



**ECHO**

Environmental influences  
on Child Health Outcomes

A program supported by the NIH

## **Measuring Positive Health Using Wearable Devices**

Ciprian Crainiceanu<sup>a</sup> and Ekaterina Smirnova<sup>b</sup>

*A Monograph from the ECHO Positive Health Outcome Working Group*

November 8, 2020

<sup>a</sup>Johns Hopkins University  
615 N. Wolfe Street  
Baltimore, Maryland 21205  
ccraini1@jhu.edu

<sup>b</sup>Virginia Commonwealth University  
1201 E Marshall St #4-100  
Richmond, VA 23298  
Ekaterina.Smirnova@vcuhealth.org



## **Table of Contents**

Summary .....	3
1. Wearable devices in health research .....	3
2. Examples of wearable devices .....	4
3. Data types and characteristics .....	8
4. Design of experiments .....	12
5. Transforming wearable device measurements into digital biomarkers .....	15
6. Data harmonization .....	18
7. Interventions .....	19
8. Specific considerations in pediatric study populations .....	20

## Summary

Wearable devices are becoming popular in pediatric research. They are easy to use, well accepted by participants, and can provide high quality data not readily obtained from other data sources. This monograph covers the basics of wearable device research and provides recommendations for best practices. It discusses the use of wearable devices in positive health research, provides examples of wearable devices, and describes the data produced by these devices. It also provides practical recommendations for study design, digital biomarker development, data harmonization, and interventions assisted by wearable devices in pediatric populations.

### 1. Wearable devices in health research

Wearable devices refer to a host of devices that are worn by individuals in controlled as well as real-world settings and can assess physiological functioning and behavioral positive health assets. They are placed on the human body and contain biosensors that provide continuous, high resolution measurements. These measurements can be used to provide real-time feedback, guide traditional and just-in-time (JIT) treatments and interventions, monitor within-person changes, and quantify population level characteristics and norms. Wearable devices complement self-reported measures of health and behavior obtained from questionnaires and diaries. Indeed, there is increased awareness that traditional self-report measurements are subject to substantial measurement error as well as recall, social-desirability, and cognitive biases (Sallis and Saelens, 2000, Washburn, 2000). In some study subpopulations it is difficult or impossible to obtain self-report information (e.g., from an infant or from a child under sedation in a pediatric intensive care unit).

At the same time self-report tools are also adapting and evolving due to the emergence of the ecological momentary assessment (EMA) technology. EMA is a new form of delivery for self-report questionnaires via mobile apps to obtain contextual (ecological), high resolution and timely (momentary), measurements (assessment) of health and behavior using apps. The combination of wearable devices, EMA, and health evaluations has the potential to substantially change health research in pediatric populations. Indeed, EMA can provide high-quality self-report health measurements that include timing and intensity of health-related events combined with contextual and environmental information. Wearable devices can quantify how, when, and for how long an EMA reported event affects or is affected by free-living activities. Health evaluations can provide context and validation and provide contextual information about the connections with positive health outcomes.

Wearable devices are also rapidly evolving with newer devices recording multiple types of measurements (e.g., physical activity and heart rate). Such devices have the potential to provide a more comprehensive and detailed quantification of the individual's health status and behavior. However, some technologies are more mature than others and much research is still needed to decide if and what wearable device measurements can be used as biomarkers for clinical decision making and clinical trials. Among the most mature and accepted mobile technologies are: EMA, video cameras, accelerometry-based assessment of physical activity, heart rate and ECG monitors, GPS trackers, and implantable glucose monitors.

## 2. Examples of wearable devices

Activity monitors are among the most ubiquitous wearable devices in health research. Activity monitors contain accelerometers, small electromechanical devices that measure acceleration in

three orthogonal axes in the reference frame of the device. Acceleration is measured internally at high resolution (sub-second level) and exported either as raw data or summaries at the minute or day level, or both. By attaching an activity monitor to the human body one can continuously measure the acceleration of that part of the body, which provides information about the intensity and timing of movement. The device can be attached to different parts of the human body depending on the particular scientific application and feasibility constraints of the study. The most popular body location is the wrist because wearing a device on the wrist is comfortable and socially acceptable, which leads to higher compliance rates. However, the hip, chest, thigh, and ankle locations have also been used extensively.

