ActiGraph

How "Future-Proof" is Your Digital Biomarker Data?

Considerations for maximizing the long-term value of wearable sensor data

White **Paper**

theactigraph.com

Table of Contents

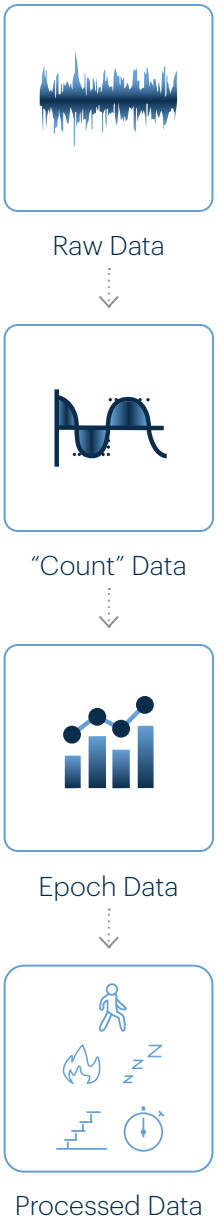
Introduction

The adoption of wearable biometric sensors to capture real-world participant data within clinical research has steadily increased in recent years. As of February 2020, there were approximately 460 wearables studies underway, with analysts predicting that 70% of trials would include a wearable device by 2025. Building upon the eSource momentum within the clinical research industry, the global pandemic has accelerated the deployment of decentralized trials (DCTs) and remote data capture technologies. This has effectively redefined the role of wearables and digital biomarkers in clinical research. Sophisticated new algorithms and analytic techniques that generate novel clinical endpoints from sensor data are also emerging at a rapid pace, promising new and more precise outcomes across more therapeutic areas and patient populations in the years to come.

There are a myriad of factors to consider when evaluating wearable motion-sensing technologies, including battery life, storage capacity, data access, validation, patient acceptance, privacy, and regulatory clearance, to name a few. However, one of the most important considerations - the longevity of the collected data - is frequently overlooked during this process. As the development of novel data analysis methods such as artificial intelligence and machine learning continues to accelerate, the ability to apply these new techniques to previously collected biosensor data is determined by how that data was collected and stored. Wearable technology platforms that generate raw *"future-proof"* data deliver long-term value by preserving comparability across multiple phases and studies, while enabling sponsors to leverage advanced methods to derive better quality outcome data.

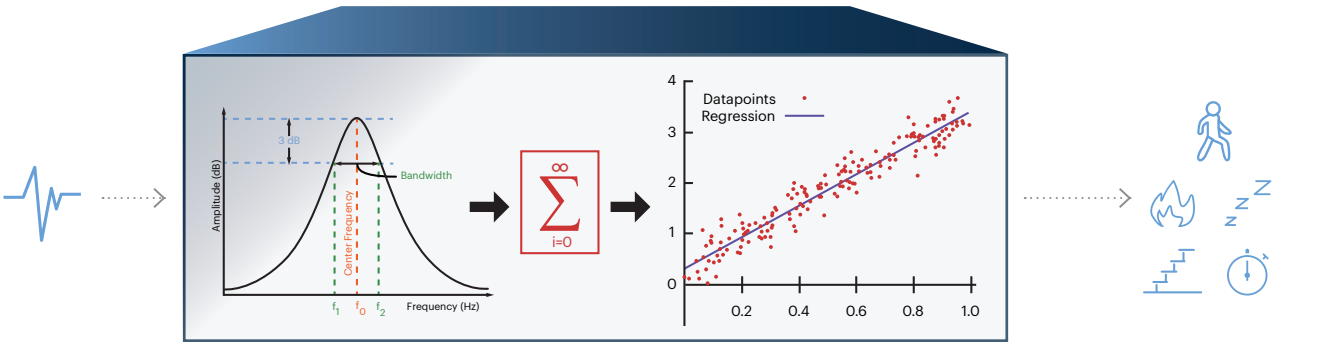
Decoding the Data: Raw Data vs Processed Data

In the context of motion-sensing wearable technologies, the term *"future-proof"* data refers to the raw accelerometer signal, which consists of gravity (G) values for each axis, sampled at a specific frequency. For example, ActiGraph's CentrePoint Insight Watch collects data at a 32 Hz sample rate, meaning it logs 32 data points per second on the x, y, and z axes. Once these data are uploaded to the CentrePoint system, calibration values captured during the manufacturing process are applied to the data, accounting for any offset of the accelerometer from the device housing. Calibrated raw data is then re-sampled and bandpass filtered to eliminate any data that falls outside the normal range of human movement. This data reduction technique aggregates the data into more manageable 1-second chunks, often referred to as *"count"* data.[2] These 1-second counts are further aggregated into 1-minute data points, called epochs, for each axis. The epoch data is used to identify activity peaks or bouts of sustained daytime activity. Activity and sleep algorithms have traditionally been developed using 1-minute epoch data to generate a variety of motion-based endpoints, including calories, MET rates, activity intensity, total activity, sleep periods, and sleep efficiency. These derived outcomes are what we refer to as processed data.



