

**The Domestic Health Index:
How Wearable Devices Can Promote the Culture of Health**

January 2019

Peter Bearman

Adam Reich

Kathryn Neckerman

Interdisciplinary Center for Innovative Theories and Empirics (INCITE)

Columbia University

Draft: Please do not distribute without authors' permission

Executive summary

The proliferation of wearable technology presents an unprecedented opportunity both to measure population health in new ways and to make it more culturally salient. We envisioned what we called a “domestic health index,” or DHI, as both a promotional tool and as a valuable dataset in its own right. Over the past two years, we have consulted experts and learned more about the rapidly evolving digital health space.

As part of this exploration, we conducted a scan of the use of wearable devices and mHealth (mobile health) apps in research, with a focus on work that would inform our planning for a DHI. Conclusions include:

- Over the decade since wearable activity trackers were introduced to the market, health researchers have begun to incorporate these devices into their work. Validation research has been a priority and will remain an ongoing need as devices are upgraded and as new functionalities (e.g. measurement of sleep stages or stress) are added.
- As the use of wearable devices in research has expanded, new digital platforms have emerged to support this research. These platforms offer tools for remote enrollment, consent, and management of study participants and for the collection and integration of streams of data from wearable devices and mHealth apps. The availability of these digital tools make a DHI far more feasible.
- Although research using wearables has expanded, very few academic researchers are using this technology for population health surveillance. A primary challenge to such use is the fact that users of wearable technology and mHealth apps are not representative of the overall population.

Next, we engaged a series of issues that bear more specifically on the design of a DHI, including what health indicators to include, how participants might be recruited and retained, and – ultimately – how this initiative might be sustained. Conclusions include:

- In order to be included in the DHI, health indicators require valid and reliable measurement via wearable devices and a robust relationship to health outcomes. In the near term, step counts, minutes of moderate to vigorous physical activity, and total sleep time are most likely to meet these criteria.

- Because wearable device users are not representative of the U.S. population, in the near term the best way to get a representative sample is to recruit and send devices to study participants.
- Privacy and data security remain a significant public concern. Privacy experts caution that technological development has outstripped the existing legal framework. As long as the DHI is implemented within an academic environment, IRB regulations should ensure participant privacy. Transparency about study design and IRB protections is essential, as privacy concerns could deter participation in DHI.
- Resources will be needed to support participant recruitment, distribution of devices, storage and processing of the resulting data, and communication efforts. To aid in identification of stakeholders who might be willing to support this work, we discuss several use cases for the DHI.

In the final section of this report, we recommend specific near-term steps for the development of a DHI.

Introduction

We began this project with a relatively simple idea: that the proliferation of wearable technology presents us with an unprecedented opportunity both to measure population health in new ways and to make it more culturally salient. Our thought was that a high-frequency index, based on the aggregation of individual data from wearable technologies, could serve as the health equivalent of the S&P 500, an ever-moving and so ever-intriguing measure of our nation's health.

We envisioned what we called a “domestic health index,” or DHI, as both a promotional tool and as a valuable dataset in its own right. As a promotional tool, we imagined it being broadcast on the nightly news, syndicated in national and local newspapers, and discussed across social media. Just as individuals orient their behavior around their wearable devices in order to “get their steps” for the day, we thought, communities might orient their attention to the DHI as a way of understanding their collective well-being. As a research tool, the DHI could allow people to answer questions that are unanswerable with current indicators—allowing researchers to explore, for instance, the health consequences of certain cyclical events (Tax Day, July 4th) or unusual events (natural disasters or terrorist attacks), or how health-related behaviors change over different hours of the day, days of the week, or weeks of the year.

We conclude our two years of exploration with continued optimism about the feasibility of such an idea, but a heightened awareness of the complex set of questions that remain unanswered—and the network of relationships that must be more deeply developed—in order to translate this idea into a reality. We are also sensitive to the rapidly evolving nature of the space within which this project sits. Over the short period during which we have conducted this exploration, several new organizations, platforms, and companies have emerged. Such a dynamic field presents much opportunity, in that there are many actors who might see the appeal of an idea like the DHI. But it also poses a fair amount of risk, in that it is difficult to predict which organizations, platforms, and companies will remain standing in the near future.

This white paper first examines the current status of health research using wearable devices and mHealth apps. (Although our focus is on wearable technology, we found studies of mHealth apps to be helpful, particularly in our discussion of user characteristics, where research is relatively sparse.) Second, we focus on a series of specific questions that must be engaged as part of the design and

implementation of a DHI. We conclude with our recommendations of near-term steps for research and development should RWJF continue to be interested in the creation of a DHI.

Part 1: Current status of research using wearable device and mHealth data

As a first step, we undertook a scan of existing research that has made use of wearable technology—both to get a sense of the types of behavioral data on which such studies rely, and to understand the types of questions health researchers are currently asking using such data.

Although wearable devices such as Fitbits have been available for only about a decade, researchers have used accelerometers to measure physical activity for much longer. Typically, these devices were distributed to study participants for short periods of time – often seven days – and then retrieved so the data could be downloaded for analysis. Although many studies still use this kind of protocol, the emergence of consumer-grade activity trackers provides new ways for researchers to interact with research subjects. Because activity data is wirelessly transferred to user accounts, researchers can now retrieve data (with consent from study participants) without physically accessing the device.

Although it is clear that academic and medical researchers are excited by the potential of wearable technology, they are proceeding cautiously in incorporating this technology into health research. A major focus over the past decade has been on validation studies, as researchers have sought to understand the quality of data from consumer-grade activity trackers compared with the instruments that are more conventionally used to measure physical activity, sleep, and other health behaviors.[1] We anticipate that validation studies will be an ongoing activity as new devices with novel capabilities continue to come to market.

Researchers are beginning to take advantage of the capability for remote retrieval of data to conduct long-term monitoring of physical activity and other indicators. For instance, Diaz and colleagues recently used Fitbit devices to collect a year’s worth of data on sedentary behavior among 79 adults; the data were correlated with daily perceived stress measures collected via smartphone.[2] Huberty and colleagues used Fitbits to measure daily physical activity and sedentary time over the course of pregnancy among 80 women.[3] Researchers can also request access to data that study participants previously collected on their own. In the current wave of data collection for the well-known National Longitudinal Study of Adolescent Health (Add Health) study, researchers are asking a subset of

participants if they use activity trackers and, if so, requesting access to tracker data to compare with survey-based physical activity measures.^a

Analysis of Fitabase Research Library

To understand this research in quantitative terms, we took advantage of the research library available through Fitabase (previously SmallStepsLab), which serves as an intermediary between the wearable device company Fitbit and academic researchers. Fitbit is the most widely used fitness tracker in the United States, and these devices are also the most widely used in research and the most extensively validated.[4] Fitabase maintains an online library that tracks published papers and conference proceedings making use of data from Fitbit devices. While Fitbit is, admittedly, only one of many wearable devices, the database provides a compelling snapshot of the state of research using this kind of data—particularly given Fitbit’s role on the leading edge of wearable research.

According to the database, the number of research studies based on Fitbit data has risen exponentially over the past several years, from only 3 papers in 2012 to 258 in 2017—far faster than the spread of Fitbit devices themselves. Of the 523 studies recorded between 2012 and 2017, the vast majority of studies (387) make use of step count data, which is far more than the number of studies that make use of other kinds of data, such as sleep time (38), heart rate (21), and energy expenditure (19).

The plurality of studies (203) are validation studies, meant to determine the accuracy of Fitbit devices at measuring health-related behaviors and comparing measures taken using different types of devices, while the second largest category of studies (176) are those that measure the effect of some intervention on health-related behaviors. In many of these studies, the wearable device is used to measure the behavioral outcome of interest. In some of these studies, the wearable device is actually a part of the intervention, as when researchers are interested in studying the impact of self-monitoring on health outcomes.

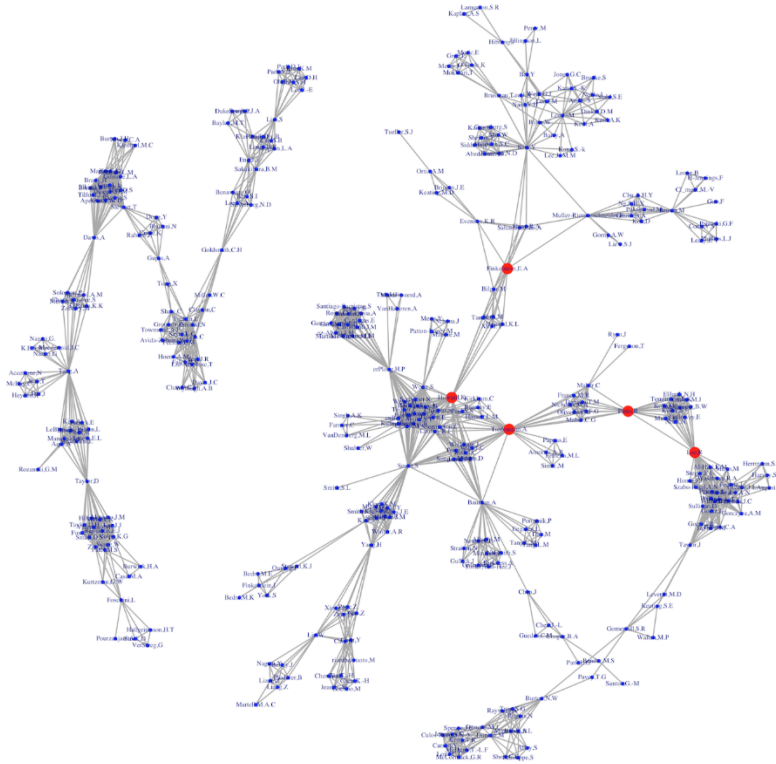
Using the Fitabase database, we created a co-author network, looking at the patterns of collaboration among those conducting research using Fitbit (and other wearable) devices. A visualization of the two largest components is below (Figure 1). Of the 2,191 authors in the database as a whole, 378 (or 17%) are a part of these two components. While validation studies are more common within the database as a whole, the authors who are most central to the co-author network (based on their betweenness centrality) tend to be doing something similar: i.e. they are studying health-related interventions

^a <https://www.rti.org/impact/national-longitudinal-study-adolescent-adult-health-add-health>

(related to elder care, diabetes, weight management, heart disease) and using measurements from wearable devices as behavioral outcomes. The top-five most central authors are highlighted in Figure 2 in red.

