Application of Rule-Based Data Quality Monitoring for Secondary Use of Healthcare Data

Zhan Wang, Ph.D.1, Meredith Zozus, Ph.D.1
1University of Texas Health Science Center at San Antonio, San Antonio, Texas, US

Description
A rule-based data quality assessment system has been developed based on design science framework to assess and monitor data quality in EHR data1,2 motivated by recent increases in national attention towards secondary use of healthcare data. The design validation was conducted through interview sessions in which the rule result reports were reviewed with physicians and information system owners. Both qualitative and quantitative methods were used to answer a key design validation question: does the system meet the design goal of finding data quality problems of importance to key healthcare stakeholders?

Methodology
A questionnaire was used to guide and document the interview sessions. Because this is early formative work, we specifically wanted to impose as little theory on the participant responses as possible to remain open to stakeholder goals and reactions that had not previously occurred to us. At the same time, we equally valued structured questions to obtain clear indication of whether the approach could identify data quality problems of importance to the organization. Thus, open-ended probes were used for each question and participant analysis and articulation of rationale was encouraged.

Result
Nine out of eleven participants (82%) stated that data quality problems were identified. The remaining two voiced that the rule results weren’t specific enough for them to tell if there were data quality problems identified. Eight out of eleven participants (73%) voiced that the data quality problems identified were in data used in clinical decision making. 54% participants choose medium to high impact to patient safety, the rest of participants is unable to judge because the data is not specific enough and there is no real cases that proves the effect. For financial and regulatory impact, 55% and 36% participants are unable to judge for the same reason. Among the participants, more participants indicated that data quality problems identified had a higher potential impact on finance (45%) than regulatory compliance (36%). Participants indicated multiple possible causes for the data errors. Manual input error, incorrect using code system and incorrect clinical practice were the most often selected potential reasons for data quality problems. 48% participants indicated that they would like to initiate an analysis to better understand the data quality problems, including the reason, correction and prevention of data quality problem. Only one participant choose no action and stated three reasons: (1) It is unlikely to find anything from large volume of data, (2) has no resources to analyze the problems and (3) the participant saw no likelihood of profit. In addition to enrichment of structured data, five themes were identified from the qualitative data: (1) Preferences to support use of rule results in practice, (2) Suggestions for additional representations of rule results, (3) Unintended uses of rule results, (4) Limitations of a rule-based approach, and (5) Organizational barriers to rule-based data quality management.

Conclusion
The empirical validation yielded evidence (1) supporting the ability of a rules-based approach to identify potentially important data quality problems, and (2) indicating potential organizational willingness to further evaluate rule-based data quality management. All participants indicated willingness to attend a follow-up meeting to review results of actions taken. Participants give constructive comments and ideas to improve the system. We gained so much useful information that a second iteration with participants to review the results of actions taken would be a logical next step for future research. This research directly evaluated the potential value of rule-based data quality assessment results for secondary use of healthcare data and found evidence to support further development and investigation of the approach.

References