Activity monitors are also routinely used for sleep research. Indeed, the raw (sub-second level) data from activity monitors is used by dedicated sleep quality algorithms, which estimate the beginning and end of the sleep period as well as sleep characteristics, such as the amount of activity during sleep. Most of these algorithms tend to be proprietary and closed source and may not be generalizable across devices, though there is increased interest in open source, reproducible software.

Heart rate and electrocardiogram (ECG) monitors are devices typically attached to the skin in the chest area, though wrist monitors have also started to be used. They collect information associated with cardiac activity and transform the signal into the number of beats per minute (heart rate) or continuous beat-by-beat electric profiles (ECG). Such devices are used both for screening in the general population and for monitoring at-risk clinical subpopulations. They can continuously measure heart rates, P waves, QRS complexes, T waves, heart rate variability, and heart rate

variability as a function of the time of day. The P wave is generated by the electrical depolarization of the right and left atrial muscles, the QRS complex is associated with right and left ventricular depolarization and activation, and the interval from the beginning of the QRS complex to the end of the T wave (QT interval) corresponds to ventricular repolarization (VR). When coupled with a wearable accelerometer, the positive health biomarkers can be measured at rest or during activity. Both heart rate and ECG features are important positive health measures that can be collected non-invasively in a large number of study participants in their free-living environment or under controlled conditions.

Global positioning system (GPS) trackers use a system of satellites in Earth's orbit combined with devices on the ground to accurately estimate the location of an individual. The GPS technology is increasingly used in health applications. Indeed, GPS devices allow quantification of where individuals spend their time (activity space), whether, when and where a person is using transportation (based on fast changes in the location), providing continuous environmental and contextual information using mapping technology. These are all potentially useful measures of the contexts of activity and other behaviors. An important problem that needs to be addressed by researchers using GPS technology is that the data contain highly identifiable protected health information (PHI), which raises substantial data protection and patient privacy. Indeed, even plotting an individual's GPS pattern with or without time stamps can pinpoint their home and work locations. Other concerns are accuracy of the device, especially indoors, and battery life, which tends to be substantially lower than in other wearable devices.

Implantable continuous glucose monitors (CGM) have been used by individuals with diabetes to monitor and manage the disease. Indeed, CGMs can be found and bought without a prescription in most pharmacies in the US. CGMs are also continuously improving by becoming more portable, having better battery life, and being less intrusive. With increasing rates of newly diagnosed cases of type 1 and 2 pediatric diabetes mellitus, the use of CGMs in pediatric research is likely to rise.

There are many devices and technologies and many more are continuously being developed leading to a rapidly evolving wearable device ecosystem. Here we enumerate just some of them, as providing an exhaustive list is not possible. Wearable environmental exposure monitors are promising, but the technology for producing continuous, calibrated and reliable measurements that do not degrade over time is not yet mature. Silicone bands that trap environmental toxicants may provide a useful, one-time cumulative exposure measurement over a period of time. GPS devices combined with interior air quality monitors could provide information about chemical and respiratory pollution exposure. Other wearable devices include, among many others, blood pressure, pulse oxymetry and respiratory rate, body temperature, and sweat composition.

As devices become smaller and battery life improves, they can be combined to address a wider variety of scientific problems. For example, activity monitoring via accelerometers could be combined with GPS information to provide contextual information about which locations or interventions are associated with increased physical activity or improved sleep. Activity and heart rate monitoring could provide information about unusual, potentially clinical, associations between activity and cardiac responses.



### 3. Data types and characteristics

All wearable devices produce high-density time series data that can be collected for days, weeks or even months at a time with or without recharging the device. For example, accelerometers typically collect data at between 10 to 100Hz (10 to 100 observations per second) along each of 3 orthogonal axes in the reference frame of the device. Similarly, chest-worn ECG monitors collect data at high resolution (100 to 250Hz), as data are used to quantify the time-trajectory of the electric cardiac system during each individual heartbeat. As heartbeats last typically between 0.3 to 1.5 seconds this resolution is necessary to capture the PQRST features of each heartbeat. These data can be processed as summaries at coarser resolutions (e.g., minute, hour, or daily level activity counts or heart variability). In general, the structure of data obtained from wearable devices is that of high-resolution time series with a natural multilevel structure induced by the repeated measures over multiple days. Depending on the level of aggregation of the data (e.g., milliseconds as in ECG monitors or five-minute intervals in glucose monitors), data structures may exhibit substantial differences even if the target of measurement is the same. For details on data organization, see, for example Leroux 2018.

