Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-303

Business Rules to Improve Secondary Data Use of Electronic Healthcare Systems

Jonathan BLAISURE[a,b,1] and Werner CEUSTERS[a,b]
[a] *Department of Biomedical Informatics, University at Buffalo, Buffalo NY, USA*
[b] *Institute for Healthcare Informatics, University at Buffalo, Buffalo NY, USA*

Abstract. The 'fit for purpose' paradigm used for data quality assessment in electronic healthcare record (EHR) systems is not so fit when assessed in the light of secondary data use. An analysis of the difficulties encountered in trying to use existing EHR data for cohort identification for prospective clinical trials and retrograde data analytics, revealed the root causes to fall in three categories: (1) issues in workflow and data registration, (2) preventable inadequacies in software configuration and personalization and (3) software development issues on the side of the vendor. By reviewing secondary data use requirements and formulating value adding business rules, development and data collection practices can be steered towards greater value in secondary data consumption.

Keywords. EHR, Secondary Use, Master Data Management, Business rules

1. Introduction

The quality of a healthcare system depends largely on the quality of data it relies upon to design, deliver, monitor and improve healthcare services that meet the needs of the targeted population. Though many definitions for data quality have been proposed, they predominantly express data quality as '*the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data*' [1] or, in summary, that datasets need to be '*fit for purpose*' [2]. Differences in definitions are then found in the amount and type of features and characteristics that relate to data quality and how they are to be measured in terms of various dimensions [3]. It is also this perspective on data quality that is typically adhered to in the healthcare domain where it forms the basis for data governance and data quality policies and procedures [4] and the translation thereof in business rules, i.e. '*statements that aims to influence or guide business processes in the organization*' [5] with the goal to ensure that the data are fit for purpose within the operational environment.

This 'fit for purpose qua intended use' perspective on datasets is in part what drives the design and use of electronic information systems within specific types of care settings such as inpatient charts, outpatient charts and practice management systems. Practice management systems deal mostly with scheduling, resource management and reimbursement. They can be standalone or integrated as part of an electronic health record (EHR). They contain operating data which strongly overlap other domain datasets

---

[1] Jonathan Blaisure, PhD student, Department of Biomedical Informatics, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo, 77 Goodell street, Suite 540, Buffalo NY 14203, USA; E-mail: jcblaisu@buffalo.edu.

such as patient and provider demographics, schedules and reimbursement details. Outpatient systems are designed for ambulatory health care settings characterized by fast patient turnaround. They keep data for billing and auditing as well as snapshots of a patient's health to support treatment and care decision making. Inpatient systems deal with more severe cases, end of life, chronic illnesses and other long term diseases and injuries. These systems deal with data about the care delivered to patients and help manage patient populations in the facility itself.

This 'fit for purpose' quality perspective on datasets managed by individual information systems results, unfortunately, from a rather narrow interpretation of what the 'intended use' of the dataset exactly is: it narrows it down to the needs of the users of these specific systems. A problem then arises when data are to be re-used or integrated in other IS, for instance to monitor public health or improve global health services and policies. This is because individual information systems, when used within similar or overlapping geographic regions, exhibit overlap in patient population and information, but in different levels of detail, such level being determined by what makes the data fit for purpose for the users of each specific system in question, and this independent from whether this level fits the purpose of the users of the other systems it is communicating with or of the system(s) which integrate the data. The inpatient EHR of one organization may, for instance, record a patient being admitted for an asthma exacerbation that required overnight intervention. Some details about that treatment will be reported to the practice management system and finally some of those details will travel to the insurance company's record system where it is processed and eventually used to decide on payment. The patient may afterwards report for a follow-up exam by their primary care provider where more data about the same incident will be collected in the outpatient EHR system. Pooling data from these systems is very complex due to proprietary and badly documented designs that are not always in line with best practices for data interoperability [6]. In the scenario sketched above, each system may have some documentation about the patient event. Data gathered from each of these systems may support or conflict with one another or be such that from the data alone it cannot even be assessed that they report on the very same event. These conflicts require reconciliation or the documentation thereof.

An ideal situation would arise when electronic information systems would not just be designed to manage datasets that exhibit a level of detail that is fit for purpose for their own operational environment, but also for any secondary use thereafter. This does not mean that said systems would need to collect more data than required for the benefits of their own users – unless, of course, there would not be an additional burden – but rather that the data points would be more precise about what they describe. Our hypothesis is that the data curation problems we are experiencing today can be partly solved by implementing appropriate business rules. The work described in this paper demonstrates the feasibility of this proposal.

