SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3150.

SAFADI HELÉNA

Magatartástudományok

című program

Programvezető: Dr. Kovács József, egyetemi tanár

Témavezető: Dr. Belicza Éva, habilitált egyetemi docens

EXPLORING THE POTENTIAL APPLICATIONS OF BIG DATA IN DIABETES CARE IN HUNGARY

PhD thesis

Helena Safadi MD

Doctoral School of Mental Health Sciences
Semmelweis University

Supervisor: Éva Belicza PhD
Official reviewers: Imre Boncz, MD, PhD
Klára Gadó, MD, PhD

Head of the Complex Examination Committee: Attila Szijártó, MD, D.Sc.

Members of the Complex Examination Committee:
Balázs Hauser, MD, PhD
Csaba Óváry, MD, PhD

Budapest
2025

TABLE OF CONTENTS

LIST OF ABBREVATIONS

BDA	Big Data Analytics
EHRs	Electronic Health Records
EMRs	Electronic Medical Records
PROs	Patient Reported Outcomes,
PROM	Patient Reported Outcome Measure
IT	Information Technology
COVID-19	Coronavirus disease 2019
NHIFA	National Health Insurance Fund Administration
SSN	Social security number
ICD	International Classification of Diseases
EESZT	National eHealth Infrastructure
EUBIROD	European Best Information through Regional Outcomes in Diabetes
NICE	National Institute of Health and Care Excellence
ADA	American Diabetes Association
BMI	Body Mass Index
eGFR	estimated Glomerular Filtration Rate
HbA1c	Haemoglobin-A1c (glycated haemoglobin)
HDL	High-Density lipoprotein cholesterol
LDL	Low-Density lipoprotein cholesterol
ECG	Electrocardiogram
TIA	Transient Ischemic Attack
ATC	Anatomical Therapeutic Chemical
PCOS	Polycystic ovary syndrome
ICHI, OENO	International Classification of Health Interventions
INS	Insulin
nonINS	Non-insulin
METF	Metformin,
SU	Sulfonylurea
HR	Hazard ratio
ref	Reference group
RHR	Ratio of hazard ratio

ns	Non-significant
NR	Non-relevant
CI	Confidence interval
PSA	Prostate-Specific Antigen
OR	Odds ratio

1. INTRODUCTION

1.1. “Big data” and Big Data Analytics

The definition of “big data” is continuously developing. Besides mentioning the fact that traditional methods are insufficient for managing big data, it is often described by its characteristics with a set of ’V’s:

- Volume: refers to the huge amount of data it contains;
- Variety: means it includes different types, formats and sources of data;
- Velocity: desribes the high rate of data inflow and changes;
- Veracity: reflects the quality and consistency of data;
- Value: relates to the knowledge that can be gained by processing the data;
- Variability: implies the potential for scalability and extensibility of the data to be processed;
- Visualization: points to the ability to interpret data in a meaningful way;
- Valence: shows it creates the possibility of linking various databases (1–6).

Big Data Analytics (BDA) is the process of extracting meaningful insights from large, complex datasets. By employing advanced analytical methods, we can identify patterns, correlations, and trends that would be difficult to achieve using traditional data analysis techniques. These findings can be used to optimise business operations, improving decision-making, and drive innovation (7).

1.2. Big data and BDA in healthcare

BDA in healthcare has the potential to ’acquire, store, process and analyze large amounts of health data in various forms, and deliver meaningful information to users, which allows them to discover business values and insights in a timely fashion’(8).

In healthcare, there are a wide variety of data types that can be used for BDA, including clinical data, administrative data, finance data, medical imaging data, laboratory test data, which can be derived from various sources such as Electronic Health Records (EHRs), Electronic Medical Records (EMRs), Patient Reported Outcomes (PROs), pharmacy prescription records, insurance records, internet of things, mobile Health or wearable devices and ’omics’ (genomics and transcriptomics) studies (2,3,5).

The main types of BDA are (1,9):

- Descriptive analytics: which is the exploration and discovery of information in order to understand and analyze healthcare decisions, outcomes and quality, and make informed decisions;
- Predictive analytics: which predicts upcoming events based on historical or summarized health data;
- Prescriptive analytics: which usually provides decision support in case of too many existing alternatives.

BDA can therefore support quality improvement, make the work of health professionals easier, facilitate scientific and research activities, reduce costs, increase efficiency and even promote sustainability (1,10). Wang et al. classified the related benefits as Information Technology (IT) infrastructure benefits (building business flexibility for current and future changes, IT cost reduction, increased IT infrastructure capability), operational benifits (cost reduction, cycle time reduction, productivity improvement, quality improvement, customer service improvement), managerial benefits (better resource management, improved decision-making and planning, performance improvement), strategic benefits (support for business growth, support for business alliance, building for business innovations, building cost leadership, generating product differentiation, building external linkages), and organizational benefits (changing work patterns, facilitationg organizational learning, empowerment, building common vision) (11).

However, meeting these achievements is not so easy, there are many challenges to overcome (2,5,6):

- Storage and processing: The large volume of data presents obvious storage challenges, especially given that the volume is growing rapidly. Although security reasons initially supported the development of local servers, it is now cleared that cloud-based solutions with adequate security can reduce costs, fast disaster recovery and easier expansion. Hybrid systems may also be useful.
- Cleaning: To gain reliable results, data collected should be cleansed to ensure accuracy, correctness, consistency and relevancy. This is difficult to achieve for large volumes, and nowadays, automazed algorithms are used alongside manual technics.
- Unified format: To analyze data on the same characteristic or event, it is necessary that they are presented in a uniform way. Although standardized coding systems

already exist in some areas, heterogeneous representation remains a major difficulty for much information.

- Accuracy: Entering or reporting data into EHRs or EMRs is often inaccurate leading to quality issues in data analysis.
- Image pre-processing: Similar to the previous point, there may be quality issues with images, largely due to noise, artifacts and careless handling and processing of images.
- Privacy and security: Ensuring data protection and cybersecurity has become an essential need for managing any health information. Security rules and measures have been recently introduced that are essential for BDA.
- Meta-data: To support research activity, information about who, why, how and when created and analyzed data should be collected so that future studies can use the same querries, make benchmarking or build on it.
- Querying: Querying can be a challenge when different datasets cannot be linked or interconnected due to the lack of or the difference in standard coding systems.
- Visualization: Presenting the complex results of BDA in a clear, easy-to-understand, quick-to-digest way can be critical to the use of the results.
- Data sharing: Compiling big data or even tracking a patient between providers often requires linking data from different sources, institutions, therefore the sharing of data. The provision of data can harm interests, so understanding the aim of BDA is crucial for success.
- Data ownership: There can be many owners of data depending on the source. Data provision sometimes creates a complex situation, e.g. at a healthcare provider, the provider manages the patient data, but the patient is the owner of the data. Therefore, it might be an interesting question who should give consent to use the data.
- Skills requirement: BDA requires skills from data collection to interpretation of the results for their users. Unfortunately, the number of professionals is not growing as fast as it should.
- Technological requirements: In the absence of an adequate IT infrastructure, BDA is not feasible, exacerbated by the fact that access to databases is usually restricted to a very limited number of professionals.
- Limited awareness and support: Without funding and awareness, one is increasingly left behind and misses out on the opportunities offered by BDA.

- Healthcare models: It is important to have adequate evidence to demonstrate the economic benefits of a particular health investment.

1.3.Big Data and BDA in diabetes care

Diabetes care is one of the areas most frequently studied by BDA. It is partly due to the large number of people affected by diabetes. According to the International Diabetes Federation's 2021 data, the prevalence of diabetes in Europe is 9.2%, which is expected to increase to 10.4% by 2045 despite a decrease in incidence. Diabetes is responsible for 7.7% of all deaths under the age of 60, as well as 8.6% of all healthcare expenditures. (12) A great number of wearable devices have been introduced to monitor diabetes-related parameters. These are not restricted to glucose monitoring but include many other influencing factors such as physical activity, sleep or nutrition. Furthermore, information about common co-morbidities can also be collected by these devices leading to an accelerating increase in the amount of data on diabetes (13). Besides these patient-held data, EHRs, EMRs, registries, surveys, smart glucose meters, insulin pumps and automated insulin delivery systems, digital images from retinal screening, social media platforms, and environmental data provide sources for BDA (14,15). A further decisive factor in the rise of BDA in diabetes care is that indicators of care have been defined for some time and in many places, uniform registers and databases have been established (16–18).

Big data can support diabetes care in many ways (15,17,18):

- Patient self-management can be facilitated by personalized education and support. Timely, well-visualized, customized information containing lifestyle advices as well as notifications and recommendations of glucose monitoring devices can strengthen empowerment, while artificial intelligence (AI) -driven systems can even administer insulin contiuously by an autonomous closed-loop system working as an ‘artificial pancreas’;
- Image analysis mainly support retinopathy screening, however, foot monitoring and even chest radiographs can also support the diagnosis of diabetes and the management of complications;
- Clinical decision support includes complication prediction, diagnostic support, personalized treatment and prescription;

- Analysing EMR data can help optimize the workflow, the patient pathway, facilitate multidisciplinary care and measure the quality of care;
- Epidemiologic reports and surveillance systems can be developed, researches can be conducted on risk factors, comorbidities, adverse events, cost-effectiveness, etc. and show trends that form the basis of health strategies and resource allocation.
- Combining different datasets can result in national diabetes cohorts that enable both aetiological and demographic analysis. Linking health data with data from different areas, such as socioeconomic or environmental databases, can lead to even deeper analyses.

1.4.Big data and BDA in Hungary

Over the past 5 years, scientific publications on the domestic application of big data have covered a wide range of topics. Studies have looked at areas such as proteomics (19), drug safety (20), coronavirus disease 2019 (COVID-19) -related population movement monitoring (21), artificial neural networks in pharmaceutical manufacturing (22), autonomous non-technical skill assessments in minimal invasive surgery (23), methods of artificial intelligence and their application in imaging diagnostics (24), building prospective big data database (25,26) or interplay between phenotypic resistance to relevant antibiotics in Gram-negative urinary pathogens (27). Although this indicates that a new era of BDA has started in our country, yet, most of the publications are related to the analysis of different registries (28–32) and most importantly, the National Health Insurance Fund Administration (NHIFA) database. Without being exhaustive, here are some of the areas that are covered by the latter studies. Researches were carried out in the fields of pulmonology (33–36), cardiology (37–39), oncology (40–47), neurology (48–50), and psychiatry (51–53). By type, there were burden of disease studies (54), screening participation studies (55–57), epidemiological studies (33,42,44,48,58), retrospective studies (40,43,59), and survival studies (41,45,60).

For diabetes, BDA researches focused on antidiabetic treatment (61–66) incidence, prevalence, mortality (58,67,68,68–71), burden of disease and economic analyses (72,73), co-morbidities (74–77), and COVID-19 infection (78).

2. OBJECTIVES

NHIFA database has a long history of serving domestic health data analysis. However, it contains data collected for financial purposes, therefore the scope of use is limited (14). On the other hand, given the volume, scale, relevance and continuity of the database, it will certainly remain an important building block for future BDA. In order to make the most of it, it is necessary to consider what areas of research the data might be suitable for. As mentioned earlier, diabetes is one of the most frequently chosen areas of BDA and has already been the focus of many different research studies in Hungary. Its importance is indisputable. Between 2001 and 2016, the incidence of drug-treated type 2 diabetes in Hungary showed an average annual decrease of 6.46%, reaching 350.7 cases/100,000 person-years by 2016. However, prevalence continued to rise until 2011, and then showed a slight decrease after a three-year plateau, reaching 7942.6/100,000 population in 2016 (67). It is noteworthy that based on data for 2016, every fourth or fifth individual between the ages of 65 and 84 was a person with Type 2 diabetes treated by antidiabetic drug (68). On the other hand, standardized total mortality showed a significant 8.35% increase between 2001 and 2016, which was more pronounced among men and younger age groups (41-60 years) (69). Based on data for 2016, the mortality rate of people with Type 2 diabetes was higher than that of people without diabetes in women and younger age groups; more than four times the mortality rate was detected in 35-39 year old women, more than three times in men, while the hazard ratio barely exceeded 1 in both sexes among those over 85 years of age (70). Similar age-dependent risks have been shown for heart attack and stroke in people with diabetes (75). Given its significance and the extend of research conducted on it, diabetes appears to be a suitable area for exploring previously unknown or unutilized analytical possibilities of the NHIFA database.

Accordingly, our goal was to evaluate the NHIFA database from an analyzability perspective, and by testing new analytical possibilities in the field of diabetes, to identify and perform them, thus enriching the body of available results on domestic diabetes care.

Our research questions were:

1. What are the diabetes research areas, if any, where NHIFA database analysis could be a valuable tool but have not been investigated in Hungary so far?

2. What specific analytical questions can be substantively addressed within the delineated diabetes research domains?
3. What novel discoveries and recommendations can be revealed regarding domestic diabetes care based on the test analysis/analyses?
4. What general conclusions can be inferred about the NHIFA database for its application in BDA, based on the outcomes of the test analysis/analyses?
5. What prospective recommendations can be formulated for the utilization of the database?

3. METHODS

3.1. The NHIFA Database and data management

Data from five sources were provided by the NHIFA for our study: data on primary care, outpatient specialty care, hospital utilization, demographic dataset of the population with a social security number (SSN, SSN dataset), and the stock of dispensed prescriptions. All datasets include individual utilization records, with individuals initially identified by their SSN. The unique personal identifiers in the datasets were converted into depersonalized identifiers by the National Infocommunications Ltd. This process ensured the maintainability of data linkages while effectively anonymizing individuals within the datasets (79).

NHIFA collects data pertaining to all publicly financed healthcare services. This data encompasses the date of service utilization, diagnoses coded according to the 10th version of the International Classification of Diseases (ICD-10), and the tests and procedures undertaken. Notably, the datasets do not include the results of diagnostic tests. The datasets provided to us encompass service utilization records from 2010 to 2021. NHIFA registers the date of birth, sex, date of death, citizenship, and SSN validity for individuals with a Hungarian SSN in its SSN dataset. The provided datasets include records for all individuals who were alive on 1st January, 2010 (79).

The database was provided by the National Research, Development and Innovation Office in Hungary and the research was approved by the Medical Research Council under number IV-1543-1/2022-EKU. The combined database from the five datasets with depersonalized identifiers was stored on an external local server ensuring compliance with national data security regulations for research and health data. Access to the server was granted to one researcher, who extracted the required data according to the research algorithms developed and transferred it to the IBM SPSS v27.0 software for analysis.