**Figure 1** provides an example of the multi-resolution nature of the data collected by accelerometers. The left column displays high resolution raw accelerometry data over five days (upper panel), a 16-hour interval from day two shown as an orange box in the upper panel (middle panel), and a six-minute interval from day two shown as an orange vertical bar in the upper and middle panels (lower panel). The middle column displays the corresponding minute-level activity summaries (activity intensity) over the same five days (upper panel) and 16-hour interval (middle panel). The lower middle panel displays cumulative activity intensity over multiple days. The right

column displays similar plots with the middle column, but data represent the proportion of time active in every minute (activity intensity greater than zero) instead of activity intensity.

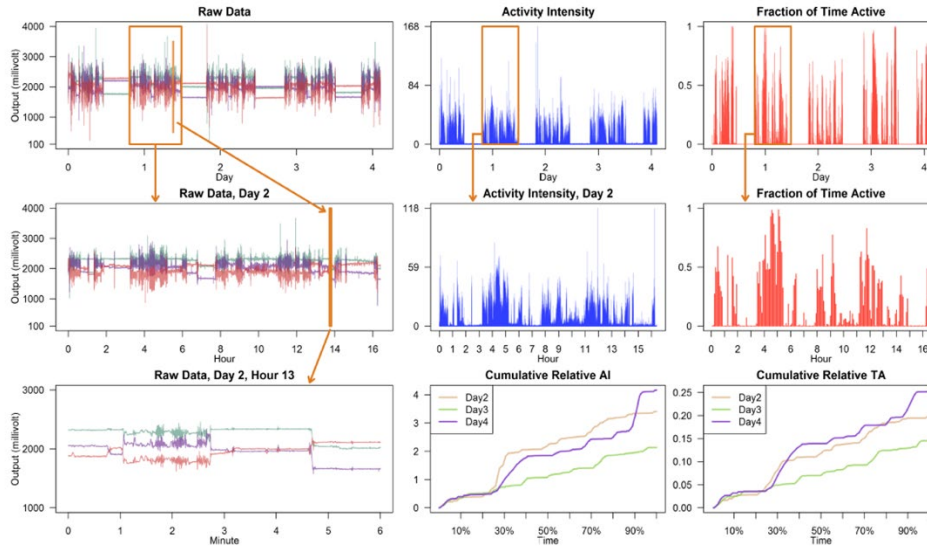


Figure 1. Example of multi-resolution data obtained from wearable devices. Left column: high resolution raw accelerometry data. Middle panels: minute-level accelerometry summaries (activity intensity). Right panels: different types of minute-level accelerometry summaries (fraction of time active during a minute).

The main differences between data produced by wearable devices and conventional time series are that: (1) a high-resolution time series is collected for each study participant, which results in populations of time series; (2) time series data have a natural correlation structure induced by the study protocol (e.g., multiple days of monitoring induces a multilevel structure whereas longitudinal monitoring induces a longitudinal structure across visits with a multilevel structure within visit); (3) the data structure is new for each type of device and level of aggregation, which requires specific protocols for analysis and quality control; (4) data are highly non-stationary with complex correlation structures both within- and between-days; and, (5) data cannot be

synchronized across days and study participants because the underlying biological rhythms are often unobserved and typically de-synchronized. Therefore, designing the experiment, collecting, curating, and analyzing wearable device data requires a set of analytic skills that is different from the one required by traditional analytic protocols. In many situations analytic protocols for wearable devices are not available and substantial effort and planning is required to ensure data quality as well as pre-processing and analytic reproducibility.

