y. Lin, B.A. Buchholz, J. S. Vogel, Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate, Biometrics. 59 (2003) 676–685, https://doi.org/10.1111/1541-0420.00078.
[58] F. Yao, H.-G. Müller, J.-L. Wang, Functional data analysis for sparse longitudinal data, J. Am. Stat. Assoc. 100 (2005) 577–590, https://doi.org/10.1198/016214504000001745.
[59] P.C. Besse, H. Cardot, F. Ferraty, Simultaneous non-parametric regressions of unbalanced longitudinal data, Comput. Stat. Data Anal. 24 (1997) 255–270, https://doi.org/10.1016/S0167-9473(96)00067-9.
[60] L. Xiao, C. Li, W. Checkley, C. Crainiceanu, Fast covariance estimation for sparse functional data, Stat. Comput. 28 (2018) 511–522, https://doi.org/10.1007/s11222-017-9744-8.

[61] M.R. Grigsby, J. Di, A. Leroux, V. Zipunnikov, L. Xiao, C. Crainiceanu, W. Checkley, Novel metrics for growth model selection, Emerg. Themes Epidemiol. 15 (2018) 4, https://doi.org/10.1186/s12982-018-0072-z.

[62] J. Goldsmith, F. Scheipl, L. Huang, J. Wrobel, C. Di, J. Gellar, J. Harezlak, M. W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, P.T. Reiss, refund: Regression with Functional Data. https://cran.r-project.org/package=refund, 2020.

[63] J.O. Ramsay, S. Graves, G. Hooker, fda: Functional Data Analysis. https://cran.r-project.org/package=fda, 2020.

[64] S. Wood, mgcv: FMixed GAM Computation Vehicle with Automatic Smoothness Estimation. https://cran.r-project.org/web/packages/mgcv/index.html, 2021.

[65] H.F. Nweke, Y.W. Teh, M.A. Al-garadi, U.R. Alo, Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges, Expert Syst. Appl. 105 (2018) 233–261, https://doi.org/10.1016/j.eswa.2018.03.056.

[66] T. Cho, U. Sunarya, M. Yeo, B. Hwang, Y.S. Koo, C. Park, Deep-ACTINet: end-to-end deep learning architecture for automatic sleep-wake detection using wrist actigraphy, Electronics. 8 (2019) 1461, https://doi.org/10.3390/electronics8121461.

[67] S. Haghayegh, S. Khoshnevis, M.H. Smolensky, K.R. Diller, Application of deep learning to improve sleep scoring of wrist actigraphy, Sleep Med. 74 (2020) 235–241, https://doi.org/10.1016/j.sleep.2020.05.008.

[68] J. Tilmanne, J. Urbain, M.V. Kotheare, A.Va. Wouwer, S.V. Kothare, Algorithms for sleep-wake identification using actigraphy: a comparative study and new results, J. Sleep Res. 18 (2009) 85–98, https://doi.org/10.1111/j.1365-2869.2008.00706.x.

[69] F. Luna-Perejón, M.J. Domínguez-Morales, A. Civit-Balcells, Wearable fall detector using recurrent neural networks, Sensors. 19 (2019) 4885, https://doi.org/10.3390/s19224885.

[70] J. Zhang, Y. Wu, A new method for automatic sleep stage classification, IEEE Trans. Biomed. Circuits Syst. 11 (2017) 1097–1110, https://doi.org/10.1109/TBCAS.2017.2719631.

[71] F. Chollet, Keras, GitHub Repos. https://github.com/keras-team/keras, 2015.

[72] F. Chollet, Deep Learning with R, Manning, 2018.

[73] V. Tresp, T. Briegel, A solution for missing data in recurrent neural networks with an application to blood glucose prediction, Adv. Neural Inf. Proces. Syst. (1998) 971–977.

[74] L. Cai, Z. Wang, H. Gao, D. Shen, S. Ji, Deep adversarial learning for multi-modality missing data completion, in: Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., ACM, New York, NY, USA, 2018, pp. 1158–1166, https://doi.org/10.1145/3219819.3219963.

[75] S. Parveen, P. Green, Speech recognition with missing data using recurrent neural nets, Adv. Neural Inf. Proces. Syst. 14 (2001) 1189–1195.

[76] M.S. Tackney, D.G. Cook, D. Stahl, K. Ismail, E. Williamson, J. Carpenter, A framework for handling missing accelerometer outcome data in trials, Trials. 22 (2021) 379, https://doi.org/10.1186/s13063-021-05284-8.

