



## Population-Scale Pervasive Health

Tim Althoff, Stanford University

Our everyday behavior is critical to our health. An estimated 60 percent of the human health condition is determined by behavioral factors—including exercise, sleep, and diet—as well as social and environmental factors.<sup>1</sup> Historically, these behavioral, social, and environmental factors have been difficult to measure and quantify. Scientists and clinicians typically relied on (guided) self-reports, which are subjective and often biased. Additionally, the assessments have been typically limited to participants recalling static health information (such as their weight or health conditions) and summarizing health-related behaviors and activities over limited time periods. These measurement limitations have led to reduced ecological validity and often a sparseness of data about highly dynamic health-related behaviors, such as physical activity.

Pervasive computing and pervasive health research could transform this landscape and fill in the measurement “gaps.” Ubiquitous sensors, both in the environment and in our personal devices, clothing, and bodies, can continuously collect data, allowing for dynamic measurement and more nuanced and robust investigation of health-related behaviors, activities, and physiological signals from our bodies. Connecting this data to health outcomes could unveil a great deal of which behaviors are predictive of, or even causally responsible for, our well-being.

### UBIQUITOUS SENSING

Current commercial mobile and wearable devices include many of the sensors and techniques used in health research (including accelerometer, gyroscope, location, heart-rate, and skin-conductance sensors and activity classification algorithms, smart notifications, and report capabilities). Furthermore, many smartphone apps let the user self-report activities and conditions that are challenging to capture automatically, such as the consumption of food, alcohol, and caffeine or the user’s emotional status. Today, smartphones are used by 69 percent of the adult population in developed countries and 46 percent of the adult population in developing economies, with adoption rates growing rapidly.<sup>1</sup> Social media posts, as well as web search queries, can also reveal a great amount about individuals’ behaviors, health, and well-being.<sup>2</sup> For example, individuals share and discuss goals, behaviors, sicknesses, diagnoses, and mental health challenges on platforms including Facebook, Reddit, Twitter, and Instagram.<sup>2–5</sup>

### REFERENCES

1. E. Anthes, “Mental Health: There’s an App for That,” *Nature*, vol. 532, no. 7597, 2016, pp. 20–23.
2. E. Yom-Tov, *Crowdsourced Health: How What You Do on the Internet Will Improve Medicine*, MIT Press, 2016.
3. T. Althoff, K. Clark, and J. Leskovec, “Large-Scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health,” *Trans. Assoc. Computational Linguistics*, 2016, pp. 463–476.
4. M. De Choudhury et al., “Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media,” *Proc. 2016 CHI Conf. Human Factors in Computing Systems*, 2016, pp. 2098–2110.
5. F. Ofli et al., “Is Saki #Delicious? The Food Perception Gap on Instagram and Its Relation to Health,” *Proc. 26th Int’l Conf. World Wide Web*, 2017, pp. 509–518.

In particular, *population-scale pervasive health* research attempts to harness such data, which has already been collected through commercial devices and web applications (see the “Ubiquitous Sensing” sidebar), to study human behaviors and the links between that data and health and well-being. Leveraging these existing datasets enables

studies of behaviors and health at an unprecedented scale (in terms of number of subjects), resolution (regarding the number and granularity of activities tracked), and duration (length of observation period) relatively inexpensively and quickly.

Population-scale pervasive health research can complement more

traditional pervasive computing and pervasive health research by highlighting the possibilities, opportunities, and challenges that arise from analyzing behavior and health data at scale. Here, I identify lessons learned from my own work and from other excellent contributions to the field and current challenges in this research area.

### LESSONS LEARNED

Although there are great advantages in leveraging large-scale datasets for individual and population health, there are limitations to this approach as well. Unlike typical “in-person” experimental studies or cohort studies, researchers often have limited or no control over what data is collected or how it has been collected. Furthermore, the data is typically observational and thus without any randomization into different conditions or treatments. This makes identification of causal relationships fundamentally challenging.

To overcome these challenges, interdisciplinary teams of researchers are developing specialized computational methods and tools, often drawing on data mining, social network analysis, and causal inference research, with the aim of obtaining unconfounded and actionable conclusions that are necessary to make any impact on healthcare and public policy. Following are lessons learned from leveraging existing devices and data for population-scale pervasive health research.

