

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Publications, UNL Libraries

Libraries at University of Nebraska-Lincoln

---

2016

## Opportunities and challenges in the use of personal health data for health research

Matthew J. Bietz

Cinnamon S. Bloss

Scout Calvert

Job G. Godino

Judith Gregory

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.unl.edu/libraryscience>



Part of the [Bioethics and Medical Ethics Commons](#), [Health Information Technology Commons](#), and the [Library and Information Science Commons](#)

---

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Matthew J. Bietz, Cinnamon S. Bloss, Scout Calvert, Job G. Godino, Judith Gregory, Michael P. Claffey, Jerry Sheehan, and Kevin Patrick

---

# Opportunities and challenges in the use of personal health data for health research

Matthew J. Bietz,<sup>1,2</sup> Cinnamon S. Bloss,<sup>2,3,4</sup> Scout Calvert,<sup>5</sup>  
Job G. Godino,<sup>2,4</sup> Judith Gregory,<sup>1</sup> Michael P. Claffey,<sup>6</sup>  
Jerry Sheehan,<sup>7</sup> and Kevin Patrick<sup>2,4</sup>

1 Department of Informatics, Donald Bren School of Information and  
Computer Sciences, University of California, Irvine, California

2 Center for Wireless and Population Health Systems, Calit2, University of  
California, San Diego, California

3 Department of Psychiatry, School of Medicine, University of California,  
San Diego, California

4 Department of Family Medicine and Public Health, School of Medicine,  
University of California, San Diego, California

5 Louise M. Darling Biomedical Library, University of California, Los Angeles,  
California

6 Department of Psychology, University of California, San Diego, California

7 Information Technology Center, Montana State University, Bozeman,  
Montana

*Correspondence* — Kevin Patrick, The Qualcomm Institute/Calit2, 9500 Gilman  
Drive, Atkinson Hall (mail code 0811), La Jolla, CA 92093-0811, USA;  
kpatrick@ucsd.edu

---

Published in Journal of the American Medical Informatics Association 23 (2016), e42-e48.

doi:10.1093/jamia/ocv118

Copyright © 2015 by the Authors. Published by Oxford University Press on behalf of the  
American Medical Informatics Association. Used by permission.

Submitted 1 May 2015; revised 8 June 2015; accepted 26 June 2015; published 2 September  
2015.

**Abstract**

*Objective:* Understand barriers to the use of personal health data (PHD) in research from the perspective of three stakeholder groups: early adopter individuals who track data about their health, researchers who may use PHD as part of their research, and companies that market self-tracking devices, apps or services, and aggregate and manage the data that are generated.

*Materials and Methods:* A targeted convenience sample of 465 individuals and 134 researchers completed an extensive online survey. Thirty-five hour-long semi-structured qualitative interviews were conducted with a subset of 11 individuals and 9 researchers, as well as 15 company/key informants.

*Results:* Challenges to the use of PHD for research were identified in six areas: data ownership; data access for research; privacy; informed consent and ethics; research methods and data quality; and the unpredictable nature of the rapidly evolving ecosystem of devices, apps, and other services that leave “digital footprints.” Individuals reported willingness to anonymously share PHD if it would be used to advance research for the good of the public. Researchers were enthusiastic about using PHD for research, but noted barriers related to intellectual property, licensing, and the need for legal agreements with companies. Companies were interested in research but stressed that their first priority was maintaining customer relationships.

*Conclusion:* Although challenges exist in leveraging PHD for research, there are many opportunities for stakeholder engagement, and experimentation with these data is already taking place. These early examples foreshadow a much larger set of activities with the potential to positively transform how health research is conducted.

**Keywords:** mobile health (mHealth), wearable sensors, Internet of Things (IoT), big data, personal data, data sharing

**Introduction**

Consumer-oriented electronic devices, apps, and services are now able to capture a variety of parameters directly relevant to human health. Advances in microtechnology, data processing and storage, wireless communication and networking infrastructure, and battery capacity have resulted in the proliferation of devices that have made it possible for individuals to produce ever-larger streams of data across the lifespan, throughout the course of health and illness, and in a geo-spatial context. Applications designed to collect, store, and analyze these personal health data (PHD) have proliferated and are increasingly being used by a wide range of individuals for self-tracking. In

early 2013, the Pew Research Center's Tracking for Health study found that 69% of Americans track some form of health-related information and 21% use a digital device to do so.<sup>1</sup> In addition to self-tracked PHD, more and more data about individuals are being captured passively as people surf the web, communicate with one another on social networks, make financial transactions, or conduct other activities that leave "digital footprints."<sup>2</sup>

Nearly all of the electronic devices, apps, and services that collect and store PHD are outside the mainstream of traditional health care or public health research.<sup>3</sup> This includes everything from small start-ups to globally active consumer electronics, telecommunications, and search-oriented or social network corporations. Concurrently, there seems to be an increasing willingness for individuals to share their PHD with others.

