# Privacy Attitudes for Emerging Health Technologies

**Logan Warberg**

**Daniel Calderón Villela**

**Jun Ma**

**Preethi Josephina Mudialba**

**Nelson Wong**

**Lorrie Faith Cranor**

Carnegie Mellon University

Pittsburgh, PA 15213, USA

warberg@cmu.edu

ddcalder@andrew.cmu.edu

junm2@andrew.cmu.edu

pjosephi@andrew.cmu.edu

nelsonw@andrew.cmu.edu

lorrie@cmu.edu

## Abstract

We conducted a preliminary survey to measure users' privacy attitudes regarding the collection and sharing of health information. Using a sample of workers from Amazon Mechanical Turk, we measured users' privacy attitudes for different health technologies across four privacy themes previously identified within the literature. We compared our results with a sample previously collected by a research team at the University of California in San Diego (UCSD). For our online sample of size n=50, we found that the responses for most items do not differ between sample populations at a statistically significant level. Participants' attitudes relating to the privacy themes largely do not correlate with privacy measures captured elsewhere in the survey. We discuss these results and consider several limitations within our methods and their potential impacts on our results.

## Introduction

The emergence of new health technologies has created unique privacy challenges for consumers. As these devices and services collect potentially sensitive health data, users are faced with difficult choices over how their data will be used. This study seeks to measure privacy attitudes regarding the collecting and sharing of health information in the context of six health technology scenarios. This study was conducted in collaboration with a research team at the University of California San Diego (UCSD), and builds upon previously collected survey and focus group data.

Our study consists of two parts. In the first, we conducted a replication of a survey previously administered by UCSD to a sample of patients, underrepresented minorities, and public-school

adolescents in the San Diego area. Our analysis compared the responses from a sample of online workers to these samples collected previously.

The second portion of our study examined four privacy themes identified by Cheung et al. and explored within focus group interviews conducted by UCSD [2]. We created closed-ended questions to capture these themes within a survey. We additionally constructed a regression model to identify potential inferences between participants' responses to these new questions and their measured IUIPC scores.

## Background

Multiple scales have been developed to measure the privacy sensitivity of users across various domains. Several of these scales are designed to measure general privacy attitudes regarding data collection – typically in an online setting. Perhaps the most prominent of these is the Internet Users' Information Privacy Concerns (IUIPC) scale by Malhotra et al [3]. Due to its use within the privacy literature, this scale is used within the UCSD survey and within our own survey.

*Health Privacy Concerns*
Existing privacy scales do not measure the privacy attitudes of users regarding health data. Recent literature regarding the privacy concerns of health data has started to fill this gap. Bietz et al. investigated the privacy concerns regarding personal health data (PHD) for three different sets of participants: adopters of health tracking devices, researchers whose work may intersect with personal health data, and companies that market and sell self-tracking health devices and services [1]. The authors administered a combination of online surveys and in-person interviews, finding that participants were largely

willing to share their anonymized PHD if it benefited the public good.

Building upon the work of Bietz et al, Cheung et al. sought to understand the privacy attitudes of early adopters of new heath technologies such as genome sequencing [2]. They investigated whether these early adopters conceptualized privacy in different ways from the general populace by conducting 11 in-person interviews along with focus groups. They found that four general themes emerged: Interviewees were willing to share personal data to contribute on scientific improvement, but they still had concerns about the purpose for which their data is used (understanding), who has access (control), and whether the use of their data would result in negative personal outcomes (discrimination). We examine these four themes within our survey.

## Methods

We constructed a survey consisting of the questions from the previous UCSD health privacy survey along with four closed-ended questions designed to capture the privacy themes from Cheung et al. These questions measured participants' agreement with statements reflecting each theme on a seven-point Likert scale. Each set of four questions was applied across six hypothetical scenarios concerning emerging health technologies identified within the UCSD focus group study: electronic health records (EHR), medical genetic tests, recreational genetic tests, biobanks, fitness devices, and social media. After each scenario, we included an open-response question that asked participants to explain what they considered to be the most important factor in answering the thematic questions. Responses were collected from a sample of 50 workers on Amazon Mechanical Turk (mTurk).

Figure 1 shows the text used within the medical testing scenario in the survey.

> Genetic testing is a type of medical test that examines the DNA in your cells. Your DNA is responsible for determining your physical characteristics such as eye and hair color, along with how your body might react to particular drugs or medical treatments. Genetic tests are often used by doctors to identify susceptibility to certain diseases and medical conditions, such as Alzheimer's disease. The results of genetic tests are only available to your doctor and hospital staff.
>
> Imagine that you visit a doctor's office for a health issue. Before prescribing a particular treatment, your doctor performs a genetic test by collecting a DNA sample (e.g. saliva) to determine how your body might react.

**Figure 1.** Text of the medical genetic testing scenario used within the survey.

*Analysis*
Our analysis consisted of both quantitative and qualitative analysis designed to answer the research questions raised by both the replication and extension portions of our study. In the quantitative analysis, we compared the survey responses collected from the mTurk sample to those collected by the UCSD study on the questions shared by both of our surveys. We ran Welch's two-sample t-tests comparing each individual question in the original survey with those from our survey to determine whether statistically significant differences existed in the responses between the two sample populations.

