Journal of Biomedical Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

The need for harmonized structured documentation and chances of secondary use – Results of a systematic analysis with automated form comparison for prostate and breast cancer

Rainer Krumm [a,*], Axel Semjonow [b], Joke Tio [c], Heike Duhme [d], Thomas Bürkle [e], Jörg Haier [d], Martin Dugas [a,1], Bernhard Breil [a,1]

[a] Institute of Medical Informatics, University of Münster, Albert-Schweitzer-Campus 1, A11, D-48149 Münster, Germany
[b] University Hospital Münster, Department of Urology – Prostate Center, Albert-Schweitzer-Campus 1, A1, D-48149 Münster, Germany
[c] University Hospital Münster, Department of Obstetrics and Gynecology, Albert-Schweitzer-Campus 1, A1, D-48149 Münster, Germany
[d] University Hospital Münster, Comprehensive Cancer Center Münster, Waldeyerstr. 1, D-48149 Münster, Germany
[e] University Erlangen-Nürnberg, Chair of Medical Informatics, Krankenhausstraße 12, D-91054 Erlangen, Germany

ARTICLE INFO

Article history:
Received 16 August 2013
Accepted 7 April 2014
Available online xxxx

Keywords:
Semantic interoperability
Automated form comparison
Common data elements
Secondary use
Standardization
Clinical documentation

ABSTRACT

Introduction: Medical documentation is a time-consuming task and there is a growing number of documentation requirements. In order to improve documentation, harmonization and standardization based on existing forms and medical concepts are needed. Systematic analysis of forms can contribute to standardization building upon new methods for automated comparison of forms. Objectives of this research are quantification and comparison of data elements for breast and prostate cancer to discover similarities, differences and reuse potential between documentation sets. In addition, common data elements for each entity should be identified by automated comparison of forms.

Materials and methods: A collection of 57 forms regarding prostate and breast cancer from quality management, registries, clinical documentation of two university hospitals (Erlangen, Münster), research datasets, certification requirements and trial documentation were transformed into the Operational Data Model (ODM). These ODM-files were semantically enriched with concept codes and analyzed with the compareODM algorithm. Comparison results were aggregated and lists of common concepts were generated. Grid images, dendrograms and spider charts were used for illustration.

Results: Overall, 1008 data elements for prostate cancer and 1232 data elements for breast cancer were analyzed. Average routine documentation consists of 390 data elements per disease entity and site. Comparisons of forms identified up to 20 comparable data elements in cancer conference forms from both hospitals. Urology forms contain up to 53 comparable data elements with quality management and up to 21 with registry forms. Urology documentation of both hospitals contains up to 34 comparable items with international common data elements. Clinical documentation sets share up to 24 comparable data elements with trial documentation. Within clinical documentation administrative items are most common comparable items. Selected common medical concepts are contained in up to 16 forms.

Discussion: The amount of documentation for cancer patients is enormous. There is an urgent need for standardized structured single source documentation. Semantic annotation is time-consuming, but enables automated comparison between different form types, hospital sites and even languages. This approach can help to identify common data elements in medical documentation. Standardization of forms and building up forms on the basis of coding systems is desirable. Several comparable data elements within the analyzed forms demonstrate the harmonization potential, which would enable better data reuse.

Conclusion: Identifying common data elements in medical forms from different settings with systematic and automated form comparison is feasible.

* Corresponding author.
E-mail addresses: Rainer.Krumm@uni-muenster.de (R. Krumm), Axel.Semjonow@ukmuenster.de (A. Semjonow), Joke.Tio@ukmuenster.de (J. Tio), Heike.Duhme@ukmuenster.de (H. Duhme), Thomas.Buerkle@imi.med.uni-erlangen.de (T. Bürkle), Joerg.Haier@ukmuenster.de (J. Haier), Dugas@uni-muenster.de (M. Dugas), Bernhard.Breil@uni-muenster.de (B. Breil).
[1] Authors contributed equally.

http://dx.doi.org/10.1016/j.jbi.2014.04.008
1532-0464/

1. Introduction

Medical documentation is a required, but time-consuming task and therefore in the focus of health informatics research. The official journal of the German Medical Association titled 2009 one article "Documentation in medicine – it is a madness" [1] and addressed the problem of documentation reality. There is high documentation workload and a huge variety of forms for different purposes with redundant data capture.

The documentation process is very complex and sophisticated especially in oncology. Nevertheless, there is a lack of standardization which results in a huge variety of different forms in hospitals even for the same disease areas. This work is focused on the analysis of documentation items[2] of widespread cancer diseases. Prostate cancer is the most common cancer (other than skin cancer) among American men and breast cancer the second leading cause of cancer-related death in women [2,3]. Both cancer documentation sets consist of many forms from different sources and contain also redundant items. It was reported previously that e.g. for breast cancer up to 65% of data elements are gathered redundantly [4]. An optimization of this documentation process is desirable because documentation should burden the physician as few as possible [5].

