Assessment of Pharmacology Costs in Diabetes Treatment Using OMOP CDM: A Nationally Representative Study

Evgeniy Krastev
Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
eck@fmi.uni-sofia.bg

Petko Kovachev
Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
az@petko.info

Simeon Abanos
Faculty of Math. and Informatics
Sofia University St. Kliment Ohridski
Sofia, Bulgaria
simeonabanos@gmail.com

Ralitsa Krasteva
Specialized Hospital for Active Treatment
of Children's Diseases
Sofia, Bulgaria
ralitsakrasteva@outlook.com

Dimitar Tcharaktchiev
Department of Medical Informatics
Medical University
Sofia, Bulgaria
dimitardt@gmail.com

Abstract—**Public diabetes reports rarely use large volume of observational health data because these data are usually heterogenous in terms of structure and semantic presentation of clinical concepts. In this paper we employ observational health data transformed to OMOP CDM database in a nationally representative study with the goal to estimate the pharmacology costs for diabetes treatment in Bulgaria during 2018. The OMOP CDM database contains health data from pseudonymized outpatient records of 501,065 patients with diabetes (45.3% male and 54.7% female). The mean age of the patients with Type 1 (42.249 patients) and Type 2 (458.816 patients) gives mean age 57.04 (CI 95%, [56.87, 57.22]) years for Type 1 and 66.38 (CI 95%, [66.35, 66.41]) years for Type 2. Drug costs are evaluated with respect to the major classes of drugs prescribed for diabetes treatment and diabetes comorbidities treatment. The annual average cost of drugs per patient with diabetes is estimated to 750 Euros. The obtained results are new and help to understand the trends and effects in using different classes of drugs for diabetes treatment. Novel drug diabetes therapies are found to be evolving in 2018, while the Metformin prescriptions prevail significantly. The costs in this study we evaluate both at patient-centric level by age groups and gender specifics and at high level in terms of cost distributions among the drug classes in each group. The results are graphically visualized, discussed, and compared in relation to existing public sources.**

Keywords- diabetes mellitus; diabetes register; nationally representative study; registry; database; OMOP CDM; Common Data Model; pharmacology; cost analysis.

I. INTRODUCTION

Observational health data (OHD) are valuable resource for the assessment of cost effectiveness and management of healthcare services. Typical examples of OHD include values of plasma glucose, glycated hemoglobin (HbA1c) levels, measurements of blood pressure as well as prescriptions for drug therapy in accordance with a medical diagnose. Such clinical data are collected routinely during health procedures under real-world conditions during the execution of most of the business processes in the healthcare system. Therefore, OHD is indispensable in the evaluation of typical key performance indicators for cost effectiveness like budget, satisfaction of healthcare standards and quality assurance. These indicators support decisions making for figuring out the reimbursement status of innovative drugs or introducing improvements in existing treatment protocols.

This paper extends the study in a related research work [1] with a deeper analysis of the drug prescriptions collected in a nationally representative study aiming to obtain accurate estimates of the pharmacology cost for treatment of patients suffering from diabetes mellitus (DM) Type 1 and Type 2. DM is a chronic metabolic illness with steadily increasing prevalence and associated to high risk for comorbidities such as blindness, kidney failure, heart attacks, stroke, and lower limb amputation [2] [3]. A large number of the patients with DM depend on life-saving insulin to control daily the glucose blood levels. Often, they need to take additionally one or more medications to avoid or treat the illness complications such as high blood pressure [4]. Therefore, DM treatment incurs significant expenses for life-long regular purchases of multiple medication products [5]. Part of the medication costs are reimbursed by health insurance funds, while the greatest part of these expenses harshly affects directly or indirectly the finances of the individuals, their families, and the society.

Recent research study [6] reports a total estimated cost of $412.9 billion in 2022 (compared to $327 billion in 2017 [7]) for treatment of diabetes in the USA, where a patient with diabetes spends on the average $19,736 annually for medical treatment. Major sources for indirect costs include reduced or total loss of productivity, respectively, due to disability ($28.3 billion) or premature death ($32.4 billion). Thus, the medical expenditures of a patient with DM are 2.3 times higher than what expenditures would be in the absence of diabetes [8]. Cost effectiveness of diabetes treatment turns out to be even more important to consider in resource constrained low-income and middle-income countries [9] [10]. Most of the existing sources focus on the study of outpatient costs for treatment of DM Type 2 where the annual medication costs in 2018 are estimated in a wide range between $15 and $500 (median is $177) depending on the country GDP [11]. It is noteworthy, that many of these studies aggregate and process data from statistical national-level surveys instead of using

original clinical documents. Therefore, the numerical results in these studies are obtained as statistical estimates from data extrapolated to the entire population of a country [12]. The thus obtained values of the numerical indicators in recent systematic reviews are suitable and sufficiently well describe the substantial financial burden imposed on the society by the diabetes illness. Apparently, management of healthcare services and improvement of medication treatment protocols of diabetes require a more accurate assessment of the pharmacology costs employing evidence-based data.

The limited availability of nationally representative data that provides patient-centric evidence for the medical treatment of diabetes appears to be the major obstacle to obtaining accurate evidence-based data for estimation of the pharmacology costs. Although huge volumes of OHD are generated in the healthcare system of each country, it is often difficult to overcome severe interoperability, methodological and organizational problems in any attempt to extend local evidence-based data processing at national level. Therefore, the number of national diabetes registers maintaining data about the health status of diabetics remains rather limited [13] [14]. The primary reason is that health data is persisted on heterogeneous platforms at multiple remote locations and besides, it is managed by disparate information models and technologies.

In practice, healthcare providers use Electronic Health Records (EHR) to persist systematically observational data as well as other medical information like prescribed medications, allergies, laboratory test results and demographics data about the patient in digital format [15]. Unlike the EHR, an Electronic Medical Record (EMR) such as the Outpatient record (OpR) provides a more detailed description of the patient's medical history than the EHR because it is maintained by a single healthcare provider (HP) [16]. Similarly to the EHR, the OpR captures rich observational data about the health status of a patient and allows the HP to follow it while prescribing treatment activities and procedures across time. In the general case, an EHR comprises the patient's EMRs from potentially different HPs. Thus, the EHR enables sharing of knowledge, skills and experience through communication between the actors in the healthcare system, serve the basis for research and education, satisfy organizational and legal requirements [17]. Nowadays, a lot of these opportunities for utilizing EHRs cannot be fully exploited. The reason is the lack of interoperability among the heterogeneous and proprietary nature of the software applications used by multiple HPs. Such interoperability problems stem from the primary distinction between EHRs and EMRs. EHRs are introduced for the purpose of sharing health data among organizations while EMRs serve the needs of a single HP. Therefore, the EMRs and in particular, the OpRs of a patient cannot be seamlessly integrated in the EHR of that patient.

