Crowdsourcing Public Opinion for Sharing Medical Records for the Advancement of Science

Chunhua Weng*a, Tianyong Haoa, Carol Friedmana, John Hurdleb

aDepartment of Biomedical Informatics, Columbia University, New York City, NY, USA
bDepartment of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

Abstract

This study used Amazon Mechanical Turk to crowdsource public opinions about sharing medical records for clinical research. The 1,508 valid respondents comprised 58.7% males, 54% without college degrees, 41.5% students or unemployed, and 84.3% under 40 years old. More than 74% were somewhat willing to share de-identified records. Education level, employment status, and gender were identified as significant predictors of willingness to share one’s own or one’s family’s medical records (partially identifiable, completely identifiable, or de-identified). Thematic analysis applied to respondent comments uncovered barriers to sharing, including the inability to track uses and users of their information, potential harm (such as identity theft or healthcare denial), lack of trust, and worries about information misuse. Our study suggests that implementing reliable medical record de-identification and emphasizing trust development are essential to addressing such concerns. Amazon Mechanical Turk proved cost-effective for collecting public opinions with short surveys.

Keywords

Crowdsourcing; data collection; privacy

Introduction

Secondary use of electronic health record (EHR) data offers promise in advancing clinical and translational research. Access to electronic patient data promises efficient research at scale and at lower cost. However, such research also faces significant barriers [1], such as controversy around ownership of medical information and difficulties in sharing information among clinical researchers or across institutions. This can result in information fragmentation and lack of transparency. Due to HIPAA regulations, the need for consent to capture patients’ willingness to share their medical records has hindered research by informatics investigators [2]. Patient-led sharing of medical records for research is seen positively in several parts of the world, but barriers to patients’ willingness to share medical...
data for research remain common. In a study conducted in Australia, 95% of participants believed that medical research with data sharing was necessary in general [3]. Only 73% of these participants, however, were willing to share their own data for research. Investigators have studied differences between those willing and unwilling to share their medical records data. Buckley et al. found in a cohort that healthy controls were less willing to share their medical data [4], while Shavers et al. found that race played a factor; for example African Americans were found to be less likely to participate in research [5].

Although several studies have been published regarding different aspects of record sharing, little is known about the feasibility of engaging the public to archive the clinical phenome for medical research [6]. To investigate what lies behind the general unwillingness to share medical data for advancing science and to learn how we can increase the public’s willingness to share, we conducted a crowdsourced survey on individuals’ willingness to share medical data. This is the first study of thematic trends in free-text survey comments relating why individuals might be unwilling to share their own or their family’s medical data.

**Methods**

We used a survey instrument previously published by a panel of natural language processing experts [7]. It included questions on demographics, willingness to share medical records for research (with and without identifiers), and willingness to share specific types of health information (also a first), such as medications, laboratory test results, and chronic illnesses. We asked respondents if they would be unwilling to share their own or family members’ clinical information for research purposes, and, if not, why not.

We surveyed the opinions of the public who live in the United States and who speak English using the Amazon Mechanical Turk (AMT) system. AMT is an online crowdsourcing marketplace that recruits human workers to perform tasks for a nominal fee ($0.10 for completing this entire survey). Only one response was allowed from each AMT worker.

For data quality control we inserted two common-sense knowledge questions: “What is the first month of the year?” and “Who is the current president of the United States?” in the survey to identify valid responses. Any respondent who failed to correctly answer these two questions was removed from the analysis. Pearson’s χ² tests were used to examine relationships between respondent characteristics and willingness to share medical records. Free-text responses were categorized thematically.

Following widely accepted guidelines for thematic analysis, two independent coders performed the analysis. Each coder first familiarized themselves with the data and then generated initial codes separately. Then the two coders met to compare the codes and reach consensus. Each coder independently searched for themes in the codes that were relevant to the research question of this study: Is the public willing to share their medical records for research? What are the barriers? Multinomial regression was used to generate probability matrices to evaluate the predictive probability of gender, age, and education on willingness to share and reasons not to share.
Results

1,774 AMT workers completed the survey, of whom 1,508 answered the common-sense knowledge questions correctly and were included in the analysis. Among these respondents, 58.1% were male, 54.5% had not completed a college degree, 41% were students or unemployed, and 84.2% were younger than 40. This cohort appears relatively young and well educated. The demographics are shown in Table 1.