[77] T. Vetrovsky, C.C.T. Clark, M.C. Bisi, M. Siranec, A. Linhart, J.J. Tufano, M. J. Duncan, J. Belohlavek, Advances in accelerometry for cardiovascular patients: a systematic review with practical recommendations, ESC Hear. Fail. 7 (2020) 2021–2031, https://doi.org/10.1002/ehf2.12781.

[78] C.B. Dillon, A.P. Fitzgerald, P.M. Kearney, I.J. Perry, K.L. Rennie, R. Kozarski, C. M. Phillips, Number of days required to estimate habitual activity using wrist-worn GENEActiv accelerometer: a cross-sectional study, PLoS One 11 (2016), e0109913, https://doi.org/10.1371/journal.pone.0109913.

[79] J.K. Urbanek, A.P. Spira, J. Di, A. Leroux, C. Crainiceanu, V. Zipunnikov, Epidemiology of objectively measured bedtime and chronotype in US adolescents and adults: NHANES 2003–2006, ArXiv. (2017) 1–19, https://doi.org/10.1080/07420528.2017.1411359.

[80] C.D. Klinker, J. Schipperijn, H. Christian, J. Kerr, A.K. Ersbøll, J. Troelsen, Using accelerometers and global positioning system devices to assess gender and age differences in children's school, transport, leisure and home based physical activity, Int. J. Behav. Nutr. Phys. Act. 11 (2014) 8, https://doi.org/10.1186/1479-5868-11-8.

[81] J. Di, A. Leroux, J. Urbanek, R. Varadhan, A. Spira, J. Schrack, V. Zipunnikov, Patterns of sedentary and active time accumulation are associated with mortality in US adults: the NHANES study, BioRxiv. 182337 (2017) 1–32, https://doi.org/10.1101/182337. http://biorxiv.org/content/early/2017/08/31/182337.abstract.

[82] P.-L. Kuo, J. Di, L. Ferrucci, F.R. Lin, Analysis of hearing loss and physical activity among US adults aged 60-69 years, JAMA Netw. Open 4 (2021), e215484, https://doi.org/10.1001/jamanetworkopen.2021.5484.

[83] C.H. Mallinckrodt, P.W. Lane, D. Schnell, Y. Peng, J.P. Mancuso, Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials, Drug Inf. J. 42 (2008) 303–319, https://doi.org/10.1177/009286150804200402.

[84] M. Powney, P. Williamson, J. Kirkham, R. Kolamunnage-Dona, A review of the handling of missing longitudinal outcome data in clinical trials, Trials. 15 (2014) 237, https://doi.org/10.1186/1745-6215-15-237.

[85] Y. Dong, C.-Y.J. Peng, Principled missing data methods for researchers, Springerplus. 2 (2013) 222, https://doi.org/10.1186/2193-1801-2-222.

[86] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B 39 (1977) 1–22, https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

[87] D.B. Rubin, Multiple Imputation for Nonresponse in Survey, Wiley, 2004.

[88] T. Raghunathan, J. Lepkowski, J. Van Hoewyk, P. Solenberger, A multivariate technique for multiply imputing missing values using a sequence of regression models, Surv. Methodol. 27 (2001) 85–96.

[89] S. van Buuren, K. Groothuis-Oudshoorn, {mice}: multivariate imputation by chained equations in R, J. Stat. Softw. 45 (2011) 1–67. https://www.jstatsoft.org/v45/i03/.

[90] N.M. Edwards, G.D. Myer, H.J. Kalkwarf, J.G. Woo, P.R. Khoury, T.E. Hewett, S. R. Daniels, Outdoor temperature, precipitation, and wind speed affect physical activity levels in children: a longitudinal cohort study, J. Phys. Act. Health 12 (2015) 1074–1081, https://doi.org/10.1123/jpah.2014-0125.

[91] F. Harrison, E.M. van Sluijs, K. Corder, U. Ekelund, A. Jones, The changing relationship between rainfall and children's physical activity in spring and summer: a longitudinal study, Int. J. Behav. Nutr. Phys. Act. 12 (2015) 41, https://doi.org/10.1186/s12966-015-0202-8.

[92] J. Honaker, G. King, M. Blackwell, Amelia II: a program for missing data, J. Stat. Softw. 45 (2011), https://doi.org/10.18637/jss.v045.i07.

[93] L. Ji, S.-M. Chow, A.C. Schermerhorn, N.C. Jacobson, E.M. Cummings, Handling missing data in the modeling of intensive longitudinal data, Struct. Equ. Model. A Multidiscip. J. 25 (2018) 715–736, https://doi.org/10.1080/10705511.2017.1417046.

[94] J. Tobin, Estimation of relationships for limited dependent variables, Econometrica. 26 (1958) 24, https://doi.org/10.2307/1907382.

