Data Validity as an Ethical Issue

Authors: Dalton George, Kendall Darfler

Summary: *Two members of a design team debate issues concerning data validity early in the design process of an application designed to collect health data to be used in research.*

Emma is a software engineer who currently works for a large tech company designing a software platform that will allow for the easy creation of smartphone apps aimed at personal health data collection. This platform allows medical researchers to recruit willing study participants more easily and collect their anonymized health data through smart phone reporting of health conditions. The smartphone reporting involves answering subjective questionnaires about one's overall health, monitoring vitals like heart rate or blood pressure, and performing tasks directly on or with the mobile device. Participant data will then be used to guide medical research on treatment protocols, patient compliance with treatment regimes, and patient health outcomes. Emma has been placed on the solutions design team to replace a member who just moved off the project, and they have been in deliberation about how to best move forward with the project, which is currently undergoing preliminary testing.

Emma, who also holds a degree in public health and has extensive experience working on public health research projects, has two specific concerns about data validity issues. Data validity refers to the accuracy of the data, that is, the degree to which the data represents the population or activities being measured. First, Emma wants to know how algorithms in the mobile health app can be designed to ensure that what users are reporting about their health is truly representative of their disease experience. Second, she also thinks that the platform that has been designed so far treats the data as being reported in universal fashion across different communities of people, and that this is another problem. Emma visits Rick, the software engineer leading her design team, in his office to report her concerns.

Tackling the first data validity problem, Emma explains, "Because patients have varying understandings and outlooks on their own health conditions, they may be biased observers. They might report overly optimistic or pessimistic health data to the apps." She continues, "Because of this, there is a high potential for observer bias to go unrecognized in our platform, as it stands."

"There is also a high potential for sampling bias," Emma continues, "That is, the differing willingness of certain individuals to participate in smartphone health studies could impact how accurate the data is. This type of smartphone reporting could over-represent specific segments of the target market. Or it could leave out people without smartphones, like older and poorer people, who also tend to have more health concerns." Rick looks troubled. "Those are very important groups to bring into health studies," he says.

Emma nods and continues, "By not accounting for these biases, the data processed will not be valid, and it could produce skewed and incomplete datasets for the research our clients are doing. Our platform might be set up to create inaccurate health studies."

Rick sits quietly at his desk, mulling over Emma's explanation. After a long pause, he states, "As it is now, our platform will function with a high degree of modular cohesion, speed, and classification accuracy. It's such an elegant design." He pauses again, thinking very carefully about how to proceed. "Also," he perks up, and seems to have an idea for how to resolve the issue, "with enough training examples the classifiers implemented on the platform will converge

upon a separation of those responses, and the system will group them as needed. This could account for any population over-representation." Emma nods along, listening, but still unsure that this will address the validity of the data.

"But," she says slowly, "We do need to meet the needs of our clients, which include providing valid data. And if we don't do that, we haven't done our jobs correctly."

Rick appears frustrated. "Look," he says, "we are only responsible for aggregating the data we obtain from their research. We can't control the quality of the subjects' responses. Interpretation of the data is up to the clients themselves."

"We can't assume that clients will understand the biases of those reporting the data, though," Emma retorts. "That could lead to data being aggregated in ways to could be potentially harmful to our clients' projects. I really think we should take a step back and reconsider these biases. I know it will slow our release date, but it might be good to consult with some of our clients to make sure we design a system that effectively reports people's health data."

"I'll consider it," says Rick. "But I don't know if that's the best plan."

Questions:

1.   Do you agree with Emma? With Rick?

2.   When in the design process should data validity concerns be addressed?

3.   Should software engineers and computer scientists be expected to address validity concerns?

4.   What are the limitations of collecting and processing health data in this way? How important is it for computer scientists and software engineers to recognize those limitations?

5.   Why might diverse design input, from multiple fields be valuable in a case like this? Can you imagine other situations where such input would be valuable in the designing of algorithms or systems using algorithms?

6.   What are some specific actions the team can take to address these issues?

Resources for Further Reading

1.   Byrnes, Nanette. (2014) "Mobile Health's Growing Pains." MIT Technology Review. https://www.technologyreview.com/s/529031/mobile-healths-growing-pains/

2.   Peterson, Carolyn, Samantha, Adams A., and Paul R. DeMurco. (2015). "mHealth: Don't Forget All the Stakeholders in the Business Case." Med 2 0 4(2). doi:10.2196/med20.4349

3.   Carter, A., Liddle, J. Hall, W. and H. Chenery. (2015). "Mobile Phones in Research and Treatment: Ethical Guidelines and Future Directions." JMIR Mhealth Uhealth 3(4):e95. DOI:10.2196/mhealth.4538

4.   Eysenbach, Gunther. (2005). "The Law of Attrition." Journal of Medical Internet Research 7(1).

Dalton George, MS, and Kendall Darfler, MS, are graduates of the Drexel University Center for Science, Technology and Society. June 2017.