This can be seen in the Quantified Self movement, where individuals meet to share insights gained from their self-tracking activities.<sup>4</sup> Additionally, many people now share their data with those who have similar medical conditions in the context of online groups such as PatientsLikeMe<sup>5</sup> or Crohnology,<sup>6</sup> to learn as much as possible about their shared health concerns. The trend for sharing extends to opening up PHD to see what insights others might see in them, as exemplified by the Open Humans Project.<sup>7</sup>

### ***New Opportunities and Challenges for Research***

The growing amount of PHD presents an opportunity to move beyond the use of population-level data for simple descriptive epidemiology to its use for making causal inferences. Fundamental principles of epidemiology pertaining to how causality should be determined were developed at a time when health-related measures were infrequently collected and expensive in terms of time, materials, and participant burden.<sup>8</sup> These barriers are now often dramatically reduced by the increasing ubiquity of PHD. We increasingly have sufficient data on a variety of determinants of health such that we may be on the cusp of a new form of establishing causality, akin to how researchers in fields like atmospheric science or economics make predictions about future events from the models they develop on ever-changing real-time data sets.

These new methods of acquiring data and approaching research raise new challenges but also familiar issues, including data access, privacy, and consent. Privacy norms and expectations are becoming more diverse, stretched in opposite directions by opposing trends. On one hand, sharing is common in an era of online communication and social networking sites like Facebook, Twitter, and Pinterest. On the other hand, there may be increased desire for attention to privacy as a result of adverse media events such as those that surrounded the National Security Agency data collection efforts.<sup>9</sup> Closely related to privacy is the need for informed consent in order to maintain public trust in the research enterprise. For researchers, data access becomes complicated when researchers acquire data from third parties rather than collect it directly. The scientific method is based on full transparency in data generation, manipulation, and analysis. Entities with a vested interest in protecting their intellectual property may refuse to open the black boxes of their proprietary software and algorithms, making it difficult, if not impossible to interpret the data, establish its validity, and replicate research. Whereas big data technologies in physics and genomics were largely developed by academics with funds from public agencies, almost all of the resources relevant to PHD are commercially developed and are subject to a variety of intellectual property and licensing restrictions.

### ***The Current Project***

In mid-2013, the Robert Wood Johnson Foundation funded the Health Data Exploration (HDE) project to gain insights into how various stakeholder groups think about PHD and its use for research. Stakeholders included individuals who track data about their health; researchers who might use the data as part of their research; and companies that market self-tracking devices, apps or services to aggregate and manage the data that are generated. For individuals, the aim was to understand experiences with health tracking, the kinds of data that are tracked, and attitudes toward data sharing and privacy. For researchers, the aim was to understand the kinds of data that would be useful for research, concerns about data quality and reliability, and perceived barriers to their use of PHD. For companies, the aim was to understand what data are collected; the legal, policy, and business concerns

around these data; and the overall willingness and ability to make data available to external researchers. This project was conducted from the perspective that research using new forms of PHD would not supplant current efforts to understand health, but rather that these new forms may complement and add value to existing medical and public health efforts to measure the environmental, social, behavioral, and medical determinants that comprise the full picture of health and society. This paper describes the results of this effort, which at a high level was to identify opportunities and challenges to using PHD for research.

## **Methods**

This project was reviewed by the institutional review board (IRB) of the University of California, San Diego via the UC Reliance Registry (UC IRB Reliance #711) and was approved by University of California, San Diego and the University of California, Irvine.

## **Overview**

An environmental scan was conducted to identify HDE-related peer-reviewed and other scientific publications, foundation reports, governmental reports, and key thought pieces in the popular media and other sources. These efforts overlapped with the deployment of an online survey to individuals and researchers that was conducted from August 1, 2013 to September 11, 2013. In addition, in-depth interviews were conducted with researchers, individuals, and company/key informants to develop a deeper understanding of themes that arose in the surveys, in discussions with advisory board members, and in the literature review. Based upon the results of these efforts, several opportunities and challenges related to progress in the field of PHD were identified and are presented here.

## **Surveys**

Surveys were designed to elucidate attitudes and experiences with self-tracking data for both individuals and researchers. Survey instruments explored a set of high-level research questions developed by

the research team. These instruments were pilot-tested for comprehensibility, logic, and time required to complete. This was done in three phases: first using think-aloud protocols as pilot participants took the survey on paper; second, deployment of an early draft of the online survey to explore usability and comprehension; and finally, deployment of a near-final version of the survey to test any changes that had been made and confirm that the survey was an appropriate length. The survey was also reviewed by an external expert in survey research design before deployment. The high-level questions and full survey instruments are included in the Supplementary Materials. Surveys were administered via the web using a local installation of LimeSurvey, an open-source survey management and analysis platform.<sup>10</sup> At the survey's conclusion, respondents who were interested in a follow-up interview could choose to provide contact information in a web form that was separate from the survey, in order to protect their anonymity.