Wearable devices either contain new technologies or older technologies applied in novel ways to health applications. However, the data produced by them should be treated as carefully as data obtained from conventional measurement instruments. In particular, replication and validation studies should be conducted for each device, while the performance and characteristics of the measurements and associated algorithms should be carefully investigated in the target pediatric population. Indeed, processing algorithms that were successfully used in a small healthy young adult sample could fail in a clinical pediatric population. Therefore, replication studies are highly recommended including: (1) using two or more otherwise identical devices and the same sampling protocol on multiple study participants; (2) quantifying measurement reliability and establishing sampling protocols (e.g., number of days of continuous monitoring) necessary to achieve a given level of measurement reliability; and, (3) quantifying the differences in measurements in a subsample of the target population given the same set of tasks performed in a supervised environment. Validation of wearable device measurements should be conducted: (1) directly using gold standard measurements, when technically possible, and direct observation, when appropriate; and (2) indirectly, by quantifying the association with known measures of health and behavior.

Missing data is an important problem for all types of measurements. However, traditional missing data refers to missed visits when it is known that data are missing. In addition to missing visits, wearable devices have their own specific missing data problems. Indeed, even if the study participant is compliant with the visit, wearable device data can still be missing because: (1) the device malfunctions or is not properly calibrated; or (2) the study participant does not or cannot wear the device during particular times. Depending on the type of device it can be possible to detect these periods of non-wear. Indeed, in the case of glucose and ECG monitors the signal is simply lost and missing data flags can be added. Things are more complicated for activity monitors because they can be removed at any time and either placed back after a particular activity (e.g., swimming or showering) or removed for an extended period of time. In these cases, it is more difficult to estimate non-wear periods as the recorded data look similar to data collected during periods of rest or sleep. The algorithms for non-wear detection are only as good as our understanding of the data and are based on strong, albeit reasonable assumptions (e.g., during sleep there are small movements that do not occur during non-wear). However, the best strategy is to plan and implement sampling protocols designed to increase protocol adherence. A few ideas that could work in practice are: (1) discuss the importance of protocol adherence and what can and cannot be done with the wearable device; (2) follow up via phone and/or apps to remind study participants about wearing the device; (3) use a low-burden diary that records important events (e.g., sleep start and end, non-wear periods); (4) use hypoallergenic devices placed on body locations that are acceptable and comfortable for study participants (e.g., study participants are much more likely to wear wrist accelerometers for long periods of time compared to hip or chest accelerometers); and, (5) provide instructions for safe removal and re-placement of the device. Other measures can also be considered, though missing data can occur even in the best designed

studies. These data should be flagged as missing and information about the reasons for missingness should be added, when available. When it is unclear whether data are missing, an explicit protocol for data quality control should be in place and quality control flags should be added to the raw or processed data.

Wearable devices produce large amounts of data (e.g., millions of observations per day if data are sampled at higher than one-second resolution or thousands of observations if data are sampled at the minute level). This raises logistical questions about data acquisition, storage, quality control, and curation. This typically requires a dedicated analytic team that has a good understanding of the wearable computing data, the potential logistic and analytic pitfalls, and the required technical skills for data analysis. Planning can substantially improve the quality of the data handling and curation processes. Whenever possible, it is recommended to run smaller sub-studies before the beginning of the main study to identify the potential data collection and quality control problems and develop manuals of operations (MOPs) and documentation. To maximize the use of the high-density data obtained from wearable devices it is critical to have a well-trained analytic team. ECHO-DAC can provide both training and analytic support for a variety of wearable devices. It is recommended to contact the ECHO-DAC team during the planning stages, though support is also provided, whenever possible, for data that was already collected under study-specific protocols.

#### 4. Design of experiments

Data collection for wearable computing is typically guided by the design of the parent study with sensors being placed on the body at pre-determined visits. Devices can be sent back via mail or, more rarely, can be removed by the research team during a subsequent visit. This is not the only

way to collect wearable device data, and specific designs could be considered to improve the quality of the data, enhance integration with other study components, and directly address the problem of interest. Planning, building a dedicated wearable computing task force, and consulting with specialists could substantially improve the design of experiments and data quality. Below we provide a non-exhaustive list of strategies to consider during the design phase.

The most important component is to carefully define the scientific questions and variables and identify the sensor or combination of sensors that provide the data that can be used to operationalize some of the variables. This phase requires substantial scientific input and discussions among the team members while accounting for the availability of resources and feasibility of options. Conducting a thorough literature review and discussing with leading experts can substantially focus discussions and eliminate sensors and platforms that are unreliable or not well suited to answer the scientific questions.