Considerable research efforts have been made in the last twenty years to resolve the interoperability issues in the exchange of clinical data [18]. Data exchange schemas and standards for reference models have been introduced for sharing EHR data across clinicians, patients and communities [19] [20] [21]. This approach allows disparate health information systems to effectively communicate, exchange data and process the exchanged data within and across the organizational boundaries. Services for accessing and sharing EHRs may accommodate their requirements with respect to three distinct levels of interoperability-foundational, structural and semantic interoperability [22]. Foundational interoperability is limited to the availability of information technology, allowing EHR data exchange. Structural interoperability upgrades foundational interoperability with requirements for representing the exchanged data in predefined syntax and thus, allowing interpretation of data at individual data field level. Most often interoperability at that level is used for exchange of observational data represented in terms of a Common Data Model (CDM) where the physical implementation could be a relational database or an XML Schema [23] [24]. The semantic interoperability level employs standard terminologies, classifications and vocabularies to encode EHR clinical data so that the receiving information systems can correctly interpret the clinical meaning of such data without human intervention [25] [26]. It is noteworthy that the clinical meaning is inferred not from the individual data values themselves rather from the way in, which such data are linked together as compound clinical concepts, hierarchically structured terms, problems or associated with preceding healthcare events. This interoperability level preserves the semantic context of the exchanged clinical data by representing clinical concepts in terms of standard reference models such as ISO/EN 13606 and HL7 FHIR. Therefore, the exchange of EHR extracts usually implements such semantic interoperability standards.

In this paper we consider a pharmacology case study that illustrates the potential of CDM to facilitate access to observational data and enhance population-based statistical research. It is motivated by the need for accumulating evidence on cost effectiveness and budget impact through Health technology assessment (HTA) [27]. The objective is to assess the burden of pharmacology costs spent for treatment of diabetes in a nationally representative dataset. The data source for this study is the Bulgarian Database Register (BDR) that is an Observational Medical Outcomes Partnership (OMOP) CDM standardized database publicly available at the EHDEN Portal [28] [29]. This database contains observational data (observation period 01.01.2018–31.12.2018) of all the outpatient records (6,887,876) issued in Bulgaria to patients with diabetes (501,065). The outpatient records are compiled by the general practitioners (GPs) and the specialists from ambulatory care for every patient encounter. In this case study the CDM appears to be the optimal solution for imposing structural interoperability in dealing with disparate data sources such as the variety of software applications employed to produce the outpatient records. Thus, the dataset of the BDR can be accessed remotely and return aggregated results by executing analytical code locally in the secure environment of the data custodian.

This paper is divided into sections as follows. In the following section, we make a brief overview of the existing CDM that enhance big medical data analytics [30] [31] [32] and elaborate on the OMOP CDM of the BDR. In Section III, we present aggregated results obtained by executing the

analytical code. In Section IV, we discuss the obtained results and compare them with existing research work [33]. Section V makes a conclusion and provides remarks on future work.

II. Methods and Materials

This paper considers a case study where the original data sources are outpatient records created by a large number of GPs and specialists from ambulatory care using heterogeneous databases and client applications with disparate programming interface for data access, management and analysis. It entails problems caused by poor data interoperability such as patient-matching with observational data, pseudonymization of records, satisfying requirements for integrity and consistency of clinical data. The development of software tools for analysis and assessment of data in distributed dataset environment is rather complicated and inefficient as well. The need for imposing some kind of unification of these disparate data sources focused our attention on using CDM in this research.

The literature review provides convincing evidence that CDM are the preferred solution in cases of poor data interoperability when simultaneous analysis of disparate data sources is required [24] [34]. There are three most widely used CDMs for observational data research, namely, the OMOP CDM, the Sentinel and the Patient Centered Outcomes Research Institute (PCORNet). Each one of these CDMs has its strengths and weaknesses.

The PCORNet CDM [30] introduces its own standard organization and representation of EHR data for a distributed network of nine population-based Clinical Research Networks of data contributors (more than 14 billion diagnoses, 2.6 billion medication orders and 9.8 billion laboratory results) [35]. Nowadays, PCORNet persists data from the healthcare encounters of more than 30 million patients across the USA. The information model of the PCORNet CDM enables users to execute to query against the health data of a large number of patients and promptly receive the result in a standardized format. Special attention is dedicated to data quality by applying a two-stage for screening PCORNet-accessible data. During the first stage data is examined for conformance (adherence of EHRs to PCORnet CDM), completeness (diagnosis codes are not missing and correctly recorded), plausibility (ensure that the values are meaningful), and persistence (ensure that source data is preserved upon running queries). At the second stage we consider data quality issues, relevant to the specific research problem. A major weakness of this CDM is the missing support for clinical outcome measures as well as data linkage, for example, queries cannot "*de-duplicate*" patients appearing in multiple networks.

The Sentinel CDM was introduced in 2007 by the Federal Drug Agency (FDA) to monitor drug safety and includes EHR and register data in the following core subject areas utilization, registration, pharmacy, demographics, lab, death and vital signs (more than 463.3 million unique patient identifiers spanning the time period from 2000 to 2023, 19.7 billion drug dispensations, 20.2 billion unique medical encounters, 67.3 million patients with at least one laboratory test results) [31] [36]. Sentinel uses a distributed data approach in, which Data Partners maintain physical and operational control locally over their electronic health data. The Data Partners are the organizations that collect routinely OHD on every patient encounter and transform it in the Sentinel CDM (SCDM) information model.

The SCDM is a data structure that standardizes administrative and clinical information across Data Partners. It is extensible to any data source because data is represented as detailed as possible avoiding embedded terminology mapping to maximize transparency and analytic flexibility. This way the network of Sentinel Data Partners allows FDA to generate larger datasets and study adverse events or drug dispensations even in small populations. Thus, the Sentinel CDM is flexible about demands for running data queries in any type of analysis. Data quality assurance practices are aligned with the FDA guidelines where three levels of quality checks are introduced to assess pre-defined data quality measures and characteristics [37]. Queries are processed in a distributed pattern as follows. Firstly, query requests are distributed to the data partners where the queries run locally. Next, query results with direct identifiers removed are returned to the central server for aggregation and final processing. It entails keeping copies of large amounts of data and time-consuming data synchronization even for simple queries. A critical system limitation is the inability to identify enough medical conditions of interest in OHD to a satisfactory level of accuracy [38]. Other weaknesses include limited data mapping, extensions of the CDM affect data usability, data granularity entails loss of information and local knowledge and finally, ongoing model refinements are driven entirely by the FDA.

The OMOP CDM was introduced about the same time as the Sentinel CDM for the purpose of studying the effects of medicinal products. Currently, it is extensively used in the US and Europe where it is underpinned by the Observational Health Data Sciences and Informatics (OHDSI) network and the EHDEN project of the EU (more than 153 EHDEN data partners, more than 1.12 billion unique patient identifiers) [39]. Similarly to Sentinel and PCORNet, the OMOP CDM maps disparate data sources to a "*patient-centric*" relational database with predefined tables linked directly or indirectly to patients. The tables correspond to the CDM core subject areas such as *PERSON*, *VISIT_OCCURRENCE*, *DRGUG_EXPOSURE*, *MEASUREMENT*, *OBSERVATION*, *DEATH*. There are also tables describing *DEVICE_EXPOSURE*, *PROCEDURE_EXPOSURE* as well as standardized vocabularies for normalizing the meaning of data within the CDM. Thus, the OMOP CDM has the potential to meet the requirements of HTA.