To determine in part whether our questions captured the four privacy themes outlined in Cheung et al, we calculated the 24 mean responses composed of each of the four thematic question responses for each of the six scenarios, and analyzed how these mean responses compared across scenarios and across themes. To determine if there was a correlation between IUIPC question responses and the responses to our thematic questions, we ran ordinary linear regressions with the given thematic question response as the dependent variable, and the 10 IUIPC questions as the independent variables.

In the qualitative analysis of our thematic question responses, we read each of the responses to our open-ended questions and searched for recurring themes from each scenario. We looked for themes that were not directly related to the themes from the paper by Cheung et al. After identifying the themes from each scenario, we compared them across each scenario and looked for themes that appeared in all six scenarios as well as other unique and interesting themes that appeared in only a small subset of scenarios.

## Results
*Comparison of Sample Populations*
A comparison of the responses between the two sample populations found significant differences ($p<0.05$) on 14 of the 70 survey questions. After controlling for the male subset within both samples (as gender between the two sample populations was found to be significantly different at the $p<0.01$ level) and
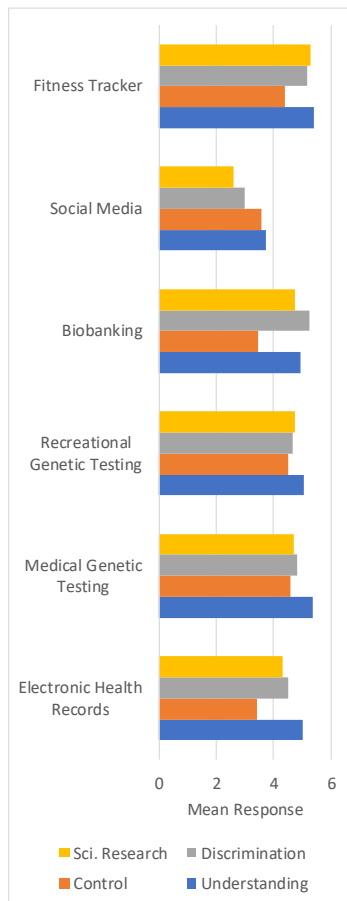
**Figure 2.** Average Agreement Across Privacy Themes

correcting for multiple testing, we found significant differences in responses for 5 questions across the survey. Thus, we see that while these two populations responded similarly, their responses in some areas are significantly different.

*Privacy Theme Questions*
In Figure 2, we calculate the mean score for each of the four privacy theme questions across all six health technology categories. Each question measured on a 7-point scale the extent to which participants agreed with statements concerning their understanding of how their data would be used, their ability to control who has access to their data, their perception that their data would be used to treat them fairly, and their comfort with contributing their data to scientific research. 1 indicates "strongly disagree" and 7 indicates "strongly agree." Figure 2 shows that participants' responses to each of the theme questions were largely internally consistent within health technology categories - with minor variations.

Participants typically agreed or disagreed with all the privacy theme statements within a category - apart from Electronic Health Records and Biobanking, where participants agreed with each of the theme statements except control. In all categories except Biobanking, participants agreed most that they understood how their data would be used. For all categories except Social Media, the average score was above 4.5, corresponding to "somewhat agree" or higher. That the responses for the four privacy theme statements are somewhat internally consistent within each of the health technology categories may indicate that the themes themselves are consistent in capturing unique but related concepts.

A qualitative analysis of our open-ended response questions revealed five prominent themes: 'privacy conscientious ', 'no privacy concern', 'stolen information', 'selling data', and 'trust'. These themes reflected the types and levels of concern participants had for outcomes relating to their potential data disclosure. Each of the themes occurred in different proportions across each of the six technology scenarios, where 'privacy conscientious' was the most prominent theme in all but the fitness tracker scenario.

*Privacy Theme and IUIPC Score Inference*
From an analysis of our regression models, we can conclude that our collected data showed significant correlations between our thematic question responses, and subsets of the IUIPC questions. It should be cautioned, however, that given the small size of our sample, these results have low statistical power, and may therefore prove spurious. To be certain of the validity of these correlations, we would need to confirm these findings with a larger sample.

## Conclusion

While an analysis of our sample data did not find a statistically significant difference in responses between the population of Mechanical Turk workers and the population surveyed by UCSD, further studies using a larger sample size may find a difference. Overall, our questions appear to capture the four privacy themes in Cheung et al. based on our quantitative and qualitative analysis, however, these themes largely did not capture the same ideas measured by the IUIPC. While we found significant correlations between our themes and two of the IUIPC questions, these possible insights would need to be measured again within a larger confirmation study.

## References

1.    Matthew J Bietz, Cinnamon S Bloss, Scout Calvert, et al. 2015. Opportunities and Challenges in the Use of Personal Health Data for Health Research. *Journal of the American Medical Informatics Association*: ocv118.

2.    Cynthia Cheung, Matthew J. Bietz, Kevin Patrick, and Cinnamon S. Bloss. 2016. Privacy attitudes among early adopters of emerging health technologies. *PLoS ONE* 11, 11: 1–12.

3.    Naresh K Malhotra, Sung S Kim, James Agarwal, Georgia Tech, and West Peachtree. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model *Information Systems Research* 15, 4: 336–355.