

Exploring Markets of Data for Personal Health Information

K. Thomas Pickard

StartCodon

Danville, CA, USA

ktpickard@startcodon.org

Abstract—To realize preventive and personalized medicine, large numbers of consumers must pool health information to create datasets that can be analyzed for wellness and disease trends. To date, consumers have been reluctant to share personal health information for a variety of reasons. To explore how financial rewards may influence data sharing, the concept of Markets of Data (MoDAT) is applied to health information. Results from a global online survey show that a previously uncovered group of consumers exists who are willing to sell their de-identified personal health information. Incorporating this information into existing health research databases has the potential to improve healthcare worldwide.

Keywords—*Health Information; Genomics; Personal Genetic Information; Personalized Medicine; Preventive Medicine*

I. INTRODUCTION

Globally, healthcare is regionalized, populations are aging, and individuals are increasingly more educated about healthcare issues. Consumers are also seeking pricing transparency to purchase value-based, rather than fee-based healthcare. Personal health information can accelerate next generation health services in response to these changing requirements. This study explores the possibility of exchanging personal health information in return for personal financial reward.

Benefits from information sharing between health agencies have been documented. For example, the Premier healthcare alliance claims that sharing outcomes data saved 92,000 lives and \$9 billion over the four and a half year study period. Nationwide, hospitals could save a projected 950,000 lives and \$93 billion over a five-year period [1]. In the U.S., government-sponsored programs are underway to develop an interoperable data infrastructure to link health providers nationally [2][3]. In June 2013, an InfoSys survey showed that only 58% of U.S.-based respondents were willing to share personal or family medical history with healthcare professionals [4]. Other surveys have found that consumers are more willing to share health data. In a March 2013 survey, Cisco found that 74% of consumers were willing to participate in remote health services and cloud-based storage of their personal health data [5]. Sharing health information is acknowledged to be more complex than in other areas such as

consumer spending and financial data, where similar privacy concerns exist [6][7]. New participant-centered models suggest that sharing is more likely to occur when individuals have the power to select the conditions under which they share [8]. International coalitions such as the Global Alliance for Genomics and Health also have a role to ensure interoperability of genomic and clinical data, whether shared or sold [9].

II. DESCRIPTION AND METHODS

The objective of this study was to investigate how consumer attitudes toward sharing health information may change if a financial reward was offered. Two key questions guided the online survey developed for this research:

1. If identifying information such as name, age, etc. is removed, will consumers share genomic, medical and health information for research purposes?
2. Are consumers more or less motivated to share their health information if a financial reward is offered?

Other questions attempted to gauge what data types and with whom consumers are willing to share. These questions were repeated after asking if a financial reward would influence information sharing preferences. Finally, participants were asked to select an approximate dollar amount in exchange for their health information.

The study was conducted using Ask Your Target Market (AYTM) an online market research tool with a worldwide panel of 20MM consumers. For this survey, a global panel of 400 participants was selected at random by AYTM. Questions were based on a previous health information sharing survey [10]. Participants were not consented, and responses were published in accordance with AYTM's consumer panel privacy policy [11]. The only inclusion criterion was self-declaration to be at least 18 years of age. All responses are anonymous and not personally identifiable.

III. RESULTS

During the survey time frame of July 19, 2014 from 08:26 to 10:06 UTC, 400 participants answered a seven question survey and provided approximately 5,700 data points in less than two hours.

A. Demographics

Responses were received from 59 countries and from all seven continents except Antarctica. Using the United Nations geoscheme, responses by continent were: Asia (40%); Europe (35%); North America (17%); Africa (5%); Oceania (2%); South America (2%) (Figure 1). Eleven countries (Croatia, India, Indonesia, Italy, Malaysia, Pakistan, Philippines, Poland, Romania, United Kingdom, United States) provided ten or more responses. Most respondents had completed at least four years of college (64%), with 39% reporting a Master's level education or higher. Top areas of employment were "other" (30%), IT (17%) and student (11%) (Figure 2). The majority of the respondents were employed full-time (53%). Over half of the respondents reported a household income of less than \$25,000 USD, and at least 20 responses were received for all household income levels up to \$100,000 USD (Figure 3).

The self-reported gender of the respondents was 61.5% female and 38.5% male (Figure 4). Age cohorts were banded in 5-year increments from 20 to 90+ years; the first cohort also included 18 and 19 year-olds (n=13). Age cohorts between 20 to 74 years received at least three responses in each age band, and the cohorts from 75 to 90+ received zero responses. The average age cohort was 30-34 years, and the pseudo median was 26.6 years (Figure 5).