#### The Power of Scale

Leveraging widely used devices and applications allows for studies at unprecedented scale, much beyond what is feasible in more traditional laboratory or cohort settings. For example, leveraging data collected through a popular smartphone activity-tracking application (Argus by Azumio), I worked with Rok Sosič, Jennifer Hicks, Abby King, Scott Delp, and Jure Leskovec to study and compare physical activity patterns at a planetary scale across 717,000 people from over 110

countries.<sup>2</sup> The scale of such datasets enables new insights based on comparisons across subpopulations of various kinds—for example, based on demographic attributes including gender, age, weight status, and country of origin.

In the past, many of these comparisons were extremely expensive if not infeasible. For example, until recently, we had no large-scale database of objective physical activity measures spanning multiple countries. However, using the data from Argus, we were able to estimate the distribution of physical activity levels within countries, on a global scale.<sup>2</sup> This analysis, for the first time, revealed patterns of worldwide activity inequality. In some countries the gap between “activity rich” and “activity poor” people was much larger than in other countries, and the size of this gap was found to be a strong predictor of obesity incidence in the respective countries. Furthermore, it had been well established that men tend to be more physically active than women, on average. However, our analyses revealed that in countries such as Sweden and Ukraine the gender gap was almost negligible, whereas in countries such as Saudi Arabia and the US the gap was substantial. Such gaps could have detrimental consequences for women’s health.

The large number of people tracking their exercise through a wearable device further enables us to study population-scale phenomena, such as the viral spread of Pokémon Go. Because many wearable users of devices such as the Microsoft Band have agreed to share their data for research, Ryen White, Eric Horvitz, and I were able to conduct a study of Pokémon Go’s impact on 83,000 people’s physical activity (without this data, such a study could have required a massive participant recruiting and data collection enterprise).<sup>3</sup> We found that playing Pokémon Go led to significant increases in physical activity over a period of 30 days, with particularly engaged users increasing their activity by 1,473 steps a day on average, a more than 25 percent

increase compared with their prior activity level. Although these activity increases were often short-lived, Pokémon Go was able to reach low-activity populations—whereas four leading mobile health apps we looked at for comparison largely drew from an already very active population.

#### Natural Experiments and Causal Inference

A fundamental challenge with observational (that is, non-experimental) studies is using correlational data to infer causality. Understanding the causes of health outcomes is necessary for improving as opposed to merely predicting such outcomes. Accurately inferring causality from observational data is difficult because different conditions likely were not assigned randomly, leading to confounding and biased estimates.

To circumvent this, researchers increasingly apply matching methods, which match every treated unit to one or more non-treated units with similar observable characteristics (exact, almost-exact, distance- or propensity-score-based).<sup>4,5</sup> By matching treated units to similar, non-treated units, matching enables a comparison of outcomes among treated and non-treated units to estimate the effect of the treatment-reducing bias due to confounding.

However, there might exist unobserved variables that we cannot control for, even when this would be desirable. Although sensitivity analyses can alleviate such concerns in part, it might be more promising to attempt to identify variation in the data that could be used as an instrument or natural experiment to overcome the limitations of observational data. For example, researchers can leverage weather variation, changes in built environments due to relocation, or other potentially exogenous events such as strikes of public transport workers or closings of parks, to disentangle various effects. In some of my own work on the social influence effects on exercising behavior, we

leveraged random variation capturing the delay in the formation of friendship connections (how long did it take for the receiver of a friendship request to press “accept”?) to disentangle intrinsic motivation to exercise more from actual social influence.<sup>4</sup> This allowed us to demonstrate a causal, positive effect of online friendship connections on offline physical activity.

Large-scale data facilitates the use of such methods. For example, events such as user relocation or strikes are quite rare. However, in large data, we might still observe enough of these events to enable statistically meaningful analyses. Identifying these instruments and natural experiments in large data typically requires a mix of domain expertise (for example, knowing what events would cause plausibly exogenous variation in the treatment assignment) as well as data mining methods.

### Population Bias

User populations of wearable and tracking devices, smartphones, and web applications might not be representative of national populations, even when they are very large. In fact, we have found that the users of popular wearables and tracking apps tend to be biased toward young, more affluent, gender-skewed populations.<sup>2–4,6,7</sup> For example, some of the datasets from fitness wearables we have studied were skewed towards male users, while mobile applications focused on weight loss were skewed towards a female user base.

