Using a text-mining approach to identify the context variables language barrier, living alone, cognitive frailty and non-adherence from electronic health records (EHRs)

Simone ten Hoope[1], Koen Welvaars[2], Mellanie Klok-Everaars[1], Sander van Schaik[3] and Fatma Karapinar-Çarkit[1]

1 Department of Clinical Pharmacy, OLVG Hospital, Amsterdam (Contact: f.karapinar@olvg.nl)

2 Department of Data science, OLVG Hospital, Amsterdam

3 Department of Neurology, OLVG Hospital, Amsterdam

Background

- Context variables such as language barrier, living alone, cognitive frailty and non-adherence have been shown to be related to medication-related hospital readmissions.
- These context variables are often documented in Electronic Health Records (EHRs) as "free text" in clinical notes.
- These clinical notes are time consuming to read, leaving health practitioners unaware of these context variables.

Aim

To analyze the accuracy of two text mining techniques for the identification of the context variables language barrier, living alone, cognitive frailty and non-adherence from EHRs.
The accuracy will be determined using the percentage agreement with the manual standard (primary outcome). Also the kappa, sensitivity, specificity, negative and positive predictive value will be determined.

Methods

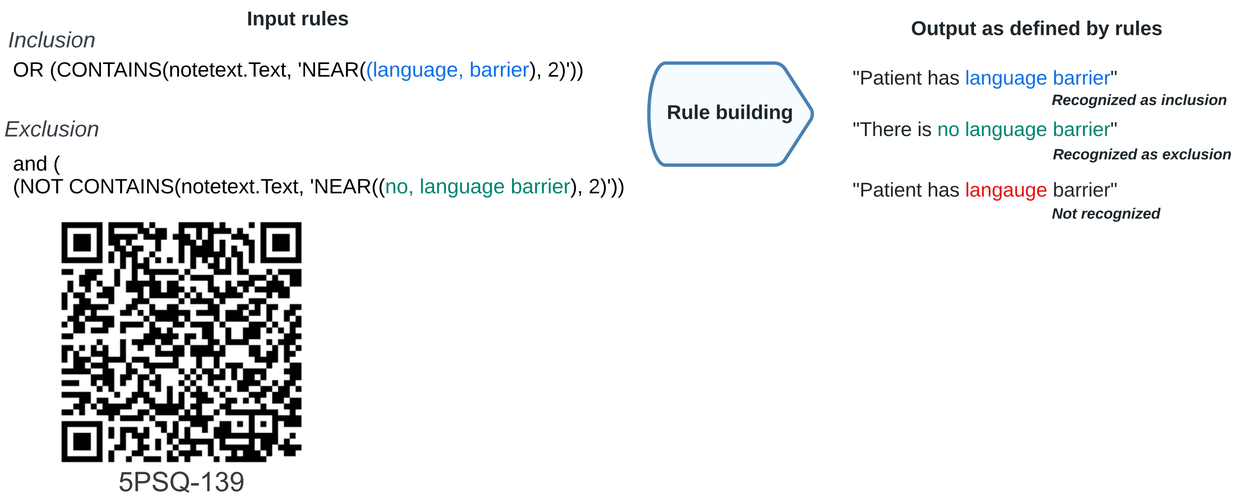
A manual standard was created from a database of 1,120 readmissions (878 patients). The EHR of each patient was manually searched to categorize a context variable as present or non-present (in duplo) and the exact free text was extracted.

Two text mining techniques (rule based algorithm (figure 1) and Named Entity Recognition (NER, figure 2)) were used to create algorithms to identify the context variables from EHRs. The algorithms were trained with the free text extracted from the EHR until a high percentage agreement with the manual standard was achieved, or for a maximum of five runs.

- For the straightforward term language barrier (with not many variations in free text descriptions) a rule based algorithm was used. In- and exclusion criteria were defined (figure 1).
- For the more complex terms living alone, cognitive frailty and non-adherence NER models were used. The models were trained based on annotated input data (figure 2).

Descriptive analyses was used (agreement, kappa, sensitivity, specificity, negative and positive predictive value).

Figure 1: Rule based algorithm, identifies text exactly as defined by the creator



5PSQ-139

Results

Context variable	Agreement (%)	Kappa	Sensitivity (%)	Specificity (%)	PPV*	NPV*
Rule based algorithm						
Language barrier (n=221)	96.8	0.90	98.6	96.3	86.9	99.7
NER algorithms						
Living alone (n=328)	91.0	0.79	85.7	93.7	84.9	94.0
Cognitive frailty (n=185)	95.1	0.83	90.8	96.7	82.0	98.0
Non-adherence (n=47)	98.0	0.63	75.0	98.7	56.3	99.5

NPV/PPV= negative and positive predictive value

- All algorithms have high agreement compared to the manual standard. The kappa for non-adherence was low.

Discussion

- The manual standard could possibly contain misclassifications when context variables are not documented or updated in the EHRs.
- The kappa was low for non-adherence, probably due to the low numbers of patients for this context factor and because non-adherence was documented sometimes only on the medication level (e.g. patient does not use metoprolol) without mentioning the broader terms for non-adherence.

Future implications

- Implement these algorithms into hospital software (EPIC).
- Be able to identify context variables at any first encounter (e.g. clinic visit, hospital admission).
- Allow for personalized care to ensure context variables are incorporated in patient counselling

Conclusion

The text-mining algorithms can identify the context variables language barrier, living alone, cognitive frailty and non-adherence.

Figure 2: Named Entity Recognition (NER) model, learns to analyze text after training