B. Willingness to Share

Seven questions explored participants' willingness to share health information. The question "If I could remove my name, age, etc. from my health information, I would share it..." received 400 responses: 50% responded "yes"; 33% responded "in some cases"; and 17% responded "no." (Figure 6). The next question asked with whom respondents were willing to share health information: my doctor or healthcare provider (50.3%); family members (48.7%); non-profit research organizations (43%); universities and academic institutions (41.2%); insurance companies (26.8%); national databanks (25%); for-profit companies (23.7%); everyone (18.9%); other (7.2%); none of the above (6.3%) (Figure 7).

C. What to Share

When asked about sharing health information, participants responded that they would share the following data types "if I could remove my identifying information": diseases and conditions (53%); medications (47%); diet (47%); exercise (47%); traits (41%); genomic data (37%); electronic medical records (35%); other (14%); none of the above (11%). The respondents' average "willingness to share" across the data types surveyed was 37% (Figure 8).

D. Exchanging Health Information for Financial Reward

When asked about being more likely or less likely to exchange health information in return for financial reward, about two-thirds were more likely, approximately one-quarter would be neither more likely nor less likely, and less than 10% would be less motivated (Figure 9).

E. Willingness to Sell with Reward

When asked about selling health information with financial reward, participants responded that they would sell data "if I could remove my identifying information" to: universities and academic institutions (37%); hospitals or healthcare systems (37%); non-profit research organizations (36%); doctors or healthcare providers (35%); for-profit companies (30%); national databanks (29%); insurance companies (27%); anyone who will pay me (26%); none of the above (17%); other (8%) (Figure 10).

F. What to Sell with Reward

When asked about selling health information with financial reward, participants responded that they would share the following data types "if I could remove my identifying information": medications (47%); diseases and conditions (45%); exercise (42%); diet (41%); traits (for example, eye color) (37%); electronic medical records (35%); genomic data (DNA) (32%); none of the above (18%); other (11%). The respondents' average "willingness to share" with financial reward across the data types surveyed was 34% (Figure 11).

G. One-time Payment

Respondents expected to receive the following one-time payment amount (USD) in exchange for their health information: \$0 – 50 (14%); \$51 – \$100 (20%); \$101 – \$200 (18%); \$201 – \$500 (16%); \$501 – \$1000 (10%); greater than \$1000 (22%); other (3%); would not sell their health information (18%) (Figure 12).

IV. DISCUSSION AND LIMITATIONS

This section discusses a number of limitations of the study, and highlights several general themes. Regarding the survey instrument, all questions received at least one response. The high survey response rate can in part be attributed to the self-selected nature of the participants, who were paid to answer the survey. The amount paid to each participant is unknown, but the cost for each completed survey response was \$0.75 USD. A number of known limitations exist surrounding the use of self-reported data, including bias that is introduced by surveying only Internet-connected consumers [12][13][14]. The greatest limitation of the study is that the number of respondents per country is relatively small, with only 11 countries providing more than ten responses. This limitation can be addressed by using "custom balancing" from AYTm. This feature ensures a minimum number of responses per desired trait. Age cohort correlations were performed against several variables where at least 20 responses were available within each age band (ages 18 – 49). Drawing conclusions for

participants who are age 50 and older should be considered unreliable due to the small number of responses in each age group ($n=8.4$ on average). Finally, with 64% of all participants reporting the completion of at least four years of college, the respondents' education level is much higher than the global tertiary education enrollment average of 25% [15]. Generalizing these results requires additional research across a broader population with more diverse educational backgrounds. In addition to these limitations, a number of general themes emerged that may influence how and when consumers share or sell their health information, discussed below.

A. Age-related Sharing

For age cohorts with more than 20 responses (ages 18 – 49), a strong correlation exists between age and receiving a financial reward in exchange for health information. The influence of financial reward decreases significantly with age (r -value = -0.93 , linear $R^2 = 0.87$) (Figure 13). A less significant correlation exists between age and the one-time expected payment. The expected payment follows an inverted parabolic curve that appears to peak around age 40 (r -value = 0.49 , parabolic $R^2 = 0.68$) (Figure 14).

B. Sharing and GDP

Gross Domestic Product (GDP) appears to have a mild predictive effect on the average expected one-time payment in exchange for health information. By using the 11 countries with at least ten responses and then eliminating the two countries with the lowest and highest one-time payment amounts (Pakistan and United States, respectively), the average one-time payment amount decreases linearly from \$250 to \$100 as GDP increases from \$56B (Croatia) to \$2,471B (United Kingdom) (linear $R^2 = 0.46$) (Figure 15) [16]. The United States confounds this analysis with an average expected payment of over \$300. This information may be useful in shaping MoDAT pricing in these countries.