Willingness to Share Medical Records

AMT workers were asked to pick an answer that best described how they felt about sharing medical records for research. This included their perception of how willing they felt the general public was to share their medical records for research, how willing they were to donate medical records of deceased family members for research, and how willing they themselves would be to share their own medical data (both with identifying information and without), as shown in Figure 1. Over 74% were at least somewhat willing to donate their de-identified records for research. 32.4% thought others “might be willing” to share medical records of deceased family members for research, while only 4.2% stated they thought that others would definitely share their medical records. 35% said they would share their deceased family member’s records for research, with 6.8% saying they would not share their family’s records. Only 13.9% expressed willingness to share their identifiable medical records for research. 20% replied they would not be willing to share their data with identifying information, whereas only 5.3% expressed unwillingness to share de-identified medical records. 50.3% indicated willingness to share their de-identified medical records for research.

Willingness to Share Specific Health Information

AMT workers were asked to indicate their willingness to share different aspects of their medical records for research (Figures 2–3). Respondents checked whether they were “willing to share,” “not willing to share,” “not sure,” or “not applicable” with regards to health information such as lab results, medications, diagnostic reports, chronic illness, mental health, cancer, disabilities, and more. Respondents were most willing to share information on their demographics (Figure 2), childhood diseases, substance abuse, alcohol and tobacco use, cancer, and surgeries (Figure 3).

The five types of health information that the respondents were most unwilling to share were domestic violence, diagnostic reports, lab test results, mental health, and disabilities. Respondents were most unsure about domestic violence, lab test results, chronic illness, cancer, and surgeries. Many respondents found diagnostic reports, mental health, medications, reproductive health, and disabilities to be “not applicable”. All descriptive statistics are in Table 2.

Reasons for Being Unwilling to Share

Respondents were asked to make one to three selections for why they would be unwilling to share their medical records for research, including an option for, “Not applicable, I am
willing to share this information.” 203 respondents marked two responses, 251 checked three, and 39 marked more than three responses (results in Figure 4).

763 (50.6%) AMT workers solely checked, “Not applicable because I am willing to share my medical information,” along with 27 who also made other selections. 488 respondents checked “I don’t trust that my information will be kept confidential.” 463 marked, “It would make me uncomfortable to share this information.” 310 respondents checked, “It may compromise my future health care or insurance,” and 279 felt, “I am afraid my information will be used by the government.”

The last question in the survey asked respondents to share any other reasons they would be unwilling to share their medical records for research, with a text box for free responses. Of the 1,508 workers who took the survey, 689 (45.7%) provided a response for the final question in the survey. Only three respondents declared that they were unwilling to share without a reason, while 105 respondents stated they were willing to share. The remaining 581 respondents indicated issues or reasons that caused them to be unwilling to share, or conditions that would make them willing to share their medical information for research. Each response was coded to match one or more of the 13 codes listed in Table 3. Respondents had the most concerns with sharing identifying information, privacy, and potential harm caused by sharing. A total of 681 codes for reasons unwilling to share were applied to these responses.

Including identifying information generated the most concern from respondents. Respondents expressed worries about the information being traced back to them. One respondent posed the idea of potential harm from unauthorized sharing of medical data, “I would only be concerned about sharing my personal, identifying information because I’d be concerned it might get shared -- even inadvertently or accidentally.” All respondents expressed some willingness to share their medical data for research if identifiers were removed, and some distinctly stated they would be unwilling to share if identifiers were not removed. Privacy and potential harm tied for being the second most concerning issues.

Of the 690 respondents who provided a response, 116 expressed a desire for privacy. A strong sensitivity to the right to privacy was expressed by several respondents, e.g., that it was “nobody’s business,” that the information was “too personal to share,” and that it would be “an invasion of privacy.” One stated, “My medical records are between me and my doctor and I don’t believe they are anyone else’s business.” Some respondents also felt that their privacy would not be guaranteed if they shared their medical data for research.

Several respondents were concerned that harm may come to them from sharing their medical data. Specific concerns included identity theft, missed job opportunities or loss of employment, and medical care and health insurance discrimination. A total of 112 respondents were unwilling to share medical data because of a perceived risk of harm from sharing their information. One respondent stated, “I can see the results being used against me for jobs or health insurance.” Another felt that not sharing medical data/records would prevent potential harm, saying he would be unwilling to share to “Protect identity [and]
prevent gossip regarding my health from people who would recognize my name.” Several other respondents likewise were concerned about being judged by their information.