End-to-End Data Transparency

To deliver maximum benefit to clinical trial participants and sponsors, technology platforms used to capture digital biomarker data should provide transparency into all algorithms and testing characteristics, including thresholds for action, sensitivity, and specificity.[3] ActiGraph's technology system provides users with end-to-end transparency and access to the raw acceleration data, epoch data, processed data, and the algorithms used. This allows customers or third party data partners to extract high resolution raw data, which can then be used to develop specialized processing methods or novel algorithms specific to a particular type of movement and/or patient population. Figure 1 shows the data reduction and processing that occurs within ActiGraph's system between raw data capture and the delivery of traditional measurements.



Raw Data | Data Reduction & Signal Processing | Outcome Measures

Figure 1: End-to-End Data Processing (ActiGraph)

Beware of the Black Box Approach

In the case of most consumer-grade wearable devices, the processing that occurs between raw data collection and endpoint delivery is referred to as *"black box."* This means that some form of raw data is collected, some type of filtering is applied, and data is run through one or more proprietary algorithms to generate an endpoint. Because the raw data, filtering method, and algorithm are all proprietary, customers are left with an endpoint and no transparency into how it was derived. Unsurprisingly, this *"black box"* approach calls into question the specificity, validity, and reproducibility of processed endpoints.

Regulatory authorities are still developing guidance around digital biomarker data submissions, and they understandably do not have all the answers yet. However, the integrity of data generated through a fully transparent system is much easier to demonstrate than that of data processed using proprietary black box methods. Additionally, a transparent system allows sponsors to identify and correct missteps, such as selecting the wrong algorithm, both during and after the study.

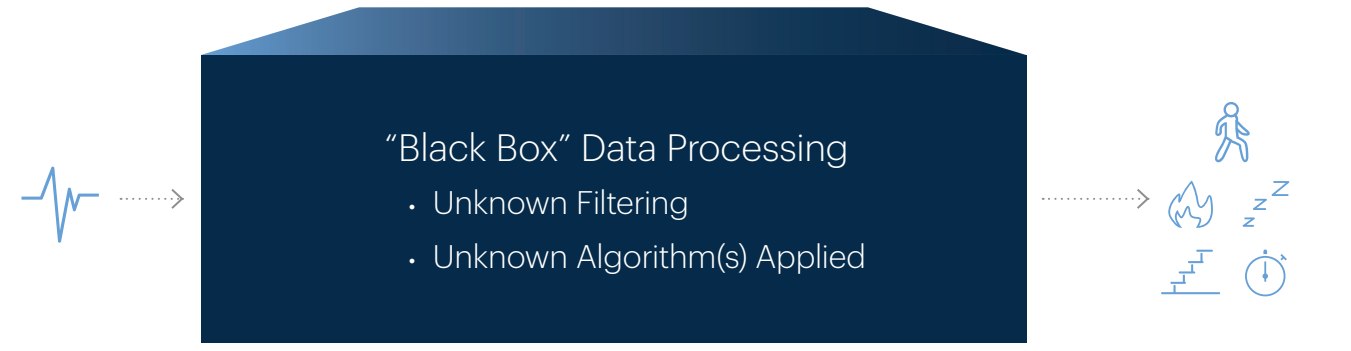


Figure 2: "Black Box" Data Processing

Data Analysis: Past, Present, and Future

Computing power and machine learning techniques have improved dramatically in recent years. Until very recently, the vast majority of algorithms used to generate motion-based endpoints were developed using 1-minute epoch data. As technology continues to advance, the development of new algorithms based on increasingly granular data, including 5-second epoch data and even raw accelerometer data, is also accelerating. While these improved methods deliver more precision and accuracy, they also introduce new layers of complexity to the data processing framework.

Complexity of Traditional vs New Methods

The intricate nature of raw data analytics is clearly illustrated by comparing traditional and new methods for determining activity intensity using cutpoints analysis. Cutpoints are a predetermined set of thresholds used to classify activity as sedentary, light, moderate, and vigorous on the per minute level. Using the old method, each minute of activity is assigned to an intensity **"bucket"** (known as data binning) based on the number of counts per minute (CPM) from a single accelerometer axis.