[95] J.A.C. Sterne, I.R. White, J.B. Carlin, M. Spratt, P. Royston, M.G. Kenward, A. M. Wood, J.R. Carpenter, Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls, BMJ. 338 (2009) b2393, https://doi.org/10.1136/bmj.b2393.

[96] J.A. Schrack, A. Leroux, J.L. Fleg, V. Zipunnikov, E.M. Simonsick, S.A. Studenski, C. Crainiceanu, L. Ferrucci, Using heart rate and accelerometry to define quantity and intensity of physical activity in older adults, J. Gerontol. Ser. A. 73 (2018) 668–675, https://doi.org/10.1093/gerona/gly029.

[97] P. Lodder, To impute or not impute: that's the question, in: G.J. Mellenbergh, H. J. Adèr (Eds.), Advis. Res. Methods Sel. Top, 2013, pp. 1–7.

[98] L.C. Benson, C.A. Clermont, E. Bošnjak, R. Ferber, The use of wearable devices for walking and running gait analysis outside of the lab: a systematic review, Gait Posture 63 (2018) 124–138, https://doi.org/10.1016/j.gaitpost.2018.04.047.

[99] D. Arvidsson, J. Fridolfsson, M. Börjesson, Measurement of physical activity in clinical practice using accelerometers, J. Intern. Med. (2019) joim.12908, https://doi.org/10.1111/joim.12908.

[100] S. Duncan, T. Stewart, L. Mackay, J. Neville, A. Narayanan, C. Walker, S. Berry, S. Morton, Wear-time compliance with a dual-accelerometer system for capturing 24-h behavioural profiles in children and adults, Int. J. Environ. Res. Public Health 15 (2018) 1296, https://doi.org/10.3390/ijerph15071296.

[101] M. Straczkiewicz, N.W. Glynn, J. Harezlak, On placement, location and orientation of wrist-worn tri-axial accelerometers during free-living measurements, Sensors. 19 (2019) 2095, https://doi.org/10.3390/s19092095.

[102] J.J. Scott, A.V. Rowlands, D.P. Cliff, P.J. Morgan, R.C. Plotnikoff, D.R. Lubans, Comparability and feasibility of wrist- and hip-worn accelerometers in free-living adolescents, J. Sci. Med. Sport 20 (2017) 1101–1106, https://doi.org/10.1016/j.jsams.2017.04.017.

[103] S.J. Faireclough, R. Noonan, A.V. Rowlands, V. Van Hees, Z. Knowles, L. M. Boddy, Wear compliance and activity in children wearing wrist- and hip-mounted accelerometers, Med. Sci. Sports Exerc. 48 (2016) 245–253, https://doi.org/10.1249/MSS.0000000000000771.

[104] B.R. Bloem, W.J. Marks, A.L. Silva de Lima, M.L. Kuijf, T. van Laar, B.P.F. Jacobs, M.M. Verbeek, R.C. Helmich, B.P. van de Warrenburg, L.J.W. Evers, J. IntHout, T. van de Zande, T.M. Snyder, R. Kapur, M.J. Meinders, The personalized Parkinson project: examining disease progression through broad biomarkers in early Parkinson's disease, BMC Neurol. 19 (2019) 160, https://doi.org/10.1186/s12883-019-1394-3.

[105] M. Haberkamp, J. Moseley, D. Athanasiou, F. de Andres-Trelles, A. Elferink, M. M. Rosa, A. Magrelli, European regulators' views on a wearable-derived performance measurement of ambulation for Duchenne muscular dystrophy regulatory trials, Neuromuscul. Disord. 29 (2019) 514–516, https://doi.org/10.1016/j.nmd.2019.06.003.

[106] J.C. Goldsack, A. Coravos, J.P. Bakker, B. Bent, A.V. Dowling, C. Fitzer-Attas, A. Godfrey, J.G. Godino, N. Gujar, E. Izmailova, C. Manta, B. Peterson, B. Vandendriessche, W.A. Wood, K.W. Wang, J. Dunn, Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for biometric monitoring technologies (BioMeTs), Npj Digit. Med. 3 (2020) 55, https://doi.org/10.1038/s41746-020-0260-4.

[107] A. Godfrey, B. Vandendriessche, J.P. Bakker, C. Fitzer-Attas, N. Gujar, M. Hobbs, Q. Liu, C.A. Northcott, V. Parks, W.A. Wood, V. Zipunnikov, J.A. Wagner, E. S. Izmailova, Fit-for-purpose biometric monitoring technologies: leveraging the laboratory biomarker experience, Clin. Transl. Sci. 14 (2021) 62–74, https://doi.org/10.1111/cts.12865.