2. Methods

The University at Buffalo's Institute for Healthcare Informatics (IHI) primary mission is to gather fully identified healthcare data sets into a centralized secure environment where the data can be studied, documented and appropriately distributed for secondary data use projects. To date, the IHI houses a fully identified outpatient EMR database that contains about 650 thousand patient records (outpatient data) and a data set from a local healthcare insurance provider (claims data) that contains about 1.2 million unique patients. The data is held under an institutional review board (IRB) protocol and data use agreements are signed by the data providers. The goal is to lower the barriers around data requests for secondary data usage since providers often do not have the resources or proper incentives to deliver data for research [7].

For the work presented here, we analyzed the last 19 secondary use data requests received (4 requests for claims data, 13 for cohort identification and 2 for data collection) to obtain more insight in the type of problems IHI staff encountered to deliver the data at the level of precision and quality expected by the researchers. A qualitative comparative analysis [8] was performed on the document trail (emails, meeting summaries, …) that resulted from the entire process consisting of (1) the initial – typically vague – data request, (2) the discussions between IHI staff and requesters to fully understand the data needs, (3) the construction of SQL queries, (4) the root cause analysis for requests that were not satisfiable, and (5) the development of business rules that could remediate the issues.

3. Results

We have identified three levels at which information gathering for secondary data use can be improved by implementing more appropriate business rules:

1. *Personnel workflow* – improving workflows that are (1a) ill-defined or contradictory to well-accepted data gathering practices that lead to data inconsistency at the practice management level or (1b) can be optimized for secondary use purposes.
2. *Software configuration* – the data collection software has certain configurations that can be modified by the system administrators. For example, a certain field in a screen that a healthcare professional uses during a patient visit can be marked as required rather than optional.
3. *Software development* – Identifying and documenting disparities between production and secondary data use requirements as a mechanism to present feedback to software vendors to improve future software releases.

4. Discussion

We provide here two examples of the analysis procedures applied and the business rules that were generated therefrom. Request A (Table 1) originally asked for a cohort of ‘young adult survivors of cancer’. ‘Young adult’ translated after clarification in criterion C1, i.e. patients with ages 15 to 39, of which satisfaction could be determined based on the patient’s date of birth in the EHR. Whereas what would count as ‘cancer’ was determined on the basis of ICD-codes, it was needed to relax ‘young adult having cancer’ (criterion C2) into ‘being diagnosed with cancer’ as it turned out not to be possible in every single case to determine on the basis of the EHR data (1) when the diagnosis was made, nor (2) when the cancer actually occurred since EHRs typically do not distinguish diagnoses from what they are about, i.e. the disease in the patient [9]. A date is always added when the diagnosis was recorded in the patient chart, but that can be a long time after the facts, for instance as the result of an anamnesis about prior disorders.

Table 1. Unsatisfiable criteria of request (A) for 'young adult cancer survivors'

	Business Rule Level	Criteria #	Category
a.	Workflow policy to require onset date of a disease recorded in the diagnosis in the provided onset field.	**C2**	**1&2**
b.	Requiring an onset date field through the application configuration.		
c.	Software development to incorporate other data sources to supplement current data.	**C3**	**3**

Table 2. Unsatisfiable criteria of request (B) for patients over 18 with childhood onset multiple sclerosis, an unaffected primary family member and living parents with available healthcare record.

	Business Rule Level	Criteria #	Category
d.	Workflow policy to encourage better family history record taking and use existing discrete data fields to record the data.	**C2**	**1, 2, 3**
e.	Configure new fields in the application to capture data in discrete fields instead of free text.		
f.	Require functionality in the application to appropriately implement the linking of family members.		
g.	Workflow policy to encourage better family history record taking and use existing discrete data fields	**C4**	**1,3**
h.	Require functionality in the application to enter information about relatives.		

