P. Scott et al. (Eds.)

doi:10.3233/SHTI220910

Pharmaceutical Feedback Loop – A Concept to Improve Prescription Safety and Data Quality

Ines REINECKE[a,1], Franziska BATHELT[a], Martin SEDLMAYR[a] and Andreas KÜHN[a]
[a] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*

Abstract. Data quality is essential for utilizing real world data (RWD) in scientific context. Based on drug prescriptions in a hospital information system (HIS), algorithms performed a mapping of unstructured drug data to ATC codes. Visualization of the resulting distribution of structured to unstructured data based on ATC codes was created and used to explore a defined limitation of the current drug prescription highlighting the example of proton pump inhibitors. As a second step, a generalization of this approach was inductively created. As result we were able to identify 4 crucial steps for a feedback loop framework: The first step being the actual use of the HIS by clinician for drug prescription, second the processing of the entered unstructured and structured data and performing automatic analyses and visualization of the resulting distributions. The third step included an interdisciplinary expert evaluation of the data distribution followed by the fourth step, consisting of feedback to the stakeholders and generating actions as teaching or re-modelling of the system incorporating the actual learning process. The presented approach represents a continuously learning system based on RWD, although it is limited by analyzing the distribution of mapped unstructured text to ATC codes and therefore does not allow to analyze free text not mapped to ATC codes (false negatives). Future work will focus on the evaluation of this approach to analyze the impact on prescription data quality and the potential improvement on patient safety in general.

Keywords. real world data, feedback loop, patient safety, data quality, interoperability, prescriptions, secondary usage of clinical data

1. Introduction

The usage of real world data (RWD) is a great chance conducting studies at large scale as they complement randomized, controlled trials restricted to a limited scope and data [1]. Observational research based on (RWD) requires highly structured data that is harmonized to internationally used terminologies such as SNOMED-CT, LOINC, ICD10 to ensure common understanding of the results around the world [2,3]. Therefore, measuring and improving RWD quality and its data structure is an important step to move forward.

In Germany, all university hospitals belong to the Medical Informatics Initiative (MI-I) funded by the Federal Ministry of Education and Research (BMBF) and are in the

[1] Corresponding Author: Ines Reinecke, Fetscherstraße 74, 01307Dresden; E-mail: ines.reinecke@tu-dresden.de

process of establishing Data Integration Centers (DIC) to integrate existing inpatient data in one place [4]. The DICs are intended to provide a foundation for retrospective research based on the MI-I core data set available as FHIR profiles across multiple sites in a standardized manner. The data quality assessment and improvement activities are important supportive tasks during the process of data integration [5]. Those activities do not include the result communication to the clinicians and they are limited to the technical improvement steps of available RWD.

To drive data quality initiatives as a whole, this paper aims to present a concept for a feedback loop process between the DIC and the clinicians to educate them by communicating existing data quality issues. Establishing feedback loops together with clinicians will foster the improvement of documentation quality through improved skills and understanding of the impact of data quality issues for further usage in research.

2. Methods

2.1. Relevant data details

In a previous work the assessment and improvement of the structuredness of drug prescription data has been implemented at the University Hospital Carl Gustav Carus Dresden (UKD) [5]. The approach assessed the ratio between structured and unstructured drug prescriptions entered by users in the hospital information system (HIS) and applied three algorithms to the unstructured data to determine the correct ingredient information as Anatomical Therapeutic Chemical (ATC) code level 5. The algorithm results have been evaluated for a sufficient number of unstructured data to ensure correct results for at least 80% of all drug prescription in the HIS. The results show data structure differs a lot between the ATC codes.

2.2. Problem-driven approach to feedback process

The mapping process performed by the three algorithms resulted in a ratio of structured to unstructured entries for each ATC code. A visualization comparing the distribution of these ratios for all mapped ATC codes was created using an interactive scatter plot with x-axis as the (log scaled) total amount of prescription entries (structured and, following the mapping algorithm, number of unstructured) and with y-axis as the ratio of unstructured elements to the total amount of entries per ATC code. The mean of this unstructured-to-total ratio of all entries was calculated and displayed in the plot.