3.2. Semi-structured interviews

Semi-stuctured individual and focused-group interviews were conducted at the beginning of our research with a total of 13 specialists including 4 diabetes expert, 4 general practitioners, 2 nephrologists, 2 anaesthesiologists and intensive care specialists, 1 neurologist, 1 ophtalmologist and 1 vascular surgeon. Diabetes experts and general

practitioners included doctors from both the capital and the countryside. The interviews focused on the following topics:

- Coding and the quality of information: diabetes diagnosis, pregnancy and conditions other than diabetes when antidiabetics can be prescribed, tests related to diabetes care, complications and co-morbidities, hospitalization;
- Diabetes care: diagnosis, status check, treatment, care and monitoring, acute and chronic complications, patient engagement, influencing factors of diabetes care, impressions on the strengths and weaknesses of care;
- COVID-19 pandemic (as it fell within the study period): effects and impact on diagnosing new cases, regular check-ups, patient flow, complication rates and patients' adherence.

The results of the interviews were used to assess the applicability of the database, formulate research objectives, design the research and explain the results.

Following the completion of the initial analyses, additional semi-structured, focused-group interviews were conducted with a narrower group of experts to ensure the rigorous interpretation of the findings.

3.3. Identifying the diabetes research area(s) to be analyzed

3.3.1. Identification of potential areas for diabetes research

In order to determine potential areas of investigation, we assessed which areas supported by the BDA for diabetes care could be explored using the available data within the database. Subsequently, we identified research gaps within the domestic literature. Through expert interviews, we refined our focus to those areas where the database data was of adequate quality and where there was a demonstrated need for further research in the field of diabetes care.

From the areas listed in chapter 1.3, patient self-management and image analysis were found to be irrelevant as neither information from or to portable devices nor images are available in the database. Those decision support or prediction studies that require genetic, molecular or patient reported data were also not feasible due to lack of such data. In comparison with the domestic research, which is described at the end of chapter 1.4, the most lacking areas were workflow, pathway and quality of care analyses. For clinical decision support, there has been some research on complication prediction, but not on

diagnostic support or personalized treatment and prescription. Epidemiologic studies have been conducted, but the scope and depth of this kind of research could be extended further.

According to the findings of the initial interviews (see Table 1), assessing the quality of care, particularly process indicators, was deemed highly pertinent. This is underlined by the lack of clarity of roles and responsibilities, lengthy waiting lists, limited availability of lifestyle therapy and patient education, a lack of multidisciplinary approach, the limited competencies of general practitioners, and the perception of specialists treating complications that patients were referred too late. While the database does not include point of care or privately conducted tests, given that the vast majority of relevant examinations were part of the publicly funded healthcare system, it remains a valuable resource for conducting meaningful analyses. Moreover, the study of complications was considered a substantial area of focus, considering that diabetes is often diagnosed only when complications arise, and the management of complications typically begins at a relatively late stage. Conversely, an analysis of diabetes complications within the NHIFA database presents significant challenges. The coding of these complications is inconsistent and often incomplete. Furthermore, most complications are not linked to specific drugs or diagnostic tests that could facilitate standardized data extraction. The only complication that appeared to be investigable in relatively detail was cancer. Prior to this research, no studies had been conducted in Hungary to examine the association between diabetes and cancer.

Table 1: The key findings of the initial semi-structured interviews

Coding and the quality of coding:
• ICD codes are unreliable to define diabetes: not every person with a diabetes code actually has diabetes (e.g. diabetes code can appear if diabetes is only suspected and then remains in the following reports, or diabetes can be coded so that certain medication can be prescribed or prediabetic conditions can be coded as diabetes), people can have diabetes without a related code (for example in primary health care data, in many cases only one diagnosis is registered even in case of multimorbidity or specialists involved in the care of diabetes complications can indicate only the complication and not the underlying diabetes) • codes are also unreliable to define the type of diabetes: the first code sticks to the patient, changes are not documented, more codes exist in parallel

- coding of complications happens occasionally and it varies how it happens, sometimes more serious condition is coded to allow prescribing newer drugs
- gestational diabetes may also be coded differently, sometimes it is coded as Type 1 diabetes and it sticks to the patient
- antidiabetic drug prescription is better to define diabetes, however it excludes those who are only on lifestyle therapy and those with polycystic ovary syndrome can take metformin without having diabetes
- tests run by general practitioners are not available in the primary health care data
- referrals often do not include that the test requested is related to diabetes care
- tests happening in private care are not seen in our database (although it has improved recently)

Diabetes care:

- there is no professional control on the coding of diabetes
- the diagnosis of diabetes often reveals when a complication appears (already late)
- roles and responsibilities in diabetes care are unclear, it is up to one's conscience to pay attention to the patient, who can easily fall out of the system
- waiting lists can be very long for the related specialities mainly for nephrology, diabetology and even laboratory care
- lifestyle therapy (including dietetics) is very elementary or even missing, personalised therapy is even more limited (the situation is worse in rural areas and more favourable in community practices)
- diabetic foot care does not have a clear organisational background
- there are no time and people for patient education
- hospital care is mainly for acute complications
- there is no real multidisciplinarity in diabetes care, providers usually do not communicate with each other and they do not or cannot follow the patients
- the National eHealth Infrastructure (EESZT) supports data-sharing, however, it is difficult to search for the results and they may be contradictory
- general practitioners' competencies in diabetes care is more limited than reasonable (for example they can only prescribe metformin and sulfonyurea) and they are incentivised to give fewer referrals
- specialties related to diabetes complications feel the patients are seen late

COVID-19 pandemic:

- not everyone noticed a decrease in the number of patients during the pandemic, and the patient flow did not really increase afterwards, nor did the number of new cases
- co-operation depended on the patients: some of them became even more cautious, some of them were afraid to seek care (in the latter case, patients were seen by a doctor in a more neglected condition)
- lifestyle-related activities deteriorated, for example more people became obese
- general use of telemedicine is the main achievement of the pandemic.

3.3.2. Setting the specific research objective(s) for assessing the quality of diabetes care in Hungary

Both incidence and prevalence, as well as various outcome indicators, are generally an integral part of regional or national level analysis of diabetes care, but by themselves are not sufficient to characterise the quality of care and identify intervention points and areas for improvement. It is no longer a question that the organisation of care should be based on IT solutions that use data generated during healthcare administration (80). Accordingly, many countries operate diabetes registries and clinical audit programs (81,82). The range of indicators used varies and covers different areas, therefore the European Best Information through Regional Outcomes in Diabetes (EUBIROD) Network has compiled a complete set of indicators as a guideline, which already allows analyses at international level. The European Diabetes Data Collection for Clinical Audit and Patient Care contains indicators for five main areas, covering the areas of epidemiology, structure, patient care process, intermediate and final outcome indicators (83). The Lancet Commission also emphasizes data-based care organisation in its study on diabetes, in which a list of recommended data content for registries is also presented. This covers medical history data, clinical examinations, laboratory results, micro- and macrovascular complications, comorbidities, oral and injectable antidiabetic drugs, and cardiovascular disease medications (84). In addition to the above, the long-standing Australian National Diabetes Audit also pays particular attention to mental health and Patient Reported Outcome Measure (PROM) surveys (85). The National Diabetes Audit, which has been operating since 2003 and covers England and Wales, focuses on 5 key questions: is every person with diabetes diagnosed and documented in the registry, what proportion of diagnosed patients receive the nine key processes of diabetes care as defined by the National Institute of Health and Care Excellence (NICE), what proportion of them achieve blood sugar, blood pressure and cardiovascular risk management targets, what proportion of them are offered and participate in structured patient education, and how do the rates of acute and long-term complications change (86)?

In Hungary, the regular activities and audit criteria for diabetes care are regulated by the current professional guidelines. The Hungarian clinical guideline for the diagnosis, antihyperglycemic treatment and care of diabetes mellitus in adults (87) in this regard incorporate a consolidated set of recommendations from the most recent 2024 Diabetes

Care Standards of the American Diabetes Association (ADA) (88). Table 2 summarizes the activities required to be performed at least annually or at every medical consultation, as stipulated by the current national guidelines. The requirements have remained consistent in the domestic guidelines since their introduction in 2009 (89–91). Table 3 details the guideline recommendations concerning quality indicators of care.

Table 2: Activities required to be performed at least annually or at every medical consultation in diabetes care, according to the current national guidelines (87)

Activities to be performed at least once a year:
• Complete physical examination: measurement of weight, height, waist circumference, blood pressure, special examination of the feet, determination of BMI • Examination of the fundus • Complete laboratory examination: HbA1c, fasting blood glucose (and postprandial if necessary), serum total cholesterol, HDL cholesterol, LDL cholesterol (measured or estimated), triglycerides, creatinine, eGFR, urine analysis (urine glucose and acetone, sediment, urine culture if necessary and quantitative albuminuria [microalbuminuria]) • Review of treatment • Checking of self-monitoring technique • Review of dietary and nutritional knowledge
In addition, the following activities should be performed at each medical consultation:
• Patient education; • Informing about blood glucose measurement data; • Assessment and reinforcement of therapy adherence.

BMI = Body Mass Index; eGFR = estimated Glomerular Filtration Rate; HbA1c = Haemoglobin-A1c (glycated haemoglobin); HDL = High-Density lipoprotein cholesterol; LDL = Low-Density lipoprotein cholesterol

A comparison of the available administrative data revealed that only a limited set of potential indicators – specifically the completion of annual laboratory tests – could be analyzed. Therapeutic target indicators could not be constructed from the existing data due to the absence of laboratory test results. The completion of physical examinations, patient education, lifestyle counseling, and therapy follow-up is not included in the data transmitted to the NHIFA. The database did not allow for the specific identification of funduscopic examinations, and the exclusion of ophtalmological examinations completed in private care would have significantly biased the results.

Consequently, the ultimate objective of this research direction became the analysis of laboratory tests among adult people with diabetes in Hungary based on the NHIFA database.

Table 3: Recommendations concerning quality indicators of care, according to the current national guidelines (87)

- Number of HbA1c determinations per year (desired: 4 measurements per year for insulin-treated patients, 2 measurements per year for non-insulin-treated patients)
- Proportion of insulin-treated patients who perform self-monitoring of blood glucose (desired: 90%)
- Complete laboratory testing (including eGFR and microalbuminuria) once a year (desired: 100%)
- Funduscopic examination once a year (desired: 90%)
- Neuropathy examination (tuning fork), foot inspection once a year (desired: 100%)
- ECG once a year (desired: 90%)
- Physical examination (anthropometric parameters), blood pressure measurement once a year (desired: 100%)
- Ankle-brachial index determination in Type 2 diabetes, over the age of 50, or in the event of another cardiovascular event (stroke, TIA, infarction, angina pectoris) regardless of age; in Type 1 diabetes, after 10 years of disease duration or in the event of another cardiovascular event (stroke, TIA, infarction, angina pectoris) regardless of diabetes duration (desired 80%)
- Provision of basic dietary therapy knowledge to newly diagnosed diabetics (desired: 100%)

HbA1c = Haemoglobin-A1c (glycated haemoglobin); eGFR = estimated Glomerular Filtration Rate; ECG= Electrocardiogram; TIA = Transient Ischemic Attack

3.3.3. Setting the specific research objective(s) to investigate the association between diabetes and cancer in Hungary

Diabetes is characterized by a range of complications, including micro- and macrovascular complications such as retinopathy, nephropathy neuropathy, and cardiovascular diseases (87). Furthermore, diabetes is associated with a broader spectrum of health issues, including impaired bone health, cognitive impairement, cancer, erectile dysfunction, non-alcoholic fatty liver disease and steatohepatitis, pancreatitis, periodontal diseases, sensory impairment, and obstructive sleep apnoe (88).

While there are shared risk factors between diabetes and cancer, this alone does not fully explain their coexistence. Numerous studies have unequivocally demonstrated that

diabetes constitutes an independent risk factor for the development of various cancers (92–94).

A meta-analysis including 151 cohorts sought to assess the causal relationship between diabetes and cancer. Strong causal associations were observed between Type 2 diabetes and the incidence of liver, pancreatic, and endometrial cancers, as well as pancreatic cancer mortality. A likely to be causal association was found with gallbladder cancer incidence. In contrast, the associations with kidney, colorectal, and thyroid cancer incidence were less robust. Finally, the analysis revealed that the association between Type 2 diabetes and leukaemia, prostate, breast, bladder, stomach, ovarian, non-Hodgkin lymphoma, melanoma, lung, or esophageal cancer is unlikely to be causal. Interestingly, Type 2 diabetes was even associated with a decreased risk of prostate cancer (RR: 0.83; 0.79, 0.88) (95).

For Type 1 diabetes, the evidence is less conclusive. While some studies within the review by Zhu et al. indicated an increased incidence of liver, pancreas, kidney, esophageal, stomach, lung, thyroid, squamous cell carcinoma, and leukaemia in individuals with Type 1 diabetes, other studies found no significant associations (96).

Contradictory findings were also observed when analysing the association between sex-related cancers and diabetes (96).

However, a meta-analysis of 121 cohorts revealed that females with any type of diabetes exhibit an approximately 6% higher overall cancer risk compared to males with diabetes. This disparity varies by cancer site, with females demonstrating a greater risk of oral, stomach, and kidney cancer, and leukaemia, but a lower risk of liver cancer (97).

Numerous studies have delved into the underlying mechanisms linking diabetes and cancer. These include genetic research (96,98,99), studies on shared risk factors such as obesity, inflammation, hyperglycaemia or hyperinsulinaemia, sex hormones, and their associated molecular patways (94,96,99–103), as well as research examining the role of antidiabetic medications (96,99,102–104) and nutrition factors (99).

Until recently, the temporal relationship between diabetes diagnosis and the onset of cancer remained relatively understudied. Lega et al. demonstrated a significantly elevated risk of most cancers in people with diabetes within a decade before and immediately after a diabetes diagnosis (105). Johnson et al. corroborated these findings, observing an increased cancer risk across various sites shortly after the diabetes onset, with only

colorectal, liver, endometrial, and pancreatic cancer risks remaining elevated in later periods (106). Furthermore, Carstensen et al. found that the rate ratio for all cancers peaked immediately following the diagnosis of diabetes, irrespective of insulin use (107). In contrast, Hu et al. observed that when comparing cohorts with and without diabetes, the hazard ratio (HR) for cancer was highest approximately eight years after the diagnosis of Type 2 diabetes, with a similar pattern observed across different cancer sites (108).
In the light of these findings, we have narrowed down the research focus in this field to two key areas. The NHIFA database did not provide sufficient data to explore the molecular mechanisms underlying the relationship between diabetes and cancer, while the absence of data on influencing factors hindered the ability to analyze the complexity of the relationship in more depth. However, we could still aim to determine the risk of developing cancer in patients with diabetes compared to the non-diabetic population. Secondly, we focused on examining how the time of cancer diagnosis is related to the time of diabetes diagnosis.