Such an analysis allows us to identify the small group of scholars central to the field who ought to be consulted in the development of a DHI. It also reveals the rather narrow focus of the field as it exists today, in that these most central scholars are all engaged in work concerning the measurement and improvement of health at the individual, rather than population, level.

Figure 1: Co-Authorship Network of Fitabase Articles, Two Largest Components



Large-scale e-cohort studies

Perhaps the most novel use of activity trackers is their deployment in large-scale “e-cohort” studies. These studies, enabled by advances in mobile technology and motivated in part by the precision medicine movement, seek to enroll hundreds of thousands of participants who agree to take surveys,

undergo health screenings, and give researchers access to electronic health information such as medical records and sensor data. The goals of these studies are distinct from that of the DHI: while the goal of the DHI is to monitor population health, these e-cohort initiatives seek to amass a very large sample to provide the statistical power to study how genetic factors, environmental exposures, and lifestyle factors interact to shape disease risk and response to treatments, and to expedite study recruitment for clinical trials. The e-cohort projects span the public, private, and nonprofit sectors, and range from small-scale to massive. Some of these projects are supported by federal funders, most often NIH, others by private-sector companies including health care or data science organizations. This field is rapidly evolving and it remains unclear how such projects will sustain themselves financially over the medium- to long-term.

These studies are supported by digital health platforms that allow researchers to enroll, consent, communicate with, and incentivize study participants and to retrieve and integrate a range of types of health-relevant electronic information, including activity and sleep data; measures from auxiliary devices such as smart scales or blood pressure cuffs; surveys or EMA (ecological momentary assessment) data; location or contact sensing; and electronic medical records.

All of Us

Funded by the NIH Precision Medicine initiative and implemented with a large network of participating universities and health care organizations, *All of Us* aims to recruit a research cohort of at least a million Americans. The large size is intended to ensure statistical power for study of how environmental, genetic, and lifestyle risk factors interact. After a pilot phase, national recruitment began in 2018 and is expected to continue for 5-6 years. Some participants will be recruited via health care provider organizations, while others will be enrolled directly. *All of Us* will use community outreach as well as a mobile engagement tool to reach out to communities and populations that are typically underrepresented in medical research.^b The intent is to follow participants for years, possibly for decades. Detailed individual-level data will be released only to approved researchers, but a de-identified dataset with selected measures will be accessible to the public through an online tool.

Participants will be asked to provide a range of types of data, including self-reported health and lifestyle information, access to electronic medical records, physiologic (heart rate, blood pressure) and anthropometric (height, weight) measures, and bio-specimen collection (biomarker and genetic

^b https://allofus.nih.gov/sites/default/files/allofus-initialprotocol-v1_0.pdf

information in blood, urine, and/or saliva). Although *All of Us* plans to integrate wearable activity trackers into the cohort data collection, the extent of wearable data collection is not yet clear. Initial plans call for distribution of 10,000 Fitbit devices to *All of Us* study participants; after a year of Fitbit data collection, researchers will make recommendations for how activity trackers might be used in *All of Us*.^c

Health eHeart and EUREKA

Heart eHealth, based at the University of California, San Francisco (UCSF), is a cardiovascular cohort that has enrolled more than a hundred thousand study participants since its inception in 2013. Recruitment and data collection are entirely online. Participants are asked to complete periodic online/mobile surveys and to provide access to electronic medical records; they are also invited to complete additional study components. UCSF researchers have used these data for a variety of studies, for instance examining the association between sleep quality and atrial fibrillation or relating e-cigarette use to smoking behavior and cardiopulmonary symptoms.[5, 6] Health eHeart study investigators have made innovative use of mobile sensors. For instance, a study published last year used smartwatch step count and heart rate data to train a neural network to detect atrial fibrillation.[7] Another used smartphone GPS and cell tower triangulation to identify study participants who may have been hospitalized.[8] Like most studies with volunteer subjects, Heart eHealth is not representative of the population as a whole. Comparing the Health eHeart cohort with the weighted 2013-14 National Health and Nutrition Examination Survey (NHANES) sample, Guo et al. found that the Health eHeart cohort overrepresented women, non-Hispanic whites, more highly educated people, people with cardiovascular disease and risk factors, people with more prevalent medical conditions but better self-rated health, and underrepresented current smokers.[9]

With NIH support, the Health eHeart investigators have begun to make their platform available to other researchers under a new initiative called EUREKA (initially called Health ePeople). EUREKA is a digital platform for health or medical research by academic, industry, or “citizen scientist” researchers. In some cases, investigators conduct secondary analyses using the existing cohort members; in others, the digital research administration tools are used to recruit and collect data from new participants. EUREKA’s default policy is to add the deidentified data from all studies to a “data warehouse” that is accessible to all. EUREKA is a nonprofit organization that charges researchers user fees to cover costs. Although the

^c <https://investor.fitbit.com/press/press-releases/press-release-details/2017/Fitbit-Selected-for-National-Institutes-of-Health-NIH-Precision-Medicine-Research-Program-with-The-Scripps-Research-Institute-TSRI/default.aspx>

Health eHeart study invites participants to link their activity trackers and other sensors (e.g. blood pressure) and donate their data, we do not know how many participants have done so. Based on the Guo et al. reported that among participants with complete baseline data, only 2% used a Bluetooth-enabled blood pressure device and provided at least one measurement.[9]

Health eHeart investigators have not, by and large, sought to use their data for population health surveillance. However, when we spoke with Dr. Jeffrey Olgin last year, we heard that he and his colleagues are exploring the population health uses of their cohort data, for instance to gauge the health effects of major public events.

Evidation

Evidation is a relatively new health and measurement company that gathers and analyzes health behavior data, mostly in service of collaborations with and analyses for pharmaceutical and health care companies. Like Health eHeart, Evidation has recruited a large pool of research participants through the Achievement app, which pays research participants for giving Evidation access to health-relevant data such as step counts, sleep, meditation, and food logging via apps such as Fitbit and Apple Health. Achievement health app users may be invited to participate in clinical trials and other research studies. Also like Health eHeart/EUREKA, Evidation has built a set of digital study management tools that can be used by clients and collaborators.[10] Evidation researchers have already published using the Achievement research subjects. For instance, one study found that people who are more diligent about using health trackers were more likely to lose weight.[11]

My Research Legacy

A more recent effort, led by the American Heart Association (AHA), aims to recruit 250,000 participants who will donate data – including activity tracker data – for studies of heart disease. As a first step, the AHA is recruiting 2,000 participants for a pilot study of people aged 21-49 who have survived a stroke or heart attack.^d The project was launched in 2016. The AHA is collaborating on My Research Legacy with the Broad Institute, a biomedical research center affiliated with MIT and Harvard University, and with the Marfan Foundation.^e The data will be housed in the AHA's Precision Medicine Platform, developed with Amazon Web Services.

^d <https://newsarchive.heart.org/my-research-legacy-invites-everyone-to-join-cutting-edge-research/>

^e https://www.huffingtonpost.com/entry/you-could-be-part-of-a-legacy-of-breakthrough-research_us_58e12564e4b0ca889ba1a719

Project Baseline

Project Baseline is a collaboration between the Schools of Medicine at Duke and Stanford Universities and Verily Life Sciences (formerly Google Life Sciences), which is funding the initiative.[12] Project Baseline will compile biomarker data collected during visits to study clinics as well as data from sleep and activity trackers, surveys, and diaries. The activity tracker is a custom-built “study watch” that collects heart rate and physical activity data but does not reveal these data to the study participants.^f For an initial pilot study, Project Baseline is using a quota sampling approach to recruit a diverse study population; a prospective participant can complete a profile and may be considered for inclusion if the study has not recruited sufficient numbers of participants in their demographic/medical category.^g Subject recruitment began in 2017; 2,000 had joined within the first 18 months.^h

Although these e-cohort projects are not intended for population health surveillance or research, they do offer a variety of models for how such large-scale data collection efforts can be conducted, and have developed infrastructure that could be relevant for the DHI.

Wearable devices and population health research

Although researchers are beginning to take advantage of mHealth and wearable devices in their research, use of these devices for population health research remains rare, as our review of the Fitabase database makes clear. We were unable to find a single research paper within the Fitabase database that pertains to population-level health outcomes.