### ***Targeted Sampling***

The study aim was to collect information regarding the use of PHD for research from individuals with actual experience in this area, in other words, individuals and researchers who were already generating or using digital self-tracking data. These are, however, essentially "hidden" populations: there are no lists of self-trackers from which we could draw a sample, and since this is a relatively new field, there are only a small percentage of these early adopters in the general population. For this reason, we chose to use a targeted convenience sampling strategy, recruiting participants through postings on self-tracking-related web pages, relevant press releases, and topical social-media channels including blogs and tweets. In order to address the possible biases resulting from this sampling strategy we asked demographic questions that enabled comparisons to the general population. In addition, we also asked questions that were included in the Pew Research Center's "Tracking for Health" study in order to calibrate our sample against Pew's national sample.<sup>1</sup> Individuals were offered participation in a drawing for an iPad or Android tablet as an incentive for participation.



## ***Interviews***

Standardized interview protocols were developed for use with each of the three stakeholder groups. Interviews with individuals and researchers were designed to complement the surveys by providing richness to the findings and eliciting data that would be difficult to collect in a survey. Company/key informant interviews included experts in the area of PHD, as well as representatives of a range of companies that provide personal health devices, apps, or services. The intent was to gauge companies' willingness to participate in collaborations with academic researchers and understand the business, technological, and social factors that affect their decision-making. Thus, semi-structured qualitative interview protocols were developed based on the same set of high-level questions that drove initial survey design, as well as preliminary analyses of the survey data that were collected.

Potential individual and researcher interviewees were selected from among those who completed the survey and were willing to be contacted for follow-up interviews. A stratified random sample of individuals was selected to ensure equal numbers of men and women and a variety of research interests. Potential participants were invited to participate by email or phone. Interviews were conducted by phone or in person and were audio-recorded and transcribed. Representatives of central organizations in the PHD economy and key informants were selected for interviews based on their personal expertise and roles as decision makers in their organizations. Detailed notes were taken for company interviews to avoid confidentiality concerns with audio-recording.

## ***Data and Statistical Analyses***

All analyses were conducted using the statistical software package R. Percentages, two-sided *t*-tests, Mann–Whitney U-tests, or chi-square tests were used to describe and compare survey responses. All *P*-values are uncorrected for multiple testing. Interview data were analyzed from a grounded theory approach using the Dedoose software package.<sup>11,12</sup>

## Results

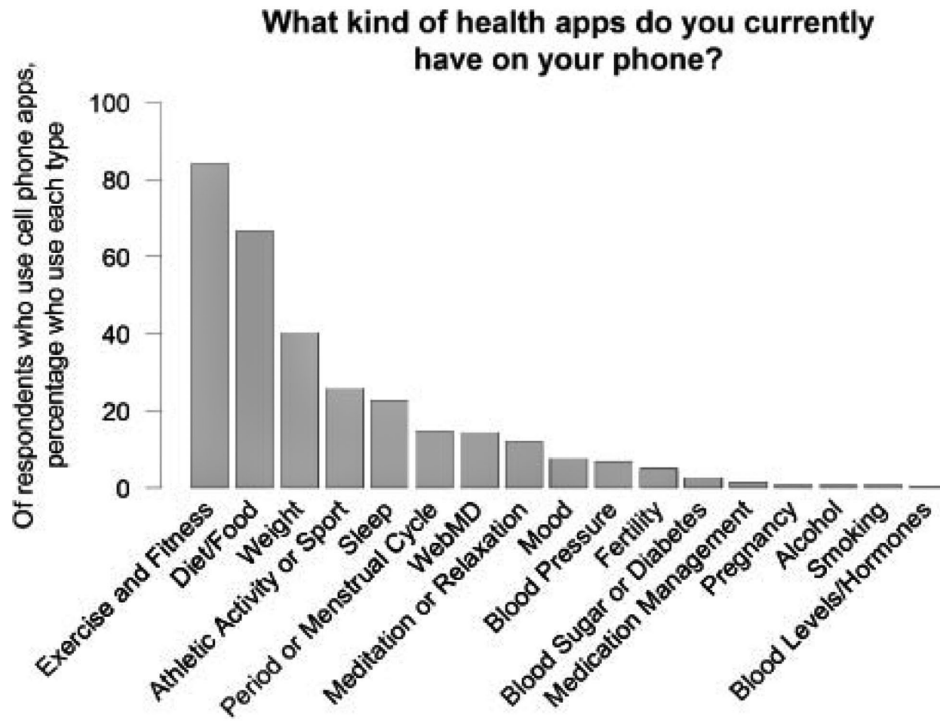
### *Individuals*

#### *Sample Characterization*

A total of 465 individuals completed the survey and 11 of these individuals were interviewed. Demographic statistics and descriptors that characterize the nature and degree of self-tracking in this cohort are provided in **Table 1**. As shown, the targeted sampling strategy was effective at recruiting a sample of individuals in which self-tracking was relatively common (91%). Although this was the goal of the project, it should be emphasized that our results are based on data from early adopters who are likely generally healthy, and thus, important additional viewpoints, including additional concerns and issues, may emerge as use of these technologies expands.