This can be prevented by modifying workflows when entering a diagnosis by requiring a field to denote the disease onset date. This may not be known so other entries besides the onset date may be entered in the case of a fuzzy temporal value. E.g., the patient may not know exactly when the disease manifested but may know that it was during childhood. The last criterion, 'having survived cancer' (C3) presents another problem because healthcare facilities are not always aware of a patient's death, unless it happened on the premises. To illustrate this, only 1% of patients are marked as deceased and another 1% that are over 100 years of age and marked as alive. Other sources of data can be leveraged to fill in this gap such as the social security death index database. Request B (

Table 2 was a cohort search for 50 patients with (C1) childhood onset multiple sclerosis, (C2) an unaffected primary family member, (C3) current age greater than 18 and (C4) living parents with available healthcare record. The purpose of the study behind the request was to determine differences in environmental variables, phenotypical variables and genotype variations between the unaffected subjects and the affected subjects. C1 presents a similar problem as C2 discussed above. Information could in this case, however, be found in free text fields of the provider notes. These onset dates were located in different sections of the note and complicated techniques were required to extract the information. C2 could not be resolved since patient/relatives relationships are either poorly or not at all documented in the EHR. The EHR system from where the data had to be obtained does allow information to be entered about primary, secondary and tertiary relations but does not have a field to link genealogy from patient to patient. C3 could straightforwardly be calculated from the birth date of the patient. Although the source EHR system has fields in the family history section to record whether a patient's parents are living or their cause of death these fields are rarely used. C4 could thus not be satisfied. Business rule management is important, yet the literature on it is not as abundant as expected.[5]. Steinke classifies business rule types differently than we did: definition, guideline, mandate, and inference [10]. But he also takes the stance that reality should define the business rules used to govern the data. His approach is harmonious with ours, yet a full evaluation of the reality surrounding the information collections systems is costly and time-consuming. Recognizing the needs of secondary data users in creating business rules to oversee data collection presents a more practical solution.

5. Conclusion

This study was conducted in the context of one EHR system, though used by over 40 practices dispersed over the wider Buffalo area. Our analysis confirms earlier studies that the potentiality for secondary data use is not enough considered when EHR systems are designed, implemented and used [11]. Data required to perform certain analyses for cohort identification were found not to be easily obtainable due to (1) the insufficient facilities offered by the data collection systems to enter, store and retrieve required information and (2) the inadequate use of available facilities due to separate workflows and business practices across providers. The analysis of the secondary use data requirements made it nevertheless possible to develop business rules that help govern data entry and quality of data partially. Ironically, several of the requests for secondary use were issued by practitioners using the EHR system, and thus in part responsible for the lack of granularity in the recorded data. It remains to be investigated whether this awareness will lead to a smooth acceptation of and adherence to the proposed business rules and, as these processes become implemented and validated, will drive software vendors to update their systems to allow data to be entered in a way that is more faithful to reality.

Acknowledgement

This work was supported in part by Clinical and Translational Science Award NIH 1 UL1 TR001412-01 from the National Institutes of Health.

References

[1] Arts, D., d.N.F. Keizer, and G.J. Scheffer, *Defining and improving data quality in medical registries: a literature review, case study, and generic framework.* Journal of the American Medical Informatics Association, 2002. **9**(6): p. 600-611.

[2] Wang, R.Y. and D.M. Strong, *Beyond accuracy: What data quality means to data consumers.* Journal of Management Information Systems, 1996. **12**(4): p. 5.

[3] Sidi, F., et al. *Data quality: A survey of data quality dimensions*. IEEE.

[4] Davoudi, S., et al., *Data Quality Management Model (Updated).* Journal of AHIMA / American Health Information Management Association, 2015. **86**(10): p. 62-65.

[5] Shouhong, W. and W. Hai, *Business Rule Management for Enterprise Information Systems.* Information Resources Management Journal (IRMJ), 2010. **23**(1): p. 53-73.

[6] Kalra, D., et al., *ARGOS policy brief on semantic interoperability.* Studies in health technology and informatics, 2011. **170**: p. 1-15.

[7] Gilbert, R., H. Goldstein, and H. Hemingway, *The market in healthcare data.* BMJ-BRITISH MEDICAL JOURNAL, 2015. **351**: p. h5897.

[8] Kane, H., et al., *Using qualitative comparative analysis to understand and quantify translation and implementation.* Transl Behav Med, 2014. **4**(2): p. 201-8.

[9] Scheuermann, R.H., W. Ceusters, and B. Smith, *Toward an ontological treatment of disease and diagnosis.* Summit on Translat Bioinforma, 2009. **2009**: p. 116-20.

[10] Steinke, G. and C. Nickolette, *Business rules as the basis of an organization's information systems.* Industrial Management and Data Systems, 2003. **103**(1-2): p. 52-63.

[11] Hersh, W.R., et al., *Caveats for the use of operational electronic health record data in comparative effectiveness research.* Med Care, 2013. **51**(8 Suppl 3): p. S30-7.