The academic scholarship most closely related to the sort of index we propose is likely the recent work of Tim Althoff and his colleagues, who used smartphone data to measure activity across more than 700,000 people in 111 countries.ⁱ They find wide variation across countries in activity levels, as well as different levels of within-country variation in activity levels.[13] Among their most interesting results, the authors find that within-country activity inequality is a better predictor of national obesity rates than average activity by country. Looking across 69 cities within the United States, they find that the walkability of a city is negatively associated with such activity inequality, highlighting the importance of city planning for public health outcomes (see Figure 2). Althoff and colleagues also used wearable devices to estimate the effects of the Pokemon GO phenomenon on physical activity.[14]

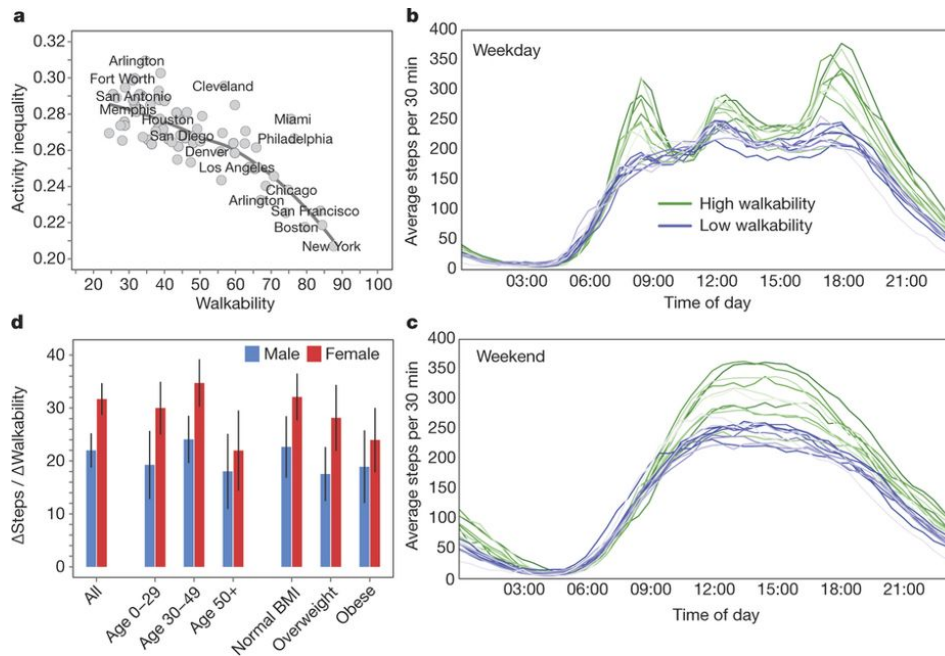
^f <https://www.projectbaseline.com/study/>

^g <https://www.projectbaseline.com/faq/index.html#am-i-eligible>

^h <https://www.nytimes.com/2018/10/18/well/live/project-baseline-cancer-prevention-heart-disease-illness.html>

ⁱ Their data is available at <http://activityinequality.stanford.edu/>

Figure 2: The Relationship Between Walkability and Activity Inequality (Althoff et al 2017)



Thus far, it has been the app and device companies that have been most active in exploiting mHealth and wearable devices for population health measurement. For instance, the mobile app Strava, an exercise app that allows people to track their running and bike riding,^j has developed a spinoff organization called Strava Metro^k that works with cities to measure and improve transportation infrastructure. Seattle’s Department of Transportation (SDOT) used Strava Metro data to examine bicycling trends in the downtown area before and after a bike lane pilot project was implemented. SDOT also used Strava data to measure bicycle volume across the city in its calculation of collision hotspots, allowing analysts to control for rider volume and thus have a more accurate understanding of risk.^l

Other app and device companies have also engaged with questions related to population health. After an earthquake hit the Bay Area at 3:20am on August 24 of 2014, the electronics company Jawbone analyzed user data to show how sleep was disrupted by the earthquake.^m Those closest to the

^j <https://www.strava.com/mobile>

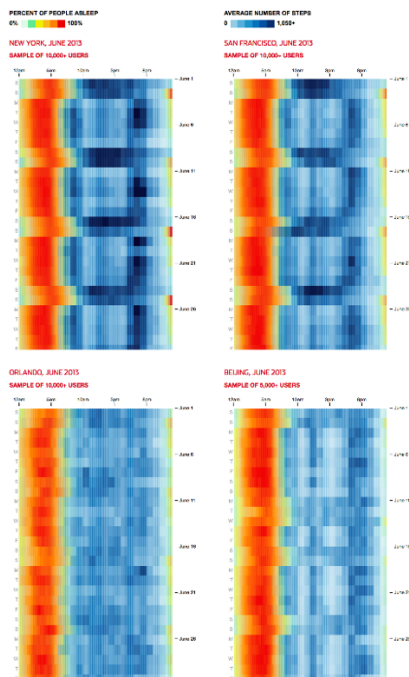
^k <https://metro.strava.com/>

^l <https://medium.com/strava-metro/seattle-dot-case-study-994545fc8117>

^m <https://www.forbes.com/sites/dandiamond/2014/08/25/how-an-earthquake-will-mess-up-your-sleep-a-jawbone-chart/#7464c6c43b3a>

earthquake’s epicenter were the most likely to be disrupted (~80% of users woke up, and many of those people did not fall back asleep that night), while those further away were less likely to be awoken. In another set of analyses quite close to the kind of index we imagine, Jawbone used wearers’ data to compare sleeping and walking patterns across major metropolitan areas.ⁿ Jawbone found that the average amount of sleep in most cities worldwide was less than the 7 hours recommended by the CDC. Users got the least amount of sleep in Tokyo, Japan (5:46 hours per night on average), and the most sleep in Melbourne, Australia (7:05 hours per night on average). In collaboration with the *Wall Street Journal*, Jawbone created graphics that display sleep and activity patterns by city, as in Figure 3. Clear from such a graphic, for instance, is that, at least among users in June of 2013, those in New York and San Francisco were engaged in healthier behaviors than those in Orlando and Beijing—they were getting more sleep, and engaged in more activity. Likewise, Fitbit has used wearers’ data to rank countries and US cities on step counts, sleep, and other health behavior indicators.^o Fitbit’s Health & Activity Index interactive tool presents averages of Fitbit-derived step counts, daily active minutes, and resting heart rate, BMI, and sleep duration for US states and major cities.^p

Figure 3: Sleep and Activity Across Four U.S. Cities, Jawbone Data



ⁿ <http://graphics.wsj.com/how-we-sleep/>

^o <https://blog.fitbit.com/fitbit-year-in-review/>

^p <https://www.fitbit.com/activity-index>

While these analyses are intriguing, we must keep their limitations in mind: they are based on data from device users and are not generalizable to the population as a whole. (Research by Althoff and colleagues is subject to the same critique.) An analysis by Evenson et al.[15] highlights this limitation. The study correlated Fitbit's state-level averages of BMI (self-reported by users), daily steps, active daily minutes, and resting heart rate with state-level health indicators derived from the 2015 Behavioral Risk Factor Surveillance System (BRFSS). Measures included BMI, maximal oxygen uptake, total minutes of physical activity per week, and minutes of vigorous physical activity per week. The authors found that correlations between the Fitbit and BRFSS state-level averages were either *fair* (Spearman rank correlation of 0.2-0.4) or *poor* (correlation of 0-0.2). These results could reflect differences in measurement (objectively-measured in Fitbit vs. self-reported in BRFSS), although it is unclear why these measurement differences would matter differently across states. It seems likely that the non-representativeness of Fitbit users contributes to these low correlations.

That said, the public health community is clearly interested in the role wearable devices could play in health surveillance. In 2014, the CDC and the American College of Sports Medicine convened a meeting to discuss surveillance of physical activity. As Fulton et al. report (p. 121), experts at the meeting suggested partnerships between public health leaders "to assist with defining how data are captured, to promote efforts to improve the validity of measures, to identify sampling strategies to improve representativeness, to develop solutions to data access and privacy concerns, and, overall, to further explore the use of these tools for surveillance." [16] Sleep researchers as well are attuned to the potential of wearable devices. Baron and colleagues note the potential of commercial activity tracker devices to contribute to our understanding of sleep patterns at the population level, although they note (p. 157) that "lack of transparency in the data processing and non-random selection of participants limits the generalizability of this data". [17]

As a signal of this interest, the CDC, the US Department of Health and Human Services, and the Public Health Agency of Canada held a Healthy Behavior Data Challenge in 2017. They solicited proposals for innovative ways to use data from wearables, social media, and mobile applications for real-time monitoring of public health. RTI International's winning submission used Amazon's Mechanical Turk to recruit subjects who used Fitbit devices and were willing to share their data; RTI also asked these subjects to take a version of the BRFSS survey, so that wearables data could be compared with survey

data on physical activity.^q Continued investment by federal funders could accelerate the use of wearable devices in health surveillance.

Characteristics of mHealth app and activity tracker users

Although activity tracker use has increased over the past decade, fewer than half of adults own these devices and even fewer use them regularly. A recent CDC study, based on a mail panel survey, reported that 12.5% of adults in the US were using a wearable activity monitor, another 12.2% had used such a device in the past, and 75.3% had never used one.[18] Adoption has been slow in other developed countries as well: in Canada and Australia, about a third of adults had ever used these devices[19, 20].