**Table 1** Demographics and self-tracking characterization (individuals,  $N = 465$ )

<i>Metric</i>	<i>HDE (%)</i>	<i>US Census (2010) (%)</i>
Gender (% Female)	65.2	
Race		
Black/African-American	2.6	12.6
Asian/Pacific Islander	7.7	5.0
Native American	2.9	0.9
White	81.4	72.4
Other	1.7	6.2
Ethnicity		
Hispanic/Latino	16.4	3.8
Annual household income		
<\$50 k	23.7	
\$50–\$100 k	29.1	
>\$100 k	47.2	
Educational attainment		
<Bachelor's degree	9.6	
4-year degree	29.3	
Graduate degree	61.1	
Track PHD for self/loved one	91	
Health tracking app on cell phone	65	
Self-identified QS member	39	
Medical reason for self-tracking	14	



**Figure 1** Individuals who use cell phone apps are most likely to use apps aimed at fitness, diet, and weight.

### Survey Results

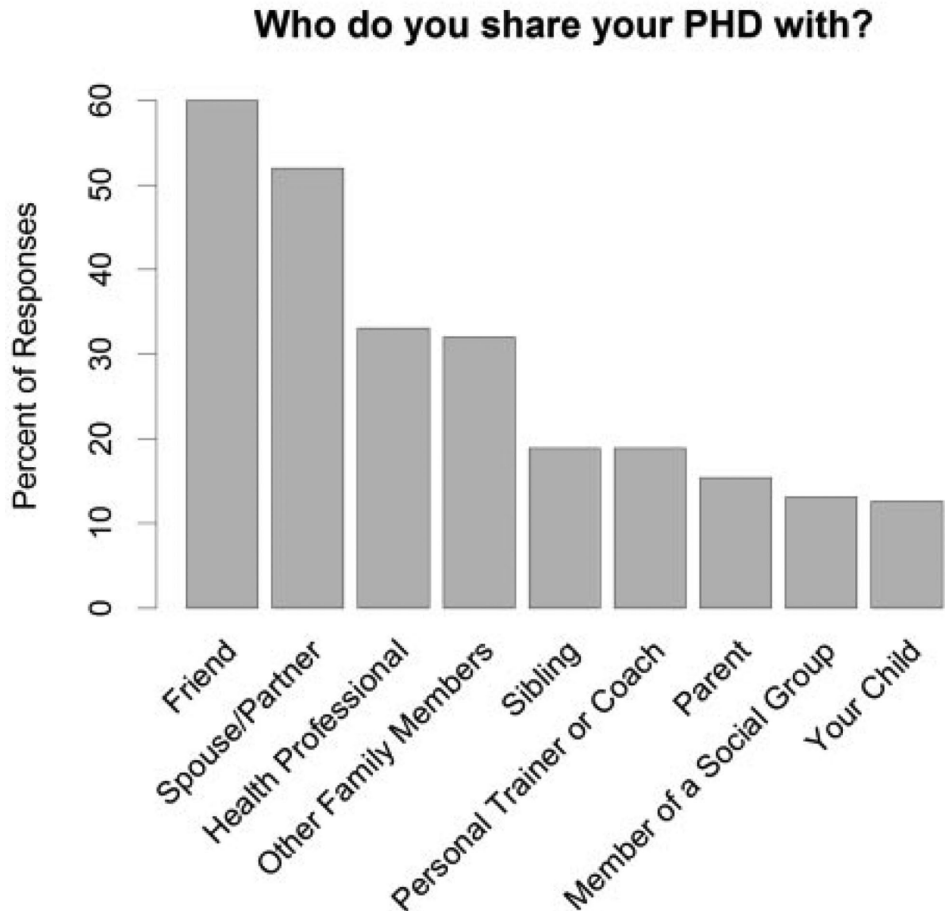
The most common types of data tracked were exercise, diet, weight, athletic activity, and sleep (**Figure 1**). Participants reported a higher frequency of tracking for general health than tracking to manage a medical condition or chronic disease. Cell phone apps were more commonly used than websites for tracking, and both were more commonly used than paper or “in your head” tracking. We also found that the use of cell-phone apps for self-tracking correlated with age, with 100% of the 18–25 year-olds who reported any self-tracking indicating that they did so using cell phone apps compared to only 18% of those over age 65 years. The use of cell phone apps for tracking health did not significantly differ by income group.

