

The Importance of Capturing Raw Sensor Data for Clinical Studies

By Geoffrey Gill

Wearable sensors are transforming clinical studies. By replacing sporadic, subjective endpoints with continuous sensor-based ones¹, wearable sensors increase accuracy, reduce participant burden, and establish common objective endpoints across trials. But clinical research demands higher-quality data than most commonly available wearables can provide. This article delves into shortcomings of sensors that are widely available, advantages that more sophisticated sensors can offer, and problems that remain to be resolved.

Overview

Most wearables process raw sensor data locally and provide only the processed outputs to the user. Essentially all consumer-grade wearables employ this method; it reduces the amount of data transferred, improves battery life, and minimizes data plan usages, all critical features for consumers. But clinical studies need more:

- **Lack of transparency** creates a black-box situation, where the clinical scientist cannot be confident that a change in an endpoint is due to a change in the patient, a change in how the wearable is used, a change in the sensors, or a change in the hidden algorithm that processes the data. Suppliers of consumer-grade wearables might "improve" algorithms at any time without informing either the user or the clinical scientist.
- **Lack of ensured consistency in the algorithm** means that a new clinical study is required to validate a wearable whenever a new algorithm is generated or even suspected.
- **Data is not re-usable.** Once data has been processed and raw source data deleted, it cannot be re-processed with better algorithms that might arise in the future.
- **Apples vs. oranges.** Since each algorithm is almost certain to produce different results from the same data, it is not possible to consolidate or compare data from different studies that used different algorithms; nor is it possible to recalibrate results.

Different consumer-grade wearables can produce wildly different results. For example, *Which* magazine studied more than 100 fitness trackers placed on marathon runners. Recorded distances showed variances of more than 40%, with the longest recorded distance

¹ See, for example, Bernard Munos, et al; "Mobile health: the power of wearables, sensors, and apps to transform clinical studies;" *Annals of the New York Academy of Sciences*, July 6, 2016; <https://nyaspubs.onlinelibrary.wiley.com/doi/full/10.1111/nyas.13117>

more than 95% longer than the shortest.² While some degree of inaccuracy might be acceptable for consumer devices, it is not acceptable for clinical research.

Benefits of Collecting Raw Data

Devices that provide raw sensor data address the above issues and yield the following major advantages:

Raw Data Can be Used to Accurately Characterize Sensors

Sensors measure specific physical phenomena like acceleration, angular velocity, electrical current and voltage. These measurements are far more precise than needed for typical consumer analyses of wearable data. Accuracy and noise characteristics can be independently measured and quantified. In some cases, ongoing calibration of signals can be performed. For example, open-source GGIR algorithms for wrist-based activity and sleep monitoring use acceleration due to gravity to calibrate measurements from accelerometers on an ongoing basis. Not only does this technology address possible sensor drift in the sensors, but it also enables accurate measurement of noise in the data. Similar calibration procedures for other types of sensors can also be implemented.

Raw Data Enables Analysis by Validated Algorithms

Of course, raw sensor data³ must be analyzed to produce meaningful results. Fortunately, many public-domain options exist for performing such analyses. Academic research has largely moved away from proprietary algorithms. Many large-scale studies have been conducted with raw accelerometer data (e.g., NHANES, UK Biobank). Literally thousands of researchers, many of whom have been working for more than a decade, are studying how to interpret raw sensor data. Much of that work is in the public domain.⁴ Even some pharmaceutical companies, including Pfizer and Novartis, are putting their proprietary algorithms and data into the public domain. In addition, broad-based collaborations have been established to develop and validate algorithms. Mobilise-D, a €50-million, five-year project to develop mobility endpoints for five diseases, may be the largest such collaboration.⁵ Another approach is the Open Wearables Initiative, a collaboration among device manufacturers, pharmaceutical companies, industry associations, and others to make algorithms and validation data more accessible to researchers and also support regulatory

² Morgan, Tom; "Marathon Training? We reveal the least accurate fitness trackers from popular brands;" *Which*, April 27, 2019; <https://www.which.co.uk/news/2019/04/marathon-training-here-are-eight-fitness-trackers-to-avoid/>

³ Raw data is, at some level, a misnomer. All data has been processed to some extent, if only to be converted from analog to digital format. We follow the common use of the term "raw data" to describe the most minimally processed data that comes directly off the hardware sensor. It generally measures physical quantities in units like millivolts, g's (acceleration due to gravity), and amps. Heart rate, steps, calories, etc. are not considered raw data in this context.

⁴ For example, Shimmer Research alone has more than 3,000 research customers in over 75 countries. Other wearables companies have similar academic client bases.

⁵ See <https://www.mobilise-d.eu/> for more information about this initiative.

approval for clinical endpoints.⁶ All these efforts require the wearables to provide raw data for use by the associated algorithms.

Algorithms Based on Raw Data Are Device Independent

Using a device that provides only a calculated value like number of steps, but no raw data, ties the researcher into that device — and often to a specific version of that device. Validating a new device is a major project. For example, a true validation of a wearable that provides only step counts requires repeating the entire validation process across each possible situation (e.g., stairs, up and down hills, and impeded versus unimpeded walking) and each possible population (e.g., elderly versus young, different diseases, and even different stages of a disease).