More importantly, these users of mHealth apps and activity trackers are not representative of the overall population. Among smartphone users in the US, Bhuyan et al.[21] and Krebs and Duncan[22] found that people who downloaded or used mHealth apps tended to be younger and better educated. In the CDC study cited above, current and past use of activity trackers were more common among people with higher self-reported activity levels, and among women, younger people, and college graduates.[18] Similar patterns have been found in other developed countries. In Hong Kong, for instance, health app users were younger, better educated, and more physically active.[23] Among study participants in Canada, Switzerland, Australia, and Hong Kong, activity tracker use was more common among younger people,[19, 24] women,[19] more highly educated people,[20] and people who were more physically active.[19, 20]

Most existing studies of activity trackers concern users in the general population who either acquired their trackers on their own or were given them by researchers. Workplace and insurance company programs could expand the numbers and change the social composition of activity tracker users. Corporate wellness programs are expected to be an area of growth for activity trackers. In 2016, ABI Research estimated that 44 million mHealth devices would be linked to U.S. corporate wellness programs over the next five years.^r The John Hancock Life Insurance company now offers discounts and rewards to insurance policy holders who agree to share their activity tracker data.^s We should note that it is unclear whether workplace wellness programs are cost-effective. Although a widely cited 2010 meta-analysis reported large savings for employees,[25] more recent evaluations have been more

^q <https://www.fitabase.com/blog/post/rti-international-wins-cdc-s-healthy-behavior-data-challenge/>

^r <https://www.prnewswire.com/news-releases/mhealth-wearables-help-employers-achieve-higher-corporate-wellness-participation-rates-300333978.html>

^s https://www.washingtonpost.com/business/2018/09/25/an-insurance-company-wants-you-hand-over-your-fitbit-data-so-they-can-make-more-money-should-you/?utm_term=.61d9a904f7fc

discouraging.[26] A 2013 RAND study reported improvements in some health behaviors/indicators but no statistically significant effects on health care costs.[27] A large randomized trial found no causal effects on medical expenditures, health behaviors, employee productivity, or health status in the first year of the program.[28]

In any event, we simply do not know how many people are using activity trackers obtained through (or purchased because of) workplace wellness programs, nor whether these users have a different demographic profile than other users who obtained their trackers privately. It is also unclear whether an employee wearing a tracker provided by his/her employer could also make the data available to the DHI. Lastly, employees are not always eager to participate in workplace wellness programs because of concerns about privacy and health-related discrimination. One point of contention in a 2018 West Virginia teachers' strike, for instance, was a requirement that teachers use a fitness tracker to earn points. Those who refused to wear the device or who didn't earn enough points would pay \$500 more in health insurance costs.[†] As a result, it remains unclear whether workplace wellness programs will have a substantial impact on the number or social composition of activity tracker users.

Patterns of device use and abandonment

Another challenge for researchers is the patterns of device use and abandonment. Many people who acquire an activity tracker use it only irregularly or simply stop using it. Concern about device abandonment emerged early – a 2013 survey by Endeavor Partners concluded that one-third of activity tracker owners had stopped using their devices within six months, and more than half had abandoned the devices within two years.[29] More recent studies also find high rates of device abandonment.[19, 30] There are a number of reasons people stop using their trackers. In some cases the device has been lost or no longer works, or a user is frustrated with the demands of recharging or maintaining the device. Others report that their priorities or leisure activities have changed or that they have learned what they could from tracking their activity.[30, 31] It is likely that people who abandon their devices differ systematically from those who do not, but to date our knowledge about these differences is very limited.[30] Most studies of device abandonment have focused on users' self-reported reasons for discontinuing use, or on device features that might encourage continued engagement.

Among current users, not all use their devices every day.[19] A study of German adults found a wide range of patterns of device use, with some participants using it every day while others wore it a few

[†] <https://www.csmonitor.com/Technology/2018/0315/Can-your-boss-make-you-wear-a-Fitbit>

days a week or only sporadically.[32, 33] A study of French adults found similar heterogeneity, as well as a dip in usage during July and August.[30] It is likely that regular and irregular users differ systematically. Among college students enrolled in a Fitbit-based study, for instance, students who slept more and were more physically active tended to wear their Fitbits more frequently.[34] Here as well, we know very little about how regular and irregular users differ.

Implications for DHI

Activity tracker users and non-users are different: users tend to be younger, more educated, and more physically active. This pattern is exactly what we would expect, and it is a major reason that researchers have not moved more quickly to take advantage of tracker data for population health surveillance. These relatively well-understood differences are complicated by abandonment and irregular use, about which we know less. Even if DHI researchers gave activity trackers to a random sample of the population, non-random attrition and missing data are likely to result in bias – and we do not know what kind of bias and how much.

Although these patterns are concerning, we note that bias due to sampling, attrition, and missing data are familiar problems for social scientists and epidemiologists, who can draw on a well-established set of tools to address them. For the DHI, the challenge is to find a solution that can be implemented quickly, to preserve the near-real-time character of the DHI.

Part 2: Designing A Domestic Health Index (DHI)

The second section of this paper engages a series of questions that bear more specifically on the design of a DHI, including what health indicators might be included, how participants might be recruited and retained, and – ultimately – how this initiative might be sustained.

What Should the DHI Measure?

The components and design of the DHI occasioned lively discussions among the experts we consulted. It is obvious that the health indicators we choose should have a robust association with health, and that wearable devices should provide valid and reliable measures of these indicators. Beyond these basic points, questions arose: should we combine the various indicators (e.g. step counts, sleep) together or present them separately? Should tracker-based measures be combined with other measures of health,

including high-frequency behavior measures such as purchases of cigarettes or produce or Google searches for health-relevant terms? What are the trade-offs between simplicity and comprehensiveness? What about structural inequalities such as poverty or environmental hazards or lack of access to opportunities for healthy behavior?

In this section, we start by reviewing the suitability of health indicators available from activity trackers; then we engage with these broader questions about the scope and design of the DHI.

Selecting health indicators for the DHI

Wearable activity trackers and mHealth apps offer an expanding array of health-relevant measures. Early trackers provided measures of step counts, distance traveled, calories burned, activity intensity and total sleep time. Devices that are currently available provide more elaborate measures of physical activity (e.g. step climbing, sedentary time), sleep stages and sleep efficiency, heart rate, oxygen saturation, and even women's menstrual cycles. More advanced devices offer sensing of measures related to fatigue, stress, illness, and mood or mental health using accelerometry, heart rate, co-location and communication patterns, respiration, or electrodermal activity.[35-42]

While these cutting-edge developments are exciting because they offer a more comprehensive, multidimensional picture of health, in the short term the newer measures are not feasible to include in a DHI, in part because they have received little validation, and in part because most people are not using wearables with these advanced capabilities. For now, we will focus on indicators that are widely available using existing trackers and for which we have at least some evidence about validity.

Walking (step counts) - Walking is the most common form of exercise. Almost everyone can walk, even among those who are in relatively poor health or who lack access to specialized fitness equipment or facilities. A great deal of research has linked walking with a variety of health outcomes. Prospective cohort studies consistently find a negative association between walking and cardiovascular disease, all-cause mortality.[43-45] Activity tracker measures of step counts have been extensively validated; walking is well measured with commercial-grade activity trackers;[1, 46] although trackers can sometimes overstate step counts, there is less error for step counts than for other tracker measures.[47]

Although inclusion of walking in the DHI is unlikely to be controversial, experts might disagree about how walking measures are expressed at a population level. There was consensus among the experts we consulted that it would be wiser to build an index around the percentage of people who meet physical activity guidelines than to present averages such as the mean daily step count. However, there is no

consensus guideline for how much people should walk. The popular 10,000-steps-a-day is not grounded in science; according to Tudor-Locke, an expert on walking behavior, this threshold originated with Japanese walking clubs.[48] Tudor-Locke and colleagues used NHANES data to estimate step count thresholds that matched the CDC guideline for physical activity: 150 minutes of moderate activity per week. They concluded that 8,000 steps per day translated to 30 minutes of moderate or vigorous physical activity (MVPA), and that 7,000 steps per day over the course of a week would yield the recommended 150 minutes of MVPA per week.[49] As a caveat: step counts can underestimate adherence to activity guidelines because running steps are counted as equivalent to walking steps; in physical activity measurement, running is classified as vigorous activity and gets double credit.[50] Step counts can also overstate adherence for people who walk slowly, as is often the case for people who are older or in poor health.[51] To count as moderate activity, walking must be brisk: 100+ steps per minute.[52]

Moderate and vigorous physical activity – The health benefits of physical activity are, of course, well documented; physical activity has been associated with lower risk of all-cause mortality, coronary heart disease, several kinds of cancer, type 2 diabetes, obesity, hypertension, and osteoporosis, as well as emotional and cognitive health.[53] In most validation studies, activity trackers provide acceptable measures of physical activity, although accuracy is lower for physical activity than it is for walking.[1, 46]