During the planning phase many questions may still remain unaddressed either because the information is unavailable or because the technical details of data collection and storage are not completely understood. In these, as in many other, cases an exploratory sub-study could be conducted on five to thirty study participants from the study population. The lessons learned from this sub-study should be analyzed and used to improve the design of the experiment, logistics, data quality and storage, and wearing protocol (e.g., adding non-allergenic straps for wrist worn devices or a skin adhesive patch on a chest worn device.) They can also be used to rule out certain wearable devices or device combinations as well as to check the reliability and validity of device

measurements. Overall, these sub-studies provide excellent training for the team, help refine the design protocol, and substantially reduce errors during the actual study.

Wearable devices, such as accelerometers, environmental sensors, or heart monitors, cannot provide contextual information. Contextual information could be obtained using different strategies. For example, in a study of sleep it could help to have a sleep diary and/or a button on the wearable device that can indicate when the person goes to sleep or removes the device. One could also add a camera, a GPS tracker, or EMA app to obtain different measurements of the context in which data were collected. If this is not possible in the entire study due to privacy, burden, or adherence concerns, these strategies can still be used in a small sub-study. While adding contextual information increases the complexity of the study design, it also improves the quality of the data and enables examinations of person-environmental interactions.

Accounting for sampling protocol and target population is important in practice. For example, in many studies that collect wearable device data there are multiple days of observations, which creates a natural multilevel structure for individual days. Moreover, certain days (e.g., Friday through Sunday) may exhibit different data patterns due to weekend effects. In multicenter studies or studies that collect data from multiple countries or in different climates (e.g., summer versus winter or northern versus southern hemisphere) there could be large center effects. These effects can be due to large differences in the environment as well as in social and behavioral characteristics of the local culture. Another important feature is that different target populations may require different study designs and processing tools. For example, algorithms and norms developed for adults may not apply to pediatric populations. For example, the definition for moderate to vigorous

physical activity (MVPA) could be different in a population of healthy 21- to 30-year-old study participants compared to a population of children with chronic kidney disease (CKD).

Irrespective of the design of experiments, a manual of operations (MoP) is crucial to describe the data sampling protocol and provide support for the team handling the implementation. The MoP can be improved and updated as new information becomes available. However, it is highly recommended that these changes are made before the main study begins to avoid changes during the study. Discussing with experts and conducting small pre-studies could substantially improve both the MoP and the integration of the data collection team. In addition, operational checklists can help team members avoid device and data handling errors and improve communication when new study members join the team. Checklists can also be used across multiple study centers to further reduce variability, batch, and site effects.

## 5. Transforming wearable device measurements into digital biomarkers

The development of digital biomarkers from measurements obtained from wearable devices should follow the same stringent protocols used for the development of any biomarker. The added complexity is that wearable devices produce highly complex and large datasets that require a specialized set of analytic skills.

Typically, a digital biomarker is a one-number summary of the high volume of data produced by the wearable device (e.g., number of minutes in sedentary time, activity volume). Many such summaries can be produced, and each may be more or less relevant to a specific scientific problem. One of the most important characteristics of a digital biomarker is to be expressed in international



measurement units or standardized relative to a well described population. This is relevant, as many devices produce highly processed output that does not respect this simple principle. For example, activity monitors can produce a variety of activity counts, which are proprietary, are not expressed in international system of measurement units and can vary substantially between and within devices. Even the number of steps should be scrutinized, as this is the result of an internal estimation algorithm that can vary with the particular software and body location.

Replication experiments are necessary to quantify both the technical and biological variability of the digital biomarker. Reliability should be estimated and the sources of measurement variability should be quantified and reported. This could be achieved by conducting a sub-study where two activity monitors are placed on the same wrist (side by side and/or stacked and taped together). The order of the monitors can be changed to check for specific biases that occur due to device positioning. Once the reliability of the digital biomarker is understood, methods and protocols for improving the reliability and mitigating measurement error and potential sources of bias should be developed.

