AUTOMATIC GENERATION OF A NATIONAL DIABETES REGISTER FROM OUTPATIENT RECORDS

Dimitar Tcharaktchiev[1], Zhivko Angelov[2], Svetla Boytcheva[3], Galia Angelova[3]

Medical University Sofia and University Specialized Hospital for Active Treatment of Endocrinology "Acad.I.Penchev" Sofia, Bulgaria[1]

ADISS Lab Ltd., Sofia, Bulgaria [2]

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria [3]

galia@lml.bas.bg

Abstract: *In this paper, we present the construction of Bulgarian National Diabetes Register, using pseudonymized outpatient records submitted to the Bulgarian National Health Insurance Fund. The automatic generation facilitates the construction because it does not burden any medical experts with additional paper work. The Register is a healthcare system integrating natural language processing in large scale and analytics functionalities that provide new views to the information concerning Diabetes Mellitus and diabetic patients in Bulgaria. This successful approach encouraged the authors to initiate a research programme in eHealth focused on collection and analysis of patient data, with the intention to assess the feasibility of secondary patient record use in evaluation of healthcare quality.*

Keywords: HEALTHCARE SYSTEMS, BIG DATA IN HEALTHCARE, AUTOMATIC ANALYSIS OF CLINICAL TEXTS

1. Introduction

Diabetes prevention and cure in Europe have improved after 2008 because less people die [1]. Patient awareness is raising, self-monitoring becomes easier, and the variety of medications is growing. However, still a very high number of diabetic patients are undiagnosed and half of the European countries cannot provide reasonably good data concerning procedure indicators. It is claimed that "as long as important data is not systematically reported and transformed into methodology, diabetes care will remain inefficient and, at worst, haphazard" [1].

The Euro Diabetes Index 2014 lists seven European countries that support diabetic registers: Sweden, Denmark, Norway, Netherlands, UK, Switzerland, and Hungary [1]. Data input to the registers is ensured either by self-registration or by burdening medical professionals with additional documentation tasks. Practically, self-registration means that a significant percent of the patients remains unregistered. Even in Sweden, which is the country with the best diabetes care delivery in Europe according to the Euro Diabetes Index 2014, the register was constructed by self-registration. During its development phase 2001-2005 the self-registration rate of patients gradually increased and reached 75%, which in 2010 still remains stable and is one of the highest in the country [2]. No information is available about the procedures for register update and maintenance.

Availability of relevant data is of primary importance in diabetes prevention and treatment ("*no data, no cure*" according to the Euro Diabetes Index 2014). However, high-quality data is hard to collect. Information about diabetic patients is often not collected nationally but rather in hospitals or at regional level, with limited comparability of collected indicators. Moreover, data often come from isolated national projects or EU-funded initiatives with fixed duration. After the project ends, no strategic plans are built by the respective political or governing institutions and in this way projects that started and proved to be successful remain feasibility studies without practical effects.

All countries in Europe have national plans for discovery, treatment and prevention of Diabetes [3] but one hardly finds information about the execution of these plans, monitoring of various plan measures and evaluation of their success. Positive health outcomes are difficult to assess too, moreover this needs to be done dynamically at national level in order to improve the treatment plans. From a technological point of view, the general impression is that healthcare authorities lack understanding about the potential of modern Information and Communication Technologies (ICT) as an enabling tool that facilitates data collection, monitoring of indicators, knowledge discovery, early alerting and automatic sending of feedbacks, evaluation of updated indicators and automatic preparation of aggregated recaps.

Surprisingly, no attempts for automatic extraction of Registers from available Electronic Health Records (EHR) repositories are mentioned in the Euro Diabetes Index 2014. A recent book about secondary use of EHR [4] lists three types of users that utilize information from patient records: clinicians searching data for their daily work; clinical researchers who need to extract patient groups or cohorts, or patients with specific diseases for their research; and finally the hospital management that needs to gather statistics and predict the future of the hospital activities. The book [4] states: "Generally when asking users what type of systems they want, they do not know". The suggested approach is to develop a prototype and show it to the users who give feedback. Thus, it somehow becomes clear that, apart from archiving purposes, the application of nation-wide EHR repositories is still limited and their potential as content repositories is not fully understood and exploited by the community of medical professionals.