As with walking, it is not a simple matter to choose a threshold for population-level measurement of physical activity. As noted above, the CDC and most international health organizations recommend at least 150 minutes per week of moderate-to-vigorous intensity physical activity. But there is a dose-response relationship between physical activity and health; health benefits are evident even for people who fail to meet this threshold. Some researchers caution that emphasizing this threshold could deter exercise; inactive people may believe they will not benefit or they have failed if they fall short of what appear to be an unrealistic standard.[54] At the other end of the spectrum, researchers caution that 150 minutes of moderate activity is not enough to gain all the health benefits that could be available from physical activity. To take one example, the authors estimated that increasing physical activity from 0 to 600 MET minutes/week reduces the risk of diabetes by 2%, while increasing physical activity from 600 to 3600 MET minutes/week further reduces the risk by 19%. The CDC guideline for minimum adult activity is roughly equivalent to 600 MET minutes per week.[55]

Sedentary time – Over the past decade, some health researchers have proposed that sedentary behavior is a distinct phenomenon from the absence of physical activity and that it is independently associated

with morbidity and mortality. Research has linked sedentary behavior to the risk of chronic disease (type 2 diabetes, cardiovascular disease) and to all-cause mortality,[56, 57] to colorectal and endometrial cancer,[58] and to cardio-metabolic biomarkers such as HDL cholesterol and insulin sensitivity.[59, 60] There is evidence as well that the health risks of sedentary time are greater among people who were less active.[61, 62] In fact one review concluded that that physical activity attenuated or even eliminated the mortality risk associated with sedentary time.[63] As the American Heart Association cautions, however, evidence about the effects of sedentary time is relatively limited compared with what we know about the benefits of physical activity.[56]

Trackers have begun to measure sedentary time only recently, and as a result the validation literature is limited and inconclusive.[64-66] One difficulty in validating tracker measures of sedentary time is an ambiguity in how sedentary behavior is defined: it can be conceptualized as a matter of posture (sitting) or as a matter of motion or energy expenditure. Not surprisingly, wearable devices are better at measuring lack of motion than they are at measuring posture.[67]

As yet, there are no consensus recommendations about limits on sedentary time. The AHA wrote in 2016 (p e271), “evidence is insufficient to determine a threshold for how much sedentary behavior is too much; a linear, dose-response pattern with no identifiable threshold is a possibility.”[56] Defining a threshold for sedentary behavior is complicated by the fact that self-reported and objective measures of sedentary time diverge so greatly; self-reported measures can underestimate objective measures by 40-60%.[61, 68] Moreover, if the risks associated with sedentary time are indeed moderated by physical activity, it may be difficult to set a single standard, because the effects of sedentary behavior will differ for active and inactive people.

Sleep – The health benefits of adequate sleep have been extensively researched. A recent systematic review and meta-analysis of prospective cohort studies found that short sleep was associated with all-cause mortality, diabetes, hypertension, cardiovascular disease, coronary heart disease, and obesity, with insufficient evidence for effects on depression and dyslipidemia.[69] There is also evidence that “long sleep,” usually defined as more than 9 hours a night for adults, has health risks, including elevated rates of mortality, incident diabetes, cardiovascular disease, stroke, coronary heart disease, and obesity.[70] Other recent reviews report similar results, with both short and long sleep are associated with all-cause mortality,[71] diabetes,[72] and coronary heart disease.[73] It is important to note that sleep studies often rely on self-reported measures of sleep duration and short/long sleep; self-reported

sleep may not match objective measures, and criteria for short and long sleep can vary from study to study.

Most activity trackers now measure total sleep time, and some also measure sleep stages (e.g. amount of REM sleep) and sleep efficiency (percentage of time in bed that is spent sleeping). Activity trackers tend to over-estimate total sleep time,[1, 17] but the error is relatively small; in one study, for instance, the Fitbit Charge 2™ overestimated sleep by 9 minutes.[74] Unfortunately, validity is lower for poor sleepers because activity trackers commonly underestimate waking periods; in another study, the Fitbit Flex overestimated sleep by 6.5 minutes among good sleepers and 32.9 minutes among insomnia patients.[75] Measurement of sleep efficiency remains inadequate.[76]

For adults, the recommended sleep time is 7-9 hours of sleep per night.[77, 78] According to the CDC, more than a third of adults get less than 7 hours of sleep a night on a regular basis.[79] The Rand Corporation estimated the annual costs of sleep deprivation at 1.23 million working days and \$411 billion in economic losses.[80] Sleep deprivation has gotten more emphasis in public health messaging than the health risks that may be associated with “long sleep.” However, given the expert consensus about healthy sleep duration, it seems reasonable for the DHI to report the percentages of people whose total sleep time falls above or below the 7- to 9-hour window.

Conclusion – Our review of evidence about health indicators suggests that step counts, physical activity, and total sleep time are currently the best measures for the DHI. Inclusion of sedentary time seems to be premature given the limited validation evidence as well as the lack of quantitative guidelines about how much sedentary time is too much.

As noted above, experts we consulted advised us to express the DHI measures not as average levels but as the percentage of people exceeding a specific threshold. Implicit in the choice of thresholds is a decision about *whose* health to prioritize. Focus on the 10,000 step-count threshold is likely to be most motivating for people not too far below this level, for whom attaining it seems realistic. At a community level, if most residents fall far below 10,000 steps per day, even a successful health promotion campaign may do little to raise the proportion attaining the 10,000-step level.

There is some evidence that population health would benefit the most from an increase in activity among relatively inactive people. While the effect of physical activity on health is dose-response – the more, the better – the effect is nonlinear, with an increase in physical activity having more effect at the low end of the continuum between inactive and highly active.[81, 82] A relatively low threshold, then,

might do more to improve population health and address health disparities as well. On the other hand, setting multiple thresholds – for instance reporting the percentage of people reaching 5,000 steps and of those reaching 10,000 steps – would make the DHI relevant to a broader range of people.

Broader questions about DHI measurement

One approach for a Domestic Health Index would be to narrowly focus on an indicator like steps or sleep—something easily understandable by a general public, and clearly related to certain public health outcomes. Given consistent findings about the relationship between moderate activity and health outcomes, we are confident that a measure derived from wearable devices, such as the percentage of people within a certain geographic area who walk more than a certain number of steps per day, would be meaningfully related to public health outcomes of interest. But of course, physical activity and sleep are partial measures of public health. Our discussion considered several alternative strategies for augmenting wearables-based measures in the DHI.

To make the DHI a more robust measure of health, several experts suggested, we might include high-frequency indicators beyond those available from wearable devices: indicators that might include point-of-sale data about produce and cigarettes from local retailers, as well as Google searches or Twitter mentions related to health outcomes. Purchase data would be more suitable for a weekly index than a daily one; many people buy produce once a week but consume it daily. But in principle, data about produce and cigarette purchases – and perhaps fast food purchases as well – are relatively high-frequency data about health behaviors and thus seem quite consistent with the goal of the DHI. If these data can be obtained quickly enough to incorporate into a near-real-time measures, they would make sense to add to the DHI.

We also discussed whether the DHI should include measures of health outcomes and of social determinants of health. Many existing health indices collapse a number of different health indicators into one commensurable health score. For instance, the Blue Cross Blue Shield Health Index^u uses health claims data from more than 40 million BCBS members, links these claims data to more than 200 health conditions that impact patients' predicted quality-adjusted life years, and aggregates individual-level data up to the level of the county. Each county, then, is given a score from 0 to 1, where the meaning of the number is associated with the extent to which people in that area are predicted to live at the maximum number of quality-adjusted life-years. The formula for the index is proprietary, created

^u <https://www.bcbs.com/the-health-of-america/health-index>.

by BCBM with research assistance from the University of Washington’s Institute for Health Metrics and Evaluation. RWJF’s County Health Rankings distill a broader range of variables, including health behaviors, health outcomes, social determinants, and health care resources, into county-level (within state) rankings.^v The County Health Rankings, based at the University of Wisconsin, are relatively transparent: one can easily disaggregate the rankings into their component parts, looking in depth at the health indicators of which the ranking consists and how the rankings are calculated.

The primary advantage of such indices is that they provide a more complete portrait of a community’s health than a single behavioral indicator like steps or sleep ever could. The primary limitation is that such a complete portrait is only partially related to proximate health behaviors that individuals and communities might change. That is, it is far easier to imagine a community orienting its behavior around an aggregate measure of steps than around trying to improve its county health ranking.

Although we acknowledge the value of existing health indices, we are not inclined to try to replicate them by adding other health measures to the DHI. Instead, we might envision correlating the DHI with the BCBS index as part of a validation exercise: if the DHI is indeed capturing community-level healthy behavior, it should be associated with the BCBS measures of health outcomes. In addition, we might envision providing county-level DHI measures for inclusion in the County Health Rankings, complementing the measure of leisure-time physical activity that is already included in the Rankings.

This relates to another question that we spent quite a lot of time discussing: namely, to what extent ought a high-frequency health index include factors important to public health that change slowly. RWJF’s County Health Rankings include measures of the physical environment, social and economic factors, and clinical care measures, almost all of which are relatively stable. The measures for which a high-frequency index might be most interesting—health behaviors we could measure using wearable devices like walking and sleeping; and health behaviors we could measure at high-frequency in other ways, like tobacco and produce purchases—are necessarily rather narrow, individual-level measures of health behavior. One might argue that, by focusing on these behavioral measures alone, one risks occluding the structural determinants of health and reinforcing the idea that health is the responsibility of the individual alone.

On the other hand, with the kind of high-frequency, behavioral index we are proposing, one might also be able to show how the structural determinants of health are meaningfully related to differences in

^v www.countyhealthrankings.org.

health behaviors both cross-sectionally (i.e. across counties) or longitudinally (i.e. within the same place after some exogenous shock). Moreover, since health behaviors are the factors that are easiest for any individual to change, it may actually make sense for a public-facing health index to privilege them.