Researchers can check their findings for robustness against these biases—for example, by stratification into subgroups (for example, by age, gender, and income level)<sup>2–4,7</sup> or by reweighting the sample to match a target population.<sup>2</sup> In fact, as long as the data provides sufficient support for all relevant subgroups, reweighting methods can approximate nationally representative populations. Being able to validate findings across many subgroups to investigate heterogeneous treatment effects can be an advantage over representative yet small study populations

(which might be statistically underpowered for such analyses).

The study population can also be compared to traditional medical research data—for example, data from the National Health and Nutrition Examination Survey ([www.cdc.gov/nchs/nhanes](http://www.cdc.gov/nchs/nhanes)) or from the World Health Organization’s Global Health Observatory ([www.who.int/gho/en](http://www.who.int/gho/en))—on key behavioral or health covariates such as the timing or length of sleep or the volume and intensity of physical activity. Such data is often available on a subpopulation level, but it’s often based on subjective survey measures, which could be vastly different from sensor-defined objective measures, limiting comparisons.

As smartphones and other mobile or wearable devices become more prevalent, we can expect population bias to decrease. Furthermore, this drawback is not specific to population-scale pervasive health studies; it also applies to all scientific studies that largely draw from WEIRD (Western, Educated, Industrialized, Rich, and Democratic) subjects. Leveraging widely used devices and applications might even help in understanding historically less-represented populations.

### Engagement and Retention

Participant engagement and retention often diminish quickly. This holds true in traditional in-lab or in-person studies but can be exacerbated in commercial device and app settings where subjects are unpaid, are not bound by a study protocol, and quickly move on when they do not perceive a clear value. These dynamics are observed widely and were at the heart of many articles about Pokémon Go. The mobile game was spectacularly successful with 28.5 million daily users shortly after release, but after a few months, approximately 80 percent of them had moved on.<sup>8</sup> These numbers highlight both the unprecedented promise of large-scale behavioral interventions and great challenges in retaining engagement levels.<sup>3</sup>

An important consequence of low user retention rates is that researchers studying the same application at different times may in fact study different user populations. These populations are all worth studying. For example, early adopters are key to developing early prototypes into mature applications, while the coming and going of short-term users may help us better understand the appeal and value propositions of our applications. Furthermore, even after a historical drop-off in engagement, millions of people around the world still play Pokémon Go, allowing us to better understand whether and when such games could lead to sustained behavior change.<sup>3,7</sup>

### Multisite Studies to Disentangle Individual Behavior and the Environment

Many existing mobile health and social media datasets cover a large number of geographical locations.<sup>9</sup> Some even have global coverage.<sup>2</sup> This means that users of the same device or application can reside in vastly different environments. While this can complicate comparative analyses between the users, it also creates an unprecedented opportunity to study the effect of different environments on human behavior. To what degree is an individual’s behavior truly individual or dictated by one’s environment?

For example, cities without prevalent and safe sidewalks and footpaths or close-by stores, schools, and parks make it much harder to be physically active. The multisite nature of large-scale datasets let us disentangle individual behavior and environmental influences (requiring adequate control of potential confounders), and might enable us to design cities more conducive to their inhabitants’ health.<sup>2</sup>

### Augmenting Large Sensor Data with Context

Although continuous sensor data paints a rich picture of our behaviors, much of this data is useless without context. If somebody records very few

steps on a given day, is it because of the person's age, weight, recent surgery, non-walkable neighborhood, scorching climate, or preference for other activities? While smaller, in-person studies let us collect such information through surveys and other measures, much of this critical context is missing in already-collected, large-scale datasets from phones and wearables. Thus, to unleash the true power of these datasets, researchers need to augment the data with contextual information.

A common approach for data augmentation is to leverage an individual's geolocation, often available through self-report, GPS, cell tower location, and IP addresses. This enables augmenting the sensor data with population-level census or health outcome data.