3.4. Identifying people with diabetes in the database

Diabetes was identified based on antidiabetic drug dispensing as diagnosis codes appeared to be unreliable (see Table 1). Drugs with Anatomical Therapeutic Chemical (ATC) code A10A, A10B were considered antidiabetic drugs. A significant number of people with diabetes diagnosis codes may not actually live with diabetes, for example people with the presence of pre-diabetic conditions, for which no specific ICD diagnosis code is available, are coded for people with diabetes. On the other hand, data on people with both diabetes and other comorbidities may only include the latter diagnosis code, especially in primary care database. Furthermore, there is no professional oversight of diagnoses, and the coding of different types of diabetes including gestational diabetes, is also inconsistent. However, antidiabetic drug dispensing necessitated the exclusion of cases where antidiabetic drugs were described without diabetes. Therefore, people with polycystic ovary syndrome (PCOS, ICD-10: E2820) documented in their primary care or specialty care records at any time during the study period were excluded. So were those with pregnancy-related events (ICD-10: "O" category and Z31-Z37), because in the case of gestational diabetes, the presence of diabetes is mostly transient, and the details of the care process differ from the general care recommendations. Another limitation of using

antidiabetic drug dispension for identifying people with diabetes is that those who receive only lifestyle therapy are also not included in the analyses. However, the primary goal in identifying diabetic patients was not to determine the incidence of diabetes, but to create cohorts of patients with diabetes who meet the criteria for diabetes care and cancer risk analyses (79,109).

3.5. Methods of the analysis of the specific researches identified

3.5.1. Analysis of laboratory tests among adult people with diabetes

We conducted a retrospective cohort study to assess the laboratory tests of patients with diabetes. We selected four cohorts for the observation of care practices. The first analyzable cohort consisted of patients who collected their antidiabetic drug for the first time in 2014. The rationale for this was that our data were available from 2010 onwards, thus first dispensing in the database may not reflect real first dispensing due to lack of data from previous years. Since the number of new antidiabetic drug dispensing decreased significantly each year until 2013 (i.e., those who had not collected antidiabetic drug back to 2010), it could be assumed that a significant proportion of patients in 2014 were indeed newly diagnosed people with diabetes. People with diabetes identified in this way were divided into subgroups based on age groups, sex, geographic location, and the type of antidiabetic drug initiated. We chose a four-year observation period, allowing us to follow patients who first collected antidiabetic drug in 2014 until 2017, and those in the fourth cohort, who first collected antidiabetic drug in 2017 until the end of 2021. Follow-up period was calculated as the number of years since the first antidiabetic drug dispensing date. Inclusion criteria were Hungarian citizenship, a valid SSN, age at inclusion date of at least 40 years and below 90 years. People with diabetes who died during the four-year follow-up period were excluded from the study. Analyses revealed that 19.6% of patients aged 40-89 years who were followed for four years only collected antidiabetic drug once during the follow-up period. Therefore, these patients were excluded from the present study, and the analyses were restricted to the population of patients who collected antidiabetic drug at least twice (Table 4) (109).

Table 4: Inclusion and exclusion criteria for laboratory tests and cancer risk analyses (109)

Analysis of laboratory tests among adult people with diabetes	Analysis of associations between diabetes and cancer
Inclusion criteria	Inclusion criteria
Hungarian citizenship with valid SSN	Hungarian citizenship with valid SSN
County of residence is known	-
First antidiabetic drug dispensing between 2014 and 2017	First antidiabetic drug dispensing in 2014 or 2015
Age of 40-89 at the first antidiabetic medication dispensing	Age of 40-89 at the first antidiabetic medication dispensing
Exclusion criteria	Exclusion criteria
PCOS is registered in the dataset	PCOS is registered in the dataset
Pregnancy-obstetric code is registered in the dataset	Pregnancy-obstetric code is registered in the dataset
Those who died during the four-year follow-up period	Those who had recorded cancer diagnosis of the studied sites back until 2010 (only for cancer rate analysis)
Those who only collected antidiabetic drug once during the four-year follow-up period	-

SSN = Social Security Number; PCOS = polycystic ovary syndrome

Among those who met the inclusion and exclusion criteria, we examined whether the laboratory tests specified in the valid clinical guideline (87) – HbA1c, blood glucose, LDL and HDL cholesterol, triglycerides, creatinine, and urine albumine, glucose and ketones – were performed each year following the first drug dispensing and in the three months preceeding the first drug dispensing. Due to the low number of LDL tests, we also included total cholesterol tests in the analyses. Laboratory tests were defined by their International Classification of Health Interventions (ICHI, OENO) codes used in Hungary (Table 5) (109).

Table 5: ICHI (OENO) codes of laboratory tests used in the analysis of laboratory tests among adult people with diabetes in Hungary (109)

Laboratory test	**ICHI code**
HbA1c	28493, 28494
Blood glucose	21310, 21312, 42144, 42145
LDL cholesterol	21422, 42148
HDL cholesterol	2142A, 42149
Total cholesterol	21420, 42146
Triglycerides	21411,42147
Serum creatinine	21141,21143,42162,42164
Urine albumine	22042
Urine glucose	22200, 22201, 22550
Urine ketones	22400, 22550

HbA1c = Haemoglobin-A1c (glycated haemoglobin); LDL = Low-Density lipoprotein cholesterol; HDL = High-Density lipoprotein cholesterol; ICHI = International Classification of Health Interventions

Table 6: Groups of people with diabetes according to drug dispensing during the follow-up period in the analysis of laboratory tests among adult people with diabetes in Hungary (109)

Name of group*	**Explanation**
INS-INS	People with diabetes who only collected insulin
INS-nonINS	People with diabetes who first collected insulin, later also collected nonINS medication
nonINS-INS	People with diabetes who first collected nonINS medication, later also collected insulin
METF-METF	People with diabetes who only collected metformin
SU-SU	People with diabetes who only collected sulfonylurea
METF-SU	People with diabetes who first collected metformin and later also collected sulfonylurea and no other antidiabetic drugs
SU-METF	People with diabetes who first collected sulfonylurea and later also collected metformin and no other antidiabetic drugs
METF-nonINS	People with diabetes who first collected metformin and later also collected other nonINS medication and are not included in group METF-SU
nonINS_other	People with diabetes who only collected nonINS medication and are not included in any of the groups METF-METF, METF-nonINS, SU-METF, SU-SU, METF-SU

*INS = insulin; nonINS = non-insulin; METF = metformin, SU = sulfonylurea
ATC codes: INS = A10A; nonINS = A10B; METF = A10BA02; SU = A10BB
ATC codes = Anatomical Therapeutic Chemical code

Based on drug dispensing, people with diabetes were grouped according to the ATC group of antidiabetic drugs they collected during the four-year follow-up period (Table 6) (109).

Descriptive statistical analysis was used to characterise the performance of laboratory tests. As part of this, we reviewed the proportion of laboratory tests performed in the three months preceding the first drug dispensing and during the four-year follow-up period. We calculated the percentage of patients who did not undergo any laboratory tests over this period. We also focused on tests performed in at least 3 different years over the four years considering that the next year's follow-up test may be performed in the same year as the first drug dispensing (for example on day 364 instead of day 366) (109).

Multivariate logistic regression analyis was used to examine the factors influencing the performance of laboratory tests in at least three different years. Due to the large sample size, the significance level was set at $P<0.001$. Independent variables included the sex, age at first drug dispensing and county of residence of the patient with diabetes, the year of first antidiabetic drug dispensing, group based on ATC codes of antidiabetic drugs dispensed during the four years, whether laboratory tests were performed in the three months preceeding the first antidiabetic drug dispensing, and the number of different years in which the patient collected antidiabetic drug during the four-year follow-up period (109).

3.5.2. Analysis of associations between diabetes and cancer

A retrospective cohort study was conducted utilizing the database to investigate the role of diabetes in cancer development. Inclusion criteria encompassed Hungarian citizenship, a valid SSN, and an age at inclusion date of at least 40 years and below 90 years. Individuals over the age of 40 were included in the study because the incidence of cancer before this age is very low. People aged 90 and over were excluded because the population size in this age group was very low. People with diabetes were identified based on the collection of antidiabetic drugs dispensed. To ensure accurate identification of incident diabetes cases, the study was restricted to patients whose first antidiabetic drug collection occurred in 2014 or 2015 (see Table 4). Since prescription records were available from 2010 onwards, we were able to verify that these patients had not received any prior antidiabetic drugs between 2010 and 2013. This approach, as in case of the

analysis of laboratory tests, allowed for a high degree of confidence that diabetes diagnosis occurred in 2014 or 2015. The date of the first dispensing was considered as the date of initial diabetes diagnosis. For individuals within the diabetes cohort, this date served as the inclusion date for the study. For people without diabetes (control group), the inclusion date was set as 1st January, 2014. As mentioned earlier, individuals with a pregnancy-obstetric code or a diagnosis of PCOS documented in their primary care or specialty care records at any time during the study period were also excluded. Finally, for cancer rate analysis only, individuals with recorded cancer diagnosis at the studied sites prior to their inclusion date (dating back to 2010) were excluded from the analysis, too (see Table 4) (79).

Six cancer sites were chosen to explore the relationship between the two diseases. We wanted to study cancer sites that were sufficiently presumed to be associated with diabetes (pancreas and liver), or which are relatively common, mostly primary cancer localisations – for accuracy of diagnosis – most of which are screenable and some of which are sex-related (colorectal, kidney, breast and prostate). The date of cancer diagnosis was determined as the date of the first relevant ICD code recorded in outpatient or inpatient specialty care. The analysis encompassed all ICD codes for cancer and specifically focused on six cancer sites (colorectal: C18-C21, liver: C22, pancreatic: C25, breast (female): C50, prostate: C61, and kidney cancer: C64).

Time to cancer diagnosis was calculated from the inclusion date, defined as the date of the first antidiabetic drug dispensing for people with diabetes and 1st January, 2014, for the control group. Study individuals were followed until death or the end of the study period, 31st December, 2021, using both administrative and mortality data (79).

In cancer rate analysis, the focus was on identifying the first diagnosis of cancers at various sites occuring after the inclusion date. Descriptive statistics were used to characterize the study population. Time to cancer diagnosis (for all cancer, and by cancer site) was analyzed using univariate Cox regression, comparing the group with diabetes to the control group. Analyses were stratified by age group (40-54 years, 55-69 years, and 70-89 years) and sex. Results were presented in a Forest plot. Multivariate Cox regression analyses were conducted to investigate time to cancer diagnosis (all cancer, and by cancer site) using diabetes status, age (as continuous variable), and sex as independent variables. Hazard ratio (HR) values for each cancer site in people with diabetes were determined,

stratified by sex and age, using the Cox proportional hazard regression model. Interactions between diabetes status and age, and between diabetes status and sex, were included in the model when statistically significant. Other potential influencing factors, such as obesity or smoking, were not included in the analyses due to lack of related data. Due to the large sample size, a significance level of 1% was adopted (consistently, 99% confidence intervals are presented). However, for interactions within the multivariate Cox model, a significance level of 5% was used (79).

For analysing the timely relationship between diabetes and cancer, the time to first cancer diagnosis was investigated both retrospectively and prospectively with respect to the inclusion date. We examined the yearly incidence of cancer diagnosis from three years before the inclusion date to six years after the inclusion date. The retrospective period was limited to three years due to data availablility constraints prior to 2011. The prospective period extended to six years, encompassing the maximum follow-up time for patients who collected prescribed antidiabetic drugs on 31st December, 2015. For the retrospective period, the denominator of cancer incidence rates included all study individuals. For the prospective period, the denominator was adjusted at each timepoint to include only those individuals who were still at risk, excluding those who had died or been diagnosed with cancer prior to that time point (79).

4. RESULTS

4.1.Results of the analysis of laboratory tests among adult people with diabetes

A total of 128,115 people with diabetes met the inclusion and exclusion criteria of the analysis laboratory tests among adult people with diabetes. Those who could be followed for four years from the first dispensing of an antidiabetic drug were primarily from the 60-69 age group (34.3%), but the 50-59 age group was also well represented (28.4%). The number of women was slightly higher, which can also be explained by the demographic composition of the basic population. It is noteworthy that the male population dominates the 40-49 age group, with a proportion of 61.8%. This may also be related to the fact that our study excluded women with a PCOS or pregnancy code and this could reduce the number of this female population (see Table 7) (109).

Table 7: Number, age distribution, and proportion of women among people with diabetes followed for at least four years and who filled at least two antidiabetic drug prescriptions (109)

First dispensing of an antidiabetic drug	Age (year)	Number of patients (people)			Age distribution (%)	Proportion of females (%)
		Male	Female	All		
2014	40-49	3 112	2 027	5 139	15.1	39.4
	50-59	5 238	5 046	10 284	30.3	49.1
	60-69	5 287	5 996	11 283	33.2	53.1
	70-79	2 142	3 666	5 808	17.1	63.1
	80-89	456	967	1 423	4.2	68.0
	Total	16 235	17 702	33 937	100.0	52.2
2015	40-49	3 084	1 940	5 024	15.9	38.6
	50-59	4 524	4 374	8 898	28.1	49.2
	60-69	4 939	5 799	10 738	33.9	54.0
	70-79	2 131	3 600	5 731	18.1	62.8
	80-89	358	881	1 239	3.9	71.1
	Total	15 036	16 594	31 630	100.0	52.5
2016	40-49	3 261	1 952	5 213	16.2	37.4
	50-59	4 516	4 236	8 752	27.3	48.4
	60-69	5 229	6 061	11 290	35.2	53.7

	70-79	2 131	3 532	5 663	17.6	62.4
	80-89	376	818	1 194	3.7	68.5
	Total	15 513	16 599	32 112	100.0	51.7
2017	40-49	3 291	1 953	5 244	17.2	37.2
	50-59	4 409	4 026	8 435	27.7	47.7
	60-69	4 973	5 624	10 597	34.8	53.1
	70-79	1 962	3 123	5 085	16.7	61.4
	80-89	312	763	1 075	3.5	71.0
	Total	14 947	15 489	30 436	100.0	50.9
Total	40-49	12 748	7 872	20 620	16.1	38.2
	50-59	18 687	17 682	36 369	28.4	48.6
	60-69	20 428	23 480	43 908	34.3	53.5
	70-79	8 366	13 921	22 287	17.4	62.5
	80-89	1 502	3 429	4 931	3.8	69.5
	Total	61 731	66 384	128 115	100.0	51.8

Figure 1 shows the distribution of patients across medication dispensing groups. It can be seen that the proportion of patients taking metformin and/or sulfonylurea is slightly above 50% in all cohorts (Figure 1) (109).