In this paper, we sketch our approach to generate automatically an anonymous Diabetic Register from outpatient records, submitted to the only health insurance fund in Bulgaria – the National Health Insurance Fund. The construction took place in 2015 and later the Register has been updated. We show how the Register is used today, together with the underlying repository of pseudonymized outpatient records. Finally, we present ideas for future work in secondary use of patient records, to be performed within the National Research Programme 'eHealth' funded by the Bulgarian Ministry of Education and Science in 2018-2021.

2. Generation of the Diabetes Register

The mandatory health insurance was introduced in Bulgaria in 1998. The National Health Insurance Fund (NHIF) was founded in 1999 with the mission to deal with the obligatory health insurance in the country. All General Practitioners and Specialists from Ambulatory Care produce reimbursement requests (Outpatient Records) whenever they contact patients and submit these requests to the NHIF. The Outpatient Records are semi-structured XML files with numerous fields containing structured and coded information about the patient and the examining medical expert, and sufficient clinical data to summarize the case. Many indicators in the Diabetic Register copy the structured data submitted to NHIF: *(i)* date and time of the visit; *(ii)* pseudonymized personal data, age, gender; *(iii)* pseudonymized visit-related information; *(iv)* diagnoses in ICD-10; *(v)* NHIF drug codes for medications that are reimbursed; *(vi)* a code if the patient needs special monitoring; *(vii)* a code concerning the need for hospitalization; *(viii)* several codes for planned consultations, lab tests and medical imaging.

The Outpatient Records contain also values of clinical tests and lab data, presented in the free text fields. Using software extractors for automatic text analysis of Bulgarian texts, which have been developed in our previous projects, we mine these values from four

free text fields: *(i) Anamnesis*: summarizes case history, previous treatments, often family history, risk factors; *(ii) Status*: summary of patient state, height, weight, body-mass index BMI, blood pressure etc.; *(iii) Clinical tests*: values of clinical examinations and lab data; *(iv) Prescribed treatment*: free text descriptions of drugs that are not reimbursed by NHIF. The focus was placed mostly on extraction of numeric values that are important for diabetes like values of blood sugar and measurement time, glycated hemoglobin, BMI, blood pressure etc. More details about the construction of the Register are given in [5, 6]. Knowledge discovery using original data mining techniques is presented in [7].

NHIF delivered pseudonymized outpatient records to the University Specialized Hospital for Active Treatment of Endocrinology (USHATE), Medical University Sofia, which was authorized by the Bulgarian Ministry of Health to host the information. The Register was constructed iteratively: at first using data for 2010-2014 and then updated twice with outpatient records for 2015 and 2016. The repository behind the Register currently contains more than 262 million pseudonymized outpatient records submitted to the NHIF in 2010-2016 for more than 7.3 million citizens (more than 5 million citizens yearly), including 483,836 diabetic patients. This repository contains also information about patients, who are not formally diagnosed with Diabetes Mellitus. This allows to study particular groups of citizens in pre-diabetes condition.

The internal Register interface in USHATE provides functionalities to authorized medical experts to track disease history of individual patients (Fig. 1). The public web interface[1] generates on the fly aggregated statistics about key factors related to the Diabetes Mellitus and provides monitoring and assessment at regional level (Fig. 2).

Fig. 1 *Diabetes Register interface in USHATE – a general view*

Fig. 2 *Diabetes Register Web interface – public generalized and aggregated data enhanced with OLAP technology*

---

[1] https://usbale.org/bg/registar-zaharen-diabet/

3. Present Use of the Register

The main objectives of the National Diabetes Register are [5]:
- Assessing Diabetes Mellitus morbidity in Bulgaria;
- Providing views to patient care in long-term;
- Improving prevention by Diabetes risk stratification;
- Improving Diabetes patient care by comprehensive analysis of comorbidities, studies of Diabetes Mellitus complications and treatment side effect.

Here we show how the objectives listed above are met in various studies performed on patient groups.

Diabetes Mellitus morbidity assessment (Fig. 3) shows the trend of a steady gradual increase of prevalence of Diabetes Mellitus in the majority of Bulgarian demographic regions. The most significant increase (Fig. 3 – dark red) is observed in the regions with the poorest population and low quality of life. The average percentage 9.47% for Bulgaria looks relatively good in the European context. The distribution of patients per ICD-10 codes of diagnosis (Fig. 4) shows that in 2016 about 93.4% of the patients have Diabetes Mellitus Type 2 and approximately 5.28 % - Diabetes Mellitus Type 1.

