Researchers identify healthcare data defects, create software for easier defect detection

May 29 2020

Credit: CC0 Public Domain

Researchers at the University of Maryland, Baltimore County (UMBC) have developed a method to investigate the quality of healthcare data using a systematic approach, which is based on creating a taxonomy for

data defects thorough literature review and examination of data. Using that taxonomy, the researchers developed software that automatically detects data defects effectively and efficiently.

The research is published in the *Journal of the American Medical Informatics Association (JAMIA)*, and is led by Güne? Koru, FAMIA, professor of information systems, and Yili Zhang, a former graduate student in Koru's lab who is now a postdoctoral fellow at Northwestern University. The paper stresses that the prevalence of defects in some of the existing healthcare data can be quite high. This must be addressed to better leverage the data to improve the quality of care, reduce costs, and achieve better healthcare outcomes. The team collaborated with an anonymous healthcare organization using real healthcare datasets.

Though many researchers today are involved in the analysis of healthcare data and are invested in its importance, there is very little research being done on the quality of the data being analyzed. Ultimately, this creates a far-reaching problem because important findings from the data may be less meaningful than assumed unless significant effort and money can be invested to deal with data quality problems with ad-hoc methods. For instance, much of the data that Koru's team analyzed contained errors of duplication, mismatched formatting and incorrect syntax.

Identifying these defects in healthcare data is deeply important when it comes to healthcare facilities providing essential services. Koru explains how healthcare facilities use the data collected. Healthcare organizations must "improve upon their services based on that data, and collect more data. If we can keep this cycle going, we can actually learn and improve more quickly, which is the main idea behind the concept of Learning Health Systems, and doing so is all the more important in the COVID-19 era," he says.

In the last decade, healthcare providers in the U.S. made a large leap from keeping patient records on paper to containing all patient information in computerized databases. This jump is significant because of the opportunity it provides for analysis, but researchers are still trying to learn how to effectively leverage the data as an asset.

Koru positions his team's research on data quality as being between the fields that are working to leverage data and the fields that are working to generate it. If the data itself—the bridge that connects the two fields—contains many inconsistencies and problems, then the relevant information cannot be used to provide better outcomes for patients and facilities.

In the future, Koru will continue to work with the partner facility's healthcare professionals to build a path forward. He will collaborate further to improve the quality of data and sustain an operation that bases much of its success on the data that it can gather from health services. His team will work with healthcare administration professionals when the software tools developed through this research are adopted in organizational settings to ensure the usability and usefulness of the tools.

"The taxonomy will help data stewards to identify, understand, and manage potential data quality problems in their future work," says Zhang.

Now more than ever, healthcare facilities are relying on strong data to support patients and the healthcare field as a whole. Koru and Zhang have found that collaborations between data researchers and healthcare organizations can generate effective solutions to the problem of data quality improvement.

More information: Yili Zhang et al, Understanding and detecting defects in healthcare administration data: Toward higher data quality to

better support healthcare operations and decisions, *Journal of the American Medical Informatics Association* (2019). DOI: 10.1093/jamia/ocz201