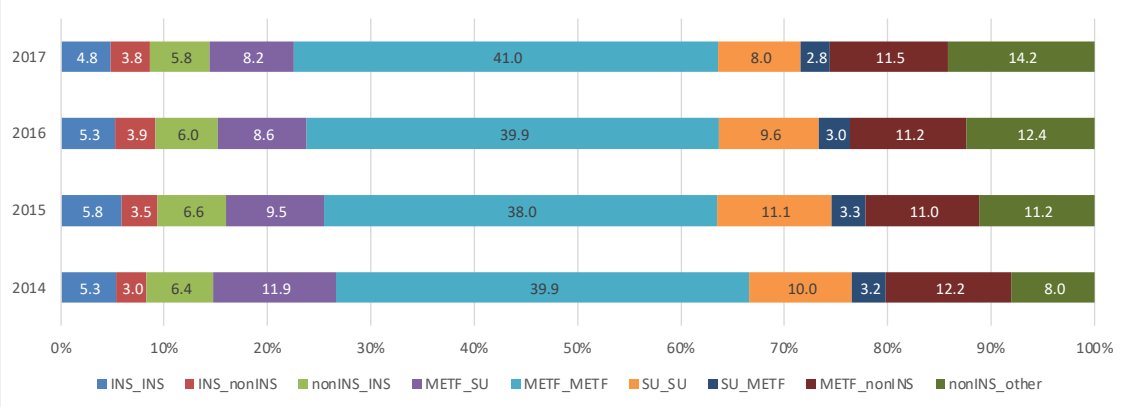


Figure 1: Distribution of included patients by medication dispensing groups based on the four years following the first drug dispensing by cohort (109)

INS = insulin; nonINS = non-insulin; METF = metformin, SU = sulfonylurea
ATC codes: INS = A10A; nonINS = A10B; METF = A10BA02; SU = A10BB
ATC codes = Anatomical Therapeutic Chemical code

Glucose testing was the most common type of test performed within the 3 months prior to the first drug dispensing (>75%), while other tests were performed less frequently. This was especially true for LDL cholesterol and urine albumin tests, which were performed with a frequency of less than 10%. The proportions were the same across the four cohorts, with the exception of HbA1c and LDL cholesterol. The use of HbA1c showed a clear increase, while the rate of LDL cholesterol testing, which was already low, decreased (Figure 2) (109).

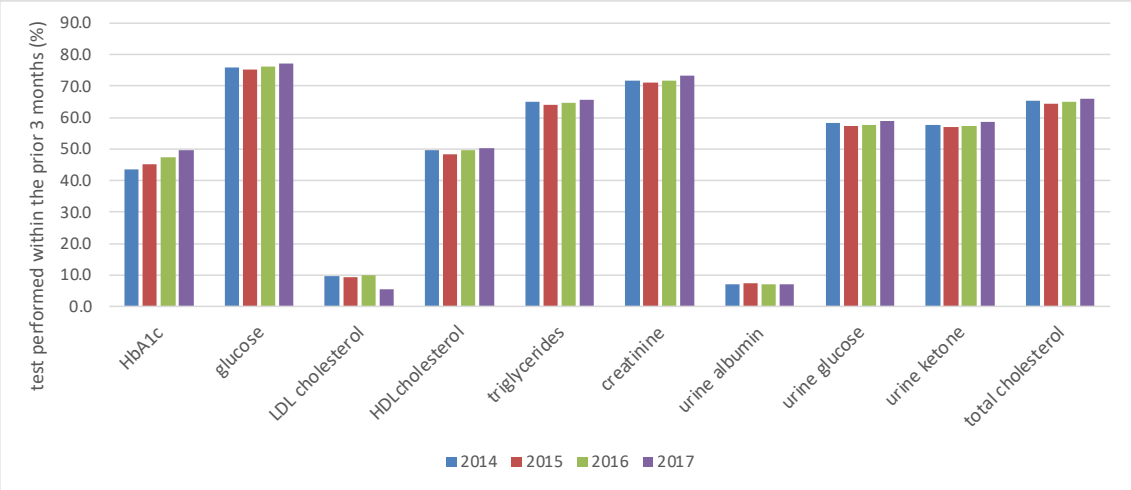


Figure 2: Proportion of laboratory tests performed within three months prior to the first dispensing of an antidiabetic drug by cohort (109)

During the four-year follow-up period, two laboratory tests – LDL cholesterol and urine albumin tests – were not performed at all in a very high proportion (>70%) of people with diabetes. In addition, there was a significant increase in the proportion of patients who did not have LDL cholesterol tests in the pandemic-affected cohorts of 2016 and 2017. The missing rates for the other tests were 15% or less and did not change significantly between the different cohorts (Figure 3) (109).

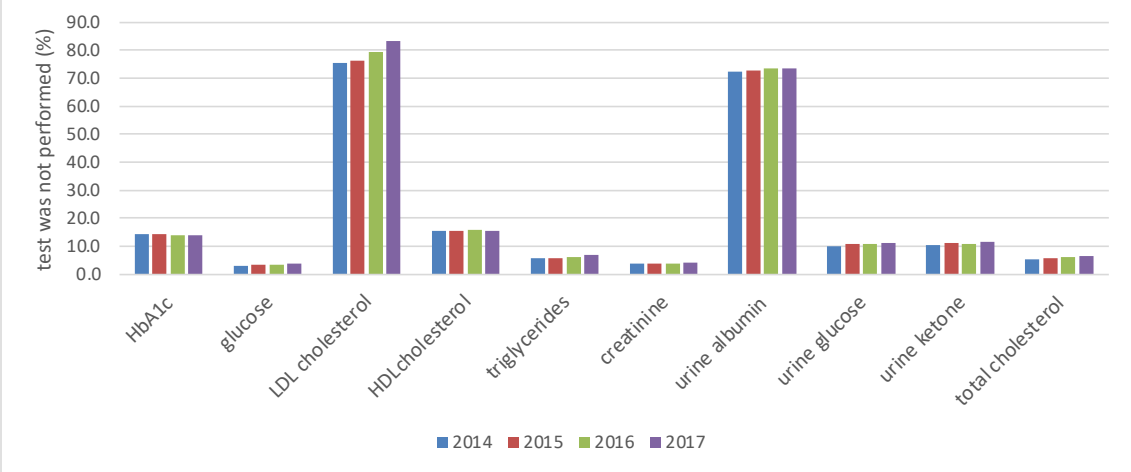


Figure 3: Proportion of people with diabetes with no laboratory tests in the four years following the first dispensing of an antidiabetic drug by cohort (109)

The occurrence of annual tests was assessed by the number of tests performed in at least three different years during the four-year follow-up period. Among laboratory tests, blood glucose and serum creatinine tests were the most frequently performed, occurring at least once in three or four different years in slightly over 70% of patients. HbA1c, urine glucose, and urine ketone tests were performed annually in 50-60% of patients. The rates of LDL cholesterol and urine albumin tests were very low. The impact of the pandemic is clearly visible in the 2016 and 2017 cohorts. The end of their four-year follow-up period is in 2020 and 2021, respectively, which may explain the decrease in test rates among them (Figure 4) (109).

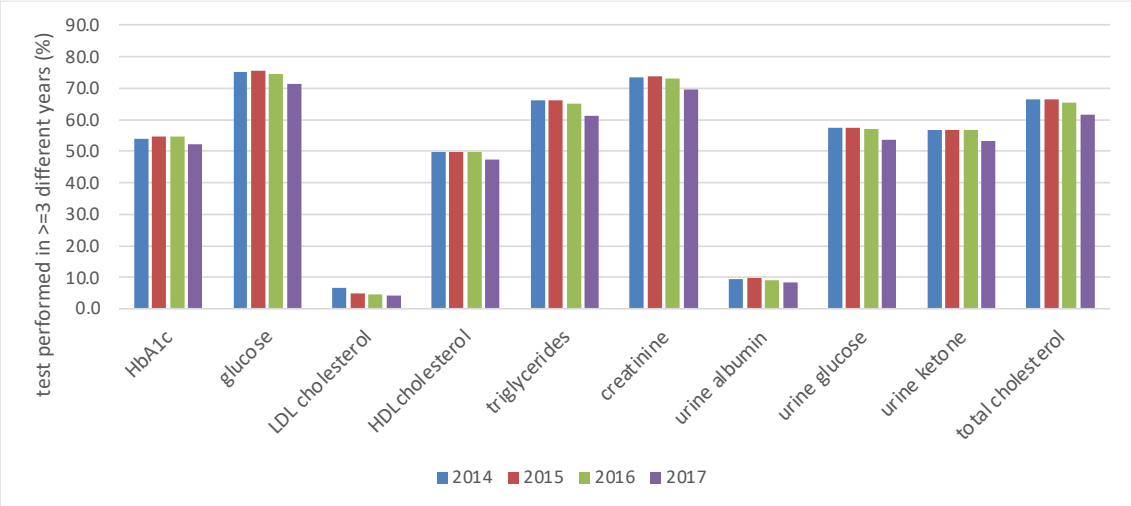


Figure 4: Proportion of people with diabetes who had laboratory tests performed in three or four different years during the four-year follow-up period by cohort (109)

We examined the drug prescription fulfillment habit of the people with diabetes included in the study. We found that approximately 13-14% of them who dispensed at least 2 prescriptions did so only in the first year. The proportion of people with diabetes who dispensed medications in all four years of follow-up showed a slight increase across the four cohorts and reached nearly 70% during the pandemic years (109).

When determining the factors influencing the performance of laboratory tests, we considered as the outcome variable those who had tests in at least 3 different years during the four-year follow-up (109).

Women were 8.1% (urine albumin) to 28.9% (blood glucose) more likely to have various laboratory tests performed. By age group, the analyses showed that the youngest age group studied had significantly worse odds than the other age groups, with the exception of three test types. For these three laboratory tests (HbA1c, LDL cholesterol, urine albumin), the oldest group had the lowest frequency of perfomance of laboratory tests, but the 40-49-year-olds also had worse rates in these cases compared to the 50-79-year-olds (Table 8) (109).

The results by county show that laboratory tests are generally performed at a higher rate in the three rural counties with universites. The weakest results are seen in Heves, Komárom, Szabolcs-Szatmár-Bereg and Somogy counties, where most laboratory tests are significantly less frequent compared to the reference county, Bács-Kiskun (Table 8) (109).

We found significant differences in the performance of laboratory tests based on the type of antidiabetic drugs people with diabetes were taking. The most favorable chances of testing compared to insulin-only therapy were in the insulin (INS) -non-insulin (nonINS), nonINS-INS, and metformin (METF) -nonINS groups, where the chances of having each test performed were significantly more favorable. There was a significantly lower chance of having tests performed if the person with diabetes was only taking metformin and/or only sulfonylurea. (Table 8). In these groups, we see significantly less favorable values for all laboratory tests compared to the reference group (insulin only) (109).

The chances of having tests performed decreased with increasing calendar year of first dispensing of an antidiabetic drug. This may be related to the COVID-19 pandemic in 2020 for the 2016 cohort and the 2020-21 pandemic years for the 2017 cohort. The Nagelkerke value, which indicates the strength of the model, was weak for all laboratory

tests, showing that other factors not included in our study had a significant impact on the performance of the tests (see Table 8) (109).

Table 8: Results of logistic regression analysis of factors influencing the performance of laboratory tests in at least three different years in the 2014-2017 diabetes cohorts with people with diabetes aged 40-89 years who who filled at least two antidiabetic drug prescriptions and were followed for at least four years (109)

Independent variables		Laboratory tests									
		HbA1c	LDL choles-terol	HDL choles-terol	Total choles-terol	Trigly-cerids	Crea-tinine	Glucose	Urine albu-min	Urine glucose	Urine ketone
	Sex (ref: male)	**1.128**	1.058	**1.161**	**1.243**	**1.239**	**1.298**	**1.289**	**1.081**	**1.208**	**1.206**
Age (year)	40-49 (ref.)										
	50-59	**1.111**	1.059	**1.153**	**1.235**	**1.233**	**1.303**	**1.283**	1.074	**1.303**	**1.301**
	60-69	**1.221**	1.033	**1.321**	**1.496**	**1.487**	**1.638**	**1.608**	**1.158**	**1.644**	**1.635**
	70-79	**1.119**	0.997	**1.365**	**1.543**	**1.527**	**1.848**	**1.772**	1.064	**1.725**	**1.727**
	80-89	**0.830**	**0.691**	1.014	1.068	1.062	**1.373**	**1.286**	**0.662**	**1.177**	**1.184**
County of residence	Bács-Kiskun (ref)										
	Baranya	**1.155**	1.162	1.103	0.991	0.997	1.130	1.017	**2.215**	**1.304**	**1.285**
	Békés	**0.830**	1.199	1.128	0.912	0.916	**0.844**	**0.802**	**2.233**	**0.637**	**0.635**
	Borsod-Abaúj-Zemplén	0.992	**0.300**	**0.543**	0.909	0.923	**0.844**	0.892	0.846	**1.197**	**1.173**
	Budapest	1.072	**0.750**	**1.309**	**1.223**	**1.245**	1.033	1.002	1.010	**1.292**	**1.282**
	Csongrád-Csanád	**1.229**	1.281	**1.251**	**1.256**	**1.273**	1.056	0.992	**1.540**	1.143	1.040
	Fejér	0.953	**1.701**	**0.756**	0.902	0.896	**0.827**	**0.753**	**2.273**	**1.235**	**1.232**
	Győr-Moson-Sopron	**1.166**	**0.665**	**1.360**	0.966	0.986	**0.849**	**0.817**	0.834	1.011	1.011
	Hajdú-Bihar	**1.336**	0.858	1.106	**1.177**	**1.188**	1.089	1.027	**3.385**	1.049	1.049
	Heves	**0.851**	**0.487**	0.877	**0.832**	**0.832**	0.857	0.893	1.046	**0.793**	**0.792**
	Jász-Nagykun-Szolnok	0.999	**0.225**	**0.704**	0.922	0.931	0.860	0.877	0.967	1.080	1.081
	Komárom	**0.723**	0.978	**0.789**	**0.764**	**0.746**	**0.720**	**0.714**	1.224	0.982	0.980
	Nógrád	**0.581**	0.984	**0.589**	**0.740**	**0.747**	**0.663**	**0.638**	**1.460**	1.082	0.984
	Pest	0.897	**0.603**	1.054	1.034	1.049	**0.877**	**0.853**	**1.273**	**1.181**	**1.124**
	Somogy	0.970	**0.210**	**0.427**	0.878	**0.819**	**0.849**	**0.796**	**0.599**	1.062	1.058
	Szabolcs-Szatmár-Bereg	1.023	**0.138**	**0.855**	1.042	1.063	1.018	0.932	1.118	**0.791**	**0.794**
	Tolna	**0.615**	**0.595**	1.139	1.009	1.036	0.936	**0.802**	**3.529**	**1.223**	1.037
	Vas	**1.342**	**0.124**	**1.344**	**1.274**	**1.282**	1.156	1.120	**2.680**	**1.287**	**1.288**
	Veszprém	0.891	**1.510**	**0.860**	**0.831**	**0.832**	**0.829**	**0.810**	**2.072**	**1.186**	1.106
	Zala	0.887	**2.304**	**1.291**	0.914	0.917	0.848	**0.820**	0.884	1.057	1.056