Results of survey items that asked participants about perceived data ownership are shown in **Table 2**. Over half of respondents believed they own all their data, and with respect to data sharing, 45% reported sharing their health tracking data with someone, either online

**Table 2** Individual Early Adopter Survey results ( $N = 465$ )

<i>Survey item</i>	<i>HDE (%)</i>
Share PHD with someone else either on/off line	45
"Probably would" or "Definitely would" be willing to share PHD for research	78
Would be "more" or "much more" likely to share PHD for research if offered monetary compensation	56
Would share PHD if privacy were assured	68
Feel that anonymity is "very" or "extremely" important	67
Mentioned privacy, anonymity, or confidentiality in open-ended question asking "under what agreements and constraints would you share your health and activity tracking data?"	63
Do you believe that you "own" – or should own – all of the data that are about you, even when these data are indirectly collected?	
I believe that I do own and should own all of my data	54
I should share ownership of my data with the company that collected my data	30
The company that collected my data owns it	4
This is not something I care about	13
Do you want to own this kind of data?	
Yes	75
No	5
This is not something I care about	20

or offline (**Figure 2**). Relevant to this finding, a theme that emerged from the interviews was that while individuals felt their self-tracking data could be useful to share with healthcare providers, they perceived that providers had little interest in it. Respondents reported a general willingness to share their data for use in research (**Table 2**). Exploring this further, participants reported that they were more willing to share data for use in a scientific study that they found interesting versus one they found uninteresting ( $P = 0.007$ ). There was no difference between general willingness to share and willingness to "donate your personal health and activity data to a scientific database." When asked about compensation, 56% of the participants said they would be "more" or "much more" likely to share data if they were compensated (**Table 2**). Many respondents also reported an aversion to commercial or profit-making use of their data.



**Figure 2** Individuals reported that they most often shared PHD with friends and partners, as well as health professionals.

In terms of privacy, 68% of respondents would only share their data “if privacy were assured” (**Table 2**). Responses also suggested that not all data are perceived as being equally sensitive, with at least one participant noting a fundamental difference between providing Global Positioning System (GPS) data capability for determining routes taken and providing data on the number of steps taken. Interviewees also often reported being more trusting of universities (vs companies), assuming ethical and regulatory processes were in place in such environments.