Fig. 3 *Prevalence of Diabetes Mellitus type 2 in demographic regions of Bulgaria in 2016*

Fig. 4 *Ratio of patients with different ICD-10 codes related to Diabetes Mellitus in 2016*

Lon-term studies of patient care have both economic and healthcare impact. They cover tasks like:
- Monitoring of treatment effect for new drugs and various combined therapies,
- Evaluation of Diabetes Mellitus glycemic control etc.

Treatment with new drugs like incretins achieves good effect on Diabetes glycemic control for most patients (Fig. 5). The statistical data show that the most popular combined treatment includes two types of Insulins and analogues for injection (Fig. 6). Three types of drug combinations are most popular for combined peroral therapy

(Fig. 7). The overall evaluation of the Diabetes Mellitus compensation per age groups is shown on Fig. 8.

Fig. 5 *Treatment effect monitoring for selected drugs – HbA1c levels before and after incretins treatment*

Fig. 6 *Patients with combined therapy of Diabetes Mellitus grouped by number of drugs included in the therapy and ATC codes of drugs per age groups*

Fig. 7 *Patients with combined therapy of Diabetes Mellitus grouped by different ATC drugs codes in the therapy, and number of drugs included in the therapy per age groups*

Fig. 8 *Study of treatment quality – percentage of Diabetes Mellitus compensation per age groups*

Prevention is focused on reducing the risk factors associated with chronic diseases. Thus prevention related applications include comprehensive studies of Diabetes Mellitus risk factors. Risk stratification for prediabetes patients was shown in [6, 7, 8], which discuss risk factors and comorbidities between several socially important diseases like Diabetes Mellitus, Mental Disorders, Cardiovascular diseases and the Chronic Obstructive Pulmonary Disease (CORD). Several artificial intelligence approaches were developed and applied in these studies – like natural language processing, data mining, and machine learning. In [6] we report about experiments with the dataset of patients who were in prediabetes condition in 2013-2014 and with Diabetes Mellitus onset in 2015. The data mining techniques identify some well-known risk factors but new interdependencies among the indicators were discovered as well. The latter will need further investigation from healthcare point of view.

One of the primary objectives of the Diabetes Registers is better treatment of diabetic patients [3]. In order to achieve this goal, research in several directions was performed:

- Comprehensive analysis of Diabetes Mellitus comorbidities [6, 7, 8],
- Study of Diabetes Mellitus complications and treatment side effect.

We investigated the side effect of Schizophrenia treatment with first and second generation antipsychotics for triggering some prediabetes conditions. There is a wide range of diseases that have comorbidities with Diabetes Mellitus. One of the most discussed relation is between Diabetes Mellitus and Cardiovascular Diseases. Fig. 9 shows the dependency between Hypertension and Diabetes Mellitus. Hypertension is considered not only as one of prediabetes conditions but also as an important complication related to patients with Diabetes Mellitus. Another interesting finding is the relation between Malignant neoplasm of breast and Diabetes Mellitus [6], or Gonarthrosis (arthrosis of knee) [7]. An analysis of comorbidities between Schizophrenia, Diabetes Mellitus and COPD is presented in [8]. Etiology examination of the comorbidity between disorders is very important for mortality prediction and other outcomes. Diabetes as a chronic disease is considered a primary factor for increase of mortality (Fig. 10). The onset of Diabetes Mellitus for males happens a little bit earlier than for females. Thus Diabetes complications for male patients are usually developed in a younger age, the mortality percentage is higher and male patients die earlier than females with the same age, especially male patients with less contacts to doctors.

Fig. 9 *Number of patients with Hypertension and Diabetes in 2016*

Fig. 10 *Statistics about age of death – patients grouped by gender, with/without Diabetes, and frequency of visits to medical professionals. Diabetic male patients with less visits die earlier*

4. Future Work: the eHealth Programme

The European Commission's eHealth Action Plan 2012-2020 [9] provides a roadmap to empower patients and healthcare workers, to link up devices and technologies, and to invest in research towards the personalized medicine of the future. The Action Plan admits the delay of introducing modern IT solutions in European healthcare and lists barriers that hamper the wider uptake of eHealth, among them:

- lack of awareness of, and confidence in eHealth solutions among patients, citizens and healthcare professionals;
- lack of interoperability between eHealth solutions;
- limited large-scale evidence of the cost-effectiveness of eHealth tools and services;
- lack of transparency regarding the utilisation of data collected by eHealth applications.