Beyond location, researchers have brought in valuable context to sensor data through combinations with web search and online social network data. For example, Web search queries allowed differentiation of Pokémon Go players from non-players in a large sample of wearable users.<sup>3</sup> In another study, web search queries enabled the non-intrusive measurement of cognitive performance from already-collected search query logs, which could then be related to wearable-based sleep measurements.<sup>6</sup> In both studies, users had connected their web search account (Bing) to their wearable device (Microsoft Band) and agreed to share their data for research purposes. Combinations with online social network data enabled estimation of social influence effects in exercising behavior—that is, whether exposure to and interactions with online friends would have an effect on someone's physical activity levels (spoiler alert: it did!).<sup>4,7</sup> Social network data might be available if the network is part of the mobile application itself, when the application imports external social network data (from Facebook, for example), or when users share their activities and behaviors on public sites such as Twitter, Instagram, or Reddit.<sup>5,9</sup>

## RESEARCH OPPORTUNITIES

There are several research directions for increasing the effectiveness and scope of population-scale pervasive health research. These relate to developing more sophisticated computational methodology, acting on inferences made, and data sharing.

### Improving Computational Tools

Large-scale behavioral data paired with powerful computational tools bring unprecedented affordances, many of which we likely have yet to uncover. Open research challenges include identification of useful signals and proxies in sensor and Web data to capture behaviors, relevant context, and outcomes. Several studies have used social media posts, messages, and badges as proxies for health outcomes including weight loss, depression, and suicidal thoughts.<sup>5,10</sup> Such proxies are potentially powerful, but we need to better understand how to appropriately identify them and to what degree they reflect clinical measures.

Many future datasets will be observational. Due to the large but uncontrolled nature of the data, it is easy to fall prey to spurious correlations (with very low p-values). Therefore, we need to develop improved computational tools for analysis as well as establish methods and protocols abiding by the highest scientific standards. This is particularly pressing in the realm of causal analyses for which, currently, few if any tools exist that are usable by non-experts. Furthermore, such tools should offer scalability and capabilities beyond simple binary treatment/control scenarios and low-dimensional covariate spaces.

### Acting on Inferences Made

The goal of population-scale pervasive health research is to translate inferences from big data into the real world to improve people's lives. This is particularly challenging for data mining researchers who can be far removed from the people they study.

Real-world impact ranges across many settings, including clinical and population health, city and community design, and commercial devices and applications. Many of the datasets currently studied are collected by large technology companies. These companies might be receptive to suggestions that will improve the lives of their users, and they often have the resources to translate ideas and prototypes into practice. However, without support of these companies (for example, due to misaligned incentives or goals or a lack of resources), it can be difficult to act on any inferences made. Still, there are great opportunities to explicitly design the online space including search and social network tools for improved health and well-being.<sup>4-7,9,10</sup>

However, significant challenges exist in moving from passive sensing to acting on data inferences and interacting with users. One might be able to detect traces of depression and self-harm,<sup>5,10</sup> cyberchondria, or fatigue levels severely increasing accident risk. But how should one act on this information? If a user did not ask for it explicitly, it might be inappropriate or unethical to bring up this information. Furthermore, well-being objectives might even conflict with short-term commercial interests of technology companies (for example, some social media posts might drive significant engagement but could negatively influence self-worth and depression). In all these cases, user experience and ethics research are vital for the appropriate implementation and complementation of big data studies.

### Data Sharing, Privacy, and Ethics

Scientific progress in population-scale pervasive health studies critically relies on the availability of data. Currently, researchers often gain access to large datasets through industry collaborations or through web scraping of social media sites. Although many of these collaborations have been fruitful, data is rarely shared with outside researchers, limiting reproducibility and scientific progress.

More data sharing is clearly desirable, but there are many challenges to share personal information about behaviors and health.<sup>11</sup> For example, there is a need for methods and best practices to successfully de-identify data without introducing noise or bias into statistical analyses. In many cases anonymization techniques are open to reidentification attacks. There are promising avenues in privacy research based on randomized responses, differential privacy, and homomorphic encryption, but these techniques are not yet used widely in practice or at scale.

Currently, behavioral and health data are largely kept in separate silos. Behavioral signals are predominantly in the hands of tech companies developing wearable devices and smartphone applications while health data typically resides within hospital insurance systems. Great value will come from developing appropriate processes to combine this data to better understand how complex behavioral patterns cause or are caused by specific health outcomes. Large cohort studies hold great promise—such as the All of Us research program (<https://allofus.nih.gov>), which seeks to collect and combine such data from over one million US volunteers. It will be important to figure out how people’s existing devices can be leveraged effectively in these studies, and how researchers can provide valuable insights or financial incentives to motivate participants to share their personal data.