The OHDSI OMOP CDM is well supported by software tools assisting the Extract-Transform-Load (ETL) process and ensuring data quality during the mapping steps. This allowed us to map to OMOP CDM health data from 6,887,876 outpatient records collected by the National Health Insurance Fund (NHIF) in Bulgaria from GPs or HPs upon the encounters of 501,065 patients with diabetes during 2018 [29] [40]. Meta data of the thus obtained OMOP CDM (v.5.3.1

[41]) of the Bulgarian Diabetes Register are published in the EHDEN Portal (Figure 1).

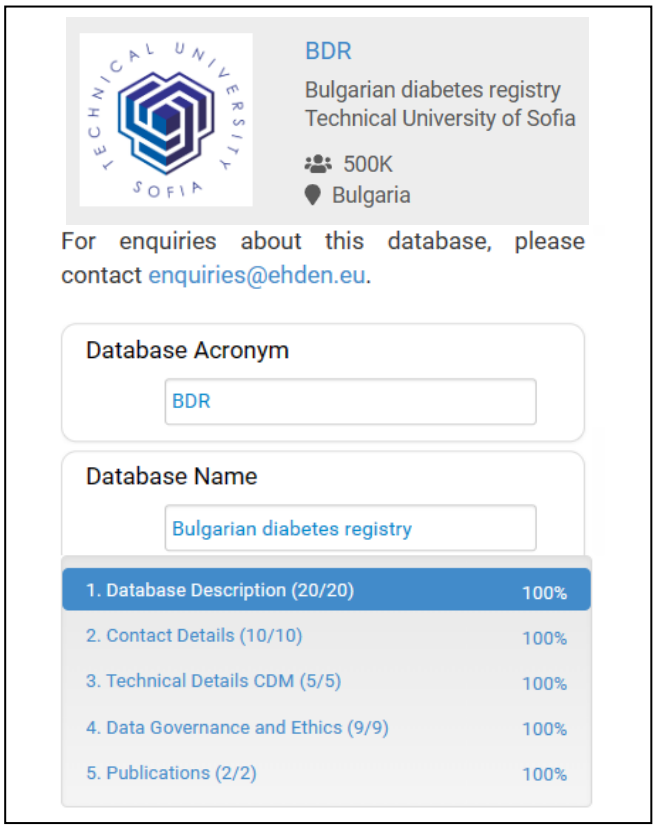

Figure 1. Link to the OMOP CDM of BDR inside the EHDEN Portal.

The distribution of diabetics (Type 1 and Type 2) relative to the population of the corresponding administrative region is displayed in Figure 2.

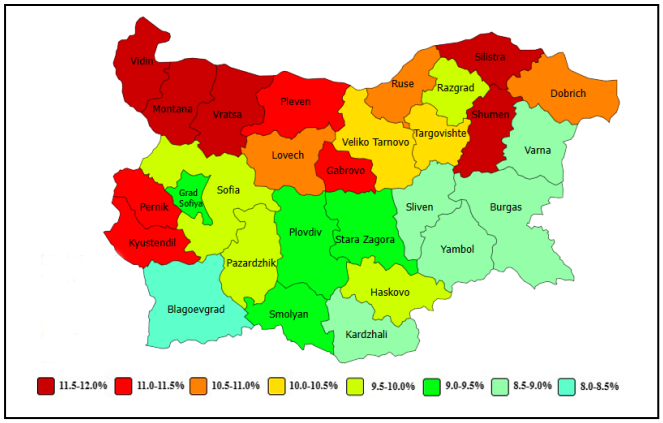

Figure 2. Distribution of patients with diabetes in Bulgaria in 2018.

This figure shows that most of the people living in the northern part of the country and especially, in the north-west part, have diabetes. These are the least populated regions of the country. It motivates us to explore the burden of costs spent for reimbursement of drugs for treatment of diabetes and its related comorbidities (cardiovascular drugs, drugs for disorders of the eyes or the nervous and urological system), for the purpose of comparing it with related research work.

The original pseudonymized outpatient records were provided by the NHIF in XML format that needed data processing for making them valid against a single XML schema. For convenience, we loaded the adapted XML instances of outpatient records in a relational database that served as a source for the ETL process (Figure 3).

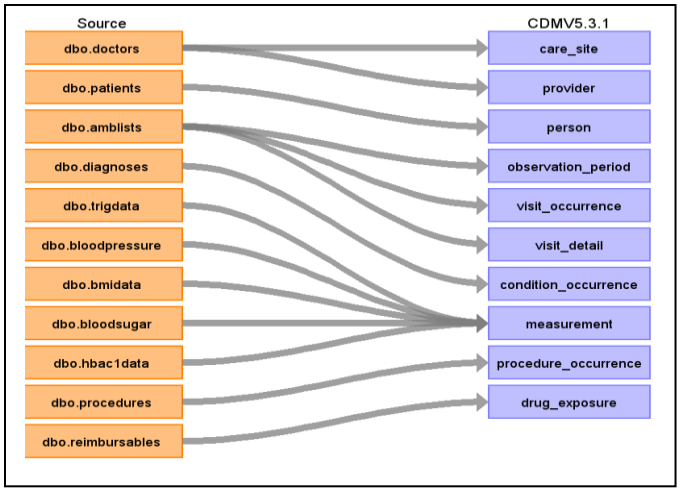

Figure 3. Mapping of outpatient records to OMOP CDM.

These records contain administrative data and coded clinical data describing health status or procedures such as:

- ✓ Date and time of the visit occurrence.
- ✓ Administrative data.
- ✓ Personal data, age, gender.
- ✓ Patient visit-related information.
- ✓ Diagnoses in ICD-10.
- ✓ ATC codes for medications reimbursed by the NHIF.
- ✓ Encodings for examinations and procedures.
- ✓ Codes describing specialized health care.
- ✓ Codes describing hospitalization need.
- ✓ Codes for planned consultations.
- ✓ Laboratory tests and medical imaging.

Observational data like patient *status*, *height*, *weight*, *Body-Mass-Index* or *blood pressure* were provided in the outpatient records as unstructured data in native language (Bulgarian text).

Special interests in this study represent the fields in the OMOP CDM table *drug_exposure* shown in Figure 4 where the field *drug_concept_i*d encodes the drugs prescribed to diabetics and reimbursed by the NHIF [42]. It is noteworthy, that the Bulgarian national drug codes are represented in the ATC hierarchical classification system. Therefore, the standard vocabularies of the BDR are linked to ATC drug codes through *drug_concept_id*.

All other medicinal products represent no interest to the NHIF, and such products are recorded as unstructured free native text in the source dataset of outpatient records using their International Non-proprietary Names for Pharmaceutical Substances (INN). The extraction of INN of individual medications and mapping the INN to ATC codes requires a lot

of efforts and currently table DRUG_EXPOSURE does not include codes of non-reimbursed medicinal products.

Therefore, in this study the assessment of the pharmacology costs is limited to evaluating only the costs of medications that are reimbursed. In addition to table *DRUG_EXPOSURE*, the analytical code in this study makes use of tables *PERSON*, *CONDITION_OCCURRENCE*, *VISIT_OCCURRENCE* and *OBSERVATION_PERIOD* of the OMOP CDM. These are the core entities in the group of tables in the OMOP CDM database referred to as "*Standardized clinical data*".

