

# Ethical Concerns for Studying Health Online

Matthew Bietz  
University of California, Irvine  
Irvine, USA  
mbietz@uci.edu

## DIGITAL HEALTH

There has been a significant interest among health researchers directed toward networked technologies and online services as platforms for research [1]. Data from wearable devices, online activities, social media interactions, environmental sensors, and the “Internet of Things” may provide insight into the day-to-day activities and lifestyle factors that are increasingly considered key health determinants. At the same time, these technologies open possibilities for interventions to modify behavior, provide feedback, or otherwise improve health outcomes. Online health communities not only provide support for individuals managing disease or working to improve their health, but they also serve as important sources of health data. PatientsLikeMe, for example, uses data from participants and forums to better understand everything from patient attitudes to drug efficacy (<https://www.patientslikeme.com/research/publications>).

While often characterized as “personal” technologies, groups and communities are almost always implicated in their use. Social media is, by definition, social. Many wearable devices and apps include social components (like activity-level competitions or the ability to post one’s progress to social media). The Internet of Things and other sensors frequently capture data not only about the individual but about any family members or others who share the same spaces or devices. These new sources of data raise ethical considerations at both the individual and community level.

Research ethics around these new forms of health data has been a key concern of the Health Data Exploration (HDE) Project (<http://hdexplore.calit.net>). HDE’s research agenda focuses on a single, large question: *given that these new forms of data may provide insight into health and other social and behavioral concerns, what are the opportunities and challenges for conducting meaningful, rigorous, and responsible research with these data?* These concerns are represented in Figure 1.

In the rest of this position paper, I will briefly outline a few concerns that arise at the group / community level when conducting research using these new forms of data.

## INCLUSION

One of the fundamental principles outlined in the Belmont Report is “respect for persons.” Typically this has been interpreted as respecting the individual’s right to make an informed decision about participating in research. Practically, this is usually operationalized as “get informed

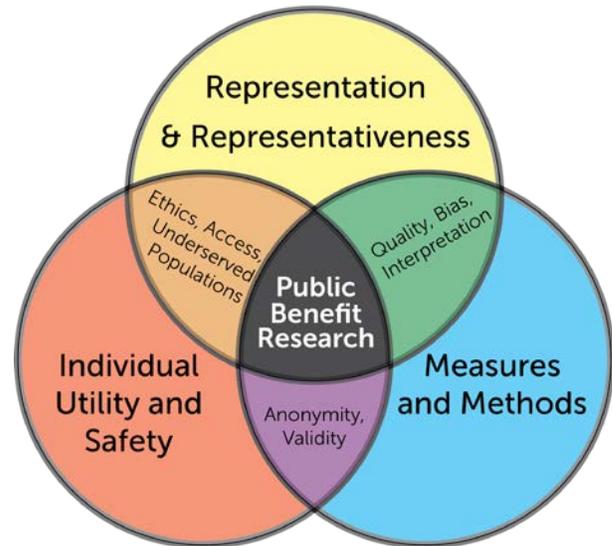


Figure 1: Health Data Exploration Research Areas

consent.” The mechanics of informed consent are significantly more complicated in research that dispenses with the traditional personal interaction between researcher and research subject. How do we ensure *informed* consent when we know most people never read terms of service? How do we get signatures from anonymous participants? How do we get consent when membership and participation change very rapidly?

These are important concerns, but there are other concerns as well. One of the challenges of these new data sources is that, while the data is very personal (in that it can reveal much about individuals), it is also often very social. Like poorly designed fishing gear, these new technologies tend to have a significant “bycatch,” the unintended data caught up in widely-cast nets. Online conversations can rarely be limited to just those individuals who have provided consent. Wearable devices and environmental sensors, especially when combined with other data like GPS location or social networks, frequently capture or imply information about anyone with whom the focal individual interacts. The assumption that an email address is used only by a single individual is frequently erroneous. Social network data can be used to make inferences based on what is known about ones’ acquaintances [3]. Given the power of search engines and statistical inference techniques, we must also remain

skeptical of any approach in which ethical concerns are assumed to be mitigated by de-identification.

