Author’s response to reviews

Title: Patients' willingness to share digital health and non-health data for research: A cross-sectional study

Authors:

Emily Seltzer (Emily.Seltzer@pennmedicine.upenn.edu)
Jesse Goldshear (Jesse.Goldshear@uphs.upenn.edu)
Sharath Guntuku (sharathchandra92@gmail.com)
Dave Grande (dgrande@wharton.upenn.edu)
David Asch (asch@wharton.upenn.edu)
Elissa Klinger (Elissa.Klinger@pennmedicine.upenn.edu)
Raina Merchant (Raina.Merchant@uphs.upenn.edu)

Version: 1 Date: 16 Jun 2019

Author’s response to reviews:

>> We thank the reviewer (and the editor!) for their comments. We have taken these comments into account and believe that the revised manuscript is much improved as a result. Our responses to individual comments are detailed below; we have addressed the comments and concerns as much as possible. We have highlighted the changes in the manuscript in blue.

Reviewer reports:

Reviewer 1

This research paper explores a very interesting and important topic, that of the relationship between patient perceptions of their respective data-traces with patient preferences and willingness to share data for research. The background requires a bit more detail, references, and explanation, and the writing style made a bit stronger, yet it nonetheless introduces the topicality of the research. The overall structure of the paper works well. The statistical analysis is sound and the tables are clear and easy to follow. The paper is moreover coherent with the larger domain and contributes to our knowledge of patient data sharing preferences.Outlined below are points that should be addressed.

Point 1: Background
Target's predictive algorithm's identification of a teen pregnancy is an often cited case and exemplifying case of the relevance of shopping data to health. However, a bit more detail of the cultural context is needed. Specifically, in which culture did this happen, and who 'criticized' Target's tracking practices? The media? The family involved? Was it critiqued internationally? Please also make stronger the statement that the sensitivity of the information and the process of communication was what was 'unacceptable' and to whom?

Response: We added more details and explanation to the Target marketing example in lines 68 - 76. The paragraph now reads as follows:

“In 2012, the retailer Target sent advertisements for baby products to a teen who had not disclosed her pregnancy to her parents. Target had concluded the teen was pregnant after she purchased items like unscented lotion and cotton balls, which figured into algorithms predicting pregnancy.1 The algorithm was allegedly accurate, but the tracking practices of Target were criticized after it was reported to the public.1,2 Individual customers reportedly complained and reported that predicting pregnancy from purchases was “creepy”.1 After the public response Target was reported to have modified its marketing practices, instead of only sending baby supply coupons to women that their algorithm deemed pregnant, Target would send baby supply coupons with other home goods items mixed in.1 “As long as a pregnant woman thinks she hasn’t been spied on, she’ll use the coupons”.2 The sensitive nature of early pregnancy makes the practice of targeted marketing seem particularly invasive. There are many regulations enacted to protect traditional clinical health information, but there is less guidance for how health related digital data should be protected. “

Point 2: Background

"Contemporary practices to safeguard the privacy of health related data…” - what 'contemporary practices' are you referring to? National practices, international practices? Guidelines or legislation? Also please reference - 'health data were largely seen as…'

Response: Contemporary practices was referring to national HIPAA guidelines. This was added to the background in lines 77 - 79 and a reference to what health data is covered by these guidelines.

Reference:

"Prior work suggests that many individuals are willing to share substantial personal information… a large genre of science fiction…." More references are needed to support your points in this and the following paragraphs.

Response: We removed the science fiction sentence from the background to be concise, as we felt the comparison with science fiction tropes would need greater explanation than available in this paper. We also added references that elucidate what individuals are willing to share. The paragraph now reads as follows:

“Prior work suggests that many individuals are willing to share substantial personal information but do not like to be surprised by how their data are used.5 The contextual integrity theory includes the idea that perceptions of privacy are based on ethical concerns that evolve over time.6 The use of proprietary algorithms to categorize individuals on the basis of behaviors or tendencies can be viewed as ‘creepy.’ In the context of health care, prior work has found that 85% of patients who reported using social media and who were willing to participate in research also agreed to share these data sources and have them linked to their electronic health record for health research.7 That consent was provided in the context of active patient care, where trust, and also perhaps perceptions of information safeguards, are typically high. Beyond social media however, little is known about what other digital traces patients would willingly share with health researchers, under what circumstances, and for what reason.”

References:


Point 4: Background

Since experience of privacy is later noted as one of the core parts of the survey, some more background to the topic of privacy would be helpful. One suggestion is to look at Helen Nissenbaum's work which explores public perceptions of data, its usage and of privacy.

Response: This is an excellent reference that describes privacy through the lens of contextual integrity! We added this reference to the manuscript. The paragraph now reads as follows:
“Prior work suggests that many individuals are willing to share substantial personal information but do not like to be surprised by how their data are used.5 The contextual integrity theory includes the idea that perceptions of privacy are based on ethical concerns that evolve over time.6 The use of proprietary algorithms to categorize individuals on the basis of behaviors or tendencies is often viewed as ‘creepy.’ In the context of health care, prior work has found that 85% of patients who reported using social media and who were willing to participate in research also agreed to share these data sources and have them linked to their electronic health record for health research.7 That consent was provided in the context of active patient care, where trust, and also perhaps perceptions of information safeguards, are typically high. Beyond social media however, little is known about what other digital traces patients would willingly share with health researchers, under what circumstances, and for what reason.”