Antidiabetic drug group during the 4-year follow-up	INS-INS (ref)										
	INS-nonINS	**1.880**	1.266	**1.316**	**1.504**	**1.498**	**1.560**	**1.654**	**1.237**	**1.321**	**1.316**
	nonINS-INS	**2.219**	**1.405**	**1.328**	**1.620**	**1.622**	**1.901**	**2.143**	1.162	**1.596**	**1.598**
	METF-SU	**0.570**	**0.308**	**0.532**	**0.535**	**0.539**	**0.552**	**0.568**	**0.298**	**0.662**	**0.681**
	METF-METF	**0.613**	**0.384**	**0.594**	**0.599**	**0.601**	**0.608**	**0.601**	**0.348**	**0.730**	**0.749**
	SU-SU	**0.423**	**0.405**	**0.520**	**0.518**	**0.520**	**0.533**	**0.522**	**0.316**	**0.651**	**0.669**
	SU-METF	**0.495**	**0.380**	**0.539**	**0.526**	**0.533**	**0.538**	**0.552**	**0.263**	**0.659**	**0.678**
	METF-nonINS	**2.311**	0.943	**1.392**	**1.706**	**1.712**	**1.794**	**1.994**	0.998	**1.633**	**1.632**
	nonINS_other	**1.661**	1.017	**1.297**	**1.536**	**1.537**	**1.627**	**1.669**	0.950	**1.482**	**1.479**
Year of first antidiabetic drug dispensing	2014 (ref)										
	2015	**0.866**	**0.649**	0.961	**0.916**	**0.915**	**0.931**	**0.915**	0.928	**0.941**	0.947
	2016	**0.825**	**0.584**	**0.931**	**0.860**	**0.864**	**0.884**	**0.854**	**0.864**	**0.917**	**0.929**
	2017	**0.699**	**0.590**	**0.819**	**0.701**	**0.704**	**0.730**	**0.695**	**0.755**	**0.763**	**0.780**
Prior laboratory test	adott laborvizsgálat volt a megelőző 3 hónapban (ref: nem volt)	**2.237**	**7.509**	**2.690**	**1.742**	**1.749**	**1.762**	**2.124**	**6.859**	**2.341**	**2.359**
Number of different years when antidiabetic drug dispensing occured during the 4-year follow-up	1 (ref)										
	2	1.070	0.919	0.987	1.028	1.031	1.045	1.041	0.956	0.993	0.988
	3	**1.819**	1.077	**1.166**	**1.269**	**1.270**	**1.314**	**1.350**	**1.236**	**1.230**	**1.222**
	4	**3.448**	**1.447**	**1.544**	**1.842**	**1.833**	**1.801**	**1.857**	**2.240**	**1.556**	**1.527**
Constant		0.395	0.078	0.478	0.848	0.826	1.107	1.095	0.049	0.416	0.409
Nagelkerke value indicating the strength of the model, R^2		0.248	0.202	0.165	0.124	0.125	0.119	0.136	0.213	0.130	0.126

INS = insulin; nonINS = non-insulin; METF = metformin, SU = sulfonylurea; HbA1c = Haemoglobin-A1c (glycated haemoglobin); LDL = Low-Density lipoprotein cholesterol; HDL = High-Density lipoprotein cholesterol; ref = reference group; **more favourable at p < 0.001 level**; **less favourable at p < 0.001 level**; non-significant

4.2. Results of the analysis of associations between diabetes and cancer

Table 9 summarizes the descriptive statistics for the cancer rate analysis, separately for the control group and the group without diabetes. A total of 3 681 774 individuals (including both groups) were involved in the analysis. Importantly, none of these individuals had a prior cancer diagnosis before 2014 (control group) or before the date of their first antidiabetic drug dispensing (group with diabetes). Of the total, 3 595 237 individuals did not fill any antidiabetic drug prescriptions during the observation period and were thus classified as the control group. The remaining 86 537 individuals filled

their first antidiabetic drug prescription in 2014 or 2015, constituting the group with diabetes. The proportion of males was comparable between the groups (45.6% in the control group and 46.7% in the group with diabetes). However, the group with diabetes exhibited a higher average age (61.4 years) and a greater proportion of individuals aged 70 years and older (23.7%) compared to the control group (average age: 58.0 years; 70+: 19.6%) (79).

Cancer diagnosis was recorded during the observation period in 8.6% of the control group (males: 9.3%, females: 8.0%), and in 10.1% of the group with diabetes (males: 11.3%, females: 9.2%). The observed difference in cancer rate between the groups was primarily driven by the younger age group (40-54 years: for people with diabetes 5.4% vs. controls 4.4%; 70-89 years: for patients with diabetes 12.7% vs. controls 12.4%). Among people with cancer diagnosis during the observation period, a higher proportion were male, and the average age was also higher in both the control group and the group with diabetes (Table 9) (79).

The findings for site-specific cancers generally mirrored the overall results. For each cancer site, the proportion of individuals with a cancer diagnosis was consistently higher in the group with diabetes compared to the control group, including when stratified by sex. Furthermore, within both groups, for cancers relevant to both sexes (pancreatic, renal, kidney, liver and colon), the raw incidence of cancer diagnosis was higher in males (Table 9) (79).

Table 9: Characteristics of the study population for the analysis of associations between diabetes and cancer (79)

Study group	Characteristics		All cancers						Total			Cancer rate (%)		
			Not present			Present								
			Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All
Control	Number of cases (people)	All	1,485,620.0	1,800,861.0	3,286,481.0	152,862.0	155,894.0	308,756.0	1,638,482.0	1,956,755.0	3,595,237.0	9.3	8.0	8.6
	Age distribution (%)	40-54	52.5	39.5	45.4	20.5	23.9	22.2	49.5	38.3	43.4	3.9	5.0	4.4
		55-69	34.4	37.0	35.8	53.3	45.8	49.6	36.2	37.7	37.0	13.8	9.7	11.5
		70-89	13.1	23.5	18.8	26.1	30.2	28.2	14.3	24.1	19.6	17.1	10.0	12.3
	Mean age (year)		55.3	59.5	57.6	62.8	63.0	62.9	56.0	59.7	58.0			

	Proportion of males (%)		45.2			49.5			45.6					
Diabetes	**Number of cases (people)**	**All**	35,883.0	41,876.0	77,759.0	4,553.0	4,225.0	8,778.0	40,436.0	46,101.0	86,537.0	11.3	9.2	10.1
	Age distribution (%)	**40-54**	34.5	24.0	28.9	14.3	14.9	14.6	32.3	23.1	27.4	5.0	5.9	5.4
		55-69	48.2	48.1	48.1	59.0	52.2	55.7	49.4	48.5	48.9	13.4	9.9	11.6
		70-89	17.3	27.9	23.0	26.7	32.9	29.7	18.4	28.4	23.7	16.4	10.6	12.7
	Mean age (year)		59.1	62.7	61.1	63.8	64.7	64.3	59.7	62.9	61.4			
	Proportion of males (%)		46.1			51.9			46.7					
Study group	**Characteristics**		**Colorectal cancer**						**Total**			**Cancer rate (%)**		
			Not present			**Present**								
			Male	**Female**	**All**	**Male**	**Female**	**All**	**Male**	**Female**	**All**	**Male**	**Female**	**All**
Control	**Number of cases (people)**	**All**	1,662,161.0	2,027,258.0	3,689,419.0	37,386.0	33,168.0	70,554.0	1,699,547.0	2,060,426.0	3,759,973.0	2.2	1.6	1.9
	Age distribution (%)	**40-54**	48.9	37.8	42.8	17.6	15.0	16.4	48.2	37.4	42.3	0.8	0.6	0.7
		55-69	36.2	38.0	37.2	51.5	44.1	48.0	36.5	38.1	37.4	3.1	1.9	2.4
		70-89	14.9	24.2	20.0	30.9	40.9	35.6	15.3	24.5	20.3	4.4	2.7	3.3
	Mean age (year)		56.2	59.9	58.2	64.1	66.3	65.1	56.4	60.0	58.3			
	Proportion of males (%)		45.1			53.0			45.2					
Diabetes	**Number of cases (people)**	**All**	41,729.0	48,917.0	90,646.0	1,226.0	957.0	2,183.0	42,955.0	49,874.0	92,829.0	2.9	1.9	2.4
	Age distribution (%)	**40-54**	31.6	22.8	26.8	13.7	9.8	12.0	31.1	22.5	26.5	1.3	0.8	1.1
		55-69	49.3	48.7	49.0	56.0	51.4	54.0	49.5	48.7	49.1	3.2	2.0	2.6
		70-89	19.1	28.6	24.2	30.3	38.8	34.0	19.4	28.8	24.5	4.4	2.6	3.3
	Mean age (year)		59.9	63.0	61.6	64.6	66.6	65.5	60.0	63.1	61.7			
	Proportion of males (%)		46.0			56.2			46.3					
Study group	**Characteristics**		**Liver cancer**						**Total**			**Cancer rate (%)**		
			Not present			**Present**								
			Male	**Female**	**All**	**Male**	**Female**	**All**	**Male**	**Female**	**All**	**Male**	**Female**	**All**
Control	**Number of cases (people)**	**All**	1,711,432.0	2,074,229.0	3,785,661.0	6,752.0	5,079.0	11,831.0	1,718,184.0	2,079,308.0	3,797,492.0	0.4	0.2	0.3
	Age distribution (%)	**40-54**	47.9	37.2	42.0	19.3	18.7	19.1	47.8	37.1	42.0	0.2	0.1	0.1
		55-69	36.5	38.1	37.4	54.8	46.9	51.4	36.6	38.1	37.4	0.6	0.3	0.4
		70-89	15.6	24.7	20.6	25.9	34.4	29.5	15.6	24.7	20.6	0.6	0.3	0.4
	Mean age (year)		56.5	60.0	58.4	62.9	64.6	63.6	56.5	60.0	58.4			
	Proportion of males (%)		45.2			57.1			45.2					
Diabetes	**Number of cases (people)**	**All**	43,350.0	50,305.0	93,655.0	335.0	187.0	522.0	43,685.0	50,492.0	94,177.0	0.8	0.4	0.6
	Age distribution (%)	**40-54**	30.8	22.4	26.3	14.0	12.3	13.4	30.7	22.3	26.2	0.4	0.2	0.3
		55-69	49.4	48.6	49.0	60.0	57.8	59.2	49.5	48.6	49.1	0.9	0.4	0.7
		70-89	19.7	29.0	24.7	26.0	29.9	27.4	19.8	29.0	24.7	1.0	0.4	0.6
	Mean age (year)		60.1	63.1	61.7	63.9	64.5	64.1	60.2	63.1	61.8			
	Proportion of males (%)		46.3			64.2			46.4					

Study group	Characteristics		Pancreatic cancer						Total			Cancer rate (%)		
			Not present			Present								
			Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All
Control	Number of cases (people)	All	1,710,135.0	2,070,193.0	3,780,328.0	7,838.0	8,598.0	16,436.0	1,717,973.0	2,078,791.0	3,796,764.0	0.5	0.4	0.4
	Age distribution (%)	40-54	47.9	37.2	42.1	21.7	14.7	18.1	47.8	37.1	42.0	0.2	0.2	0.2
		55-69	36.5	38.1	37.4	50.4	43.1	46.6	36.6	38.1	37.4	0.6	0.5	0.5
		70-89	15.6	24.7	20.6	27.9	42.2	35.4	15.6	24.7	20.6	0.8	0.7	0.7
	Mean age (year)		56.5	60.0	58.4	62.9	66.6	64.9	56.5	60.0	58.4			
	Proportion of males (%)		45.2			47.7			45.2					
Diabetes	Number of cases (people)	All	43,073.0	49,887.0	92,960.0	454.0	438.0	892.0	43,527.0	50,325.0	93,852.0	1.0	0.9	1.0
	Age distribution (%)	40-54	30.9	22.5	26.4	15.0	8.9	12.0	30.7	22.4	26.2	0.5	0.3	0.4
		55-69	49.4	48.7	49.1	56.4	42.2	49.4	49.5	48.7	49.1	1.2	0.8	1.0
		70-89	19.7	28.8	24.6	28.6	48.9	38.6	19.8	29.0	24.7	1.5	1.5	1.5
	Mean age (year)		60.1	63.1	61.7	63.9	68.4	66.1	60.1	63.1	61.7			
	Proportion of males (%)		46.3			50.9			46.4					
Study group	Characteristics		Breast cancer						Total			Cancer rate (%)		
			Not present			Present								
			Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All
Control	Number of cases (people)	All	NA	1,972,493.0	1,972,493.0	NA	49,271.0	49,271.0	NA	2,021,764.0	2,021,764.0	NA	2.4	2.4
	Age distribution (%)	40-54	NA	37.8	37.8	NA	29.9	29.9	NA	37.7	37.7	NA	1.9	1.9
		55-69	NA	37.7	37.7	NA	44.8	44.8	NA	37.8	37.8	NA	2.9	2.9
		70-89	NA	24.5	24.5	NA	25.3	25.3	NA	24.5	24.5	NA	2.5	2.5
	Mean age (year)		NA	59.9	59.9	NA	61.1	61.1	NA	59.9	59.9			
Diabetes	Number of cases (people)	All	NA	47,366.0	47,366.0	NA	1,197.0	1,197.0	NA	48,563.0	48,563.0	NA	2.5	2.5
	Age distribution (%)	40-54	NA	22.9	22.9	NA	16.1	16.1	NA	22.7	22.7	NA	1.7	1.7
		55-69	NA	48.3	48.3	NA	55.6	55.6	NA	48.5	48.5	NA	2.8	2.8
		70-89	NA	28.8	28.8	NA	28.3	28.3	NA	28.8	28.8	NA	2.4	2.4
	Mean age (year)		NA	63.0	63.0	NA	63.7	63.7	NA	63.0	63.0			
Study group	Characteristics		Prostate cancer						Total			Cancer rate (%)		
			Not present			Present								
			Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All
Control	Number of cases (people)	All	1,668,741.0	NA	1,668,741.0	28,904.0	NA	28,904.0	1,697,645.0	NA	1,697,645.0	1.7	NA	1.7
	Age distribution (%)	40-54	49.0	NA	49.0	9.3	NA	9.3	48.3	NA	48.3	0.3	NA	0.3
		55-69	36.3	NA	36.3	54.8	NA	54.8	36.6	NA	36.6	2.5	NA	2.5
		70-89	14.7	NA	14.7	35.9	NA	35.9	15.0	NA	15.0	4.1	NA	4.1
	Mean age (year)		56.1	NA	56.1	66.2	NA	66.2	56.3	NA	56.3			
Diabetes	Number of cases (people)	All	42,086.0	NA	42,086.0	864.0	NA	864.0	42,950.0	NA	42,950.0	2.0	NA	2.0