A newer method involves using 15-second periods of raw acceleration data from three accelerometer axes and extracting different mathematical features to classify the activity. Table 1 shows the cutpoint thresholds for activity intensity classifications using the traditional epoch method and the new raw data method.

Method	Light	Moderate	Vigorous	Sedentary	Non-Sedentary
Hip Cutpoints[4] (1998) Epoch (1-minute)	100-1951 CPM	1952-5724 CPM	5725 CPM or higher	0-99 CPM	
Wrist Cutpoints[5] (2015) RAW (30Hz)	SDVM≤0.79 and MANGLE>-52	SDVM≤0.26 and MANGLE≤-52, or 0.26<SDVM ≤0.79 and MANGLE>-53	0.26<SDVM ≤0.79 and MANGLE≤-53, or SDVM>0.79	SDVM≤ 0.098 and p625≤0.138, or SDVM≤0.062 and p625>0.138, or 0.098<SDVM ≤0.148 and p625 ≤0.118	0.062<SDVM ≤0.098 and p625>0.138, or 0.098<SDVM ≤0.148 and p625>0.118, or SDVM>0.148

Table 1: Activity intensity cutpoint thresholds using 1-minute epoch data and 15 second raw data. CPM=Counts per Minute; SDVM=Standard Deviation of Vector Magnitude; MANGLE=Mean angle of acceleration relative to the vertical axis; p625=Percentage of the power of the vector magnitude that is in 0.6–2.5 Hz.

While this type of raw data-based intensity classification has become drastically more complex and sophisticated, it also yields more precise and nuanced outcomes. In order to take advantage of these new analytic techniques, the raw

sensor data must both be captured and accessible from within the software system. It's important to note, however, that not all raw data is created equal. There are a variety of factors that directly impact the quality of the raw data collected, and consequently, the validity of the processed outcomes generated from these data. These include the sampling rate required to detect specific movements, the type of anti-aliasing filter applied and how it is affected by the sampling rate, backwards compatibility of hardware devices when algorithms and processing techniques are based on previous models, and device calibration and testing methods used to ensure intra-device reliability. Figure 3 below illustrates the results of ActiGraph backwards compatibility testing, which shows the accelerometer values across different devices and models compared against a known truth value.

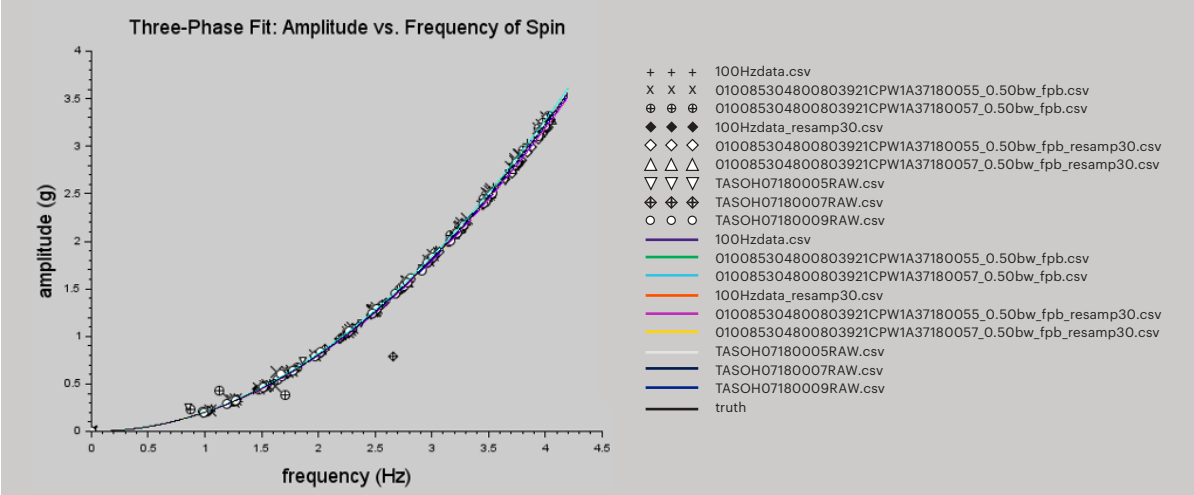


Figure 3: Demonstration of the accelerometer values across different ActiGraph devices and models, including re-sampling values, compared to a known truth value.