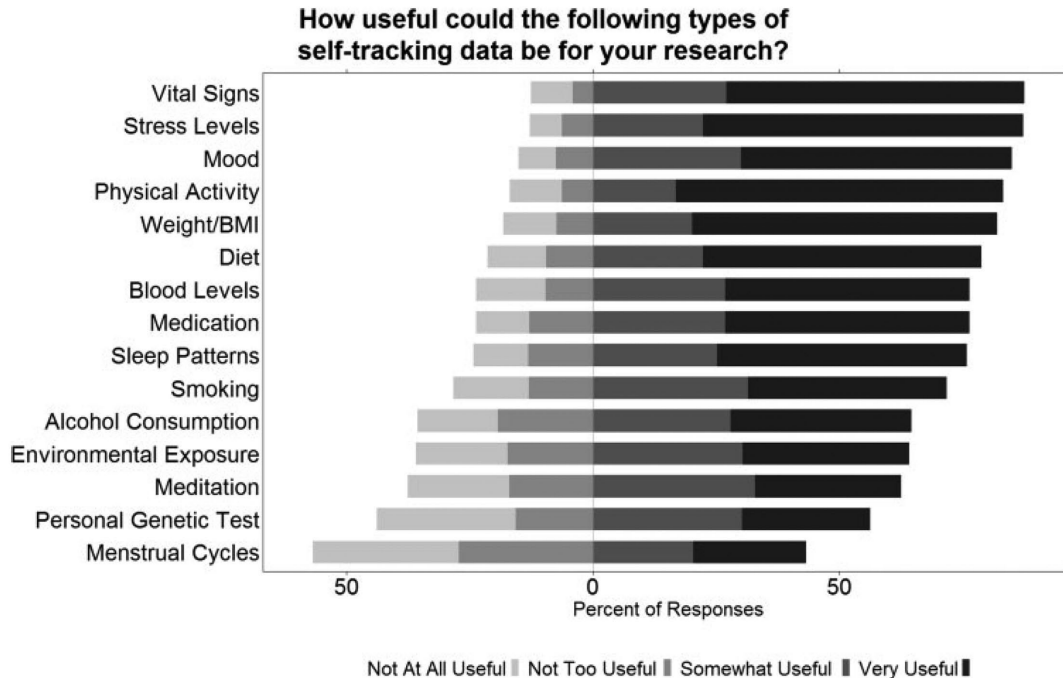
## Researchers

### Sample Characterization

A total of 134 researchers completed the survey and 9 of these researchers were interviewed. Demographic and other descriptive statistics are shown in **Table 3**. *Survey Results* Generally, the categories of data that were found to be most frequently tracked by individuals were types of data that were reported to be useful by researchers. Importantly, however, some of the data types considered most useful by researchers (e.g., vital signs, stress levels, and mood; **Figure 3**) were less likely to be self-tracked by individuals. The potential usefulness of these data was echoed in the interviews, with many researchers detailing the ways that these data can fill in gaps in more traditional clinical data collection. A clear theme that emerged was that self-tracking data could provide better measures of everyday behavior and lifestyle. Researchers also reported that aggregating data from multiple sources would be beneficial, including linking of PHD with clinical data.

**Table 3** Demographics and sector/expertise characterization (researchers,  $N = 134$ )

<i>Metric</i>	<i>HDE (%)</i>
Gender (% Female)	51.0
Domain	
Health Sciences	69.3
Social Sciences	31.6
Engineering and Technology	19.3
Life Sciences	7.9
Arts and Humanities	4.4
Sector	
Academia	74.5
Non-profit	10.9
Government	8.2
Industry	2.7
Other	2.8
Research Engagement	
Research Leader	73.8
Research Staff	19.6
Not currently doing research	6.5



**Figure 3** Researchers generally reported that PHD was valuable for research; however, the three most useful types of research data (i.e., vital signs, stress levels, and mood) were less likely to be tracked by individuals.

Researchers were also open to non-traditional data and data sources (**Table 4**). Cited barriers to the use of PHD for research were intellectual property concerns, licensing, and establishing the legal agreements necessary when collaborating with companies. Researchers also reported being concerned with the kinds of data they may get

**Table 4** Researcher survey results ( $N = 134$ )

<i>Survey item</i>	<i>HDE (%)</i>
Agree or strongly agree that self-tracking data would be useful in their own research	89
Believe that PHD could answer questions that other data could not	95
Have used publicly available data sets in their research	57
Have purchased data for use in their research	19
Have already used PHD in their research	46
Have collaborated with application, device, or social media companies	23

from companies, including the lack of standardization, potential problems with proprietary algorithms, and that most of the consumer-level health devices have not gone through a validation process. There was also concern about potential biases in datasets of PHD due to self-selection of early adopter users who can afford new technologies, or may otherwise not be representative of a general population.

### ***Companies and Key Informants***

A total of 15 interviews were conducted with companies and key informants. The companies ranged from small start-ups to very large publicly traded entities in the online media and health information space. A major theme that emerged from these interviews was that for companies, advancing research is a worthy goal, but not a primary concern. Thus, any collaboration with researchers or sharing of research data needs to respect the company's business model and goals. Furthermore, despite their technically advanced nature, some seemed reluctant to devote the resources necessary to support data export unless it serves a direct business purpose. Companies were also very concerned about customer relationships, and sharing data outside of the company presents a risk of loss of customer trust. Some companies acknowledged that collaborating with an academic institution can help provide credibility in the marketplace, but others emphasized the difficult and lengthy process required to develop such relationships.

While there was no consensus on the best approach, company and key informants, more than the other cohorts, highlighted the complexity of privacy, informed consent, and personal data. What became clear was that there is a deep intertwining of data privacy, IRBs, informed consent, licensing agreements, network and database security, HIPAA and other legal frameworks, user interface design, corporate policies, and customer relations.

### **Discussion**

Taken together, there appear to be many opportunities for, and considerable enthusiasm about, the potential for leveraging PHD for



health research. Challenges to the use of PHD for health research were identified in the following six areas: data ownership; data access for research; privacy; informed consent and ethics; research methods and data quality; and issues related to an evolving ecosystem of devices, apps, and other services that leave “digital footprints.” Nonetheless, many of the individuals interviewed considered these challenges surmountable and viewed them as opportunities for stakeholder engagement to improve knowledge, practice and policy-level efforts that support PHD use. While some of these issues overlap with one another, what follows are suggested approaches to begin to address each area.

### ***Data Ownership***

Important differences exist with respect to how individuals and companies view ownership of PHD. In the survey of individuals, we found that some did not care who owned the data they generate, although a clear majority wanted to own or at least share ownership of the data with companies. Importantly, many thought that they actually did own these data, even though many have almost certainly entered into “click-through” agreements in which they have given those rights away to companies. While this difference of opinion and/or perception does not appear to be a major barrier at present to growth in use of self-tracking technologies, it may foreshadow a deeper divide between public attitudes and corporate practices that would benefit from future policy-making in this area. Such policies may become increasingly important as researchers move to combine PHD with more traditional forms of health data, such as electronic health records.