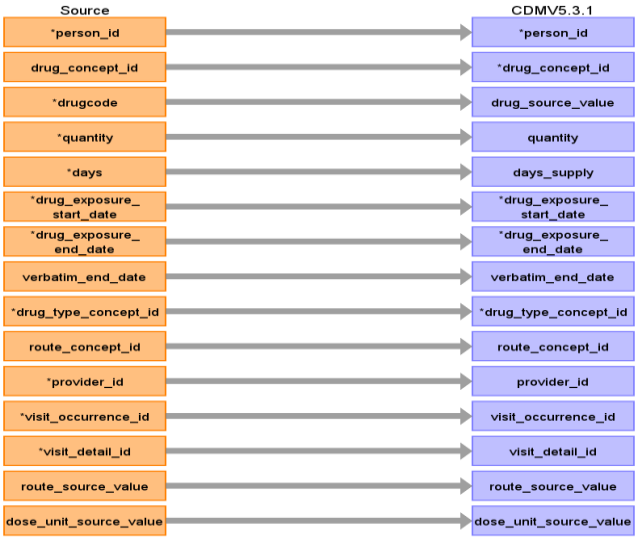

Figure 4. Mapping to table *drug_exposure* of the OMOP CDM.

Table 1. Drug classes for treatment of Diabetes.

Code	Drug class	International Nonproprietary Name (INN)
T1	Insulin	Insulin unique analogues and combination regimens
T2	Sulfonylureas	Glyburide, Glipizide, Glimepiride, Gliclazide, Tolbutamide,Chlorpropamide, Tolazamide, Vildagliptin
T3	Biguanides	Metformin
T4	Alpha-Glucosidase Inhibitors	Acarbose, Miglitol, Voglibose
T5	Thiazolidinediones	Troglitazone, Rosiglitazone, Pioglitazone
T6	Incretin-Dependent Therapies	Incretin , Exenatide, Liraglutide, Dulaglutide, Albiglutide, Lixisenatide, Semaglutide, Sitagliptin, Saxagliptin, Linagliptin,Alogliptin
T7	Meglitinides	Nateglinide, Repaglinide
T8	Sodium-Glucose Cotransporter Type 2 Inhibitors	Canagliflozin,Apagliflozin,Empagliflozin, Ertugliflozin, Dapagliflozin
T9	Statin-Dependent therapies	Simvastatin, Lovastatin, Ravastatin , Fluvastatin, Atorvastatin, Cerivastatin, Rosuvastatin, Pitavastatin

The existing literature distinguishes several distinct classes among the drugs for diabetes treatment [33] [43]. These classes are shown in Table 1, where the custom *Code* is introduced for the purpose of referencing the obtained results in the following section.

It is noteworthy, that currently, the drug class denoted as T8 in Table 1 is considered in the literature as the most modern and promising for DM treatment [33]. This is another reason to find out what is the share of prescriptions of these drugs. Similar interest represents the distribution of drugs prescribed for treatment of diabetes comorbidities.

The drug encodings displayed in Table 2 are introduced as shortcuts for referencing the drugs used for treatment of the most frequently encountered comorbidities of diabetes. By means of the respective drug *Code* in Table 2, it will be easier to quote these classes of drugs in the obtained results.

Table 2. Drug classes for Diabetes comorbidity treatment.

Code	Drug class for comorbidity treatment	ATC code prefix
A	Cardiovascular drugs	C01-C10
A1	Antithrombotic agents	B01
N	Nervous system disorders	N01-N07
G	Urological disorders	G04
S	Ophthalmolotical disorders	S01
L	Endocrine disorders	L02
M	Ttreatment of bone diseases	M05
R	Asthma drug categories	R03

Let's consider now the methods for evaluating the costs of the drug classes outlined in Table 1 and Table 2. The specification of the OMOP CDM v. 5.3.1. dedicates table COST to persist the cost of any medical case recorded in one of the OMOP tables such as DRUG_EXPOSURE, PROCEDURE_OCCURRENCE, VISIT_OCCURRENCE, VISIT_DETAIL or MEASUREMENT (Figure 4). The COST table belongs to a group of two tables entitled as "*Standardized health economics*" and it is not related referentially to any of the tables displayed in Figure 4. Once a payer (patient) completes a payment for a medicinal product or health service it triggers an event to record the payment details in the COST table. Such a scenario assumes the interaction of the OMOP CDM database with a kind of "claims" database and it is a use case that is feasible in a hospital environment. The outpatient records comprising the source dataset used in this study don't contain any payment details. Besides, ePrescriptions didn't exist in 2018. In view of the circumstances, instead of retrieving payment data from the COST table, we explored the publicly available price list for drugs reimbursed in 2018 by the NHIF [42]. By means of a custom developed Python script each of the prescribed drug codes was related to its price extracted from that price list.

Without loss of generality, in this study we assume that all the drugs prescribed to a patient are purchased in the current year. Based on this assumption, the "*Standardized clinical data*" set of tables in the OMOP CDM database allows to

group drug prescriptions and find totals per each prescribed drug that is reimbursed by the NHIF for patients suffering DM. Moreover, the "patient-centric" architecture of the OMOP CDM allows us to tally these numbers by type of diabetes, gender, age or any other property of the dataset. The results of executing these tasks are presented in the following section.

III. RESULTS

The BDR contains huge amounts of data that can provide rich information for treatment of diabetes. First of all, we get an accurate estimate for the diabetes prevalence (9.77%) in Bulgaria in 2018 (501,065 patients, 45.3% male and 54.7% female). Unlike other public data [12] [43], the diabetes prevalence is computed accurately taking into consideration the total number of individual patients with encounters registered by GPs or HPs and not by statistical estimates or extrapolation over the total population of the country.

Once we know the diabetes prevalence, it is important to learn what is the cost for diabetes treatment. The available data in the BDR allows to get detailed information on this issue from different perspectives. For shortness, here we present summary results that demonstrate the potential of HTA by limiting the scope of our research to drugs that are reimbursed by the National Health Insurance Fund as they are described in Table 1 and Table 2. The total cost (TC) of all the drugs prescribed for treatment of a patient with DM diagnose in Bulgaria in 2018 is 178,537,010 euros, where 96,201,239 euros is the amount for diabetes treatment with prescribed drugs from Table 1 that are reimbursed by the NHIF. These prescriptions cost the NHIF on the average about 356 euros annually per diabetic patient (Figure 5). Accordingly, 53.88% of the TC are for drugs prescribed for diabetes treatment (Table 1), where 61.55% is the share of the insulin class of drugs.