Confidentiality, handling of medical data by the researchers, and the likelihood of information being accessed and shared by people other than researchers was also an issue. Respondents expressed concerns over the researchers’ capabilities to keep the data secure and whether their information really would be kept confidential. 33 respondents indicated that they were concerned about confidentiality, and 23 were unwilling to share due to uncertainty in how their medical data would be handled and kept safe. One respondent posted, “Most importantly, I have no control over what is done with it, and question how securely the data is protected.” 53 respondents believed it was possible that their medical data could be hacked, leaked, or accidentally shared. One stated, “I would be afraid my personal information would get out and I’m not comfortable with that.”

3.8% of participants demonstrated concerns about accessibility, applicability, and the necessity of sharing their medical records for research. Some felt that too much information on individuals was already available for others to see (“I feel that people can find out enough about anyone. Why give out more?”), or that sharing medical records was not necessary (“I feel that this information is very personal and doesn’t need to be shared for researching.”). Others stated that they could not share because they did not have access to their own records (“I don’t even have access to them.”), or that they were unwilling to share information for specific kinds of research, such as chronic disease, since they did not have a chronic disease or information believed to be pertinent to the research. “I do not have cancer. There is no domestic violence, nor alcohol, nor substance abuse. If any of this applied to me, I would share that info.”

Sharing specific types of health information was another concern. A variety of respondents indicated that they were unwilling to share some types of health data. A few respondents demonstrated that they would be unwilling to share a given type of information specifically because sharing made them feel embarrassed. One respondent stated, “I’d be willing to share my medical data/records except for things that are personally embarrassing.”

Respondents also shared that their willingness to share was dependent on specific knowledge of the research study, such as the research purpose and who the researchers and associated academic institutions were. Some participants also stated that they were unwilling to share because they were unable to contribute to the assessments being made by researchers.

32 respondents were unwilling to share because they believed their medical records may be used for research they disapproved of or would be used for marketing purposes or for profit. Some respondents also felt that their records may be used unethically. Less than 4% of respondents were unwilling to share due to distrust in the government or because they felt they would not receive any benefit by sharing. 15 respondents were concerned about government involvement in research and healthcare, feeling the government had too much access and involvement with personal data already. A respondent said, “NSA already has all this info,” while another stated “I’m not comfortable trusting the government with my
medical information.” For some, being compensated was a significant part of being willing to share medical data for research. One respondent stated, “If the price is right then I would gladly share my information. That’s the only reason. I still wouldn’t trust that my information would be kept confidential, but it wouldn’t be the motivating factor. I’d want to be compensated.”

Barriers to sharing medical records included the inability to track the uses and users of their information and a lack of trust in researchers’ intent. Explanations for their answers were provided by 960 (54%) respondents. The primary concerns on this point were (1) lack of tracking of the users and uses of shared information; (2) fear of being harmed, including loss of medical care or insurance, identify theft, being discriminated against by future employers, and being selected for targeted advertising; and (3) lack of trust in the data collectors.

**Discussion**

In this study, Amazon Mechanical Turk proved to be a cost-effective method of collecting public opinions for biomedical research. Using AMT was also more representative of a broader population than conducting surveys among medical center personnel, although respondents did have to be computer literate. Also, the AMT survey took only a couple of days and $20 to collect responses from 1,774 respondents, while it took months for our surveys in medical centers to collect 400 responses.

The majority of participants were willing to share their medical data for research in some manner. We discovered a variety of reasons why patients may be unwilling to share their medical data for research. Some of these reasons are addressable. For example, some respondents stated they would be willing to share if they had more information on the research study, researchers, or the associated academic institution. If respondents were provided with more information and were guaranteed that their data would be protected, kept confidential, and not used by the government, willingness to share increased. Factors that are hard to address but that affected willingness to share were desire for privacy and discomfort sharing specific types of medical data. Reliable de-identification methods for medical records and trust development are critical for addressing these concerns.

Responses to the last, open-ended question of the survey were similar to responses in Weitzman et al.’s focus group study [8]; however, due to the large number of respondents who took the survey and the application of codes to the responses, we were able to determine the strongest and most prevalent concerns with sharing medical data for research. “Identifying information” was the most frequent code. Removing patient identifiers, such as name, address, and social security number, would cause the largest increase in willingness to share. This was confirmed by the increase in respondents indicating they would be willing to share medical data without identifiers over the number of those willing to share with identifiers. This pattern was also seen in a study comparing attitudes towards sharing of medical data in healthcare settings between U.S. and Japanese populations [9].

