OMOP CDM Database

Contact details

Application of OMOP Common Data Model and Language Technologies on Patient Records for Better Treatment of Bulgarian Diabetic Patients

Database features:

Contains **pseudoanonymized records of 503 753 patients with diabetes** created in ambulatory and specialized care units during 2018 - 2019.

Big Data Analytics is applied after the semi structured Outpatient records are transformed to a data base repository. This enables exploration of the tendency in process developments and personal control of the health status of particular patients e.g. HbA1c levels and BMI for patients suffering from Diabetes Mellitus.

Partners

- **Medical University _Sofia**
- **Sofia University "St. Kliment Ohridski"**
- **Technical University Sofia**

Research team:

    Prof. Dimitar Tcharaktchiev, MD, PhD

    Medical University – Sofia

    dimitardt@gmail.com

    Prof. Evgeniy Krastev, PhD

    Sofia University

    eck@fmi.uni-sofia.bg

    Prof. Ivan Ivanov, PhD

    Technical University Sofia

    iei@tu-sofia.bg

Automatic Construction of Diabetic Register

Source Data

Automatic Text Analysis of Bulgarian Clinical Narratives

A pseudonymised Register of diabetic patients was generated in 2015 from the Outpatient records, collected by the Bulgarian National Health Insurance Fund (NHIF), in compliance with all legal requirements for safety and data protection. Usual patient registration process was kept without burdening the medical experts with additional work.

Many indicators of the Register copy structured data submitted to NHIF but others contain values that are extracted automatically from the free texts in the Outpatient records. Integration of text analysis in large scale is a real novelty of the approach. Now various findings can be delivered to decision makers in order to improve the healthcare management.

The data repository currently contains more than **380 million** pseudonymised reimbursement requests (***Outpatient records***) submitted to the National Health Insurance Fund (NHIF) in **2011-2019** for more than **7 million** citizens (more than **5 million** citizens yearly), including **503 753 diabetic patients**.

Outpatient records in Bulgaria are produced by the General Practitioners and the Specialists from Ambulatory Care for every contact with the patient. Despite their primary accounting purpose these records summarize sufficiently the case and motivate the requested reimbursement. They are semi-structured files with predefined XML-format. The following structured XML fields provide data for the Diabetic register:

- Date and time of the visit,
- Pseudonymised personal data, age, gender,
- Pseudonymised visit-related information,
- Diagnoses in ICD-10,
- NHIF drug codes for medications that are reimbursed,
- A code if the patient needs special monitoring,
- A code concerning the need for hospitalisation,
- Several codes for planned consultations, lab tests and medical imaging.

Outpatient records' fields with free text were used too:

XML field name	Content
Anamnesis	Case history, previous treatments. Often *Family history, Risk Factors*
Status	Summary of patient state, height, weight, BMI, blood pressure … etc.

Clinical tests	Values of clinical examinations and lab data listed in arbitrary order
Prescribed treatment	Codes of drugs reimbursed by NHIF, free text descriptions of other drugs

Text mining finds entities of interest in free texts by focusing on important words and phrases that trigger shallow analysis in the local context. Several text analysis tools are implemented, called extractors.

```
……. Поддържа кр. налягане 130-140/85-90…….
… Sustainable bl. pressure 130-140/85-90.…
…… Сърце-РСД; СЧ-72/мин; 125/90 ……
……  Heart – NHR; FR-72/min; 125/90…….
……  тегло: 87 кг.; weight: 87 kg …….
……  Боди Мас Индекс 28 ……. ВМI 28. …….
```

They recognize with high accuracy numeric values of specific clinical examinations and lab data including the values of glycated hemoglobin (HbA1c), body mass index (BMI), weight, blood glucose and blood pressure. The aim is to extract numeric values that are central for the monitoring of most frequent chronic diseases. A drug extractor finds drug names, daily dosages, frequency and route of admission in the texts.

```
……. ТГ-1.67; HbA1c-5.9%/добър к-л/ …….
……. гликиран нв-6,4-21.07.2015 …….
…… 07.01.2015. изследв глик хемогл6.9  … …….
ХбА1С от април 2015г-9,47% …….
…….01.22 Гликиран хемоглобин-8,10 …….
```