Quality Defects Detection in Obstetric Data

Tal Aboudy

Supervised by Dr. Adir Even
Department of Industrial Engineering and Management,
Ben-Gurion University of the Negev, Beer-Sheva, Israel

In collaboration with Dr. Yoav Brezinov and Prof. Edi Vaisbuch
Department of Obstetrics and Gynecology,
Kaplan Medical Center, Rehovot, Israel

Abstract

Today, with the rapidly-increasing data resources and the growing reliance on those resources for supporting business operations and decision-making, data quality (DQ) is broadly recognized as a major issue - in information systems in general (Madnick et al., 2009; Panahy et al., 2013) and particularly in the healthcare sector (Kahn et la., 2012; Wechsler et al., 2013). DQ defects in clinical data might endanger patients' safety, introduce major risks to their well-being and decrease the quality of care (Ammenswerth and Spötl, 2009; Bowman and Rhia, 2013).

This research was conducted in collaboration with the Obstetrics & Gynecology department in Kaplan Medical Center, Rehovot. Patients with pregnancy complications often visit the department a few times during their pregnancy period for consulting and medical evaluation, and their visits are documented in the hospital's electronic medical records (EMR) system. The department's EMR offers a built-in "Copy-and-Paste" utility that enables replicating a previously-stored record that describes a patient's visit, to be used as a template for recording the next visit of the same patient. Since the majority of details collected during a visit will not change between two consecutive visits of the same patient (e.g., patient's height, demographic details, or permanent medical conditions), this utility facilitates faster data collection.

Alongside with its benefits, this "Copy-and-Paste" utility has the potential to harm DQ, due to incorrect or insufficient update of the copied values (Siegler and Adelman, 2009; O'Donnell et al., 2009). Ideally, older diagnostics should be deleted while new ones added; however, in reality, the information is not always updated properly due to lack of time or attention. As an illustration for such scenario (Table 1), the first record reflects a patient's visit that was recorded correctly. However, the second record reflects some apparent errors, which can be attributed to a record replication that was not updated carefully. In the "Diagnosis" field "Headache" is still written, while the "Visit Summary" no longer indicates headache complaints.

On the other hand, the "Visit Summary" still suggests the patient is "Not taking any medications", while the "Medications" field indicates that a new medication is taken.

Table 1: Incorrect Data Replication in a Sequence of Electronic Medical Records

Date	**Gestational Age**	**Medications**	**Diagnosis**	**Visit Summary**
28.3	32+2	-	Vertex; Headache	Pregnancy age 32+2. Not taking any medications. Complains on strong headaches in the last few days.
20.5	37+1	Zinnat, 500mg	Vertex; **Headache**; Reduced Fetal Movements	Pregnancy age 37+1. **Not taking any medications**. Complains on reduced fetal movements.

This study develops an analytical model for detection of DQ defects in obstetrics data that describes patients' state in different visits. This model can be generalized to accommodate other scenarios in which the same entity is describes at different points of time with possible record replications. The model assigns each record in a sequence with a DQ grade between 0 (low quality) and 1 (high quality) that reflects estimation of correctness. The model alerts of high probability of DQ defect in a record, when the DQ grade declines below a certain predefined threshold. The model evaluates each alpha-numeric attribute independently, considering its characteristics, e.g. its level of structure and its level of volatility (the likelihood of their values to transition over time), the distance between consecutive attribute values, and the likelihood of a state-transition over time. The DQ grade of the entire record is a weighted-average of all the record attributes' grades.

Preliminary evaluation of the model used a real-world dataset that reflects repetitive visits of patients at the department at different time points during their pregnancy period. The preliminary evaluation was conducted by using a subset of this data for model training, while another subset of 100 records were used to test the model. In this preliminary evaluation the model's accuracy was 85% and model's sensitivity was 61%. These results highlight the potential contribution of the model; however, they suggest that some development is needed before it can be offered as an operational solution.

Currently, the model is still undergoing some analytical development and enhancement. The intention is to extend both the training and the test datasets, so that they reflect a much larger sample of real-world records. Additionally, the model is planned to be further developed and

may include other forms of distance metrics, which reflect not only syntactic comparison, but also semantic comparison, based on relationship between attribute values.

References

Ammenwerth, E., and Spötl, H.P. 2009. The time needed for clinical documentation versus direct patient care. Methods of Information in Medicine 48 (1): 84-91.

Bowman, S.M.J, and Rhia, C.C.S. 2013. Impact of electronic health record systems on information integrity: Quality and safety implications. Perspectives in Health Information Management: 1c.

Kahn, M.G., Raebel, M.A., Glanz, J.M., Riedlinger, K. and Steiner, J.F. 2012. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. Medical Care 50 (Suppl. Jul): S21-9.

Madnick, S.E., Wang R., Lee, Y., and Zhu, H. 2009. Overview and framework for data and information quality research. Journal of Data and Information Quality 1 (1), Article no. 2.

O'Donnell, H.C., Kaushal, R., Barrón, Y., Callahan, M.A., Adelman, R.D., and Siegler, E.L.. 2009. Physicians' attitudes towards copy and pasting in electronic note writing. Journal of General Internal Medicine 24 (1): 63-8.

Panahy, P., Shariat, H., Sidi, F., Affendey, L.S., Jabar, M.A., Ibrahim, H., and Mustapha, A. 2013. A framework to construct data quality dimensions relationships. Indian Journal of Science and Technology 6 (5): 4422-31.

Siegler, E.L., and Adelman, R. 2009. Copy and paste: A remediable hazard of electronic health records. The American Journal of Medicine 122 (6): 495-6.

Wechsler, A., Even, A., and Weiss-Meilik, A. 2013. A model for setting optimal data-acquisition policy and its application with clinical data. In the proceedings of the 2013 International Conference on Information Systems (ICIS), Milan, Italy.