Reference:


Point 5: Background

The shift between talking about a public perceptions of data usage and then patient perceptions of data usage needs to be made clearer. To signal this distinction, Line 87, you could add something along the lines of "in the context of health care, prior work has found…"

Response: This is a helpful distinction, particularly because we evaluated the “health relatedness” of the various data types. We have now added the following in lines 92 - 93:

In the context of health care, prior work has found that 85% of patients who reported using social media and who were willing to participate in research also agreed to share these data sources and have them linked to their electronic health record for health research.7

Point 6: Methods

Why was a deception design deemed as the most appropriate? What other methods were considered? Also how did you come to determine the 5 core components of the survey?

Response: We settled on a study design that relied on deception because we wanted to simulate the actual response a participant would say when asked to share the various data types on the spot. We had considered just asking participants if they would share the various digital data types with researchers without saying that they would be sharing today, but we felt that participants might only say yes because it’s what they thought we wanted to hear, but in practice they would not actually agree to share their data.
The 5 core components were based on the aims of the Social Mediome Registry study (Padrez et al.) where participants are sharing various types of social media and EMR data with our research team. In the Social Mediome study, we are using natural language processing to evaluate the language individuals use on social media and how it correlates with health care outcomes and utilization. It was in this context that we sought to evaluate perceived health relatedness. The other components of the survey were to test for confounding variable that may influence a person’s willingness to share, for example if they had past issues with an online data breach. This has now been clarified in the revised manuscript (lines 123 - 124).

“In an IRB-approved deception design, participants were asked if they would consider donating any of the 19 different data types to health researchers, and were told that if they selected “Yes” that they would be directed to do so immediately, to simulate an actual real time response.”

Point 7: Methods

19 different types of data. How did you come up with this list of 19 data types?

Response: Originally, we were focused on people’s willingness to share various social media data with health researchers, but as we completed background research into privacy and online data sharing we felt it would make the survey richer to include other types of digital data. The various data types were crowdsourced from our broader team, questions from the social mediome study, and the Fox S reference below with the goal of including digital data from a wide spectrum of industries.

Reference:

Fox S, Jones S. The Social Life of Health Information, Pew Research Center’s Internet & American Life Project 2011.

Point 8: Discussion

Line 221. Please explain how and why transparency could be beneficial for a more patient-centric approach.

Response: We updated this sentence to better reflect the sentiment of maintaining transparency with research participants. The updated sentence now reads as follows:

“As researchers gain greater insight into the relationship between online activity and an individual’s health, transparency of these findings is essential to maintain trust. A better understanding specifically of health related digital footprints is important for being able to provide guidance to patients about their use of digital platforms and sharing practices.” Further, we reference (Wong et al, JAMA) regarding a need to return research results to patients. The
manuscript now includes the following statement: Increasing focus on returning research findings to patients is evident in the digital era where there is a movement toward open science and better patient engagement.

Reference:


Point 9: Discussion

The limitations of this study are outlined sufficiently. However what are directions for further research?

Response: We added the following sentence to the revised manuscript to discuss future directions:

“Future work should be directed towards understanding the contexts in which patients are most likely to donate data for research use, and how they would want insights shared with them.”

Point 10: Discussion

"Regulations protecting the privacy of health information…” - please provide a reference.

Response: We added the following reference to the manuscript:

Reference:


Point 11: Conclusion

The conclusion is very short. Please make more explicit what this research paper contributes to the larger domain.

Response: We updated the conclusion as follows:
Patients use a variety of digital applications that generate large amounts of data. Our work demonstrates that participants would be willing to donate some of their digital data to researchers and clinicians in pursuit of health-related insights. This work adds to the larger domain of privacy and health research by connecting various digital data with perceived health relatedness. Both the willingness to share data and the perceived relatedness of those data to health do not follow conventional divisions on which health information privacy policies are built. Future work should be directed towards understanding the contexts in which patients are most likely to donate data for research use, and how they would want insights shared with them.

And added this section to the discussion which addresses possible future directions:

While providing data back to patients would be a first step, future work would also focus on the utility of this data being provided to healthcare providers via an EMR. Less defined is how this data would be interpreted, or used, or if it would even be welcomed. Regular reports of patients’ steps walked, calories consumed, Facebook status updates, and online footprints might create overwhelming expectations of regular surveillance of questionable value and frustratingly limited opportunities to intervene even if strong signals of abnormal patterns were detected.30 This future work could assess healthcare providers use of digital data incorporated in an EMR and focus on issues related to the accuracy, interpretability, meaning, and actionability of the data.31-35

References:


Reviewer 2

## SUMMARY

This paper describes the results of a survey designed to evaluate (ER) patients' willingness to share digital data for health research. Data sources (19) included clinical (e.g. prescription, EHR), social media (e.g. Facebook, Twitter), financial data (e.g. taxes, credit card data), and Google searches. Results of the survey indicate that a relatively large proportion of the people surveyed where willing to share digital data.