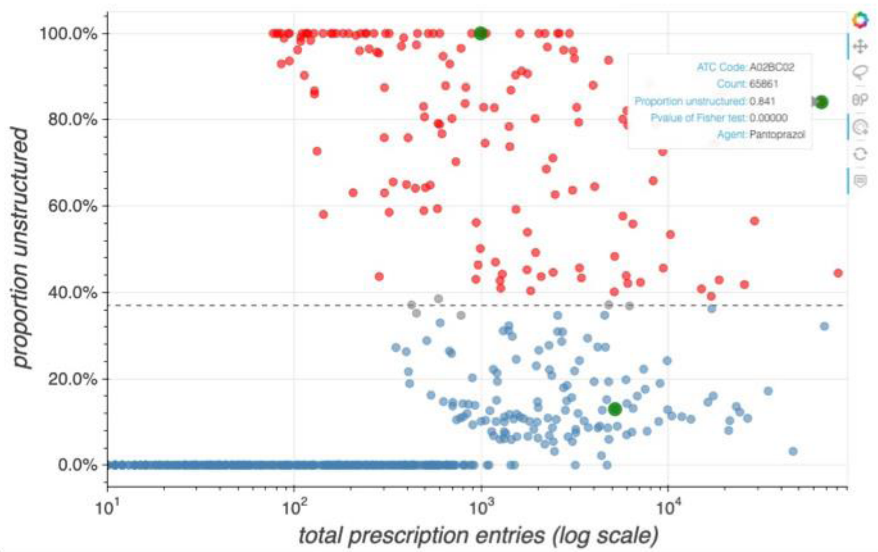
A Fisher test for each ratio of unstructured to structured entries per ATC code against the unstructured-to-structured ratio of all other entries not associated with this code was performed. A P-value < 0.05 was considered as significant deviation and the respective dot for the ATC-code was colored red, if the unstructured to structured ratio was significantly higher as the ratio for all other entries, and blue, if significantly lower. Calculations were performed with Python 3.9.10, for Fisher test SciPy 1.8.0 was used. To improve the understanding of data, an interactive scatterplot was created by using Bokeh (2.4.2).[6] The possibility to highlight certain levels of ATC codes allowing to compare related agents was given. Further, a free text search for highlighting (parts of) ingredient names was created. Consequently, a table listing all main information of highlighted ATC codes was implemented for comparison. An outlier code with high prescription rates was manually selected and explored interdisciplinary by pharmacists, clinical and medical informatics professionals.

Based on this exploration, a generalized feedback loop that can be applied to any other drug (based on ATC code) was set-up to foster data quality improvement of drug prescriptions. Main goals of this feedback loop were defined as:

1. Potential positive impact on data quality for utilizing drug data for research
2. Potential positive impact on drug prescription safety
3. Continuously "learning" of prescription system and it's actors

The developed framework was visualized for building up an evaluation concept for this feedback loop approach.

3. Results



Characteristics of highlighted ATC codes

#	ATC code	agent	N total	Proportion unstructured	Pvalue Fisher Test
0	A02BC02	Pantoprazol	65861	0.8407251636021318	0
1	A02BC05	Esomeprazol	5180	0.12953667953667955	0
2	A02BC01	Omeprazol	991	1	0

Figure 1. Visualization of ATC code distribution by proportion of unstructured entries

Figure 1 shows the visualization of ATC codes with a distribution based on the total amount of prescriptions and the ratio of unstructured to structured entries in the HIS. We initiated the step of expert evaluation for one outlier agent: "pantoprazole" (German ingredient names) with a total count of 65,861 entries. As ATC codes starting with "A02BC" represent drugs with proton pump inhibitors as ingredients we filtered these codes resulting in shown distribution (green dots). Interestingly, although pantoprazole already had a proportion of 84% unstructured entries, omeprazole had only free text prescriptions whereas esomeprazole prescriptions were mostly structured. Although the selection list of the hospital pharmacy included oral pantoprazole drug products, no structured data was available for omeprazole. First direct insight by data therefore was a miss-mapping of original structured data to ATC code which the algorithm was apparently able to restore. Further, Omeprazole was missing in the structured list provided by the hospital. From a clinical perspective, with the high amount of entries proton pump inhibitors are common drugs, considering the possibility of misinterpreting a free text entry “Omeprazole” with e.g. “Esomeprazole” there is a need to reduce the unstructured entries. As a next step, generated “actions” would be to fix the miss-mapping of pantoprazole as well as e.g. adding “omeprazole” to the structured list provided by the hospital. If major changes in architecture of prescription would be necessary, another action item might be to teach the clinical staff in correct structured prescription. Ultimately, the “learnings” of both, the actors, as well as the prescription module of the HIS might then lead to a lower proportion of unstructured entries for pantoprazole and omeprazole in a follow-up analysis, and close the feedback loop.