The Action Plan concludes that "the important issue concerning the lack of health data exchange can only be tackled by addressing in a coordinated way fragmented legal frameworks, lack of legal clarity and lack of interoperability".

The Bulgarian National Research Programme eHealth[2] has been initiated by the Bulgarian Ministry of Education and Science with the aim to address most of the barriers listed above by development of specialized research prototypes and demonstrators. Briefly, collecting pseudonymized patient records from various sources we aim to show that the EHRs of Bulgarian patients can be generated almost automatically from the collections of electronic documents that are exchanged between the National Health Insurance Fund, General Practitioners, medical specialists in Primary Care, Hospitals, Clinical Labs that perform tests and send lab results via internet, and so on. Each of these document is related to certain individual patient using his or her unique civil identifier called EGNumber. Therefore, it is relatively easy to build the patient record collection itself given that in Bulgaria many important patient-related documents are standardized because they are all submitted electronically to the National Health Insurance Fund for reimbursement. Further our aim is to demonstrate, using the big data base of anonymous patient records, how the quality of healthcare procedures and treatments can be assessed at national level together with its costs.

It is expected that the project will deliver a methodology and technological solutions for providing semantic operability among the heterogeneous data generated by the healthcare sector and will demonstrate how the patient-related data can be integrated in complex clinical systems. These large-scale evidences will increase the confidence in information technologies and big data processing by healthcare systems. Last but not least, the public demonstration how to achieve improvements of patient cure and quality of life, together with optimization of costs, will pave the way to real application and practical adoption of the EHR in Bulgaria. We plan to apply in the forthcoming research efforts all lessons learned during the construction of the national Diabetes Register which is now supported by USHATE in Medical University Sofia.

Acknowledgements

References

1.Garrofé, B., Björnberg, A., Phang, A.Y.: Euro Diabetes Index 2014. Health Consumer Powerhouse Ltd., ISBN 978-91-980687-4-0, 2014.

2. Hallgren Elfgren, I.M., Törnvall, E., Grodzinsky, E.: The process of implementation of the diabetes register in Primary Health Care. International Journal of Quality Health Care 24(4), 2012, 419-424.

3.Richardson, E. (ed.). National Diabetes Plans in Europe: what lessons are there for the prevention and control of chronic diseases in Europe? Policy Brief of the Joint Action on Chronic Diseases and Promoting Healthy Ageing across the Life Cycle, WHO Regional Office for Europe, ISSN 1997-8065, 2016.

4. Dalianis, H. Clinical Text Mining: Secondary Use of Electronic Patient Records. Springer Open, 2018, publicly accessible at https://www.springer.com/gp/book/9783319785028 .

5.Tcharaktchiev, D., Zaharieva, S., Angelova, G., Boytcheva, S., et al. Building a national register of Diabetes. Social Medicine 1(2), Research Society of Social Medicine, Informatics and Health Management, 2015, 19-21. ISSN:1310–1757, http://journals.mu-varna.bg/index.php/sm/article/view/1345

6. Boytcheva, S., Angelova, G., Angelov, Z., Tcharaktchiev, D.: Integrating Data Analysis Tools for Better Treatment of Diabetic Patients. In: Kalinichenko, L., Manolopoulos, Y., Skvortsov, N., Sukhomlin, V. (eds.) Selected Papers of the XIX International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2017), CEUR Workshop Proceedings Vol. 2022, 2017, 230-237, http://ceur-ws.org/Vol-2022/

7.Boytcheva, S., G. Angelova, Z. Angelov and D. Tcharaktchiev. Mining comorbidity patterns using retrospective analysis of big collection of outpatient records. Health Information Science and Systems, 5(3), Springer Int. Publ. 2017, ISSN: 2047-2501, https://link.springer.com/article/10.1007/s13755-017-0024-y

8.Boytcheva, S., Nikolova, I., Angelova, G., Angelov, Z.: Identification of Risk Factors in Clinical Texts through Association Rules. Proceedings of the Biomedical NLP Workshop associated with RANLP 2017. (pp. 64-72). doi:/10.26615/978-954-452-044-1_009, http://acl-bg.org/proceedings/2017/RANLP_W4%202017/pdf/BioNLP009.pdf

9. eHealth Action Plan 2012-2020 - Innovative healthcare for the 21st century. Communication from the Commission COM(2012) 736, https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century

---

[2] https://ez.mu-sofia.bg/home (in Bulgarian)