An example of an important hardware consideration that can negatively impact the quality of raw data generated is accelerometer clocking. Most accelerometers have an on-board clock, however these do not typically meet the reliability and consistency requirements for accurate raw data capture. In order to mitigate this issue, ActiGraph sensors include a separate piece of hardware that accurately clocks and records precise timestamps for each data point. Without this additional hardware, the raw data could include inconsistencies between samples and missing time periods.

Looking Ahead: Functional Testing and Novel Endpoints

Another advantage raw sensor data provides is the ability to make direct calculations from various traditional functional tests that apply to a specific therapeutic area. For example, a functional test performed from a study participant's home, such as a 6-minute walk test or sit-to-stand test, would yield a distinct time period of raw data. This raw data can then be used to calculate additional measures, including the average velocity, peak velocity, or precise time to completion. The benefits of integrating functional testing with raw data analytics are twofold. It provides sponsors with novel insights on functional ability within specific therapeutic areas, while also supporting decentralized trial designs through reduced site visits and the collection of more frequent and comprehensive home-based test data.

Raw data is also essential for the development of new motion-based digital biomarkers and novel endpoints. Highly nuanced movements such as scratch or tremor can be identified by applying machine learning techniques to raw acceleration data. Specific raw data features, which are individual measurable properties or characteristics associated with the movement in question, are identified using a set of training data. With enough training, algorithms capable of identifying these activities or events based on the raw data features can be developed, enabling the capture of novel data endpoints.

Data Quantity Considerations

Raw data is a virtual treasure trove of information, but accommodating the vast quantities of data produced can present a challenge. The smallest unit of measurement logged by an ActiGraph accelerometer is a 12-bit individual sample. Consider that each sample is then multiplied by three axes, then 32 measurements per second, then 60 seconds per minute, and it becomes clear how these data grow exponentially. Within the ActiGraph system, a single research participant generates approximately 8 MB of raw data per day. However this is a compressed version of the data. When expressed in a readable format, such as a .csv file, the raw data amounts to approximately 200 MB per participant per day.

To put this in perspective, Table 2 shows the amount of data generated by a two-year study with 50 participants, with three months of total data per subject.

Daily File	Minute File	Stored Raw data	Raw .csv Files
4.5 MB	**562.5 MB**	**36 GB**	**2 TB**

Table 2: Data file types and sizes (ActiGraph)

This is not a trivial amount of data, and some clinical trial sponsors may not have the resources to store or analyze this quantity of data. A reputable and experienced technology partner will have the ability to store this raw data for an extended period of time and provide access as needed for future analysis. As new data analysis methods or algorithms become available, sponsor organizations that are able to access and reprocess raw biosensor data will maximize their technology investment and the quality, accuracy, and long-term value of their data.

Conclusion

Wearable biosensors are becoming a standard assessment tool within clinical research, and advances to the algorithms and analytic methods used to interpret the collected data are accelerating at a rapid pace. The development of novel therapeutic area and population specific algorithms and endpoints provides the clinical research community with exciting new opportunities to identify and monitor real world participant behaviors with greater precision. By selecting a proven wearable sensor platform that generates raw, future-proof data, clinical trial sponsors will maximize the long-term value of their digital biomarker data and reap the benefits of analytic innovations on the horizon.

About ActiGraph

ActiGraph's mission is to bring life to digital data. Built on more than twenty years of remote data capture expertise, ActiGraph is the leading provider of medical-grade wearable motion sensors for the global scientific community. ActiGraph's FDA-cleared accelerometry sensors and flexible technology ecosystem deliver high quality, continuous digital data, providing valuable insights into the real world behaviors of clinical trial participants. Appearing in more than 17,000 published scientific papers to date, ActiGraph is the industry's most experienced, knowledgeable, and trusted wearable technology partner.

theactigraph.com

References

1. Jansen, Y., & Thornton, G. (2020, February 25). *Wearables & Big Data In Clinical Trials — Where Do We Stand? Clinical Leader.* https://www.clinicalleader.com/doc/wearables-big-data-in-clinical-trials-where-do-we-stand-0001

2. ActiGraph Support. (2018, November 8) *What are counts?* https://actigraphcorp.force.com/support/s/article/What-are-counts

3. Coravos, A., Khozin, S. & Mandl, K.D. *Developing and adopting safe and effective digital biomarkers to improve patient outcomes.* npj Digit. Med. 2, 14 (2019). https://doi.org/10.1038/s41746-019-0090-4

4. Freedson PS, Melanson E, Sirard J. *Calibration of the Computer Science and Applications, Inc. accelerometer.* Med Sci Sports Exerc. 1998 May;30(5):777-81.

5. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. *Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements.* J Appl Physiol. 2015;119(4):396-403.