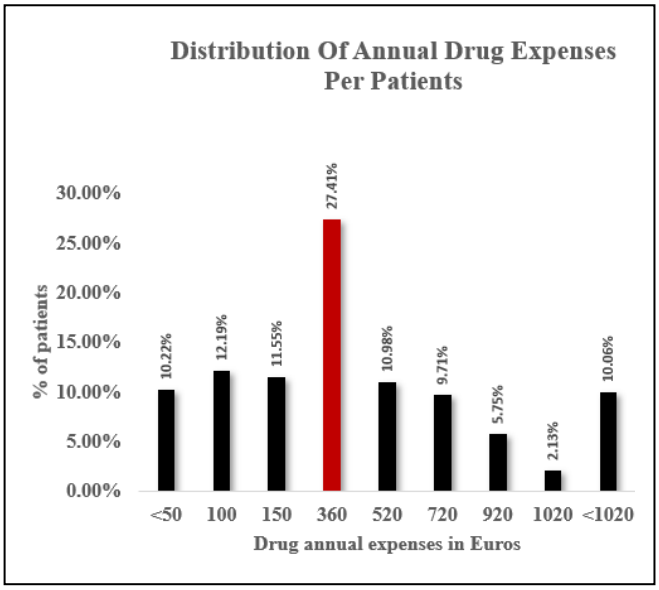

Figure 5. Distribution of annual pharmacology expenses for DM treatment.

A related research study [44] shows that the diabetes illness reaches its peak in the years after the age of 40. The expenses of patients with DM (Type 1 and Type 2) increase on the average up to 442 euros annually for this age group. Moreover, patients with diabetes undergo therapy for other illnesses like Hypertensive heart disease (ICD-10 code I11.0) or Tachycardia (ICD-10 code I48). Treatment of rare diseases or disorders caused by immune deficiency is even more expensive. These cases represent a huge burden in the overall amount reimbursed to patients for treatment of diabetes equivalent to 187,821,407 euros. Thus, the total average amount per patient that is paid annually by the NHIF for drug treatment of diabetes reaches 750 euros.

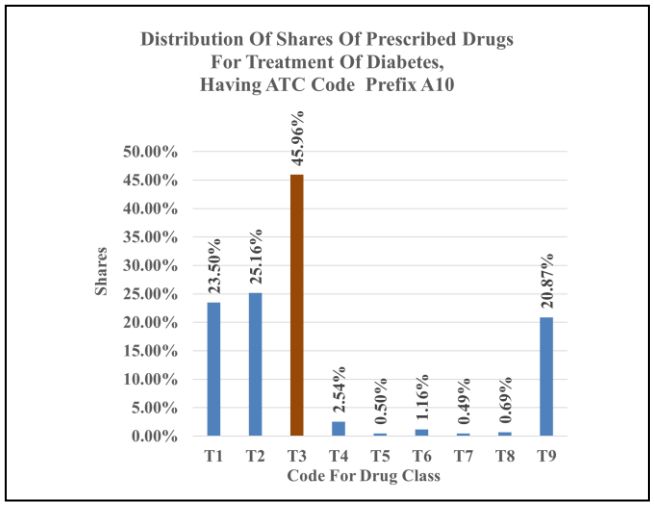

Figure 6. Shares of prescriptions for diabetes treatment.

Once we estimated the average cost for diabetes treatment, let's explore what is the share of modern drugs for diabetes treatment among all the prescribed drugs for diabetes treatment. Such are, for example, the drugs encoded as T8 in Table 1. Figure 6 shows that these drugs are rarely prescribed for diabetes treatment in Bulgaria during 2018 (0.69% of all the prescribed drugs from Table 1). Metformin drugs are the most frequently prescribed (T3 in Table 1). These kinds of drugs are usually prescribed for initial treatment of Type 2 diabetes and besides, the number patients with Type 2 diabetes (458,516; Male 45%, Female 55%) prevails significantly (91.5%) over the patients with Type 1 diabetes (42,249; Male 51%, Female 49%). This explains the peak value in the prescriptions for Metformin drugs (T3 in Table 1).

In terms of costs of the shares of the drugs in Table 1 are distributed as it is displayed in Figure 7. We notice that the largest expenses are attributed to the insulin class of drugs (T1 in Table 1) although it is the third most prescribed class of drugs in Figure 6. Drugs of that class are used for treatment of both DM with Type 1 and Type 2. For instance, the expenses for insulin drugs used for treatment of patients with DM Type 1 reach 99.13% of the total cost of drugs from Table 1 prescribed to these patients. Note, that the average price in Bulgaria for the insulin drug class has been about 60 euros against 16 euros for the Metformin drug class in 2018.

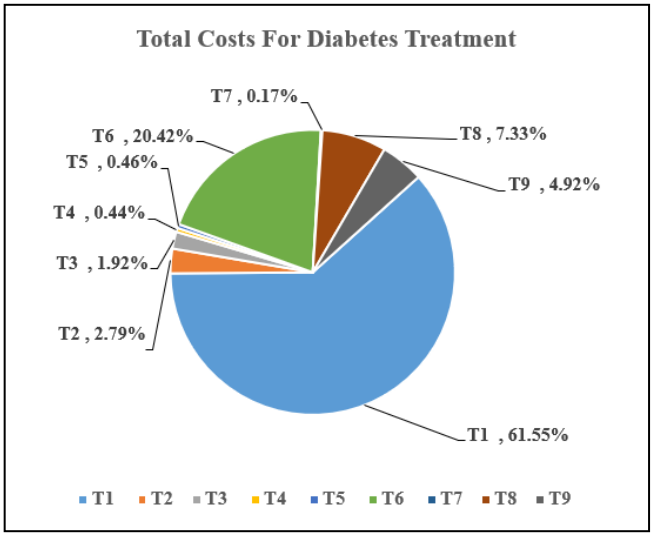

Figure 7. Total costs of drugs for diabetes treatment.

The above results provide evidence that the treatment of comorbidities accompanying the diabetes illness is almost as expensive as the treatment of the diabetes itself. Therefore, it is important to understand what the costs for treatment are of the most often encountered comorbidities.

In the existing literature there is enough evidence that the cardiovascular diseases, the disorders of the nervous system and the ophthalmological disorders are some of the most frequent comorbidities of diabetes. At the same time, little is known about the relative shares of these disorders with respect to the overall expenses for treatment diabetes comorbidities.

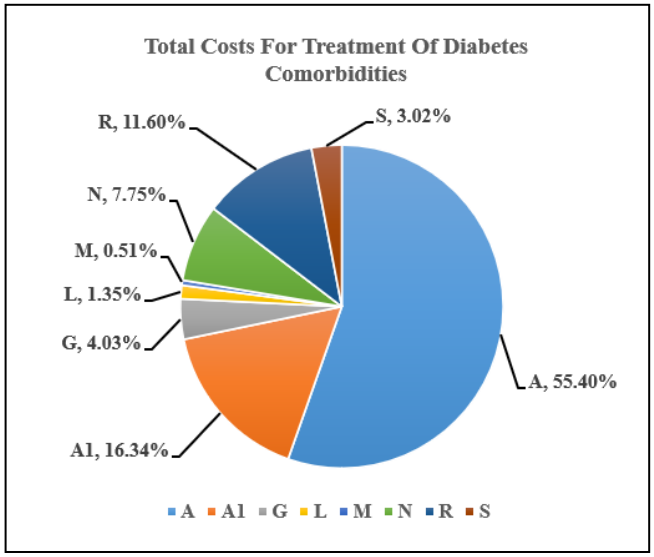

Figure 8. Total costs of drugs for treatment of diabetes comorbidities.

Figure 8 confirms that drugs for cardiovascular disorders and drugs with antithrombotic agents (code A and A1 in Table 2) have the greatest weight (71.74%) in the TC for treatment of comorbidities. The drugs for treatment of asthma of (code R in Table 2) are at the second place (11.60%) in the TC with average price of about 51 euros in 2018, where most of the prescriptions are for medical products costing above the average value.