Privacy, potential harm, and unauthorized access to or sharing of information were the next most common concerns. These, along with compromised confidentiality, improper or
Unauthorized use of data and sharing specific types of health information stem from sharing identifying information. This concurs with the findings of a study evaluating the effect of authorization forms used by hospitals on likelihood of consent which found that requesting social security numbers negatively effected the return rate of authorization forms, while distinguishing the hospital name on the forms increased the return rate [10]. The present study disagrees with prior reports of biases in willingness to share specific types of medical data. Respondents were least willing to share information about diagnostic reports, mental health, and domestic violence, while being most willing to share demographics, childhood diseases, and substance abuse.

Comparison to Survey Results from Academic Centers

In a previous study, we conducted a similar survey of 2,140 highly-educated professionals, students, and staff in two academic medical centers [7]. 56% of respondents were “somewhat/definitely willing” to share clinical data with identifiers, while 89% of respondents were “somewhat” (17%) or “definitely willing” (72%) to share without identifiers. Results were consistent across gender, age, and education, but there were some differences by geographical region. Individuals in that study were most reluctant (50–74%) to share mental health, substance abuse, and domestic violence data, but remained fine sharing diagnostic data. Mental health and domestic violence seem to be sensitive areas not likely to be shared in either study cohort. The public in the present study were more willing to share substance abuse information than health professionals.

Limitations

There were several limitations to this study. Presenting most questions with predefined answer options may have prevented more freestyle responses. 54.3% of respondents did not provide an answer to the free-text question of the survey. Results of our study should be applied with care, as respondents were not asked where they currently live nor what their citizenship is. A research group at the University of California-Irvine surveyed MTurk HIT workers to evaluate their demographics and MTurk habits. It was found that 57% claimed to live in the U.S., while 32% were in India [11]. Asking for nationalities of our AMT respondents would have helped us understand how we could overcome barriers in willingness to share medical data in specific regions. Finally, people dealing with serious illnesses often have different conceptions of privacy than those who are not and, therefore, may be more willing to share health information. Our survey did not ask for health status of the respondents. In the future, we would like to study which patients are willing to share and which are not by collecting a richer variety of respondent characteristics.

Conclusions

This study is the first to leverage a crowdsourcing approach to efficiently collect the public’s preferences and concerns around sharing medical records for the advancement of science. This study sheds light on the opportunities and challenges when engaging the public on the subject of donating their medical records to support research, as well as on the need to balance privacy and specific patient needs for health information in the age of participatory health, the “measured self,” and social media.
Acknowledgments

This work was supported by National Institutes of Health grants R01LM009886, 5UL1TR001873–02, R01LM010981, UL1 TR000040, and R01LM008635. We thank Drs. William Hersh and Justin Fletcher for their assistance and input on the survey. We thank Elhaam Borhaniana for assistance with the analysis.

References

[1]. Grande D, Mitra N, Shah A, Wan F, and Asch DA, Public preferences about secondary uses of electronic health information, JAMA Intern Med 173 (2013), 1798–1806. [PubMed: 23958803]

[2]. Ness RB and Societies of Epidemiology Joint Policy Committee, Influence of the HIPAA Privacy Rule on health research, JAMA 298 (2007), 2164–2170. [PubMed: 18000200]

[3]. King T, Brankovic L, and Gillard P, Perspectives of Australian adults about protecting the privacy of their health information in statistical databases, Int J Med Inform 81 (2012), 279–289. [PubMed: 22306206]

[4]. Buckley BS, Murphy AW, and MacFarlane AE, Public attitudes to the use in research of personal health information from general practitioners’ records: a survey of the Irish general public, J Med Ethics 37 (2011), 50–55. [PubMed: 21071570]

[5]. Shavers VL, Lynch CF, and Burmeister LF, Racial differences in factors that influence the willingness to participate in medical research studies, Ann Epidemiol 12 (2002), 248–256. [PubMed: 11988413]

[6]. Corn M, Archiving the phenome: clinical records deserve long-term preservation, J Am Med Inform Assoc 16 (2009), 1–6. [PubMed: 18952936]

[7]. Weng C, Friedman C, Rommel CA, and Hurdle JF, A two-site survey of medical center personnel’s willingness to share clinical data for research: implications for reproducible health NLP research, BMC Med Inform Decis Mak 19 (2019), 70. [PubMed: 30943963]

[8]. Weitzman ER, Kaci L, and Mandl KD, Sharing medical data for health research: the early personal health record experience, J Med Internet Res 12 (2010), e14. [PubMed: 20501431]

[9]. Kimura M, Nakaya J, Watanabe H, Shimizu T, and Nakayasu K, A survey aimed at general citizens of the US and Japan about their attitudes toward electronic medical data handling, Int J Environ Res Public Health 11 (2014), 4572–4588. [PubMed: 24776721]

