Author’s response to reviews

Title: Replicating Medication Trend Studies Using Ad Hoc Information Extraction in a Clinical Data Warehouse

Authors:
Georg Dietrich (dietrich@informatik.uni-wuerzburg.de; georg.diet@gmail.com)
Jonathan Krebs (jonathan.krebs@uni-wuerzburg.de)
Leon Liman (leon.liman@uni-wuerzburg.de)
Georg Fette (georg.fette@uni-wuerzburg.de)
Maximilian Ertl (Ertl_M@ukw.de)
Mathias Kaspar (Kaspar_M@ukw.de)
Stefan Störk (Stoerk_S@ukw.de)
Frank Puppe (frank.puppe@uni-wuerzburg.de)

Version: 1 Date: 09 Oct 2018

Author’s response to reviews:

Editor Comments:
> This is a very well written paper. The approach and structure is useful.

We are very pleased that the editor considers our work that way.

Andreas Holzinger, PhD, MSc, MPh, BEng, CEng (Reviewer 1):
> The paper is well written, tackles an important problem and is a useful contribution to the field.

> The reviewer would also like to emphasize that the authors could even replicate one study completely with all main findings as well as the sub findings.

We thank the reviewer for his encouragement.
Generally, the paper is useful and is a novel contribution, hence this reviewer would argue accepting this paper.

The authors should do some minor spell checks and language polishing.

Thank you for that remark. We corrected the spelling and reformulated some sentences to improve the language.

Mathias Brochhausen (Reviewer 2):

> I think this is a very interesting paper of high relevance to the journal.

We thank the reviewer for his valuable remarks, which have led to a substantial revision of our paper. In the following we outline our changes in response to his comments.

> The reviewer assumes that the topic of the paper is the medication that the patient received during a hospital stay, not prescriptions that the patient received. If that is incorrect, this should be made more clear.

2.1) Thank you for this information. We use the medication at discharge, which represents the prescribed medication. In a few cases, the medication at admission is also described in the text. That is also included, although it has no relevant impact to the evaluations. We add the following explanation to section 3.5.2 “Study replication”:

The drugs were extracted from the medication section of the discharge letter. That contains in almost every case the medication at discharge representing the recommended / prescribed medication. Additionally the medication at admission is described in 18% (Med1: 13%) of all cases. At discharge from hospital, patients receive 8% (Med1: 19%) more medication than at admission, while nearly all medications from admission were continued at discharge. (Tested for the main drug agent groups for hypertension.) We used the whole medication section with all medication descriptions as data source to identify whether a drug is taken or not.

> p.2: The following sentence seems a little imprecise from the perspective of knowledge representation: "Ad hoc IE is described as extracting the existence of any concepts (e.g. chronic kidney disease) or any numbers (e.g. the LVEF value) from textual sources in real-time." Is only existence extracted (so, yes no to the mentioning of that concept in the free text) or is actual text extracted? The reviewer assumes that the numbers are actually extracted, not only assessed of
whether they are present. Also, the phrasing is a great example of the deviousness of the term "concept". The reviewer assumes that if anything is extracted it is the name or the label of a concept, but it would be neat to clarify that.

2.2) Yes, the assumption of the reviewer is correct. The real value is extracted.

We described these statements in more detail and added explanations to make that more clear. The text in section 1 was changed as follows:

A novel way to retrieve information from plain text is ad hoc IE.

Ad hoc IE is described as extracting the existence of any concepts (e.g. chronic kidney disease) or any numbers (e.g. the LVEF value) from textual sources in real-time.

The Boolean ad hoc IE queries the existence (yes/no) of a medical concept. A medical concept is a named entity that may have a feature/property or a numeric value. Examples of Boolean concepts are single findings or assessments (e.g. moderate mitral insufficiency, severe aortic stenosis), drugs (e.g. Aspirin, beta blocker) or diagnoses (e.g. appendicitis, myocardial infarction).

Numeric IE extracts the value as number of a numerical concept. That could be for example the value of a laboratory finding (e.g. cholesterol, glucose, LVEF) or a derived values/indexes (e.g. BMI, age). A numerical condition can be defined optionally, like LVEF < 45, matching all mentions of LVEF with a value lower than 45. In some finding reports, the exact value of a concept is not given but there is a formulation indicating an interval or an inequality of a value (e.g. "LVEF lower than 45"). These statements can be queried in conjunction with numeric ad hoc IE exploiting both qualitative and quantitative information from textual reports e.g. for checking inclusion or exclusion criteria of studies.

In addition to count queries, which only asses the presence of a concept or the validity of constraints (e.g. BMI > 25), the actual values can also be returned for further processing.

This technique showed good results and requires little developmental effort, since the text is indexed efficiently and can be queried with powerful features [18].

> p.3: the following phrasing "Not every piece of information belongs to the patient" doesn't ring right in English. I assume that the intended meaning is that not all information is about the patient? (but the negated statements are about the patient...). I think the authors need to differentiated between statements about other persons (medical family history, if applicable) and negated statements. The statement "John Doe does not suffer from lung cancer" is about the
patient, but doesn't ascribe the property of having lung cancer to the patient, it specifically states that the patient does not bear that property.

2.3) That is a good point. We simplified that explanation in section 3.2.2:

The context of information in a discharge letter is an important topic. Many pieces of information are negated [24] (e.g. “no fever”, “dizziness is denied”) or they relate to other persons (e.g. within the context of family history). Some information like medications within in the discharge letter have a temporal context and may not be valid any longer (e.g. medication might have been stopped at hospital entry or during hospitalization, like Ramipril in Figure 8).

Depending on the application or evaluation, different types of information are relevant or must be excluded. In most cases, physicians are interested in the confirmed and current findings of a patient.

> It seems the paper uses "University Hospital of Würzburg" and "University Clinic of Würzburg". If those two expression are referring to the same entity, I'd suggest sticking to one label (I would suggest "Hospital" over "Clinic". In some healthcare systems clinics are places that provide ambulatory care.)

2.4) Thanks again for this note. We use the official name (University Hospital of Würzburg) and have modified the text accordingly.