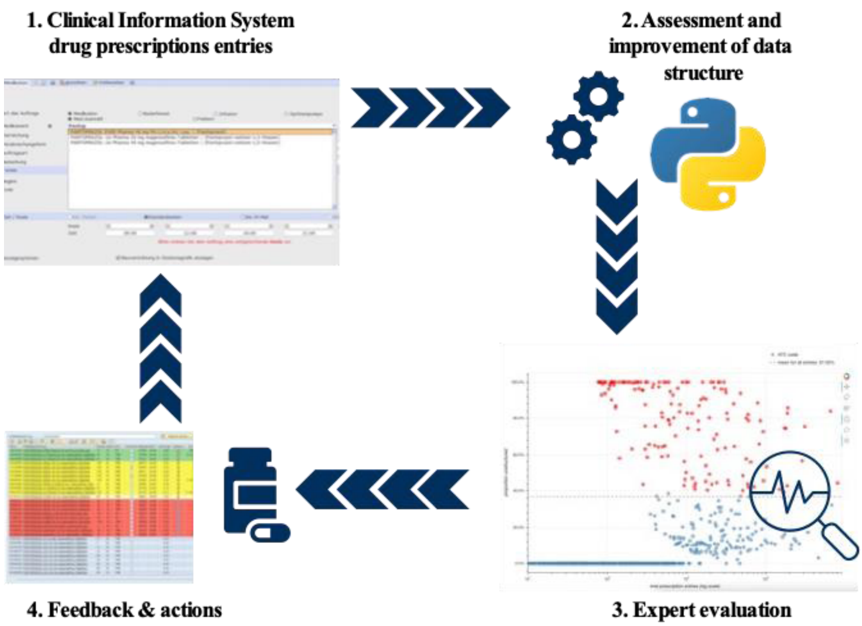


Figure 2. General concept of clinical feedback loop for the data quality of drug prescriptions

A more generalized form of this feedback framework was created and is displayed in Figure 2. When a clinician creates the prescription in the HIS, the structured list provided by the pharmacy as well as the prescription mask is used. The ability to use free text or edit the drug prescription is provided. Step 2 includes the pharmacy-intern processes and mapping to ATC codes as well as the algorithms analyzing the free text, and evaluation algorithms as the one presented by Reinecke et al.,[5] as well as evaluation algorithms as presented in this publication. In the crucial step 3, the interdisciplinary experts analyze the provided visualizations based on their domain-specific expertise to understand the current data. Subsequently, feedback is given to the stake holders of the prescriptions process and action items are generated leading to changes in software, data structure and/or teaching of the actors. Implementing these changes lead to step 1 and close the feedback loop, now with a system that have learned and allow to re-analyze the effects of these changes in next iterations.

4. Discussion and Conclusion

In this paper we developed a concept for a feedback loop for a structured drug prescription process in a problem-driven interdisciplinary approach to generate a potential positive impact on data quality. Although the idea of subsequently, computerized drug prescription data improvements is promising for retrospective data [5,7] it does not address the data structure improvement during documentation process in the source system and does not investigate root causes of data structure issues for affected ATC codes. The feedback loop starts exactly at this point in order to improve the data quality directly in the HIS at the time of creation. Thus, it might avoid costly downstream improvements leading to a better data quality (goal 1). Considering the protective effect of structured data as well as the potential ability to find weaknesses in the daily routine of the users of the prescription it also can help to increase medication safety, fulfilling goal 2. Furthermore, the feedback loop leads to a continuously learning system not only on the user's but on the application side, as well, covering goal 3.

Two main strengths characterize our concept: first, its interdisciplinary link between different stake holders. And second, the system learning approach is about both training users of the prescribing process and software evolution through decisions derived directly from RWD in a continuous feedback loop.

The limitation is the lack of knowledge of potential drug prescriptions without correct ATC code mappings, which would require a different approach. Though our concept focusses on drug prescription data, it can be extended to other RWD. In a next step we will assess the effectiveness of the presented feedback loop through an evaluation study and determine whether establishing the continuous feedback process will lead to data quality improvement in the HIS.

In conclusion, our pharmaceutical feedback loop provides a concept to continuously improve data quality based on RWD by a learning approach on both dimension – human factor and the underlying machine.

References

[1] Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation—challenges—success factors. Onkologe. 2018;24:91–98.
[2] Zhang J, Symons J, Agapow P, et al. Best practices in the real-world data life cycle. McGinnis RS, editor. PLOS Digit Health. 2022;1:e0000003.
[3] Cave A, Kurz X, Arlett P. Real-World Data for Regulatory Decision Making: Challenges and Possible Solutions for Europe. Clin Pharmacol Ther. 2019;106:36–39.
[4] Semler S, Wissing F, Heyder R. German Medical Informatics Initiative: A National Approach to Integrating Health Data from Patient Care and Medical Research. Methods Inf Med. 2018;57:e50–e56.
[5] Reinecke I, Siebel J, Fuhrmann S, et al. Assessment and improvement of the drug data structuredness from electronic health records to enable secondary usage and ensure semantic interoperability (Preprint) [Internet]. JMIR Medical Informatics; 2022 [cited 2022 Jun 20]. Available from: http://preprints.jmir.org/preprint/40312.
[6] Bokeh Development Team. Bokeh: Python library for interactive visualization [Internet]. 2018. Available from: http://www.bokeh.pydata.org. [Accessed June 2022]
[7] Kapsner LA, Kampf MO, Seuchter SA, et al. Moving Towards an EHR Data Quality Framework: The MIRACUM Approach. Stud Health Technol Inform. 2019;267:247–253.