For comparison, the drugs for treatment of disorders of the nervous system (code N in Table 2) are at the third place with 7.75% share in the TC with average price of about 130 euros. Unlike the drugs prescribed for asthma treatment, most of these prescriptions are for medical products with prices significantly below the average for all the products with code N in Table 2. Such an increase in the costs for drugs prescribed to diabetics for treatment of accompanying asthma disorders is observed for the first time and it should be taken in consideration in regulatory decision making.

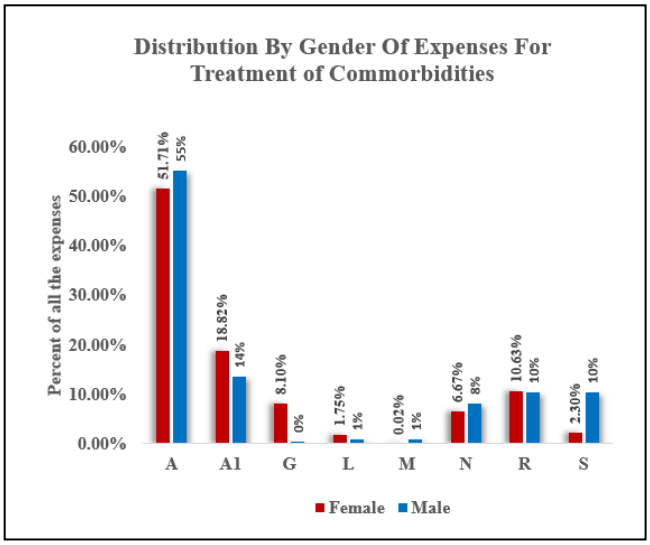

Figure 9. Distribution by gender of costs for Type 2 diabetes treatment.

The "patient-centric" architecture of the OHDSI CDM allows to investigate in greater detail the distribution of the pharmacology cost in terms of attributes gender and age in table PERSON (Figure 3).

For example, it is interesting to compare these costs for treatment of DM comorbidities subject to the gender of the patient. In Figure 9, we observe some notable differences in the distribution of these costs depending on the patient's gender. Most significant differences are discovered in the expenses for treatment of comorbidities of classes A (cardiovascular), A1(antithrombic), G (urological) and S(asthma) disorders. The expenses for treatment of male patients prevail classes A and S and that is a signal that comorbidities of DM are more specific to male patients. Similarly, we can conclude that comorbidities of classes A1 and G are more typical for female patients.

The analysis of the mean age of patients with Type 1 (42.249 patients) and Type 2 (458.816 patients) gives mean age 57.04 (CI 95%, [56.87, 57.22]) years for Type1 and 66.38 (CI 95%, [66.35, 66.41]) years for Type 2. The distribution in Figure 10 of the total costs of prescriptions per age groups of patients with Type 1 and Type 2 shows that the largest share of costs belongs to the age group 65-70 years. It coincides with the average age of patients with diabetes Type 2 that is the largest groups of patients receiving reimbursement by the NHIF for drugs used for treatment of diabetes.

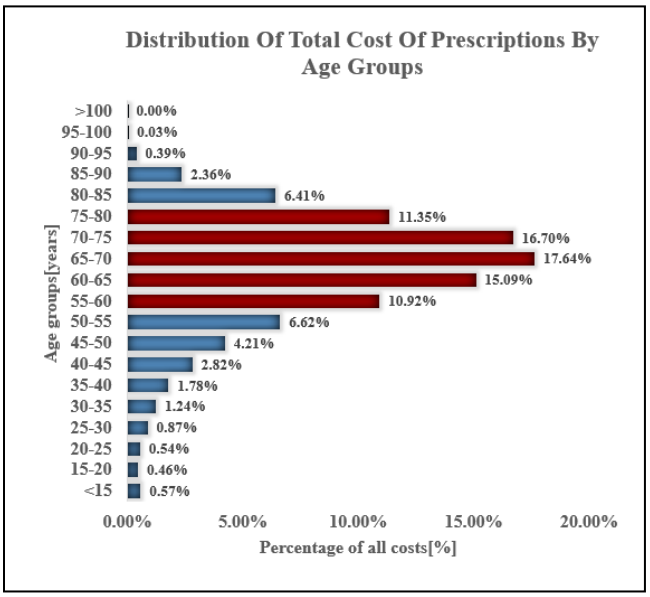

Figure 10. Distribution of costs by age.

IV. DISCUSSION

This paper reports results that are obtained by processing nationally representative data mapped to an OMOP CDM. The BDR is a physical implementation of that CDM with meta data published on the EHDEN Portal. It allows transparency in accessing data and verifying the integrity and consistency of these results. The BDR contains huge amount of pseudonymized observational data that allows to investigate diabetes treatment from different views through health assessment technologies.

A distinct feature of this study is that the obtained numerical results are obtained from evidence based OHD. The drug prescriptions are related to a large number of individual patients with a DM diagnose recorded in original clinical documents, the outpatient records issued by GPs and HPs on every patient encounter. Another important feature of the study is the nationally representative scope of the dataset of pseudonymized outpatient records used to extract data about the prescribed drugs. Unlike most statistical reviews, OHD for individual patients is not duplicated or extrapolated in this study. Data about diabetes prevalence coincide with recent reports. The here reported accurate value 9.7% of diabetes prevalence in Bulgaria in 2018 matches the statistically extrapolated prevalence data of the illness in Cyprus and Finland in 2021 [45]. A similar share of Diabetes Type 2 cases (91.5%) is established in more than 183 countries and territories [12].

The here considered pharmacology case study is just one example of the potential for exploring OHD mapped to an OMOP CDM. Without a restriction, data exploration could be extended to provide details with different level of granularity about the prescription of selected drugs or to group drug prescription by age and gender. In this regard, we must outline the following limitations that have to be taken in consideration.

First, it is rather difficult to find public literature with numeric data from population-based studies evaluating the burden of costs in diabetes treatment. In one such rare publication [43] we found evidence that matches close with our findings. Although this publication refers to data from 2014 and involves 312,223 patients from Italy, we established close correlation at several issues. For example, the share of costs on insulin drugs (T1 in Table 1) reported in that publication is 58.90% against the above quoted percentage 61.55%. Another match is established in the reported share of class A drug costs with respect to all drug costs 21.80% against 21.25 % found in our study with respect to the total of costs for drugs, prescribed for treatment of DM diagnose. There is, however, a significant difference in the average cost per diabetic patient, 1066 euros against 750 euros established from data in the BDR. This difference could be attributed to the known differences in the standard of life (and price levels) between both countries at that time.

Another issue that must be taken in consideration is that the NHIF does not reimburse always the full costs for prescribed drugs, while the amounts above quoted refer to the full drug costs. Moreover, in this study the analysis of costs considers data for patients that have health insurance.

