

SIGNANT HEALTH
THE EVIDENCE GENERATION COMPANY™

The Signal

Dealing with Missing Data from Sensors and Wearables

The potential of using sensors and wearables in developing clinical trial measures to better understand the effects of new treatment interventions is huge. As an industry, we have developed consensus recommendations and approaches on how to select wearables and sensors that provide data that are robust and reliable enough for regulatory decision making. We've also have developed implementation considerations and best practices. These can be examined in the work of the Digital Medicine Society (DiMe)[1], the ePRO Consortium[2] the Drug Information Association[3], and in the recent draft guidance on digital health technologies from FDA.[4]

One gap area in our knowledge when using sensors and wearables is how to adequately deal with missing data. This is a particularly important question when we consider data from continuous streaming devices like activity monitors and continuous glucose monitors (CGMs). For CGM data, where "time in range" is one of the key derived measures describing glycaemic control, common approaches to missing data include excluding data for patients providing less than 70% of monitoring days across 14 consecutive days. For activity monitoring data, a similar approach is commonplace: Patients providing insufficient "valid days" (days of at least a minimum amount of wear time – e.g., 12 hours or more) are excluded from the endpoint calculation and analysis. The rationale for this is that without this quantity of data, the estimates of activity or glycaemic control are considered unreliable.

Reasons for Missing Clinical Data

Discarding data always seems unsatisfactory, but in the absence of other approaches it is not surprising that this is typically the way chosen to deal with missingness. On the surface, only including data exceeding a certain threshold of completeness sounds intuitively sensible, but we need to consider the reasons that data are missing. Statisticians talk about three reasons for missingness: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). In the context of wearables data, MCAR might occur due to a device malfunction or an error in data transfer resulting in loss of data. MAR might occur if, for example, missing data were more frequently observed in female rather than male participants – perhaps due to the size or form factor of the device affecting its use. MNAR might occur because the patient elects not to use the device at times when they are feeling unwell. The way we deal with missing data may introduce bias depending on the reason for missingness.



An Example of How to Handle Missing Activity Data

Let's consider that in the context of rules for inclusion or exclusion of data based on changing the threshold for accepting data. Catellier et al[5] report an interesting illustration. In their study of continuous activity data collected over seven consecutive days among school children, they estimated the time spent in moderate to vigorous physical activity (MVPA) based on including all data, and a number of different valid day definitions by excluding data without at least eight hours daily wear time, ten hours daily wear time, and twelve hours daily wear time. The overall dataset was shown to underestimate MVPA as it contained some days with only a few hours of wear time. However, there were also differences in the MVPA estimated using the "valid days" rules and the authors inferred that excluding invalid days may introduce bias due to differences in activity between valid and invalid days. This would certainly be the case if data were missing not at random.

Emerging Methodologies

In our paper[6], we explore emerging statistical approaches for addressing missingness in continuously streamed sensor data. We look at the use of within-patient imputation techniques that use information from complete segments of each day's time series profile to estimate values in incomplete segments. These may be suitable when data are missing at random as we can assume that the values in the missing data segment are from the same distribution as the data from complete segments. Not so, however, if data are missing not at random. We also explore emerging approaches including functional data analysis, and deep learning methods with an aim to generate more discussion and research on the optimal ways to deal with missing data when estimating endpoints derived from continuous sensor or wearable data.

Dealing with missingness requires a thoughtful statistical approach to generate robust and reliable inferences. As we design trials to collect continuous data from sensors and wearables, limiting missing data must be an important consideration. This can be affected by a number of factors including the chosen device, placement location, wear interval, and use of real-time reminders and nudges to drive wear compliance. In addition, collecting the reason for missingness is an important parameter as this helps to identify when missingness is at random and not at random, which enables us to correctly classify intercurrent events and account for these in the approaches we adopt.

Read the full Elsevier article.



References

1. Goldsack, JC, Coravos, A, Bakker, JP et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). npj Digital Medicine 2020; 3:55. https://doi.org/10.1038/s41746-020-0260-4
2. Byrom B, Watson C, Doll H et al. Selection of and Evidentiary Considerations for Wearable Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations from the ePRO Consortium. Value in Health 2018; 21: 631-639.
3. Walton M, Cappelleri J, Byrom B et al. Considerations for development of an evidence dossier to support the use of mobile sensor technology for clinical outcome assessments in clinical trials. Contemporary Clinical Trials 2020; 91: 105962.
4. Food and Drug Administration. Digital Health Technologies for Remote Data Acquisition in Clinical Investigations: Guidance for Industry, Investigators, and Other Stakeholders. 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations
5. Catellier DJ, Hannan PJ, Murray DM et al. Imputation of Missing Data When Measuring Physical Activity by Accelerometry. Med Sci Sports Exerc. 2005; 37(11 Suppl): S555–S562.
6. Di J, Demanuele C, Ketterman A et al. Considerations to address missing data when deriving clinical trial endpoints from digital health technologies. Contemporary Clinical Trials 2022. https://www.sciencedirect.com/science/article/pii/S1551714421003979?via%3Dihub

---



Bill Byrom, Ph.D.

VP Product Intelligence and Positioning

Interested in reading more blogs from *The Signal*?

Click to Subscribe