### **EXCLUSION**

While many ethical hands are being wrung over issues of inclusion (privacy, consent, etc.), we must also seriously consider the ethical consequences of exclusion from research datasets. There is a history of under-representation of certain communities in health research (e.g. even with significant attention to this disparity in recent years, more health research is still done on men than women [2]). As we move online, we must remember that “Billions of people worldwide remain on big data’s periphery. Their information is not regularly collected or analyzed, because they do not routinely engage in activities that big data is designed to capture” [4]. While traditional demographic categories (gender, age, socio-economic status, etc.) are important determinants, we must adopt not just a social but a sociotechnical understanding of exclusion. For example, while researchers may want to use data from fitness trackers and “smart” scales to understand obesity, the design of these devices (short straps on wrist-worn activity monitors, low weight limits on scales) and their algorithms (accelerometer interpretation tuned to athletic gaits) may actually exclude obese people from the studies. Buying and eating habits are harder to track in immigrant neighborhoods where transactions are more likely to use cash instead of electronic payment. Privacy risks may be felt more keenly by some groups (e.g. immigrants, those with certain chronic conditions, etc.) than others.

Online research has the potential not only to collect data but also to intervene. Frequently our interventions are designed to bring benefit to individuals or communities – we create new technologies, modes of interaction, or feedback mechanisms that may have positive outcomes. It is not uncommon for research prototypes to find their way into extended use beyond the study. It is important to ask who gets to benefit from these interventions.

When we wish to conduct research in online communities, we must be more cognizant of the ways that some groups of people are left out and interrogate the ethical consequences of those exclusions.

### **HARM**

Many of our ethical calculations depend on understanding the potential harms that might occur from various situations. Medical research has gone to great lengths to quantify and characterize the kinds and severity of harms that might come from a new drug or medical procedure. However, calculating the potential harm to participants in these new forms of online and digital research presents a challenge. We often simply don’t know or have models that can predict how often a significant harm comes from using an online forum or wearable device. The kinds of harms that might occur may be more diffuse, less visible, and slower to take form (e.g. increasing anxiety, changes in social standing, etc.). We also are faced with situations in which the research poses no

significant individual harm but may actually have consequences for non-participants or the community as a whole.

At the same time, many of our research methods probably present less harm in online spaces than one may be exposed to in the course of a normal day. Our interactions are constantly algorithmically influenced, many major companies continuously expose their users to A/B testing and other experimentation, and huge amounts of data about us are being collected and aggregated behind the scenes with little ethical or safety review. Given this state of play, how do we characterize the increase in harm potential that comes from participation in research?

### **NEW STAKEHOLDERS**

Finally, it is important to point out that most research conducted in online communities or with new digital data sources involves, at some level, for-profit corporations, government entities, or other organizations that may not be held to the same ethical standards as academic researchers. Finding an ethical common ground is an important first step in conducting research across organization and sector boundaries.

I look forward to discussing these and other issues at the workshop.

### **REFERENCES**

1. Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., et al. Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association* (2015). DOI=<http://dx.doi.org/10.1093/jamia/ocv118>.
2. Glezerman, M. *Gender Medicine: The Groundbreaking New Science of Gender- and Sex-Based Diagnosis and Treatment*. Overlook Press, 2016.
3. Jernigan, C., & Mistree, B. F. T. *Gaydar: Facebook friendships expose sexual orientation*, 2009.
4. Lerman, J. Big Data and Its Exclusions. *Stanford Law Review Online*, 66, 55 (2013), 55-63.

### **BIOGRAPHICAL SKETCH**

Matthew Bietz is an Assistant Research Professor in the Department of Informatics at the University of California, Irvine. His research focuses on the infrastructures and sociotechnical conditions that support data sharing and reuse. He is a co-Investigator on the Health Data Exploration Project.