One idea we have considered, in conversation with Sammy Zahran, is to create a fast-moving health index that might “nowcast” slower-moving health outcome measures like those captured in the County Health Rankings. In other words, the goal would be to have a high-frequency estimate of measures that we typically capture irregularly (i.e. predicted annual poor physical health days or poor mental health days within a particular geographic unit). While Zahran is interested primarily in nowcasting crime rates, he pointed us towards literatures regarding nowcasting economic indicators like GDP [83, 84] as well as a range of other social phenomena including disease outbreaks.[85] The idea, most basically, is to identify which high-frequency variables (including but not limited to wearables) are most strongly associated with changes in a slow-moving outcome of interest between two previous time periods, and use these models, in turn, to “nowcast” the slow-moving outcome between the latter period and the present. The advantage of this approach would be that such a model could use high-frequency data in the service of an index that is understood to be a robust measure of health. The downside is, again, that such a measure might be difficult to interpret and that, and that it might not show much variation given the structural drivers of health.

Because of its relative simplicity, and our anticipation of its ability to orient the attention of both the public and policymakers, we continue to believe a simple set of behavioral indicators, such as the percent of population taking 5,000 and 10,000 steps, is the right place for a DHI to start.

Whom Should the DHI Measure?

Ideally, DHI wearable data ought to be collected from a representative sample of adults in the United States.

As discussed in Part 1, current users of wearable devices are not representative of the population. They are younger, more affluent and educated, and more physically active. Over time, of course, patterns of wearable use may change. A drop in price might make wearables more attractive to lower-income populations, for instance. Today’s digital natives are tomorrow’s senior citizens. Until patterns of wearable use change, however, the best way to elicit activity tracker data from a representative sample of adults is to distribute wearable devices to a randomly selected sample of the population.

Distribution of wearable devices, while more expensive than tapping data from privately-purchased devices, would have a significant benefit: all study participants would be using the same device, enhancing the comparability of the measures. Validation studies find differences across different device brands and models in the mean and standard deviation of health indicators.

Such an approach would face at least two challenges. First, the wearable device might alter the behavior of the study participant. Activity trackers are commonly used in intervention studies; in some cases, the device itself is the only intervention. That said, the impact of the device on behavior may not be large. Some researchers suggest that devices by themselves have a modest effect at best, and that social or material incentives are needed to engage users and motivate behavior change.[86, 87] But even a small effect, especially one moderated by demographic characteristics, might bias results. One possible solution is to suppress feedback to the user, as Project Baseline does, though it is unclear how that would affect user motivation to wear the device, and pilot work is needed to assess study participant incentives.

The other challenge is the high rates of device abandonment and the patterns of irregular device use discussed earlier. Participant incentive schemes are likely to be part of the solution to this problem. Because technical problems often discourage device use, user-friendly and multimodal tech support for study participants is also likely to be crucial. This is especially important for those who are new to the technology.

Privacy and data security

Attention to privacy and data security are likely to be critical to study participant recruitment and retention. Surveys indicate that people *are* concerned about privacy and data security, but do not often take steps to protect their digital lives – a situation called the “privacy paradox.” People report being more willing to share their data for research,[88, 89] which may reflect altruism or greater trust in academic institutions. But privacy protection is essential: In a survey about data sharing, 52% said they would share their data if they were assured privacy, compared with only 7% who said they would share their data if they were compensated for it (15% said they would not share data unless both conditions were met.)[88]

Legal experts and privacy advocates have raised concerns about the implications of widespread health monitoring. As Montgomery and colleagues write, advertisers already use “Big Data” to collect and integrate information about individuals from multiple sources. Wearable devices offer new tantalizing

new possibilities to data brokers and marketers: “According to a recent survey conducted for a leading digital e-commerce marketing firm, the ‘key benefit of wearables will be as a source of very granular data insights and also new types of behavioral and usage data.’”[90] As long as the DHI is implemented within an academic environment, IRB regulations should ensure that participant privacy is preserved. It is vital that the DHI consent provide rigorous privacy protection and transparent communication with study participants.

Potential stakeholders and use cases

A DHI using wearable activity trackers is now technologically feasible; the largest hurdle is likely to be cost. Resources will be needed to support participant recruitment, device distribution, storage and processing of the resulting data, and communication and dissemination. To help identify stakeholders in addition to RWJ who might be willing to support this work, we briefly discuss use cases for the DHI.

Federal health surveillance

Federal agencies spend millions a year on health surveillance efforts, including the BRFSS, the National Health Interview Survey (NHIS), and NHANES, along with many other federal or federally-funded surveys that include information on health behavior, health status, and disability. Although these surveys provide extraordinarily valuable information about health, for the most part they provide only partial measures of physical activity. While activity trackers collect activity information across life domains and activities such as work, home, leisure, and transportation, many survey questions ask only about one. For instance, BRFSS asks about the two kinds of physical activity participants engage in most often, outside of work, and then asks for the frequency and average time spent in these activities.[91] The National Household Transportation Survey asks about the number of walking trips (i.e. transportation). Moreover, self-reported measures of physical activity are often inflated.[92]

As noted above, public health researchers have already expressed interest in the potential of wearable devices for health surveillance. A 2015 Call to Action from the Surgeon General’s office is an example of this continuing interest; one goal of this Call to Action was to “fill surveillance, research, and evaluation gaps related to walking and walkability.” [92] If the CDC became committed to the use of wearable devices for health surveillance, it seems likely that device distribution and data collection would be folded into an existing survey – for instance, a sample of BRFSS study participants could be recruited into wearable data collection, leveraging the existing research infrastructure for these large surveys. In this instance, DHI researchers might reach an agreement with CDC to analyze and disseminate the health

index using its data. In the short term, there is a role for university-based researchers to conduct pilot work on the research design and logistical challenges discussed above.

State and local health surveillance

There is keen interest in community levels of health and wellbeing, as evidenced by the success of the County Health Rankings. In recent years, cities and states have become more active in gathering data on their populations, using surveys and compiling data from the Census Bureau's American Community Survey. For example, with support from Bloomberg Philanthropies Mayors Challenge, Santa Monica, California, worked with RAND Corporation and the New Economics Foundation to launch the Wellbeing Project, which collected and compiled a variety of data on health, economic opportunity, community life, education, and wellbeing in the city.^w As part of this initiative, the City of Santa Monica is working with RAND and Fitabase on a pilot project in which 200 residents of one of the city's zip codes were given a Fitbit device. The purpose of this project is to learn more about patterns of physical activity in the city, although it may, of course, also prompt city residents to walk or exercise more.^x Likewise, an excellent candidate city would be Richmond, California, where Jason Corburn (a former RWJ HSS scholar and past RWJ Health Investigator) has led a novel program that requires all city decisions (from building sidewalks, to repairing potholes, to providing shelter for the homeless) to be made against the backdrop of their population health impacts.

As part of our exploration of potential stakeholders, we reviewed the websites of city and county health departments in a sampling of small cities, those about 100,000 in population. These health departments engaged in a variety of health surveillance efforts, sometimes estimating city or county health statistics from national surveys such as BRFSS, sometimes conducting their own small-scale data collection. Monitoring of infectious disease and vital statistics was routine, but local health officials lacked information on physical activity, even while they noted the importance of diabetes, heart disease, and other chronic diseases linked to physical activity. These small city and county governments may be at the "sweet spot" for a local-level DHI: while they lack the resources to do much of their own public health monitoring, they are at the nexus of a myriad of decisions that shape population health, including those about public health screenings, health promotion campaigns, and land use, transportation, and parks and recreation. Access to DHI data, especially with enough spatial resolution for cross-

^w <https://wellbeing.smgov.net/>

^x <https://www.fitabase.com/blog/post/fitabase-santa-monica-project/>

neighborhood comparisons, could be an important tool to identify needs and assess the impact of interventions. For these reasons, small city and county governments may be particularly strategic partners for DHI pilots.

Private-sector partnerships

The most obvious private-sector partners are device manufacturers such as Fitbit and Garmin, who would, it would seem, gain valuable publicity from supporting the DHI. Such partners would bring a depth of technical experience as well as financial resources. Although we would be open to working with other private-sector partners, we anticipate that some would be motivated primarily by the prospect of access to user data. For reasons discussed above, we feel it is crucial to protect user privacy.

3. Next steps

We envision the development and dissemination of the DHI in several phases, each of which answers specific questions about the design, feasibility, and value of this kind of index.

(1) Phase 1: First pilot

The primary intent of this pilot is to answer questions about the recruitment methods, study management tools, types of devices, and incentives needed to attract and retain a representative sample of DHI participants at scale. We envision a series of trials that test alternative methods of initial contact, device distribution and subject training, ongoing subject contact, incentive levels and disbursement, and so on. Many DHI participants will be “low cost” because they are tech savvy and/or because their daily routines are relatively stable. Others, who are less familiar or comfortable with technology, or whose lives are more chaotic, will need more support. As part of these pilots, we envision testing different ways of messaging about the social value of donating data and about the ways that people can represent and contribute to knowledge about their communities.