There are great opportunities for cross-pollination between pervasive computing and population-scale pervasive computing research. For example, laboratory-based and in-person pervasive computing research can inform data mining researchers about what can be measured and correlated in more controlled settings, highlighting the human factors at play. Conversely, large-scale data mining can inform pervasive computing about what

types of inferences are possible, informing our understanding of potential public health applications of pervasive health technology at scale. Furthermore, population-scale pervasive computing might change how epidemiological and population health research is conducted by enabling continuous, objective measurements of dynamic behaviors and environmental factors.

Advances and proliferation of mobile and sensor technology are driving the creation of large behavioral datasets. Population-scale pervasive health research leverages these datasets to enable a better understanding of the relationships between our behaviors, environment, and health outcomes, with great opportunities to impact health and well-being at population-scale. ■

### ACKNOWLEDGMENTS

I would like to thank Stephen Intille, Jesus Favela, Michele Catasta, Emma Pierson, and Rok Sosič for their feedback on the manuscript, and Jure Leskovec, Rok Sosič, Eric Horvitz, Ryen White, Scott Delp, Jen Hicks, Abby King, Munmun de Choudhury, Emre Kiciman, Elad Yom-Tov, Ingmar Weber, and Yelena Mejova for countless enlightening discussions. This work was supported by a SAP Stanford Graduate Fellowship and the NIH BD2K Mobilize Center (U54 EB020405).

### REFERENCES

1. J.M. McGinnis, P. Williams-Russo, and J.R. Knickman, “The Case for More Active Policy Attention to Health Promotion,” *Health Affairs*, vol. 21, no. 2, 2002, pp. 78–93.
2. T. Althoff et al., “Large-Scale Physical Activity Data Reveal Worldwide Activity Inequality,” *Nature*, vol. 547, no. 7663, 2017, pp. 336–339.
3. T. Althoff, R.W. White, and E. Horvitz, “Influence of Pokémon Go on Physical Activity: Study and Implications,” *J. Medical Internet Research*, Dec. 2016, p. e315.
4. T. Althoff, P. Jindal, and J. Leskovec, “Online Actions with Offline Impact: How Online Social Networks Influence Online and Offline User Behavior,” *Proc. 10th ACM Int’l Conf. Web Search and Data Mining*, 2017, pp. 537–546.

5. M. De Choudhury et al., “Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media,” *Proc. 2016 CHI Conf. Human Factors in Computing Systems*, 2016, pp. 2098–2110.
6. T. Althoff et al., “Harnessing the Web for Population-Scale Physiological Sensing: A Case Study of Sleep and Performance,” *Proc. 26th Int’l Conf. World Wide Web*, 2017, pp. 113–122.
7. A. Sharneli et al., “How Gamification Affects Physical Activity: Large-Scale Analysis of Walking Challenges in a Mobile Application,” *Proc. World Wide Web Companion*, 2017, pp. 455–463.
8. S. Arif “The Number of Pokémon Go Players Still Logging in Every Day Has Plummeted by over 80%,” *VG 24/7*, 3 Apr. 2017; [www.vg247.com/2017/04/03/the-number-of-pokemon-go-players-still-logging-in-every-day-has-plummeted-by-over-80](http://www.vg247.com/2017/04/03/the-number-of-pokemon-go-players-still-logging-in-every-day-has-plummeted-by-over-80).
9. F. Ofli et al., “Is Saki #Delicious? The Food Perception Gap on Instagram and Its Relation to Health,” *Proc. 26th Int’l Conf. World Wide Web*, 2017, pp. 509–518.
10. T. Althoff, K. Clark, and J. Leskovec, “Large-Scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health,” *Trans. Assoc. Computational Linguistics*, 2016, pp. 463–476.
11. K. Caine, “Privacy Is Healthy,” *IEEE Pervasive Computing*, vol. 15, no. 4, 2016, pp. 14–19.



**Tim Althoff** is a PhD candidate in computer science at Stanford University. Contact him at [althoff@cs.stanford.edu](mailto:althoff@cs.stanford.edu).

Read your subscriptions through the myCS publications portal at <http://mycs.computer.org>.