This becomes even more relevant as physicians spend more than 25% of their daily working time for documentation tasks and hereby almost as much time as for direct patient care [6,7]. These numbers only consider routine documentation within a hospital. In addition, parallel and redundant documentation for quality management, registries and clinical research increase the physicians' workload.

The documentation for these different purposes is in large part done within different systems and structures. "The existence of parallel independent documentation systems leads to a tremendous workload, and thus hinders acceptance among physicians" [8]. Patient care forms usually consist of many free-text elements, whereas research and registry-forms are highly structured.

Studies show that there is a re-use potential for these purposes if the original information is documented in a structured way [9]. Prokosch reviewed the potential of reusing the electronic health record (EHR) [10] and Kush introduced documentation according to the single-source concept to reduce the physician's workload [11]. Further projects in this context demonstrate the high potential of this research field [12].

Without secondary use and structured documentation a lot of resources are needed to fulfill the documentation requirements for research and registries. This documentation is up until today usually performed by manual review of clinical free-text. This is an inefficient and error-prone process because the same information needs to be documented two or three times (e.g. writing of physician's letter and extracting information from it to complete quality management forms). A similar aspect can be observed in the context of clinical studies, as Getz reported that the "protocol designs are becoming more demanding and burdensome on investigative site personnel". Between 1999 and 2005 the workload for research increased by 10.5% per year [13].

At the moment many studies are dealing with optimization of electronic medical records (EMR) or electronic health record (EHR) systems but there is still a lack of documentation standardization which leads to heterogeneity in data representation [14].

Future work has to focus not only on the improvement of EHR systems but on the standardization of medical forms and their data items and to provide documentation standards that meet the different purposes by a single source documentation system. This need is emphasized by Ries et al., who analyzed oncological documentation in Germany and concluded that "none of the existing German cancer datasets (e.g. ADT or GEKID) meets clinical documentation reality" [8].

Therefore, a systematic analysis is needed to identify common concepts and elements based upon documentation reality. To identify redundant documentation items it is necessary to compare and harmonize medical data items. Semantic enriched data items can enable automated comparison. Systematized Nomenclature of Medicine (SNOMED) [15] or Unified Medical Language Systems (UMLS) [16] are very important in this context [17].

This work applies and extends **compareODM**, a recently published method to compare coded forms [18]. In contrast to the previous paper, this research goes beyond comparison of two forms and addresses a systematic analysis of the documentation landscape from two major disease entities at two university hospital sites including routine documentation, quality management, certification, research and cancer registration on the basis of real forms. Therefore, new methods to compare and visualize groups of forms were applied, for example spider charts. It was required to extend the **compareODM** method to conduct these analyses: the comparison of code lists (value lists) was integrated and the output of **compareODM** was enhanced for statistical analysis. With the compareODM Paper [18] the feasibility to compare two given ODM-forms was demonstrated. In the current research common data elements in tumor documentation for two chosen diseases were analyzed and identified.

To our knowledge, an analysis of such amount of forms from clinical documentation is not available in the literature. Results of this paper are precise lists of data elements and can contribute to improved information systems.

The scope of this work is an assessment of standardization opportunities based on a systematic and automated comparison of oncological forms in the context of prostate and breast cancer. It is well known that standardization can lead to better interoperability [19].

2. Objectives

The overall objective of this work is to analyze the current documentation landscape for two cancer entities, breast and prostate cancer. Specifically, we want to address the following research questions:

1. Is it possible to quantify the amount of forms and data elements?
2. Is there a chance of reusing elements for secondary use purposes?
3. Which similarities exist between documentation for quality management, registries, research and clinical routine?
4. What are the differences between clinical routine documentation in two German university hospitals?
5. What are the top 30 common data elements for each disease entity?

3. Methods and materials

3.1. Form analysis process

3.1.1. Form collection and coding of data elements

The documentation landscape of two common cancer entities (breast and prostate cancer) was analyzed to determine currently used data elements. For both diseases there exist structured forms.

Medical forms for these cancers entities were collected from two university hospitals (Erlangen and Münster), e.g. forms for medical history, forms regarding surgery and different types of

[2] Items and data elements are treated as synonyms.

Please cite this article in press as: Krumm R et al. The need for harmonized structured documentation and chances of secondary use – Results of a systematic analysis with automated form comparison for prostate and breast cancer. J Biomed Inform (2014), http://dx.doi.org/10.1016/j.jbi.2014.04.008