This pilot should also consider alternative ways to address the dilemma noted earlier about subject reactivity: the concern that giving people fitness trackers will change their behavior. For instance, pilot subjects could be given trackers that suppress self-monitoring for an initial period of time, allowing researchers to compare activity levels before and after subjects have access to their data; such a test would also provide information about whether people will wear a device even if they don’t get any

information from it. DHI participants might, of course, change their behavior because they know they're being monitored, even if they have no access to their data; we know of no way to avoid this.

Although it is difficult to estimate costs for such a pilot, since the number of distinct trials is unknown (and would best be estimated off of preliminary data), we could expect that substantial funds, at least \$350,000, would be required.

(2) Phase 2: Second pilot

The goals of the second pilot phase are to test a near-final version of the recruitment and study management procedures and to begin to publicize the DHI and identify stakeholders for taking the index to scale. We envision a competition for city and county governments who wish to host a pilot, with the idea of selecting 6-8 sites across a diverse range of communities. A key priority for site selection is to find creative and impactful proposals for using the data to promote a culture of health and develop local interventions. Even if the DHI is implemented nationally rather than on a city-by-city or county-by-county basis, one motivation for this index is to give communities actionable data about the health of their residents. In addition, we would want sites to commit to (1) work with the research team to get community buy-in, (2) provide extensive and candid evaluation of the roll-out, and (3) serve as ambassadors about the DHI to other communities within their state or region. For 8 sites, we would anticipate a budget of at least \$500,000.

(3) Phase 3: Planning for launch

The final phase of the process is primarily an exercise in planning, budgeting, and building a team of partners and stakeholders. The hope is that the pilots would yield a workable model for DHI data collection, including staffing, technology, and participant incentives, allowing researchers to develop an efficient plan for implementation. The second round of pilots should highlight the value of the DHI for health surveillance and intervention, and more broadly for promoting a culture of health, helping researchers identify public and/or private entities who will make a commitment to supporting the index.

References

1. Evenson, K.R., M.M. Goto, and R.D. Furberg, *Systematic review of the validity and reliability of consumer-wearable activity trackers*. International Journal of Behavioral Nutrition and Physical Activity, 2015. **12**(1): p. 159.
2. Diaz, K.M., et al., *The Influence of Daily Stress on Sedentary Behavior: Group and Person (N of 1) Level Results of a 1-Year Observational Study*. 2018. **80**(7): p. 620-627.
3. Huberty, J.L., et al., *Trajectories of objectively-measured physical activity and sedentary time over the course of pregnancy in women self-identified as inactive*. Preventive Medicine Reports, 2016. **3**: p. 353-360.
4. Henriksen, A., et al., *Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables*. J Med Internet Res, 2018. **20**(3): p. e110.
5. Christensen, M.A., et al., *Sleep characteristics that predict atrial fibrillation*. Heart Rhythm, 2018.
6. Wang, J.B., et al., *Cigarette and e-cigarette dual use and risk of cardiopulmonary symptoms in the Health eHeart Study*. PLOS ONE, 2018. **13**(7): p. e0198681.
7. Tison, G.H., et al., *Passive detection of atrial fibrillation using a commercially available smartwatch*. JAMA Cardiology, 2018. **3**(5): p. 409-416.
8. Nguyen Kaylin, T., et al., *Smartphone-Based Geofencing to Ascertain Hospitalizations*. Circulation: Cardiovascular Quality and Outcomes, 2017. **10**(3): p. e003326.
9. Guo, X., et al., *Volunteer Participation in the Health eHeart Study: A Comparison with the US Population*. Scientific Reports, 2017. **7**: p. 1956.
10. Quisel, T., et al., *Collecting and Analyzing Millions of mHealth Data Streams*, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017, ACM: Halifax, NS, Canada. p. 1971-1980.
11. Pourzanjani, A., T. Quisel, and L. Foschini, *Adherent Use of Digital Health Trackers Is Associated with Weight Loss*. PLOS ONE, 2016. **11**(4): p. e0152504.
12. Manos, D., *Rocking the Baseline: Verily, Duke, and Stanford Aim to Make Medicine More Predictive with a New Baseline Study*. Clinical OMICs, 2017. **4**(3): p. 3-4.
13. Althoff, T., et al., *Large-scale physical activity data reveal worldwide activity inequality*. Nature, 2017. **547**: p. 336.
14. Althoff, T., R.W. White, and E. Horvitz, *Influence of Pokémon Go on Physical Activity: Study and Implications*. J Med Internet Res, 2016. **18**(12): p. e315.
15. Evenson, K.R., F. Wen, and R.D. Furberg, *Assessing Validity of the Fitbit Indicators for U.S. Public Health Surveillance*. American Journal of Preventive Medicine, 2017. **53**(6): p. 931-932.
16. FULTON, J.E., et al., *Strategic Priorities for Physical Activity Surveillance in the United States*. 2016. **48**(10): p. 2057-2069.
17. Baron, K.G., et al., *Feeling validated yet? A scoping review of the use of consumer-targeted wearable and mobile technology to measure and improve sleep*. Sleep Medicine Reviews, 2018. **40**: p. 151-159.

18. Omura, J.D., et al., *National physical activity surveillance: Users of wearable activity monitors as a potential data source*. Preventive Medicine Reports, 2017. **5**: p. 124-126.
19. Macridis, S., et al., *Consumer physical activity tracking device ownership and use among a population-based sample of adults*. PLOS ONE, 2018. **13**(1): p. e0189298.
20. Alley, S., et al., *Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey*. BMJ Open, 2016. **6**(7): p. e011243.
21. Bhuyan, S.S., et al., *Use of Mobile Health Applications for Health-Seeking Behavior Among US Adults*. Journal of Medical Systems, 2016. **40**(6): p. 153.
22. Krebs, P. and T.D. Duncan, *Health App Use Among US Mobile Phone Owners: A National Survey*. JMIR mHealth uHealth, 2015. **3**(4): p. e101.
23. Shen, C., et al., *Health App Possession Among Smartphone or Tablet Owners in Hong Kong: Population-Based Survey*. JMIR Mhealth Uhealth, 2017. **5**(6): p. e77.
24. Seifert, A., et al., *The use of mobile devices for physical activity tracking in older adults' everyday life*. DIGITAL HEALTH, 2017. **3**: p. 2055207617740088.
25. Baicker, K., D. Cutler, and Z. Song, *Workplace Wellness Programs Can Generate Savings*. Health Affairs, 2010. **29**(2): p. 304-311.
26. McIntyre, A., et al., *The Dubious Empirical and Legal Foundations of Wellness Programs*. Health-Maxtric: The Journal of Law-Medicine, 2017. **27**(1): p. 59-80.
27. Mattke, S., et al., *Workplace Wellness Programs Study: Final Report*. 2013, RAND Corporation: Santa Monica, CA.
28. Jones, D., D. Molitor, and J. Reif, *What do workplace wellness programs do? Evidence from the Illinois Workplace Wellness Study*, in *Working Paper*, NBER, Editor. 2018.
29. Endeavor Partners, *Inside Wearables: How the science of human behavior change offers the secret to long-term engagement*, in *White Paper Series*. 2014.
30. Hermsen, S., et al., *Determinants for Sustained Use of an Activity Tracker: Observational Study*. JMIR Mhealth Uhealth, 2017. **5**(10): p. e164.
31. Maher, C., et al., *Users' experiences of wearable activity trackers: a cross-sectional study*. 2017. **17**(1): p. 880.
32. Meyer, J., et al. *Exploring Longitudinal Use of Activity Trackers*. in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. 2016.
33. Meyer, J., et al., *Identification and Classification of Usage Patterns in Long-Term Activity Tracking*, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, ACM: Denver, Colorado, USA. p. 667-678.
34. Faust, L., et al., *Exploring Compliance: Observations from a Large Scale Fitbit Study*, in *Proceedings of the 2nd International Workshop on Social Sensing*. 2017, ACM: Pittsburgh, PA, USA. p. 55-60.
35. Madan, A., et al., *Social sensing for epidemiological behavior change*, in *Proceedings of the 12th ACM international conference on Ubiquitous computing*. 2010, ACM: Copenhagen, Denmark. p. 291-300.
36. Erik, R. and D.C. Gari, *A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses*. Physiological Measurement, 2018. **39**(5): p. 05TR01.
37. Ghandeharioun, A., et al. *Objective assessment of depressive symptoms with machine learning and wearable sensors data*. in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. 2017.
38. Gjoreski, M., et al., *Monitoring stress with a wrist device using context*. Journal of Biomedical Informatics, 2017. **73**: p. 159-170.
39. Hwang, S. and S. Lee, *Wristband-type wearable health devices to measure construction workers' physical demands*. Automation in Construction, 2017. **83**: p. 330-340.