	Age distribution (%)	40-54	31.7	NA	31.7	5.9	NA	5.9	31.2	NA	31.2	0.4	NA	0.4
		55-69	49.5	NA	49.5	60.2	NA	60.2	49.7	NA	49.7	2.4	NA	2.4
		70-89	18.8	NA	18.8	33.9	NA	33.9	19.1	NA	19.1	3.6	NA	3.6
	Mean age (year)		59.8	NA	59.8	66.1	NA	66.1	60.0	NA	60.0			
Study group	Characteristics		Kidney cancer						Total			Cancer rate (%)		
			Not present			Present								
			Male	Female	All	Male	Female	All	Male	Female	All	Male	Female	All
Control	Number of cases (people)	All	1,704,542.0	2,069,533.0	3,774,075.0	9,193.0	6,250.0	15,443.0	1,713,735.0	2,075,783.0	3,789,518.0	0.5	0.3	0.4
	Age distribution (%)	40-54	48.0	37.2	42.1	24.9	20.8	23.2	47.9	37.2	42.0	0.3	0.2	0.2
		55-69	36.5	38.1	37.4	49.6	47.2	48.6	36.5	38.1	37.4	0.7	0.4	0.5
		70-89	15.5	24.7	20.6	25.6	32.0	28.2	15.6	24.7	20.6	0.9	0.4	0.6
	Mean age (year)		56.5	60.0	58.4	62.0	63.9	62.7	56.5	60.0	58.4			
	Proportion of males (%)		45.2			59.5			45.2					
Diabetes	Number of cases (people)	All	43,198.0	50,095.0	93,293.0	303.0	228.0	531.0	43,501.0	50,323.0	93,824.0	0.7	0.5	0.6
	Age distribution (%)	40-54	30.8	22.4	26.3	20.1	17.5	19.0	30.7	22.4	26.2	0.5	0.4	0.4
		55-69	49.5	48.6	49.0	55.4	53.5	54.6	49.5	48.6	49.0	0.8	0.5	0.6
		70-89	19.7	29.0	24.7	24.4	28.9	26.4	19.8	29.0	24.7	0.9	0.5	0.6
	Mean age (year)		60.1	63.1	61.7	62.7	63.5	63.1	60.1	63.1	61.7			
	Proportion of males (%)		46.3			57.1			46.4					

Figure 5 depicts the HRs for time to cancer diagnosis in the group with diabetes compared to the control group, stratified by subgroup (based on univariate Cox proportional hazards model). All point estimates for the HRs exceed 1.0. Furthermore, for most cancer sites and subgroups, even the lower limit of the 95% confidence interval (CI) also exceeds 1.0, stongly suggesting an increased risk of cancer diagnosis in the group with diabetes across most investigated subgroups (defined by sex and age). Regarding age, the presence of diabetes demonstrated the most pronounced increase in cancer risk in the youngest age group (40-54 years) for all cancer sites except breast cancer. Consistently: with the exception of breast and pancreatic cancer, for each cancer site, the increase of cancer risk associated with diabetes was least pronounced in the oldest age group (≥ 70 years). When examining sex-specific differences within cancer sites that affect both sexes, there is no consistent pattern for HR within the males vs within the females. The HRs were similar for colorectal cancer, higher in males for liver (and pancreatic) cancer, and higher in females for kidney cancer (Figure 5) (79).

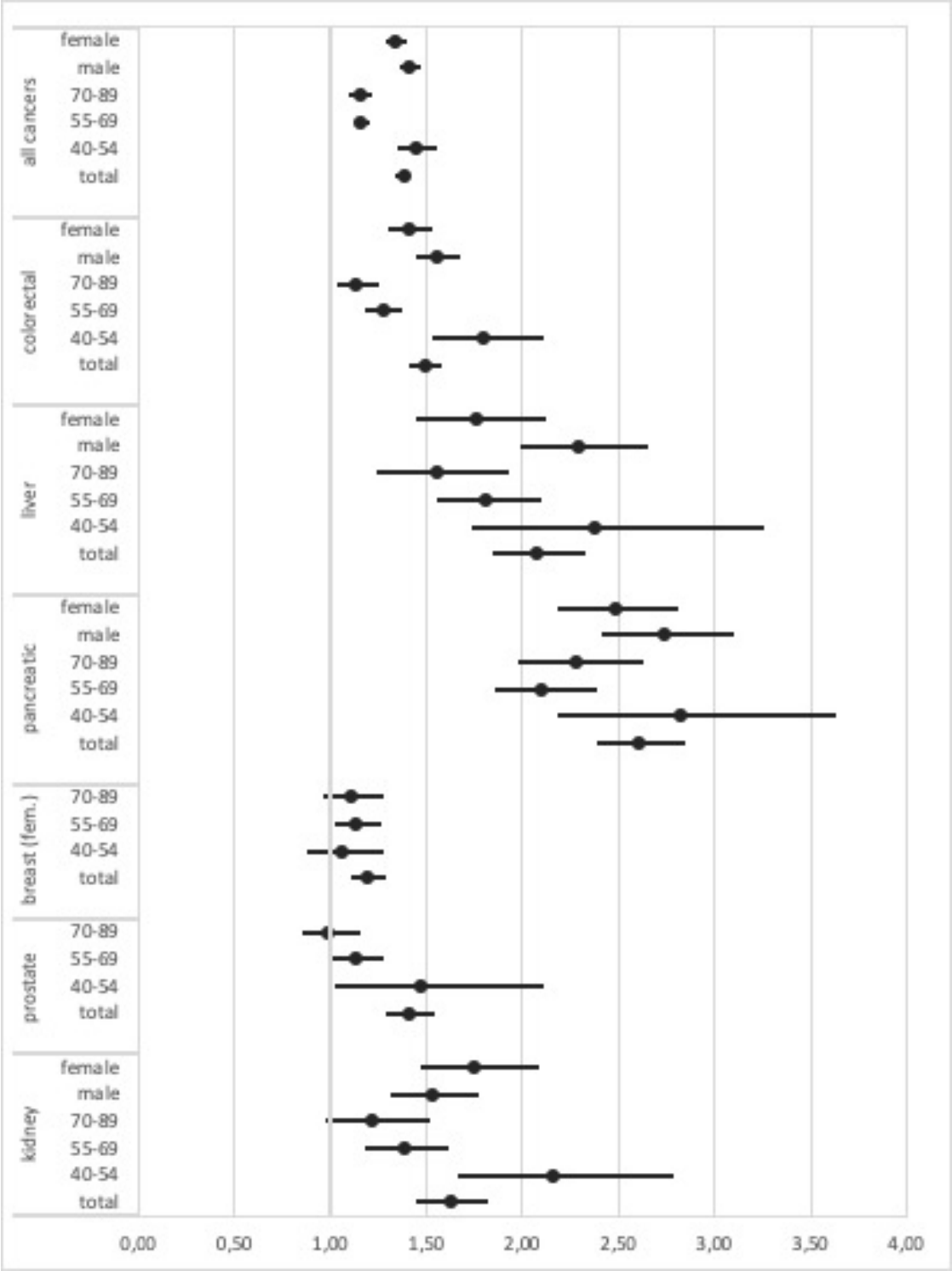


Figure 5: Forest plot: The role of diabetes status in the development of cancer by cancer sites and subgroups using univariate Cox proportional hazard model (HR and 99% CI, based on univariate Cox regression separately for each cancer site and subgroup; modelled time to cancer diagnosis with diabetes status as sole independent variable (79)

.

In the multivariate analysis, in all cancer sites, all investigated independent variables (diabetes, age, sex [where applicable]) significantly influenced the time to cancer diagnosis. The presence of diabetes, higher age, and sex of male (where applicable) are risk factors in developing cancer. The highest risk associated with diabetes was observed for pancreatic cancer (HR=2.294, 99% CI: 2.099; 2.507) and liver cancer (HR=1.830, 99% CI: 1.631; 2.054); with the lowest but still significant risk was observed for breast cancer (HR=1.137, 99% CI: 1.055; 1.227) and prostate cancer (HR=1.171, 99% CI: 1.071; 1.280). Females exhibited a significantly lower risk of cancer development compared to males: the HRs for sex range from 0.470 (kidney cancer) to 0.718 (pancreatic cancer). Age significantly increased the risk of cancer diagnosis: one year of age increment increases the risk of cancer diagnosis with 1.7% (breast cancer) to 8.3% (prostate cancer) (Table 10) (79).

Table 10: Results of the Cox proportional hazard regression analysis[a] (79)

Cancer site	Independent variables											
	Sex (ref: male)				Age (continuous)				Presence of diabetes (ref: control)			
	Hazard ratio	Sig.	Conf.int (99%)		Hazard ratio	Sig.	Conf.int (99%)		Hazard ratio	Sig.	Conf.int (99%)	
			lower	upper			lower	upper			lower	upper
All sites	0.697	***	0.690	0.703	1.045	***	1.045	1.045	1.223	***	1.189	1.258
Colorectal	0.568	***	0.557	0.579	1.059	***	1.058	1.060	1.300	***	1.229	1.375
Liver	0.503	***	0.479	0.527	1.048	***	1.046	1.050	1.830	***	1.631	2.054
Pancreatic	0.718	***	0.690	0.748	1.055	***	1.053	1.057	2.294	***	2.099	2.507
Breast (female)	NA				1.017	***	1.016	1.018	1.137	***	1.055	1.227
Prostate	NA				1.083	***	1.082	1.084	1.171	***	1.071	1.280
Kidney	0.470	***	0.451	0.491	1.043	***	1.041	1.044	1.442	***	1.287	1.616

[a]Based on multivariate Cox regression separately for each cancer site; modelled time to cancer diagnosis with independent variables of diabetes status, age (as continuous variable), and sex (where applicable).

*** p<0,001

The interaction of diabetes with sex (if applicable) and with age was also investigated for each cancer site and overall. At a 1% significance level, no significant sex-specific differences were observed in the influence of diabetes on cancer risk across any cancer site. However, at the 5% significance level, for kidney cancer, diabetes conferred a higher risk increase in females (ratio of hazard ratio [RHR] = 1.257; at age 40, HR females: 2.478, HR males: 1.972). Conversely, for liver cancer, the presence of diabetes decreased

the risk of cancer diagnosis less for females (RHR = 0.831; at age 40, HR females: 2.126, HR males: 2.559). Regarding age, at a 1% significance level, the presence of diabetes increased the risk of cancer diagnosis less as age increases for kidney, prostate, and colon cancer. This effect was most prominent for kidney cancer, with an RHR of 0.843 per 10-year increase in age (e.g., for females, HR at age 40: 2.478, at age 50: 2.090). In contrast, the presence of diabetes increased the risk of diagnosis of pancreatic and breast cancer similarly across all ages (Table 11) (79).

Table 11: Interaction of sex and age with the presence of diabetes[a] (79)

Cancer site	**sex * diabetes**				**age * diabetes**			
	ratio of HR	Sig.	95,0% CI for ratio of HR		ratio of HR	Sig.	95,0% CI for ratio of HR	
			Lower	Upper			Lower	Upper
All sites	ns				0,994	***	0,992	0,996
Colorectal	ns				0,991	***	0,987	0,995
Liver	0,831	*	0,691	0,999	0,989	*	0,981	0,998
Pancreatic	ns				ns			
Breast (female)	NR				ns			
Prostate	NR				0,989	**	0,982	0,995
Kidney	1,257	*	1,053	1,499	0,983	***	0,975	0,991

* p<0,05; ** p<0,01; *** p<0,001

reference sex: male

HR: hazard ratio

ns: non-significant

NR: non-relevant

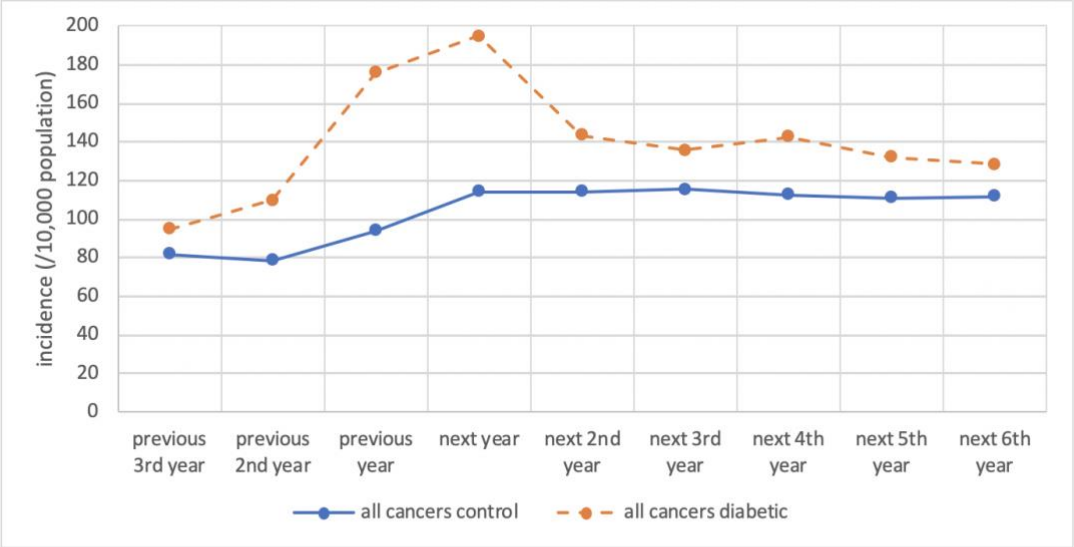
[a]Based on multivariate Cox regression separately for each cancer site; modelled time to cancer diagnosis with independent variables of diabetes status, age (as continuous variable), and sex (where applicable); interaction of diabetes status with age and with sex (where applicable) was included, unless not statistically significant.

As for the temporal relationship between diabetes and cancer, the year of the first appearance of cancer diagnosis relative to the inclusion date has also been analyzed. This analysis also included individuals who had a pre-existing cancer diagnosis prior to the inclusion date, provided they were still alive at the inclusion date. As in the previous

analyses, the control group comprised individuals who did not fill any prescribed antidiabetic drug prescriptions during the observation period. The group with diabetes included those who filled their first antidiabetic drug prescription between 1st January, 2014 and 31st December, 2015. Consistently with the previous analysis, the inclusion date was set as 1st January, 2014 for the control group, and the date of the first antidiabetic drug dispensing for the group with diabetes. We investigated the yearly incidence of cancer diagnosis from three years prior to the inclusion date to six years after the inclusion date (79).