Validation of digital biomarkers should be conducted both internally in the study and externally in new studies. Cross-validation (leave-one-out and multi-fold) using accepted measures of prediction performance (e.g., AUC, C-index, R-square) should be routinely used for the development of prognostic digital biomarkers. Studies to evaluate the validity should also be conducted against gold standard measurements. For example, the number of steps can be measured by a human observer or calculated from a video recording either in a controlled or the free-living environment. In some situations, a gold standard may be unavailable or difficult to obtain. For

example, continuous glucose monitoring (CGM) could be validated using blood samples analyzed in the lab, but obtaining multiple blood samples that are temporally synchronized with the CGM is quite difficult. In these situations, one could consider other measurements that are known correlates of the gold standard. External validation should be conducted using other datasets and through collaboration with other research teams. The increased availability of publicly available (e.g., NHANES and UK Biobank) datasets makes them particularly well-suited for external validation (Troiano et al., 2008; Schrack et al., 2013; Smirnova et al., 2020; Leroux et al., 2020). Moreover, if the digital biomarker pipeline is open source, other research teams can replicate it and provide scientific and technical validation. Ultimately, the research community will attempt to replicate and validate promising digital biomarkers in various studies.

The association and predicted performance of digital biomarkers should be quantified both in unadjusted (single predictor) and adjusted (digital biomarker plus traditional risk factors) models for the outcomes of interest. When multiple digital biomarkers are obtained, their mutual correlation matrix and association plots (scatterplots) should be presented. Similar plots could be produced to evaluate the individual associations of digital biomarkers with health outcomes. The improvement in cross-validated prediction performance (AUC, C-index, or R-square) should be quantified when the digital biomarker is added to traditional risk factors. One could also investigate whether a combination of digital biomarkers outperforms individual biomarkers in terms of individual and added prognostic performance.

All steps involved in the process of obtaining and validating the digital biomarker should be transparent and reproducible. Ideally, a detailed vignette with human-readable software should

accompany the description of the pre-processing pipeline. This is necessary because of the complexity of the data, the many choices that are made during data processing, and the need to reproduce the same derivations on different datasets. Indeed, in many situations, published papers are insufficient to address the need for transparency and reproducibility.

Ultimately, the digital biomarker needs to be interpretable and accepted by the research community. It also needs to be easy to explain in plain terms to a wider audience to ensure that findings are translatable and applicable to wider populations. Thus, even if sophisticated methods are used for data summarization (e.g., PCA, factor analysis, clustering) simple proxies that have the same or similar prognostic performance should be explored and, if possible, used instead.

## 6. Data harmonization

Data harmonization within and between studies is necessary when a study has multiple data collection centers or multiple studies are analyzed together. This is especially important with new, abundant and unfamiliar data, such as produced by wearable devices. This requires data normalization, common study protocols and MoPs. For example, ECHO-DAC developed a MoP for Actigraph, a wearable activity monitor. Studies that plan new data collection can adopt the manual and discuss with ECHO-DAC investigators specific issues associated with their studies. As discussed before, it is important for digital biomarkers to be expressed in the same units and be accompanied by explicit definitions, open source software and instructional vignettes. Without normalization before data collection and during pre-processing, data harmonization across studies is not possible. An additional step for harmonization is to develop protocols for common data structures and ontologies (variable names and definitions). In general, collecting, pre-processing,

and storing data should be done with data integration as a goal not as an afterthought. This simple change in stated goals is a crucial step in improving efficiency and achieving the stated goals of individual studies: reproducibility, transparency, and translatability of findings.

## 7. Interventions

Wearable devices are particularly well suited for interventional studies due to their continuous, dynamic measurements. For example, in a clinical trial, wearable devices can be worn continuously before and after treatment initiation. This provides objective measures with high temporal fidelity that can be used to: (1) objectively quantify the effect of treatment at any time point after treatment initiation; (2) quantify the temporal shape of the population and individual treatment effects; (3) provide insights into subgroups that are more or less likely to have a treatment effect; and (4) quantify treatment heterogeneity in terms of recovery curves at the individual level. This level of detail is made possible by wearable devices, it substantially enhances the analytic toolbox of traditional clinical trials, and can provide more sensitive endpoints in a variety of treatment/endpoint scenarios. Consider, for example a clinical trial for back surgery. Using EMA one could continuously assess the pain level of the individual over the course of the follow up while wearable accelerometers can objectively quantify the individual physical activity intensity at any time point during months of follow up. Similar approaches can be used for many types of interventions including for birth and growth outcomes, weight loss, and academic performance. The pairing of EMA and wearable devices is poised to change the way clinical trials and interventional studies are conducted.