Since the finance reports of NHIF are public [46], we managed to calculate the amounts really reimbursed by the NHIF for diabetic drugs (Table 1) to be 67,208,241 euros in 2018. As expected, this amount is about 30% less than the amount reported in the above section (96,201,239 euros). Here we must take in consideration that only a fraction of all the prescribed drugs in 2018 are dispensed to patients in the same year. Besides, the quantities of the prescribed drugs are usually greater than the quantitates of the reimbursed drugs. Thus, we can conclude that the results reported in this paper are consistent with the real-life practice.

V. CONCLUSION AND FUTURE WORK

This paper demonstrates the potential of the OMOP CDM to facilitate access to observational data accumulated from heterogenous datasets and extract knowledge using standard statistical tools. The assessment of the burden caused by the pharmacology costs on the healthcare system is important for regulatory decision making as well as for drug suppliers in planning their market strategies. Even though the assessment of the pharmacology costs in this study considers only the drugs reimbursed by the NHIF and data for patients with health insurance, the study produces an evidence-based estimate of the financial burden of DM on the society from different points of view including gender and age group distributions. The obtained results help to understand the trends and effects in using different classes of drugs for diabetes treatment and especially, the trends in applying novel drug therapies for diabetes treatment. Public diabetes surveillance reports with such results are rather rare to find in the existing literature primarily because most often the datasets are heterogenous in terms of structure and lack of interoperability of the data sources. Unlike most regularly published reports in the public space, this paper reports results obtained from a population-based study rather than applying aggregated statistical estimates.

The BDR implements an open-source OMOP CDM that allows overcoming poor interoperability among heterogeneous and often, incompatible data providers. It contains the latest and complete dataset of outpatient records issued to 501,065 distinct patients with diabetes in Bulgaria at every encounter to GP or HP in 2018. Among other CDM briefly reviewed in this paper the OMOP CDM proves the best potential for applying health assessment technology in obtaining reliable, transparent and verifiable results though analysis of observational data.

The pharmacology case study makes public lot of new results that help understand better the burden of costs generated in the process of prescribing drugs for diabetes treatment. Two major groups of drugs are considered, drugs for treatment the diabetes and drugs for treatment of diabetes comorbidities. This study presents numerical evidence that the cost for treatment of diabetes comorbidities is as expensive as treatment of diabetes itself. It emphasizes the need for developing better strategies and policies for prophylactic and control of the diabetes illness in order to minimize deterioration of the patient health status and respectively, the minimize the cost burden on the society.

Numerical evidence shows that novel drug therapies of diabetes in this country are just beginning to evolve in 2018, while the prescriptions of Metformin drugs prevail significantly among all the rest. Contrary to the expectations, the costs of prescribed drugs for treatment of comorbidities in diabetes caused by asthma surmount the costs of prescribed drugs for therapy of the nervous system or urological disorders. The costs are evaluated both at patient-centric level as well as at high level in terms of cost distributions among the drug classes in each one of the two groups. The results are graphically visualized, discussed and compared in relation to existing public sources.

In our future work we focus on exploring the trends in using novel drug therapies for diabetes in Bulgaria. Preliminary results based on new public data sources during 2018-2021 show a significant and rapid increase in prescriptions of novel drug class therapies (T8 in Table 1), decrease in other prescriptions (T7 in Table 1) and stable interest in other (T3 in Table 1). Moreover, we work on updating the BDR with fresh data once it becomes available.

ACKNOWLEDGMENT

This research is supported by grant 4.80-10-59/09.04.2024 of the Scientific Research Fund of Sofia university St. Kliment Ohridski.

REFERENCES

[1] E. Krastev, P. Kovachev, S. Abanos, and D. Tcharaktchiev, "Common Data Model for Interoperability of Observational Health Data: Bulgarian Diabetes Register Pharmacology Case Study", in GLOBAL HEALTH 2023: The Twelfth International Conference on Global Health Challenges, pp. 25-31, 2023. Available: https://www.thinkmind.org/articles/global_health_2023_1_50_70084.pdf. [Accessed 15 March 2024].

[2] WHO, "Diabetes”10 November 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/diabetes. [Accessed 10 March 2024].

[3] International Diabetes Federation, "IDF Diabetes Atlas” 2021. [Online]. Available: https://diabetesatlas.org/. [Accessed 1 March 2024].

[4] Centers for Disease Control and Prevention, "National Diabetes Statistics Report website. Estimates of Diabetes and Its Burden in the United States” 15 March 2023. [Online]. Available: https://www.cdc.gov/diabetes/data/statistics-report/index.html. [Accessed 15 March 2024].

[5] R. Turner, C. Cull, V. Frighi, and R. Holman, "Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). UK Prospective Diabetes Study (UKPDS) Group” JAMA, vol. 281, no. 21, pp. 2005-2012, 1999.

[6] E. D. L. Parker, Janice, T. Mahoney, N. Ume, G. Yang, R. A. Gabbay, N. A. ElSayed, and R. R. Bannuru, "Economic Costs of Diabetes in the U.S. in 2022” Diabetes Care, vol. 47, no. 1, pp. 26–43, 2024.

[7] R. Bosetti, L. Tabatabai, G. Naufal, T. Menser, and B. Kash, "Comprehensive cost-effectiveness of diabetes management for the underserved in the United States: A systematic review” PloS one, vol. 16, no. 11, pp. e0260139, 2021.

[8] American Diabetes Association, "Economic Costs of Diabetes in the U.S. in 2017” Diabetes Care, vol. 41, no. 5, pp. 917-928, 2018.

[9] M. Kamusheva, M. Vassileva, A. Savova, M. Manova, and G. Petrova, "An Overview of the Reimbursement Decision-Making Processes in Bulgaria as a Reference Country for the Middle-Income European Countries” Frontiers in Public Health, vol. 6, 2018.

[10] D. Flood, J. Seiglie, M. Dunn, S. Tschida, M. Theilmann, M. Marcus, et al., "The state of diabetes treatment coverage in 55 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 680 102 adults” Lancet Healthy Longev, vol. 2, no. 6, pp. e340-e351, 2021.

[11] C. Moucheraud, C. Lenz, M. Latkovic, and V. J. Wirtz, "The costs of diabetes treatment in low-and middle-income countries: a systematic review” BMJ Glob Health., vol. 4, no. 1, p. e001258, 2019.

[12] GBD 2021 Diabetes Collaborators, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021” Lancet, vol. 402, pp. 203–234, 2023.

[13] J. C. G. Bak, E. H. Serné, M. H. H. Kramer, M. Nieuwdorp and C. L. Verheugt, "National diabetes registries: do they make a difference?” Acta diabetologica, vol. 58, no. 3, pp. 267–278, 2021.

[14] University Specialized Hospital for Active Treatment in Endocrinology, "Diabetes Register” USBALE, 2024. [Online]. Available: https://usbale.org/bg/registar-zaharen-diabet/. [Accessed 11 January 2024].

[15] K. Dipak, "Electronic Health Record Standards, "Yearbook of medical informatics, vol. 45, no. 01, pp. 136-144, 2006.

[16] L. Spooner and K. Pesaturo, "The Medical record” in Fundamental Skills for Patient Care in Pharmacy Practice, Jones & Bartlett Learning, 2014.

[17] B. Blobel, "Interoperable EHR Systems–Challenges, Standards and Solutions” European Journal for Biomedical Informatics, vol. 14, no. 2, pp. 10-19, 2018.