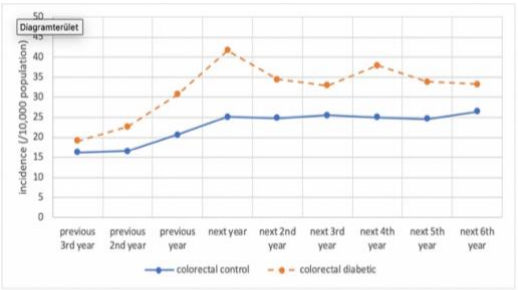
Figure 6a-g illustrates the annual incidence of cancer diagnosis for each year before and after the inclusion date. In both the control group and the group with diabetes, regardless of the specific cancer site, the incidence exhibited an initial increase prior to the inclusion date, culminating in a peak in the year following the inclusion date. The rate of increase over time was notably higher in the group with diabetes. In the year following the inclusion date, the incidence reached 114 / 10,000 population in the control group compared to 195 / 10,000 population in the group with diabetes. Subsequently, the incidence in the control group remained relatively stable (109 / 10,000 – 113 / 10,000), while the incidence in the group with diabetes declined towards the control group level, but remaining consistently higher (128 / 10,000 -143 / 10,000). This pattern was observed across most cancer sites, with the peak incidence in the year following inclusion being most pronounced for pancreatic and liver cancer. Breast cancer exhibited a distinct pattern, with no pre-inclusion increase in the control group and a subsequent concergence of incidence rates between the two groups. Similarly, in prostate cancer, the incidence rates in the group with diabetes converged with those of the control group starting from four years after the inclusion date (79).

6a All cancer sites



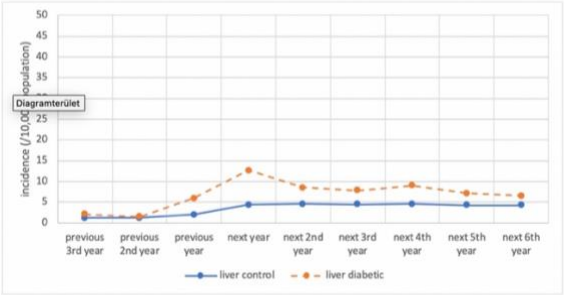
Population size (No. of individuals)									
Control group	3,799,953	3,799,953	3,799,953	3,799,953	3,718,316	3,633,190	3,550,737	3,464,744	3,379,120
Group with diabetes	94,328	94,328	94,328	94,328	93,222	90,549	87,461	84,612	81,603

6b Colorectal cancer



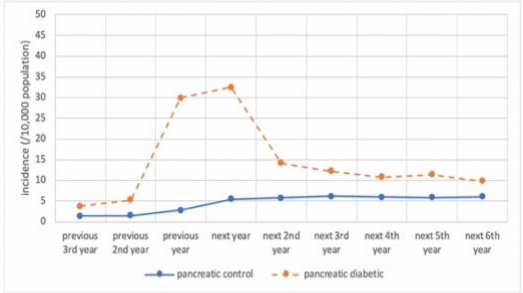
Population size (No. of individuals)									
Control group	3,799,953	3,799,953	3,799,953	3,799,953	3,718,316	3,633,190	3,550,737	3,464,744	3,379,120
Group with diabetes	94,328	94,328	94,328	94,328	93,222	90,549	87,461	84,612	81,603

6c Liver cancer



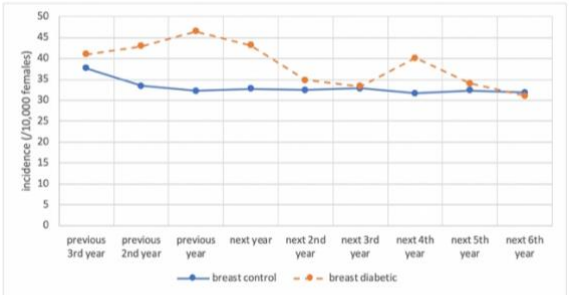
Population size (No. of individuals)									
Control group	3,799,953	3,799,953	3,799,953	3,799,953	3,718,316	3,633,190	3,550,737	3,464,744	3,379,120
Group with diabetes	94,328	94,328	94,328	94,328	93,222	90,549	87,461	84,612	81,603

6d Pancreatic cancer



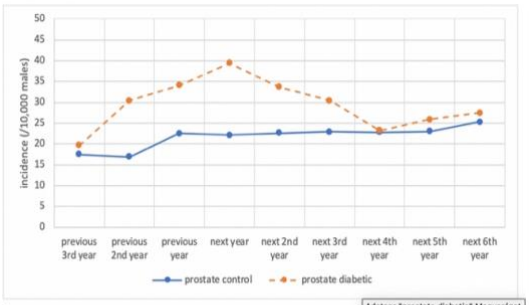
Population size (No. of individuals)									
Control group	3,799,953	3,799,953	3,799,953	3,799,953	3,718,316	3,633,190	3,550,737	3,464,744	3,379,120
Group with diabetes	94,328	94,328	94,328	94,328	93,222	90,549	87,461	84,612	81,603

6e Breast cancer



Population size (No. of individuals)									
Control group	2,080,471	2,080,471	2,080,471	2,080,471	2,040,955	1,998,437	1,957,917	1,914,500	1,871,274
Group with diabetes	50,554	50,554	50,554	50,554	50,036	48,706	47,225	45,867	44,427

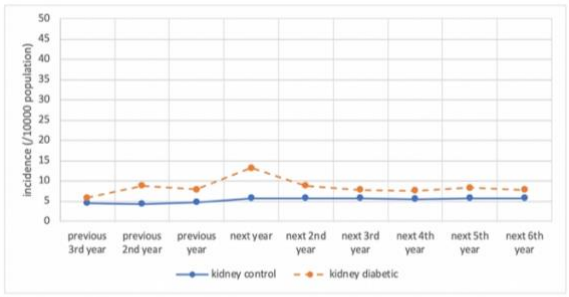
6f Prostate cancer



Adatsor "prostate diabetic" Magyarázat

Population size (No. of individuals)									
Control group	1,719,482	1,719,482	1,719,482	1,719,482	1,677,361	1,634,753	1,592,820	1,550,244	1,507,846
Group with diabetes	43,774	43,774	43,774	43,774	43,186	41,843	40,236	38,745	37,176

6g Kidney cancer



Population size (No. of individuals)									
Control group	3,799,953	3,799,953	3,799,953	3,799,953	3,718,316	3,633,190	3,550,737	3,464,744	3,379,120
Group with diabetes	94,328	94,328	94,328	94,328	93,222	90,549	87,461	84,612	81,603

Figure 6a-g: Incidence of cancer for 10,000 individuals in the years before and after the inclusion date – for all cancer sites and by cancer site (79)

5. DISCUSSION

5.1. Analysis of laboratory tests among adult people with diabetes

Using the NHIFA database, we conducted a retrospective cohort study to analyze the laboratory test utilization among people with diabetes, in accordance with the recommendations of the national professional guidelines in effect during the study period (87,89–91). In this context, we determined the proportion of laboratory tests performed during the observation period, and identified the factors influencing the performance of laboratory tests. Our results were interpreted in collaboration with general practitioners and representatives of various professions involved in diabetes care. A total of 128,115 patients were included in the study. Over 70% of them underwent blood glucose and serum creatinine testing in at least three different years over the four-year period, while HbA1c, urine glucose, and urine ketone testing were performed in 50-60% of patients. Less than 30% underwent LDL cholesterol and urine albumin testing within four years following the first antidiabetic drug dispensing. The impact of the pandemic was evident in a decrease in testing rates during 2020-2021. Testing was less likely in males, younger age groups, and those taking only metformin and/or sulfonylureas. The results showed a declining trend over time (109).

The laboratory tests related to diabetes care that we were able to examine cover four of the nine key care elements recommended by NICE (86) and only a fraction of the corresponding physical examination and laboratory evaluation elements that can be matched to these in the ADA standards (88). Among the laboratory tests examined, the low performance rates of LDL cholesterol and urine albumin tests are striking. This is probably not just a national peculiarity, as the registration of albuminuria and LDL cholesterol shows the lowest rates in the Swedish National Diabetes Register (even so, albumin is above 64% in primary care and above 80% in specialist care every year between 2014 and 2018, and LDL is above 70% and 78% accordingly) (110) and the urine albumin in the National Diabetes Audit of England and Wales (65.4% in 2014/15, which has been steadily decreasing, falling below 50% by 2017/18) (86), as well as in a large-scale Italian study covering 2018 (34.3% of all diabetics, but 62.6% of those requiring specialist care) (111). However, even these values are much better than the Hungarian data, according to which neither urine albumin nor LDL cholesterol tests were

performed even once in the four years of follow-up in nearly 70% of people with diabetes (Figure 3). One explanation for the low rate of LDL cholesterol tests could be that it is not one of the tests that can be ordered by general practitioners (112), but this is not the case for ordering urine albumin tests (109).

The low rate of albuminuria screening, compared to the databases described above, deserves particular attention for three reasons. Firstly, diabetic kidney disease, which is a precursor to cardiovascular events and renal replacement therapy, affects approximately 40% of people with diabetes. Secondly, diabetic kidney disease can only be detected in a third of cases through abnormal albuminuria. Thirdly, new treatment options have emerged in recent years (SGLT2 inhibitors, finerenone, GLP-1 agonists) that can improve the prognosis of these patients (87). Based on these, screening for abnormal albuminuria, early detection and treatment of diabetic kidney disease can lead to improved patient survival and avoidance of dialysis. It should be noted that two additional parameters potentially indicating kidney disease, the estimated glomerular filtration rate (eGFR) and the urine albumin-creatinine ratio, which is an extension of the urine albumin test, could not be examined in our study due to the lack of specific OENO codes in the available administrative data. However, the professional guidelines used in our analysis mention the determination of both the eGFR and microalbuminuria as necessary and non-optional tests. For the other tests, the proportion of people with diabetes who did not have a single test in the four years of follow-up remains below 15%, but the proportion of those who had the HbA1c test performed at least in three different years, which is necessary for the prescribing and monitoring of blood sugar-lowering therapy, is only between 50-60% (Figure 4). In comparison, the Swedish and English audits show results above 90% every year since 2014 in this regard (86,110). This is also true for those participating in specialist care in the aforementioned Italian study (91.1%), but the proportion is only 62.7% when considering all people with diabetes (111). It can be seen that blood glucose tests are performed in nearly 70% of cases in the three months preceding the initiation of antidiabetic therapy (Figure 2), and the proportion of those who undergo at least three blood glucose tests during the four years of follow-up is also around 70% (Figure 4). However, this only shows the rate of testing, and we cannot be sure whether the test was performed as part of diabetes care or whether its result had an impact on antidiabetic therapy. Despite the fact that clinical guidelines clearly recommend considering HbA1c

values for both therapy initiation and monitoring, our data does not include blood glucose measurements performed locally by providers (as point of care tests) or by patients themselves, nor does it include tests performed in the private sector (109).

The primary care performance mesurement, which now includes indicators for blood lipids, HbA1c, and since 2023, microalbumin tests as part of diabetes care and management, seems to support the performance of laboratory tests to a greater extent (113). It is likely that the indicator system in the financing of general practitioners, which has been given a much stronger emphasis since 2023 and which also includes some of the tests for the care of people with diabetes, will also improve the compliance of diabetes care. However, this also requires the settlement of the laboratory fund, as without this, the increased demand for additional tests, which mainly occurs in outpatient care, can cause serious financing difficulties for outpatient clinics, which can lead to longer waiting times and reduce the chance that the patient will attend the test despite the general practitioner's referral. As an alternative option, the patient could have the test performed in the private sector, but this would not be reflected in the publicly funded administrative data, and it is doubtful whether it could be included in the general practitioner's indicator values. The ability to prescribe drugs in the EESZT, despite its many advantages, can reduce the number of doctor-patient encounters, and thus the chance of prescribing or performing various tests (109).

Taking the performance of each laboratory test in at least three different years as the outcome variable, we further analyzed the measurable factors influencing the performance of the tests using logistic regression, which revealed three significant associations (Table 8). In terms of age, our results highlighted the vulnerability of the youngest age group studied, the 40–49-year-olds. In fact, with the exception of the oldest age group, the tests are performed the least in this age group, which is particularly important information given that excess mortality compared to the control population is highest in younger age groups and the trend of increasing overall mortality is also most pronounced among them (68,69). The same association is also true for the risk of heart attack and stroke (75), as well as certain cancers (79). All of these results should also be examined and evaluated in the light of the fact that for the working-age, younger age group, an important screening point falls out of the patient pathway, as the previously

mandatory occupational health examinations were abolished with the amendment of Act XCIII of 1993 at the end of 2023 (109).

In terms of the antidiabetic drug groups used, it can be seen that laboratory tests are most often omitted among those treated with metformin and/or sulfonylureas. Since these two drug groups are the ones that general practitioners can prescribe without specialist consultation, it can be assumed that the omission of laboratory tests is more pronounced among people with diabetes treated in primary care, and draws attention to the need to strengthen the role of primary care in diabetes care. One of the elements of this is the need to clarify the division of tasks between primary care and specialist care, as the practice of this varies depending on the provider and the district - this is also evident in the variability of the county-level results. The current national clinical guideline recommends specialized diabetological care for Type 1 diabetes and those with advanced complications (87), but even in these cases, the nature of general practitioner involvement, coordination of cooperation with specialists involved in complications, and the care of cases that do not have advanced complications but require antidiabetic drugs prescribed by a specialist are still questionable (109).

All of this actually means the development of integrated care instead of the current fragmented care, for which there have been numerous initiatives worldwide in the last decade and even currently. Integrated care includes 16 components in five main areas: service delivery, decision support, self-management support, information technology and technology, and social and community resources (114). Numerous examples of its implementation are available at European level on the European Union's portal on this topic (115), in Diabetes UK care (116,117), and in the publication of The Lancet Commission in terms of strategy and policy development (84), (109).

Among the factors influencing the performance of laboratory tests, it is also worth mentioning the deteriorating trend over time. Taking the results of the 2014 cohort as a reference, a continuous deterioration can be detected in the results of the following years. This can only be partially explained by the COVID-19 pandemic, as the deterioration was already present in the 2015 cohort, and the impact of the pandemic can only be projected onto the fourth year of the 2016 cohort and the third and fourth years of the 2017 cohort. The COVID-19 pandemic has particularly affected people with diabetes, with higher rates of hospitalisation and death compared to those without diabetes, and in this, good blood

glucose control was of particular importance (12,84,118,119). The challenges of caring for people with diabetes have been felt in several areas, including the suspension of outpatient care, reduced inpatient capacity, human resource shortages, drug shortages, unaffordable drugs, delayed seeking medical attention, limited self-management practices, transportation difficulties, and undiagnosed cases and events (120). A decrease in the number of face-to-face doctor-patient encounters has been observed worldwide (118,121,122), and this may have also affected the performance of laboratory tests. This is clearly evident in the Hungarian data for the 2016 and 2017 cohorts. A decrease of 14.2-15.2% was detected in Japan (122) and 1.9-8.8% in the southeastern United States (121), while the Swedish Diabetes Register data shows a slight decrease in the 2020s (110), and in the National Diabetes Audit of England and Wales, no decline is observed in the performance of tests other than the already declining performance of urine albumin (86). We note that the impact of the COVID-19 pandemic on access to antidiabetic drugs was not as limited as access to laboratory tests due to the e-prescription of medication, which became a common practice during this period (109).