### ***Data Access for Research***

Although individuals expressed concern about maintaining their privacy, they conveyed considerable willingness to have their PHD shared with and used by researchers. Their main concerns related to commercial uses of their PHD, to which many had an aversion. Both researchers and companies noted that even when there is general willingness to share PHD, accomplishing this can be an arduous task due to regulatory and legal constraints. IRBs must be willing to accept

data collection practices of third parties, something that has come under scrutiny after high-profile cases like the Facebook emotional contagion experiment.<sup>13</sup> Creating the right contract language, material transfer agreements, or other documentation that satisfies both corporate counsel as well as the research partners is challenging. There is a need for new technology and policy solutions that ease the movement of data between companies and researchers while protecting the rights of individuals. Other strategies include advancing and fostering the adoption of language for data use agreements and terms of service that make it easier for companies to respond if a customer desires to make their data available for research. This latter approach, used in the recently released Apple ResearchKit,<sup>14</sup> allows researchers interested in PHD to recruit participants into a study as long as participants are willing to release their data for study purposes. Finally, the notion of some form of data repository or data commons surfaced in several interviews and may be worthy of further attention.

### **Privacy**

Policies and practices that relate to privacy of health information that emerged in the era of medical records, clinical trials, and periodic public health surveys may be insufficient at this time when more and more PHD are being produced. Users of self-tracking technologies are frequently unaware of the details of data access to which they agree in the context of clicking “accept” to terms of use. Even with an awareness of data access issues and permissions, it is often difficult to predict effects on privacy. For example, while data may be anonymized before being shared, there is a very real risk of revealing a person’s identity if two or more sources of personal data are combined.<sup>15</sup> In this unsettled policy and technology environment, there is also little understanding of the nature and degree of actual risks, if any, associated with re-identification and/or other breaches of PHD privacy.

Several activities that specifically address recommendations about how to handle privacy issues for PHD might help protect the availability of these forms of data for research aimed at improving the public good. First, additional research is needed to help unpack and understand user expectations regarding the privacy of their PHD. This

understanding can then help inform conversations aimed at establishing norms of use. Second, there is a need to develop appropriate education and outreach materials to help in discussions about the realities and challenges of digital anonymity. Third, tools need to be developed to enhance user control of data, awareness of sharing, and notification of findings derived from the use of PHD in research. These controls are an essential condition for establishing the trust needed to assure that data donation is not a one-time occurrence.

### ***Informed Consent and Ethics***

Just as these new forms of data raise new questions about data privacy, they also create new questions for the ethics of research. Most of the current framing of research ethics comes from a predigital era. The very characteristics that make PHD valuable for research also make it ethically challenging. PHD provides a high level of detail about the everyday activities of individuals. Large amounts of data can be collected at relatively low cost, and many of the sensors and digital traces are generated without active engagement (or even awareness) by participants. The same devices and apps that generate PHD are also platforms for delivering information to users, providing an opportunity for intervention experiments with a sample size that was previously impractical, if not impossible. While some academic communities have considered these issues and developed ethics guidelines for internet research,<sup>16</sup> there does not appear to be broad awareness or adoption of such recommendations in health research or by IRBs. There is a need for high-level, interdisciplinary efforts to revisit fundamental ethics principles, consider how they apply to these new modalities of research, and update the procedures and recommendations that guide researchers and IRBs. Simultaneously, there is an opportunity for experimentation with new models and technologies of informed consent, de-identification, and trusted sharing that can balance respect for the individual with the scientific potential of PHD.

### ***Research Methods and Data Quality***

Several informants identified obstacles that relate to PHD research methods. One of the most common concerns is about validity and

reliability of the data given the wide variety of sensors and devices that are now in use to capture PHD. Unlike medical devices that undergo a rigorous FDA approval process, consumer-level health devices and apps only need pass the test of the marketplace to become widely used. For some types of research such as population level monitoring of general trends in physical activity, consumer-level pedometers, or wearable activity trackers may be acceptable. However, if PHD is to be coupled with more traditional forms of health data (e.g., clinical trial data) and then used to improve health interventions, more will need to be known about how well PHD devices and apps represent the underlying constructs they aim to measure. A related concern is the potential bias in PHD that derives from who uses personal health devices and who does not. Along these lines, it should be emphasized that our results are based on data from early adopters who are likely generally healthy. Important additional viewpoints, including additional concerns and issues, may emerge as use of these technologies expands. It remains to be seen how generalizable these data are to the general population.