C. Data Markets for Health Information

Ohsawa and co-authors envision a Market of Data (MoDAT) where data can be “sold, opened free, or shared after negotiation.” Negotiation between parties is based on physical or virtual “data jackets” that contain metadata, sample data and other instructions to facilitate transactions between buyers and sellers [17][18]. An emerging dimension in data selection may involve the use of prediction markets, which can beat polls or experts in terms of prediction accuracy [19]. CrowdMed is an example of prediction markets used in healthcare. Patient “case solvers” place stock market-like bets that raise and lower probabilities of potential solutions for consumers with complicated health issues [20].

Finding companies who are willing to pay consumers directly for their health information remains a challenge. In the U.S., companies such as Miinome had planned to build a platform to allow consumers “to possibly cash in on the value of their DNA by selling the data back to marketers and researchers” [21]. Instead, the company now markets “genetic

enlightenment” for wellness and lifestyle choices based on genetic profiles [22]. Other projects, such as the Hub of All Things (HAT) aims to develop “a platform for securely storing the totality of data and providing a safe place in which to trade it.” By connecting previously disparate data elements, HAT believes that businesses can create new experiences for consumers in exchange for their “time, attention, endorsement and data” [23].

On the other hand, U.S.-based companies such as Axiom, Epsilon and Experian broker the sale of consumer information on a large scale. In these cases, consumers do not receive any financial reward. In two recent contemporary accounts, expectant mothers uncovered how companies traded health information about their pregnancies, demonstrating a clear market for health information [24][25]. Maintaining privacy and interacting with the Internet was difficult, and remains one of the largest obstacles to health information sharing.

D. Privacy and De-identified Information

The *privacy paradox* is the idea that consumers are worried about their privacy, but choose to disclose information despite their reservations. Acquisti and co-authors present evidence that privacy concerns follow a U-shaped distribution that cluster around extreme focal points [26]. Although some researchers believe that a growing privacy movement will require stricter controls [27], privacy preference research in social network sites shows that consumers become more relaxed with sharing over time [28][29]. The results from the current study indicate a high willingness to share health information with physicians, family members, and non-profit research institutions, despite privacy concerns. Similarly, the survey results suggest that certain consumers are willing to sell their health information to universities and academic institutions; hospitals and healthcare systems; and non-profit research organizations. In both cases, the questions in this survey assume that health information has been de-identified. Although providing an assurance of privacy has been shown to be important, it cannot be guaranteed [30][31].

E. Motivation / Value Proposition

For participants, both motivation and the amount of data shared appear to increase when a financial reward is offered (Figure 16). Monetizing the sharing process produced a surprising result, demonstrating that an alternative source of health information may exist. If true, creating a pool of health data from individuals willing to sell their health information could produce results for many.

V. CONCLUSIONS

Sharing health information is critical to the realization of preventive and personalized medicine. Key challenges include maintaining privacy and permission-based sharing so consumers can determine how and when health information will be shared. The results from this study indicate a strong willingness to share and sell personal health information in the

population surveyed. A critical mass of these consumers will enable next generation healthcare services, ultimately leading to improved health outcomes for all.

SUPPLEMENTAL INFORMATION

1. *Exploring-Markets-of-Data-for-Personal-Health-Information-Questions.pdf* – PDF document with survey questions
2. *Exploring-Markets-of-Data-for-Personal-Health-Information-Responses.csv* – Comma-delimited file with all response data
3. *Exploring-Markets-of-Data-for-Personal-Health-Information-Analysis.xls* – Microsoft Excel document with analysis data
4. *Exploring-Markets-of-Data-for-Personal-Health-Information-Figures.pdf* – PDF document with all figures

ADDITIONAL INFORMATION AND DECLARATIONS

K. Thomas Pickard has no conflicts to declare.

ACKNOWLEDGMENT

The author wishes to thank Professor Yukio Ohsawa for the opportunity to present this work.