In contrast, with algorithms that run off raw data, any sensor may be used as long as it provides adequate raw data (i.e., a sufficient sample rate, dynamic range, and bit resolution for the algorithm's purpose), freeing researchers from dependence on a specific device. In general, sensors provide much more precision and better noise characteristics than algorithms require, enabling use of many off-the-shelf sensors. If there is a question, adequacy of the raw data can be determined by looking at performance characteristics of the sensor, which are often published. If not, independent testing of sensor performance can be completed without repeating the entire validation study because levels can be characterized based on known physical phenomena.

Raw Data Enables Determination of Measurement Accuracy

With raw data, noise in the system can often be quantified, making it much easier to determine the accuracy of the results.

Anomalies Can Be Investigated Using Raw Data

In any data set, there will be anomalies. For example, if a study participant does not record any steps in an hour, is that because they are sedentary or because they are not wearing the sensor? Patients have been known to place activity trackers on their dogs to please their doctor with a higher recorded level of activity. One researcher accidentally put her activity tracker through the laundry *twice*, where it recorded 5,000 steps. If a wearable provides only a calculated measure but no raw data, identifying and resolving such anomalies can be difficult to impossible. But high-quality raw data can be used to tease out answers to complicated problems. For example, raw three-axis accelerometer data would show a distinctive pattern of movement only achievable by tumbling in a clothes dryer. Similarly, a dog will have a different movement profile than a person. It should even be possible to use raw data to verify that the actual study participant was the person using the wearable, based on the characteristics of their stride and the swing of their arm.

Study Data Can Be Re-Analyzed with Updated Algorithms

Raw sensor data can be re-analyzed using the most up-to-date algorithms. If, for example, a problem appears during the course of a study, the algorithm can be modified (as specified by an appropriate protocol). After the study, the data set can be processed with multiple algorithms that might not have existed when the study started.

⁶ See www.owear.org for more information about this initiative.

Study Data Can Be Used for Computations that a Wearable Cannot

For example, raw data can be used to train a machine learning (AI) system or be fed into such a system for analysis.

Raw Data Enables Combining Data Across Multiple Studies

Some of the greatest benefits of using raw data can be realized by analyzing data across multiple studies. Since sensors that produce raw data share similar characteristics, the data they produce can be integrated (with adjustments), even when it comes from different devices, different studies, and different organizations.

Sage Bionetworks has created a model for this type of work.⁷ Several organizations⁸ are already collating data sets that can be used across organizations. The Open Wearables Initiative has established a forum and a mechanism to share algorithms and data from many organizations and studies. It is also initiating a number of Sage DREAM Challenges to help crowdsource and benchmark algorithms for certain digital endpoints.

Once a database with ground truth has been established, it becomes possible to test new algorithms and analyses without collecting new data. Because collecting the data is the most time-consuming and expensive part of the process by far, such databases can dramatically accelerate validation and improvement of algorithms and analysis. For this to happen, though, appropriate contextual information like population characteristics and study design must be collected. In most cases, several data sets, for different populations and other contexts, will be required.

With the right raw sensor data and contextual information, highly reliable metadata analyses can be performed to determine the relevant value of different treatments.

Over time, as these databases are populated with sensor and contextual information, they can be used to establish norms for the relevant populations. At some point, it may be possible to create synthetic control arms⁹ from these data.

Managing the Data

There are two common concerns about collecting vast amounts of raw data:

- **The data will be too voluminous to transmit, store and manage.** The amount of data involved is tiny compared to widespread consumer applications. For example, a full month of 3-axis accelerometer data collected at 25Hz is well under 500MB. In comparison, a typical standard-resolution digital movie file that is streamed daily to millions of homes is 1-2GB in size. Web hosting services make this scale of data collection accessible and affordable for developers of wearable systems.

⁷ See, for example, Ellrott, K., Buchanan, A., Creason, A. *et al.* Reproducible biomedical benchmarking in the cloud: lessons from crowd-sourced data challenges. *Genome Biol* 20, 195 (2019). <https://doi.org/10.1186/s13059-019-1794-0>

⁸ See, for example, Sage Bionetworks' Synapse (<https://www.synapse.org/>) and University of California, Irvine's Machine Learning Repository(<https://archive.ics.uci.edu/ml/index.php>)

⁹ Synthetic control arms use existing data to establish a baseline rather than collecting new data from participants who are on a placebo.

- **Nobody can actually look at that much data.** Nobody needs to look at the data. It still gets processed by algorithms. Humans see the same results, in the same form. However, they also have access to additional algorithms, e.g., to look at the data in different ways and to assess the accuracy of the results.

Conclusions

Given the state of technology when the first consumer wearables were created, it is understandable that they generated only calculated results. But these limited results have no life beyond the study from which they were generated. In contrast, raw sensor data lives on. It is vital for achieving the potential of wearables in clinical studies. It offers many benefits, perhaps the most important being that it renders data reusable for creating and validating new algorithms and comparing data across studies.

Now the old barriers are falling. The advent of high-data-rate, low-power wireless technologies, such as Bluetooth 5.0, advanced Wi-Fi, and eventually 5G, is nullifying technical limits. And collaborative ventures including the Open Wearables Initiative and Mobilise-D are working to accelerate development of validated clinical endpoints. For such efforts to succeed, raw sensor data is essential.

Author

Geoffrey Gill is President of Shimmer Americas. Contact him at 1.617.945.2628 or ggill@shimmersensing.com.