[10]. Bolcic-Jankovic D, Clarridge BR, Fowler FJ Jr., and Weissman JS, Do characteristics of HIPAA consent forms affect the response rate?, Med Care 45 (2007), 100–103. [PubMed: 17279027]

[11]. Ross J, Irani L, Silverman MS, Zaldivar A, and Tomlinson B, Who are the crowdworkers?: shifting demographics in mechanical turk, in: CHI ‘10 Extended Abstracts on Human Factors in Computing Systems, ACM, Atlanta, Georgia, USA, 2010, pp. 2863–2872.
Figure 1 –
Willingness to Share Data
Figure 2 –.
Willingness to Share Specific EHR Data Types
Figure 3 –
Willingness to Share Data about Specific Conditions
Figure 4 –
Concerns Behind Unwillingness to Share
| Variable  | Value                        | n (%)    |
|-----------|------------------------------|----------|
| Gender    | Female                       | 623 (41.31) |
|           | Male                         | 885 (58.69) |
| Age       | 18–25                        | 668 (44.3) |
|           | 26–40                        | 603 (40.0) |
|           | 41–55                        | 176 (11.7) |
|           | 56 or older                  | 61 (4.0)  |
| Education | Less than high school        | 10 (0.7)  |
|           | High school/GED              | 159 (10.5) |
|           | Some college                 | 646 (42.8) |
|           | Bachelor’s degree or college graduate | 547 (36.3) |
|           | Graduate or professional degree | 146 (9.7) |
| Employment| Religious                    | 3 (0.2)   |
|           | Nonprofit organization       | 31 (2.1)  |
|           | Government and public administration | 58 (3.8)  |
|           | Education                    | 89 (5.9)  |
|           | Health care and social assistance | 89 (5.9)  |
|           | Homemaker                    | 107 (7.1) |
|           | Scientific or technical      | 142 (9.4) |
|           | Not employed                 | 223 (14.8) |
|           | For-profit business          | 354 (23.5) |
|           | Student                      | 403 (26.7) |
Table 2 – Descriptive Statistics of Participant Willingness

| Willing to Share Records [1=not at all, 5=definitely] | Mean | SD  |
|------------------------------------------------------|------|-----|
| What other people think                              | 2.83 | 1.04|
| Expired family member                                | 3.77 | 1.22|
| Identified                                            | 2.98 | 1.34|
| De-identified                                         | 4.09 | 1.15|

| Willing to Share Conditions                          | Freq | %   |
|------------------------------------------------------|------|-----|
| Demographics                                         | 1,291| 85.61%|
| Childhood disease                                    | 1,291| 85.61%|
| Substance abuse                                      | 1,138| 75.46%|
| Alcohol and smoking                                  | 1,123| 74.47%|
| Cancer                                               | 1,120| 74.27%|
| Surgeries                                            | 1,119| 74.20%|
| Chronic illness                                      | 1,086| 72.02%|
| Vitals                                               | 1,056| 70.03%|
| Medications                                          | 1,013| 67.18%|
| Reproductive health                                  | 1,007| 66.78%|
| Lab results                                          | 998  | 66.18%|
| Disabilities                                         | 987  | 65.45%|
| Domestic violence                                    | 905  | 60.01%|
| Mental health                                        | 825  | 54.71%|
| Diagnostic results                                   | 799  | 52.98%|

*Stud Health Technol Inform. Author manuscript; available in PMC 2019 November 13.*
### Table 3 –
The Thirteen Themes Behind Unwillingness to Share

| Theme                                | Total | % of those unwilling to share |
|--------------------------------------|-------|------------------------------|
| Risks with identifiable information | 155   | 26.70%                       |
| Potential harm                       | 116   | 20.00%                       |
| Risks with privacy                   | 116   | 20.00%                       |
| Unauthorized access to or sharing of information | 53   | 9.10%                        |
| Lack of knowledge of research study  | 49    | 8.40%                        |
| Improper/unauthorized use of information | 33   | 5.70%                        |
| Compromised confidentiality          | 32    | 5.50%                        |
| Uncomfortable sharing medical data/records | 29   | 5.00%                        |
| Beliefs on information sharing       | 28    | 4.80%                        |
| Specific health information          | 26    | 4.50%                        |
| Medical data handling                | 23    | 4.00%                        |
| Distrust in government               | 13    | 2.20%                        |
| Insufficient compensation/no benefit to participant | 8 | 1.40%                      |