REFERENCES

- [1] B. Monegain. "Data-sharing initiative reduces deaths," *Healthcare IT News*. March 19, 2013.
- [2] M. McKinney. "HHS to develop data-sharing plan for outcomes research," *ModernHealthcare*, 2013.
- [3] U.S. Department of Health and Human Services. "States prepare for seamless exchange of health records after disasters: Ten state HIE programs have established infrastructure for secure exchange of health information," 2013.
- [4] InfoSys. "Digital consumer study," 2013. <http://www.infosys.com/digital-consumer-study>
- [5] B.T. Horowitz. "Cisco says 74 percent of consumers open to telehealth sessions," *eWeek*, 2013. <http://www.eweek.com/networking/cisco-says-74-percent-of-consumers-open-to-telehealth-sessions/>
- [6] R. Agarwal and C. Anderson. "The complexity of consumer willingness to disclose personal information: Unraveling health information privacy concerns," *eHealth Initiative's 5th Annual Conference*, Washington, DC, 2008.
- [7] S. Balam et al. "Studies on understanding individual willingness to disclose genetic information to public and private stakeholders," Thesis. University of Maryland, College Park, MD, 2013.
- [8] J. Kaye et al. "From patients to partners: participant-centric initiatives in biomedical research," *Nature Reviews Genetics*, 2012. doi:10.1038/nrg3218
- [9] Global Alliance for Genomics and Health. 2014. <http://genomicsandhealth.org>
- [10] K.T. Pickard and M. Swan. "Big desire to share big health data: A shift in consumer attitudes toward personal health information," *In 2014 AAAI Spring Symposium Series*, 2014.
- [11] AYTm. "Umongous, Inc. consumer panel privacy," 2014. http://aytm.com/pages/consumer_privacy
- [12] M. Swan. "Crowdsourced research studies," *J Med Internet Res*. vol. 14, iss. 2, e46, 2012.
- [13] M. Swan. "Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen," *J. Pers. Med.* 2(3), 93-118; doi:10.3390/jpm2030093, 2012.
- [14] M. Swan. "The quantified self: Fundamental disruption in big data science and biological discovery," *Big Data*. 1(2): 85-99, 2013.
- [15] UNESCO Institute for Statistics in EdStats. "Tertiary education – A global report," 2012. <http://datatopics.worldbank.org/education/files/EducationAtlas/TertiaryEd.pptx>
- [16] United Nations Statistics Division. "GDP and its breakdown at current prices in US Dollars," 2013. <http://unstats.un.org/unsd/snaama/dnllist.asp>
- [17] Y. Ohsawa, C. Liu, Y. Suda, H. Kido. "Data jackets for synthesizing values in the market of data," *In 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems – KES2013*, 2013.
- [18] Y. Ohsawa, H. Kido, T. Hayashi, C. Liu. "Innovators marketplace on data jackets for externalizing the value of data via stakeholders' requirement communication," *In 2014 AAAI Spring Symposium Series*, 2014.
- [19] M. Swan. "Prediction markets round-up," July 14, 2014. <http://futurememes.blogspot.fr/2014/07/prediction-markets-round-up.html>
- [20] CrowdMed. 2014. <https://www.crowdmed.com/>
- [21] D. Hernandez. "Selling your most personal item: You," *Wired*, March 27, 2013. <http://www.wired.com/2013/03/miinome-genetic-marketplace/>
- [22] Miinome. 2014. <http://miinome.com>
- [23] L. Phillips. "A market for personal data," 2014. <http://hubofallthings.com/market-for-personal-data/>
- [24] N. Holt. "Bump tracker: Nine months of big data," *The Atlantic*. August 12, 2014. <http://www.theatlantic.com/technology/archive/2014/08/the-bump-effect-nine-months-of-big-data/375820/>
- [25] S. Gray. "One woman's attempt to hide her pregnancy from big data — it's more difficult than you'd expect," *Salon*. April 28, 2014. http://www.salon.com/2014/04/28/one_womans_attempt_to_hide_her_pregnancy_from_big_data/
- [26] A. Acquisti, L. John, G. Loewenstein. "What is privacy worth?" 2010 future of privacy forum's best "Privacy Papers for Policy Makers" competition, 2010.
- [27] K. Lewis, J. Kaufman, N. Christakis. "The taste for privacy: An analysis of college student privacy settings in an online social network," *Journal of Computer-Mediated Communication*, 14(1), 79-100, 2008.
- [28] S. Barnes. "A privacy paradox: Social networking in the United States," *First Monday*, vol. 11, no. 9, 2006.
- [29] S. Utz and N. Kramer. "The privacy paradox on social network sites revisited: The role of individual characteristics and group norms," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 3(2), article 1, 2009.
- [30] Health Data Exploration Project. "Personal data for the public good," 2014. <http://www.rwjf.org/content/dam/farm/reports/reports/2014/rwjf411080>
- [31] M. Gymrek, A.L. McGuire, D. Golan, E. Halperin, Y. Erlich. "Identifying personal genomes by surname inference," *Science*. vol. 339 no. 6117 pp. 321-324, 2013. doi: 10.1126/science.1229566