40. Marzano, L., et al., *The application of mHealth to mental health: opportunities and challenges*. The Lancet Psychiatry, 2015. **2**(10): p. 942-948.
41. Mohr, D.C., M. Zhang, and S.M. Schueller, *Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning*. Annual Review of Clinical Psychology, 2017. **13**(1): p. 23-47.
42. Garcia-Ceja, E., V. Osmani, and O. Mayora, *Automatic Stress Detection in Working Environments From Smartphones and Accelerometer Data: A First Step*. IEEE Journal of Biomedical and Health Informatics, 2016. **20**(4): p. 1053-1060.
43. Hamer, M. and Y. Chida, *Walking and primary prevention: a meta-analysis of prospective cohort studies*. British Journal of Sports Medicine, 2008. **42**(4): p. 238.
44. Dwyer, T., et al., *Objectively Measured Daily Steps and Subsequent Long Term All-Cause Mortality: The Tasped Prospective Cohort Study*. PLOS ONE, 2015. **10**(11): p. e0141274.
45. Klenk, J., et al., *Objectively Measured Walking Duration and Sedentary Behaviour and Four-Year Mortality in Older People*. PLOS ONE, 2016. **11**(4): p. e0153779.
46. Ferguson, T., et al., *The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study*. International Journal of Behavioral Nutrition and Physical Activity, 2015. **12**(1): p. 42.
47. O'Connell, S., G. ÓLaighin, and L.R. Quinlan, *When a Step Is Not a Step! Specificity Analysis of Five Physical Activity Monitors*. PLOS ONE, 2017. **12**(1): p. e0169616.
48. Tudor-Locke, C. and D.R. Bassett, *How Many Steps/Day Are Enough?* Sports Medicine, 2004. **34**(1): p. 1-8.
49. Tudor-Locke, C., et al., *Accelerometer steps/day translation of moderate-to-vigorous activity*. Preventive Medicine, 2011. **53**(1): p. 31-33.
50. McConnell, M.V., et al., *Mobile Health Advances in Physical Activity, Fitness, and Atrial Fibrillation: Moving Hearts*. Journal of the American College of Cardiology, 2018. **71**(23): p. 2691-2701.
51. White, D.K., et al., *Walking to Meet Physical Activity Guidelines in Knee Osteoarthritis: Is 10,000 Steps Enough?* Archives of Physical Medicine and Rehabilitation, 2013. **94**(4): p. 711-717.
52. Tudor-Locke, C., et al., *How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review*. British Journal of Sports Medicine, 2018. **52**(12): p. 776.
53. U.S. Department of Health and Human Services, *Physical Activity Guidelines for Americans*. 2018, U.S. Department of Health and Human Services.; Washington, DC.
54. Warburton, D.E.R. and S.S.D. Bredin, *Health benefits of physical activity: a systematic review of current systematic reviews*. Current Opinion in Cardiology, 2017. **32**(5): p. 541-556.
55. Kyu, H.H., et al., *Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013*. BMJ, 2016. **354**: p. i3857.
56. American Heart Association, *Sedentary Behavior and Cardiovascular Morbidity and Mortality: A Science Advisory From the American Heart Association*. Circulation, 2016. **134**(13): p. e262-e279.
57. Koster, A., et al., *Association of Sedentary Time with Mortality Independent of Moderate to Vigorous Physical Activity*. PLOS ONE, 2012. **7**(6): p. e37696.
58. Kerr, J., C. Anderson, and S.M. Lippman, *Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence*. The Lancet Oncology, 2017. **18**(8): p. e457-e471.
59. Powell, C., et al., *The cross-sectional associations between objectively measured sedentary time and cardiometabolic health markers in adults – a systematic review with meta-analysis component*. Obesity Reviews, 2018. **19**(3): p. 381-395.

60. Brocklebank, L.A., et al., *Accelerometer-measured sedentary time and cardiometabolic biomarkers: A systematic review*. Preventive Medicine, 2015. **76**: p. 92-102.
61. Stamatakis, E., et al., *Is the time right for quantitative public health guidelines on sitting? A narrative review of sedentary behaviour research paradigms and findings*. British Journal of Sports Medicine, 2018: p. bjsports-2018-099131.
62. Biswas, A., et al., *Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: A systematic review and meta-analysis*. Annals of Internal Medicine, 2015. **162**(2): p. 123-132.
63. Ekelund, U., et al., *Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women*. The Lancet, 2016. **388**(10051): p. 1302-1310.
64. Sanders, J.P., et al., *Devices for Self-Monitoring Sedentary Time or Physical Activity: A Scoping Review*. J Med Internet Res, 2016. **18**(5): p. e90.
65. Gomersall, R.S., et al., *Estimating Physical Activity and Sedentary Behavior in a Free-Living Context: A Pragmatic Comparison of Consumer-Based Activity Trackers and ActiGraph Accelerometry*. J Med Internet Res, 2016. **18**(9): p. e239.
66. Dominick, G.M., et al., *Physical Activity Assessment Between Consumer- and Research-Grade Accelerometers: A Comparative Study in Free-Living Conditions*. JMIR Mhealth Uhealth, 2016. **4**(3): p. e110.
67. ROSENBERGER, M.E., et al., *Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices*. Medicine & Science in Sports & Exercise, 2016. **48**(3): p. 457-465.
68. Ku, P.-W., et al., *A cut-off of daily sedentary time and all-cause mortality in adults: a meta-regression analysis involving more than 1 million participants*. BMC Medicine, 2018. **16**(1): p. 74.
69. Itani, O., et al., *Short sleep duration and health outcomes: a systematic review, meta-analysis, and meta-regression*. Sleep Medicine, 2017. **32**: p. 246-256.
70. Jike, M., et al., *Long sleep duration and health outcomes: A systematic review, meta-analysis and meta-regression*. Sleep Medicine Reviews, 2018. **39**: p. 25-36.
71. Yin, J., et al., *Relationship of Sleep Duration With All-Cause Mortality and Cardiovascular Events: A Systematic Review and Dose-Response Meta-Analysis of Prospective Cohort Studies*. Journal of the American Heart Association, 2017. **6**(9): p. e005947.
72. Shan, Z., et al., *Sleep Duration and Risk of Type 2 Diabetes: A Meta-analysis of Prospective Studies*. Diabetes Care, 2015. **38**(3): p. 529.
73. Wang, D., et al., *Sleep duration and risk of coronary heart disease: A systematic review and meta-analysis of prospective cohort studies*. International Journal of Cardiology, 2016. **219**: p. 231-239.
74. de Zambotti, M., et al., *A validation study of Fitbit Charge 2™ compared with polysomnography in adults*. Chronobiology International, 2018. **35**(4): p. 465-476.
75. Kang, S.-G., et al., *Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers*. Journal of Psychosomatic Research, 2017. **97**: p. 38-44.
76. Mantua, J., N. Gravel, and M.R. Spencer, *Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography*. Sensors, 2016. **16**(5).
77. Badr, M.S., et al., *Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society*. Journal of Clinical Sleep Medicine, 2015. **11**(06): p. 591-592.
78. Hirshkowitz, M., et al., *National Sleep Foundation's sleep time duration recommendations: methodology and results summary*. Sleep Health, 2015. **1**(1): p. 40-43.

79. Liu, Y., et al., *Prevalence of Healthy Sleep Duration among Adults — United States, 2014*. Morbidity and Mortality Weekly Report, 2016. **65**(6): p. 137-141.
80. Hafner, M., et al. *Why Sleep Matters-The Economic Costs of Insufficient Sleep: A Cross-Country Comparative Analysis*. Rand health quarterly, 2017. **6**, 11.
81. MANDIC, S., et al., *Characterizing Differences in Mortality at the Low End of the Fitness Spectrum*. 2009. **41**(8): p. 1573-1579.
82. Matthews, C.E., et al., *Accelerometer-measured dose-response for physical activity, sedentary time, and mortality in US adults*. The American Journal of Clinical Nutrition, 2016. **104**(5): p. 1424-1432.
83. Aastveit, K.A., et al., *Nowcasting GDP in Real Time: A Density Combination Approach*. Journal of Business & Economic Statistics, 2014. **32**(1): p. 48-68.
84. Giannone, D., L. Reichlin, and D. Small, *Nowcasting: The real-time informational content of macroeconomic data*. Journal of Monetary Economics, 2008. **55**(4): p. 665-676.
85. Schmidt Charles, W., *Trending Now: Using Social Media to Predict and Track Disease Outbreaks*. Environmental Health Perspectives, 2012. **120**(1): p. a30-a33.
86. Patel, M.S., D.A. Asch, and K.G. Volpp, *Wearable devices as facilitators, not drivers, of health behavior change*. JAMA, 2015. **313**(5): p. 459-460.
87. Sullivan, A.N. and M.E. Lachman, *Behavior Change with Fitness Technology in Sedentary Adults: A Review of the Evidence for Increasing Physical Activity*. Frontiers in Public Health, 2017. **4**(289).
88. Chen, J., A. Bauman, and M. Allman-Farinelli, *A Study to Determine the Most Popular Lifestyle Smartphone Applications and Willingness of the Public to Share Their Personal Data for Health Research*. Telemedicine and e-Health, 2016. **22**(8): p. 655-665.
89. Seifert, A., M. Christen, and M. Martin, *Willingness of Older Adults to Share Mobile Health Data with Researchers*. GeroPsych, 2018. **31**(1): p. 41-49.
90. Montgomery, K., J. Chester, and K. Kopp, *Health Wearables: Ensuring Fairness, Preventing Discrimination, and Promoting Equity in an Emerging Internet-of-Things Environment*. Journal of Information Policy, 2018. **8**: p. 34-77.
91. Watson, K.B., et al., *U.S. Adults' Participation in Specific Activities: Behavioral Risk Factor Surveillance System—2011*. Journal of Physical Activity and Health, 2015. **12**(6 Suppl 1): p. S3-S10.
92. Centers for Disease Control and Prevention, *Status Report for Step It Up! The Surgeon General's Call to Action to Promote Walking and Walkable Communities*. 2017, Centers for Disease Control and Prevention, US Dept of Health and Human Services: Atlanta GA.