This is a well-written paper with some interesting results. The topic is of substantial interest given current debates on (health) information privacy.

Response: Thank you!

Comments below.

## COMMENTS

* In 64-69. Transition between first and second paragraph would benefit from more fleshing out. Response: We added the following sentence to help improve the transition between the paragraphs:

“There are many regulations enacted to protect traditional clinical health information, but there is less guidance for how health related digital data should be protected.”

* In 64. Target example could be expanded on in terms of implications. Also, finding an additional example may be Beneficial.

Response: We added more details and explanation to the Target marketing example in lines 68 - 76. The paragraph now reads as follows:

“In 2012, the retailer Target sent advertisements for baby products to a teen who had not disclosed her pregnancy to her parents. Target had concluded the teen was pregnant after she purchased items like unscented lotion and cotton balls, which figured into algorithms predicting pregnancy.1 The algorithm was allegedly accurate, but the tracking practices of Target were criticized after it was reported to the public.1,2 Individual customers reportedly complained and reported that predicting pregnancy from purchases was “creepy”.1 After the public response Target was reported to have modified changed its marketing practices, instead of only sending baby supply coupons to women that their algorithm deemed pregnant, Target would send baby supply coupons with other home goods items mixed in.1 “As long as a pregnant woman thinks she hasn’t been spied on, she’ll use the coupons”.2 The sensitive nature of early pregnancy makes the practice of targeted marketing seem particularly invasive. There are many regulations
enacted to protect traditional clinical health information, but there is less guidance for how health related digital data should be protected. “

* In 80. "perceived value" suggest "perceived and actual value"
Response: Thank you for the edit, the text has been updated.

* In 84. Consider elaborating on discussion of science fiction (include example/s).
Response: We removed this part of the background based on the comments from the other reviewer.

* In 100. "Patients seeking care in an academic Emergency Department". It would be helpful to say a bit more about the ER context (e.g. urban/rural/suburban)
Response: We added the following to the manuscript:

“Patients seeking care in a high volume, urban, academic Emergency Department from July to November 2017 were approached by research assistants for study participation.”

* Suggest looking at papers by Glenn Coppersmith (ourdatahelps.org).
Response: This is an excellent reference that strongly relates to the work we do, particularly the work utilizing natural language processing. We added a reference to a relevant paper and the revised manuscript paragraph reads as follows:

“A growing literature addresses correlations between digital data and health outcomes and health care utilization.14–21 Much of this research relies on participants sharing personal data with researchers. Less is known however about patients’ perceptions about how connected these data are with their health.”

Reference

* p12. Some elements of the table are not aligned correctly
Response: Thank you, the table alignment has been updated!
* Suggest fleshing out some of the core privacy/ethics issues more (e.g. Solve - understanding privacy; Nissenbaum - privacy in context).

Response: Thank you. We added more details to the manuscript and added a reference as follows:

“Prior work suggests that many individuals are willing to share substantial personal information but do not like to be surprised by how their data are used.5 The contextual integrity theory includes the idea that perceptions of privacy are based on ethical concerns that evolve over time.6 The use of proprietary algorithms to categorize individuals on the basis of behaviors or tendencies is often viewed as ‘creepy.’ In the context of health care, prior work has found that 85% of patients who reported using social media and who were willing to participate in research also agreed to share these data sources and have them linked to their electronic health record for health research.7 That consent was provided in the context of active patient care, where trust, and also perhaps perceptions of information safeguards, are typically high. Beyond social media however, little is known about what other digital traces patients would willingly share with health researchers, under what circumstances, and for what reason.”

Reference:


--------------------Editorial Policies-----------------------

Declarations

- Ethics approval and consent to participate
- Consent to publish
- Availability of data and materials
- Competing interests
- Funding
- Authors' Contributions
- Acknowledgements

>> We added the following to the revised manuscript.
Ethics approval and consent to participate: This study was approved by the University of Pennsylvania’s Institutional Review Board (#827652). A written consent was obtained from the participants.

Consent for publication: Not applicable

Availability of data and material: The datasets generated and/or analyzed during the current study are not publicly available due to the IRB guidelines but are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests

Funding: This project was supported by a Robert Wood Johnson Foundation Pioneer Award (72695). No sponsor of funding source played a role in: “study design and the collection, analysis, and interpretation of data and the writing of the article and the decision to submit it for publication.” All researchers are independent from funders

Authors' contributions: E. S, J. G, and R. M. M originated the study. E. S, J. G, S. C. G, D. G., and R.M.M developed methods, interpreted analysis, and contributed to the writing of the article. E. K, D.G., S. C. G., and R. M. M assisted with the interpretation of the findings and contributed to the writing of the article. All authors read and approved the final manuscript

Acknowledgements: We thank research assistants Janice Lau, Molly Casey, and Justine Marks for their assistance with data collection. We also thank the patients for sharing their responses to the survey.