[18] HIMSS, Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations, 5th ed., CRC Press, 2019, p. 75.

[19] HL7, "HL7 FHIR release 5” HL7, 26 March 2023. [Online]. Available: https://www.hl7.org/fhir/. [Accessed 11 March 2024].

[20] ISO/TC 215, "ISO 13606-1:2019 Health Informatics-Electronic health record communication-Part 1: Reference model" ISO, 2019. [Online]. Available: https://www.iso.org/standard/67868.html. [Accessed 6 March 2023].

[21] openEHR, "The EHR Information Model" The openEHR Foundation, 17 November 2022. [Online]. Available: https://specifications.openehr.org/releases/RM/latest/ehr.html#_the_ehr_information_model. [Accessed 14 March 2024].

[22] HIMSS, "Interoperability in Healthcare" Healthcare Information and Management Systems Society, 2023. [Online]. Available: https://www.himss.org/resources/interoperability-healthcare. [Accessed 11 March 2024].

[23] J. Weeks and R. Pardee, "Learning to Share Health Care Data: A Brief Timeline of Influential Common Data Models and Distributed Health Data Networks in U.S. Health Care Research" eGEMs (Generating Evidence & Methods to improve patient outcomes), vol. 7, no. 1:4, pp. 1-7, 2019.

[24] S. Kent, E. Burn, D. Dawoud, P. Jonsson, J. T. Østby, N. Hughes, P. Rijnbeek and J. C. Bouvy, "Common Problems, Common Data Model Solutions: Evidence Generation for Health Technology Assessment" Pharmacoeconomics, vol. 39, no. 3, pp. 275-285, 2021.

[25] SNOMED International, "SNOMED CT" 2023. [Online]. Available: https://www.snomed.org/. [Accessed 14 March 2024].

[26] WHO, "Anatomical Therapeutic Chemical Classification System" WHO Collaborating Centre for Drug Statistics Methodology, May 2023. [Online]. Available: https://www.whocc.no/. [Accessed 11 June 2023].

[27] R. Claire, J. Elvidge and D. Dawoud, "EHDEN-D2.7-Workshop for HTA community report" Zenodo (https://doi.org/10.5281/zenodo.7550919), 2023.

[28] EHDEN, "Bulgarian Database Register (BDR)" EHDEN Portal, 2023. [Online]. Available: https://portal.ehden.eu/. [Accessed 5 March 2024].

[29] E. Krastev and E. Markov, "Bulgarian Diabetes Register-mapping onto OMOP CDM (CC#10)" January 2023. [Online]. Available: EHDEN Community Calls, https://www.ehden.eu/bulgarian-diabetes-register-mapping-onto-omop-cdm-cc10/. [Accessed 15 March 2024].

[30] PCORnet, "PCORnet Common Data Model" The National Patient-Centered Clinical Research Network, 22 October 2020. [Online]. Available: https://pcornet.org/news/pcornet-common-data-model/. [Accessed 10 March 2024].

[31] Sentinel, "Sentinel Common Data Model" Sentinel Operations Center, 3 June 2022. [Online]. Available: https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model/sentinel-common-data-model. [Accessed 11 March 2024].

[32] Observational Health Data Sciences and Informatics, "OMOP Common Data Model" ohdsi.org, 2022. [Online]. Available: https://ohdsi.github.io/CommonDataModel/. [Accessed 10 March 2024].

[33] A. D. Dahlén, G. Dashi, I. Maslov, M. M. Attwood, J. Jonsson, V. Trukhan, et al., "Trends in Antidiabetic Drug Discovery: FDA Approved Drugs, New Drugs in Clinical Trials and Global Sales" Front Pharmacol., vol. 12:80754, pp. 1-16, 2022.

[34] H. Liyanage, S. Liaw, J. Jonnagaddala, W. Hinton and S. de Lusignan, "Common Data Models (CDMs) to Enhance International Big Data Analytics: A Diabetes Use Case to Compare Three CDMs" in Volume 255: Decision Support Systems and Education, Studies in Health Technology and Informatics, IOS Press, 2018, pp. 60-64.

[35] C. B. Forrest, K. M. McTigue, A. F. Hernandez, L. W. Cohen, H. Cruz, K. Haynes, et al., "PCORnet® 2020: current state, accomplishments, and future directions" Journal of clinical epidemiology, vol. 129, p. 60–67, 2021.

[36] European Medicines Agency, "Common Data Model for Europe? -Why? Which? How?" 11-12 December 2018. [Online]. Available: https://www.ema.europa.eu/en/documents/report/common-data-model-europe-why-which-how-workshop-report_en.pdf. [Accessed 10 March 2024].

[37] Sentinel Operations Center, "Sentinel Data Quality Assurance Practices" 27 February 2017. [Online]. Available: https://www.sentinelinitiative.org/sites/default/files/data/distributed-database/Sentinel_DataQAPractices_Memo.pdf. [Accessed 22 March 2024].

[38] J. Brown, J. Maro, M. Nguyen and R. Ball, "Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system" JAMIA, vol. 27, no. 5, pp. 793-797, 2020.

[39] Observational Health Data Sciences and Informatics, The Book of OHDSI, https://ohdsi.github.io/TheBookOfOhdsi/, 2021.

[40] National Scientific Program eHealth, "National Scientific Program "Electronic Healthcare in Bulgaria" (e-health)" Sofia University. Faculty of Mathematics and Informatics, 2024. [Online]. Available: https://ehealth.fmi.uni-sofia.bg/. [Accessed 10 January 2024].

[41] Observational Health Data Sciences and Informatics, "OMOP Common Data Model v5.3" 2021. [Online]. Available: https://ohdsi.github.io/CommonDataModel/cdm53.html. [Accessed 22 March 2024].

[42] National Council on Prices and Reimbursement of Medical Products, "Appendix 1. Medicinal products intended for the treatment of diseases, which are paid under the Health Insurance Act (in Bulgarian)" 2 January 2019. [Online]. Available: https://www.nhif.bg/upload/12053/Izmenenie_Pril1_PLS-NZOK_01_01_2019.xls. [Accessed 22 March 2024].

[43] L. Scalone, G. Cesana, G. Furneri, R. Ciampichini, P. Beck-Peccoz, V. Chiodini, et al, "Burden of Diabetes Mellitus Estimated with a Longitudinal Population-Based Study Using Administrative Databases" PLoS ONE, vol. 9, no. 12, pp. 1-11, 2014.

[44] E. Krastev, S. Abanos, P. Kovachev and D. Tcharaktchiev, "Diabetes Prevalence and Duration Data Extracted from Outpatient Records Representative for the Bulgarian Population" in Stud Health Technol Inform., IOS Press, 2023, pp. 230-233.

[45] statista, "Prevalence of diabetes in adult population in Europe 2021, by country" 2021. [Online]. Available: https://www.statista.com/statistics/1081006/prevalence-of-diabetes-in-europe/. [Accessed 22 March 2024].

[46] NHIF, "Report on quarterly expenses for 2018. National Health Insurance Fund (in Bulgarian)" 4 November 2019. [Online]. Available: https://www.nhif.bg/bg/medicine_food/quarter-payments/2022. [Accessed 12 March 2024].