A different type of emerging intervention is the dynamic, or just-in-time (JIT), intervention. JIT interventions can be used to provide interventions in response to changes in the data monitored (e.g., lower physical activity, unusual cardiac signals, or sudden changes in mood) and they can be triggered either by the monitoring system (with or without a human in the loop) or by the individual who is monitored (e.g., asking for help during a mental or physical health crisis). JIT interventions can change to adapt to the needs of the individual or to improve overall health outcomes. For example, in a weight loss study, appropriate interventions can be chosen from a menu of behavioral-modification options including advice on physical activity, eating habits, and peer support.

## 8. Specific considerations in pediatric study populations

Deploying wearable devices in pediatric populations require tailored approaches informed by what is technically feasible as well as socially and behaviorally acceptable. For example, wearable activity monitors may be perceived to interfere with children's social activities, may not be allowed during particular school activities (e.g., during physical education training), or may be inappropriate to wear on the wrist for very young children. Similarly, EMA and detailed diaries may be inappropriate for young children and may need to be filled by their legal guardian instead.

Some considerations will be obvious and can be addressed before the study starts, while others will occur during the study. For example, in a pediatric study, students were not allowed to wear a wrist accelerometer during physical activities at school. This was partially addressed by the research team by providing a letter explaining that the student is part of a study and asking the school staff to allow wearing of the device whenever safety was not affected. Another example is a study of physical activity in babies under one year of age using accelerometers. Wearing the

device on the wrist was not possible or acceptable and there were important concerns about potential dermatologic reactions and interference of the device with bathing. To address some of these problems, the research team used an ankle accelerometer wrapped in a hypoallergenic waterproof material.

In many of studies privacy is an important concern and handling of personal health information (PHI) should be done carefully. Indeed, some wearable devices can collect data that can directly identify the individual. For example, GPS devices provide extraordinary levels of detail that can easily identify the home and school address of a child. Some devices may also inadvertently reveal information about the behavior of the children that parents and children may not want to be recorded. This can happen especially when the team deploying the device does not have enough experience with a specific technology. This is a particular concern as technology can change quickly, while building expertise takes substantial time and effort. Again, to mitigate some of these problems, it is recommended to run small initial sub-studies with the following goals: (1) familiarize the research team with the technology and its use in pediatric populations; (2) obtain feedback from study participants and their legal guardians; (3) change protocols accordingly.

## 9. References

Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. *Research quarterly for exercise and sport*. 2000; 71(sup2): 1-14

Washburn RA. Assessment of physical activity in older adults. *Research quarterly for exercise and sport*. 2000; 71(sup2): 79-87

Schrack JA, Zipunnikov V, Goldsmith J, et al. Assessing the “physical cliff”: detailed quantification of age-related differences in daily patterns of physical activity. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*. 2013; 69(8): 973-979.

Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Medicine and science in sports and exercise*. 2008; 40(1): 181

Leroux A.: rnhanesdata: NHANES Accelerometry Data Pipeline. <https://github.com/andrew-leroux/rnhanesdata>. R package version 1.0. Accessed June 1, 2018.

Smirnova E, Leroux A, Cao Q, Tabacu L, Zipunnikov V, Crainiceanu C, Urbanek JK. The Predictive Performance of Objective Measures of Physical Activity Derived From Accelerometry Data for 5-Year All-Cause Mortality in Older Adults: National Health and Nutritional Examination Survey 2003-2006. *J Gerontol A Biol Sci Med Sci*. 2020 Sep 16;75(9):1779-1785.

Andrew Leroux, Shiyao Xu, Prosenjit Kundu, John Muschelli, Ekaterina Smirnova, Nilanjan Chatterjee, Ciprian Crainiceanu. Quantifying the Predictive Performance of Objectively Measured Physical Activity on Mortality in the UK Biobank, *The Journals of Gerontology: Series A*