### ***An Evolving Ecosystem***

Finally, the field of PHD collection and use is an area in flux. In many ways this is more of an opportunity than a challenge, as it gives all stakeholders involved an opportunity to impact the landscape as it evolves. One area of significant change will be in the area of self-tracking technologies themselves. Some of the issues researchers highlighted around the validity of the data and lack of standardization will be addressed as the consumer health device, apps, and services market matures. We also expect that as policies are developed, laws are written, and standard practices emerge, some of the uncertainty around data ownership, privacy, and ethics will lessen.

### **Conclusion**

Creative solutions must be found that allow individual rights to be respected while providing access to high-quality and relevant PHD for research, that balance open science with intellectual property, and

that enable productive and mutually beneficial collaborations between the private sector and the academy. A great deal of experimentation is taking place that is working toward these goals. Findings from this project suggest that the public good can be served by these advances, but that there is also work to be done to ensure that policy, legal, and technological developments enhance the potential to generate knowledge out of PHD, and ultimately, improve health and well-being.



**Acknowledgments** We thank the many individuals, researchers, company representatives, and key informants who shared with us their perspectives on personal health data. We also acknowledge our Health Data Exploration Project National Advisory Board members: Linda Avey, Hugo Campos, Robert M. Kaplan, Sendhil Mullainathan, Tim O'Reilly, Larry Smarr, Martha Wofford, and Gary Wolf. We further thank Robert Wood Johnson Foundation program officers Stephen Downs and Lori Melichar, as well as Ramesh Rao, Alexandra Hubenko, Tiffany Fox, and Jemma Weymouth, all of whom assisted with aspects of the project.

**Funding** This work was supported by Robert Wood Johnson Foundation grant number 71693, entitled "The Health Data Exploration (HDE) project" (PI: Patrick, #71693, 2013-2017).

**Competing interests** The authors have no competing interests to declare.

**Contributors** M.J.B., J.S., and K.P. designed the study. J.G., S.C., and M.P.C. assisted with survey and study design. M.J.B., J.S., S.C., J.G., and K.P. conducted study interviews. M.J.B. and K.P. oversaw data acquisition and management. M.J.B. and S.C. analyzed the data. M.J.B., C.S.B., and K.P. drafted the manuscript and edited the manuscript for intellectual content. J.G.G. and J.S. edited the manuscript for intellectual content.

**Supplementary Material** Supplementary material — HDE Individuals and Researchers Codebooks — is attached to the archive record for this article.

## References

1. Fox S, Duggan M. *Tracking for Health*. Pew Research Center's Internet & American Life Project; 2013. <http://pewinternet.org/Reports/2013/Tracking-for-Health.aspx>. Accessed July 30, 2015.
2. Welser HT, Smith M, Fisher D, Gleave E. Distilling digital traces: Computational social science approaches to studying the internet. In *The Sage Handbook of Online Research Methods*. London: Sage; 2008:116–140.

3. Clarke M, Bogia D, Hassing K, Steubesand L, Chan T, Ayyagari D. Developing a Standard for Personal Health Devices based on 11073. In: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Cité Internationale, Lyon, France, August 23–26, 2007. New York: IEEE; 2007:6174–6176.
4. Wolf G. Know thyself: tracking every facet of life, from sleep to mood to pain. In *Wired Magazine* 2009:365.
5. *patientslikeme*. <http://www.patientslikeme.com/> . Accessed April 28, 2015.
6. *Crohnology*. <https://crohnology.com/> . Accessed April 28, 2015.
7. *Open Humans*. <https://www.openhumans.org/> . Accessed April 28, 2015.
8. Hill AB. The environment and disease: Association or causation? *Proc R Soc Med*. 1965;58:295–300.
9. Agency NS. *Surveillance Techniques: How Your Data Becomes Our Data*. <https://nsa.gov1.info/surveillance/> . Accessed May 1, 2015.
10. Team LP, Schmitz Carsten. *LimeSurvey: An Open Source Survey Tool*. Hamburg, Germany: LimeSurvey Project; 2012.
11. Charmaz K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Thousand Oaks, CA: Sage; 2006.
12. *Dedoose Version 5.0.11, Web Application for Managing, Analyzing, and Presenting Qualitative and Mixed Method Research Data*. Los Angeles, CA: SocioCultural Research Consultants, LLC; 2014.
13. Kahn JP, Vayena E, Mastroianni AC. Opinion: learning as we go: lessons from the publication of Facebook’s social-computing research. *Proc Natl Acad Sci USA*. 2014;111(38):13677–13679.
14. *Apple ResearchKit*. <https://www.apple.com/researchkit/> . Accessed April 28, 2015.
15. Sweeney L, Abu A, Winn J. *Identifying Participants in the Personal Genome Project by Name*. Harvard University: Data Privacy Lab; 2013.
16. Markham A, Buchanan E. *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (version 2.0)*. Association of Internet Researchers, 2012. <http://aoir.org/reports/ethics2.pdf> . Accessed July 30, 2015.