Care delivery and treatment in accordance with clinical guidelines improves healthcare outcomes (123,124), and monitoring based on registers and clinical audits also provides an opportunity to develop appropriate quality improvement measures and, at a higher level, strategies and programs related to care. Although in our study we were only able to analyze a few elements of the adequacy of diabetes care based on administrative data, the results can still be used to formulate numerous development proposals. Strengthening the role of primary care in diabetes care can be supported by the development of integrated care, but its individual elements can also be forward-looking in themselves. Through automated solutions integrated into the general practitioner's information system, notifications can be sent about when the necessity of regular tests of a person with diabetes becomes timely or when test results exceed the treshold. Clarification of roles and responsibilities, even through the development of care organisation guidelines, would result in coordinated work between the professions involved and fewer patients falling out of the regular care process. In this context, it is worth considering the review of the range of tests and the range of antidiabetic drugs that can be prescribed by general practitioners, at least in cases where the general practitioner has other relevant specialist qualifications or licenses. Active involvement of people with diabetes and their relatives

in the care process, increased responsibility, and the development of various self-management opportunities would result in a patient-centered, individualized and effective care process. This naturally also presupposes a complex reconsideration and renewal of patient education activities (109).

Another way to improve diabetes care in Hungary is to introduce a comprehensive assessment system. There is currently no systematic quality assessment in Hungary to monitor diabetes care, and there is also no registry of adults with diabetes. However, using the possibilities of the EESZT, it would be possible to carry out real-time assessments and feed back the results to the providers, the profession and the decision-makers, solely from the data generated during patient care, i.e. without any additional administrative burden. As can be seen, even with the existing administrative data, a closer picture can be gained of diabetes care in Hungary, which can be developed into a complete registry and even a national audit by making clinical parameters analyzable and registering interventions related to care that go beyond laboratory tests by coding them. This would allow not only the analysis of the performance of individual tests, but also the assessment of the effectiveness of care activities, with the possibility of evaluating therapeutic target values.

However, when evaluating the results of our study, a number of limitations must also be taken into account. We worked with administrative data, but these did not include events that took place in private care and "point of care" tests performed at providers. We also have no information about care elements prescribed by providers but not ultimately performed due to capacity shortages. Similarly, we could not identify missed tests resulting from non-adherence and comoliance issues of newly diagnosed patients. The antidiabetic drugs examined did not include those that are not publicly fund, but this did not weaken the results from the point of view of the research objective. Those identified as people with diabetes did not include those who only collected antidiabetic drugs once, but this would only be a real problem when reporting incidence data. Those who died during the four-year follow-up period were excluded from our study, so our research focused more on healthier, better-conditioned people with diabetes. The range of indicators that could be examined was limited by the fact that we only saw the performance of laboratory tests, not the results themselves (109).

5.2. Analysis of associations between diabetes and cancer

We analyzed the relationship between diabetes and cancer for the whole Hungarian population using the NHIFA database from the period of 2010-2021.

Multivariate analysis for each cancer site demostrated a significantly higher risk of cancer in the group with diabetes compared to the control group. This association was most pronounced for pancreatic and liver cancer, the two cancer sites in which diabetes has been shown to play the greatest role (92). While Tsidilis et al. reported significant association between diabetes and cancer only for breast and colorectal cancer (125), numerous other reviews and national studies have supported an increased cancer risk associated with diabetes across cancer sites. Although these studies vary slightly in their research design and statistical analyses, their findings generally seem to be more or less comparable (79).

Our findings for liver cancer align with that of Ling et al. (126), although other studies reported higher HR values than ours (93,95,108,127–129). For kidney cancer, our results are not substantially different from that of previous studies (93,95,108,126–129), as are our findings for colorectal (93,95,108,126–129) and breast cancer (93,95,108,126–130), (79).

Our analysis revealed the strongest association between diabetes and pancreatic cancer among all cancer sites investigated. While the majority of previous studies support a significant association (93,95,108,126,128,129), the UK Biobank Study reported a considerably lower risk (127), whereas Zhang et al. observed a substantially higher risk (131), (79).

Most studies have shown an inverse association between diabetes and prostate cancer, suggesting reduced risk (93,95,108,127–129,132,133). This finding contradicts our results, which demonstrate a significant positive association, though the HR for prostate cancer was indeed the lowest of all cancer sites (HR 1.17, 99% CI: 1.07 ;1.28). Xu et al. also reported a decreased risk of prostate cancer, the relative risk was a little bit stronger for low grade and localised disease (132). Interestingly, Hong et al. even observed a significant positive association between diabetes and high-grade prostate cancer, which disappeared when analyzing the association between diabetes and low-grade prostate cancer (134). The Ohsaki Cohort Study similarly found an increased risk of advanced prostate cancer among people with diabetes (135). These divergent findings suggest that

the strength and direction of the association between diabetes and prostate cancer may vary depending on factors such as cancer grade and stage. Notably, Kasper et al. observed a weaker inverse association between diabetes and prostate cancer in the pre-Prostate-Specific Antigen (PSA) testing era compared to the PSA era (136), suggesting that screening practices may influence these findings. Our results may also be due to a possibly higher proportion of advanced prostate cancer cases in our cohort. It is important to note that one previous study, conducted in an Asian population, demonstrated a positive association between diabetes and prostate cancer, aligning with our findings (137), (79). Based on the multivariate analyses, across all cancer sites, age and sex (where applicable) significantly influenced time to cancer diagnosis: besides the presence of diabetes, higher age, male sex (if applicable) are risk factors in developing cancer. Age as a risk factor for cancer is generally known (138,139) and although the mechanism is not clear, the higher risk of male sex is also long known evidence (140–142), (79).

Our study also investigated sex-specific differences in the diabetes-cancer association. No significant sex-specific intercations were observed for most cancer sites. Specifically, for pancreatic and colorectal cancer, there was no difference between males compared to females, suggesting that the presence of diabetes influences the risk of cancer diagnosis the same way for males and females. For kidney and liver cancer, we found significant difference (only at 5% significance level): for kidney cancer, the presence of diabetes conferred a significantly higher risk increase in females, whereas for liver cancer, the diabetes-associated risk reduction was less pronounced in females. These findings partially align with those of Ohkuma et al., who observed a slightly higher excess cancer risk associated with diabetes in females, although the direction and magnitude of sex differences varied across cancer sites. Notably, our findings for kidney and liver cancer, concur with their observations (97). For liver cancer, similar result was also shown by Fang et al. (143), (79).

One of our key findings pertains to the significant interaction between age and diabetes in influencing cancer risk. Multivariate analysis, incorporating age as continuous variable, revealed that for kidney, prostate, and colon cancers, the presence of diabetes increases the risk of cancer diagnosis less and less as age increases. This effect was most pronounced for kidney cancer, where the RHR for ten years of increase in age is 0.843 (eg, for females, HR at age 40: 2.478, at age 50: 2.090). A marginal interaction (at 5%)

was observed for liver cancer. The univariate analysis supported these findings, demonstrating that diabetes conferred the highest cancer risk in the youngest age group (40-54 years) for these four cancer sites, and the lowest risk in the oldest age group (≥70 years). This aligns with previous research. Tseng et al. observed the strongest association between diabetes and prostate cancer in the youngest age group (40-64 years) (144), while Yang et al. reported the highest relative risk for overall and gastrointestinal cancers among patients diagnosed with Type 2 diabetes before the age of 50 (145). In contrast, for pancreatic and breast cancer, there was no significant interaction observed between the presence of diabetes and age as continuous variable. For pancreatic cancer, the univariate analysis did not reveal a linear relationship between age and risk, although the highest HR was observed in the youngest age group. For breast cancer, there is no relevant difference in HR across the 3 age groups. To further investigate the impact of age on breast cancer risk, future studies may benefit from analyzing pre- and postmenopausal women separately (79).

To investigate the temporal relationship between diabetes and cancer, we analyzed cancer diagnoses occurring between 2011 and 2021. We defined 2011 as the earliest possible date of cancer diagnosis inclusion, as all cancer cases recorded in 2010, the first year for which we had data, appear as new cases. Across all cancer sites, the incidence of cancer was consistently higher in the group with diabetes throughout the study period. In the period before the reference year, the incidence of cancer is lower than the actual observed incidence rate because only patients who were still alive in the reference year were included. The difference in cancer incidence between the control group and the group with diabetes might be influenced by factors such as differences in age distribution between the two groups, among other factors. As a result, the rate increases from 2011 until the year after the reference year, nevertheless, it is worth mentioning that the rate of increase over time is relevantly higher in the group with diabetes until the year after the reference year (79).

Our findings on temporality are consistent with those of Lega et al. as they also observed an elevated cancer risk in the group with diabetes prior to diabetes diagnosis. While Lega et al. assessed risk over a 10-year pre-diagnosis period, our year-by-year analysis highlighted a prominent increase in cancer incidence in the years immediately preceding diabetes diagnosis. Although our follow-up period was shorter, both our study and that of

Lega et al. demonstrated a higher cancer incidence in the period immediately following diabetes diagnosis for pancreatic, liver, colorectal, and prostate cancers (105). Supporting these findings, a Danish population-based study observed a substantial decrease in cancer risk for most cancer sites within the first two years after diabetes diagnosis for most cancer sites (107). Similarly, a Swedish national study reported higher HRs for cancer incidence in case of those with one year of the diabetes diagnosis compared to the overall cohort, with exceptions of prostate and liver cancer of the sites we have also examined (129). Johnson et al. found significantly higher cancer incidence among people with diabetes within three months of diabetes diagnosis, and this elevated risk persisted for up to 10 years in many cases, suggesting a complex relationship between diabetes and cancer beyond mere detection bias (146). Ballotari et al. observed an increasing risk of cancer development in the 10 years following diabetes diagnosis in a Northern Italian population, followed by a subsequent decrease to a moderately elevated risk (147). Our findings align with these observations, demonstrating an increased cancer risk detectable as early as two years prior to diabetes diagnosis, with elevated risk persisting for up to six years after diagnosis in our study (79).

Given the high mortality rate and strong association with diabetes, investigating pancreatic cancer is a critical priority in terms of timeliness. Our findings align with those of Zhang et al., demonstrating that the highest risk of pancreatic cancer occurs within the first two years after diabetes diagnosis, specifically within the first year according to our analysis. Notably, we observed a similarly elevated risk in the year preceding diabetes diagnosis. (131). Chari et al. reported that 56% of patients diagnosed with pancreatic cancer within three years of diabetes onset were diagnosed within six months of diabetes diagnosis, further emphasizing the critical period shortly after diabetes onset (148), which is consistent with our results. They also found that approximately 1% of patients with new-onset diabetes aged 50 years and older develop pancreatic cancer within three years, representing an eight-fold increase in risk compared to the general population (148). In our study, the pancreatic cancer incidence in the group with diabetes was tenfold higher than in the control group in the year preceding diabetes diagnosis and approximately sixfold higher in the year following diagnosis, considering that these results have not been adjusted for age. These findings underscore the importance of heightened surveillance for the other disease, whichever comes first (79).

Based on our findings, a crucial recommendation is to re-evaluate and refine cancer screening strategies for people with diabetes. Ling et al. emphasized the potential benefits of earlier or more systematically screening in this population, a notion supported by our observation of a stronger diabetes-cancer association at younger ages (126). This recommendation is particularly relevant given that individuals with diabetes are less likely to undergo recommended cancer screenings (149). Notably, Lao et al. found that females with diabetes had a higher probability of being diagnosed with advanced-stage (stage III-IV) breast cancer (odds ratio [OR] 1.14) and were more likely to be diagnosed with screen-detected breast cancer compared to females without diabetes aged 45-69 years (OR 1.13). The high risk of pancreatic cancer in individuals with newly diagnosed diabetes underscores the need for heightened surveillance in this population. A risk-scoring system, such as that proposed by Sharma et al., could be valuable in identifying high-risk individuals (150). As the prevalence of diabetes continues to rise, due to better care of the classic complications of diabetes, the implementation of more rigorous screening strategies for individuals with diabetes can potentially reduce the overall burden of cancer, leading to earlier diagnosis, improved outcomes, and potentially reduced healthcare costs associated with the management of advanced cancers (151), (79).

The observation of cancer occurrences preceding or shortly following diabetes diagnosis supports the potential role of hyperinsulinemia in cancer development (94,96,99–101). It suggests that screening for hyperinsulinemia, rather than solely focusing on diabetes, may be a valuable consideration in certain cases. Furthermore, exploring the relevance of existing cancer screening programs for individuals with hyperinsulinemia warrants further investigation (for example, those for which screening is already available or a programme is in place, but which might be less frequent or would be carried out at a later age). To effectively implement such strategies, raising awareness among healthcare providers and patients regarding the cancer risk associated with diabetes is crucial. Notably, many current diabetes care guidelines do not explicitly include cancer screening as a key component of diabetes management. Establishing clear roles and responsibilities for implementing these screening strategies within different healthcare systems is essential. Furthermore, prioritizing primary prevention remains paramount in mitigating the individual and societal burdens of both diabetes and cancer. In many cases, even the

guidelines for diabetes do not include cancer as a complication of diabetes, and screening for cancer (or at least the idea of it) is not really included in the monitoring activities in diabetes care (79).

While the cost-effectiveness of expanding population-level cancer screening for individuals with diabetes requires further investigation, several practical steps can be taken immediately. Routinely inquiring about cancer screening history and results at the time of diabetes diagnosis (and even during hyperinsulinemia screening) is crucial. Additionally, educating patients about the importance of cancer screening and available resources within the public and private healthcare systems is essential (79).

A key strength of our study lies exactly in its utilization of a large, population-based administrative health data covering more than ten years. This enabled the creation of substantial cohorts, with the diabetes cohort including all individuals above the age of 40 with incident diabetes diagnosed in the years of 2014 and 2015. The large sample size allowed for statistically robust analyses with significance level of 1%. However, certain limitations should be acknowledged. The lack of detailed clinical data, including prognostic factors and confounders allowed only age- and sex-adjusted calculations. Furthermore, the database did not allow for a distinction between Type 1 and Type 2 diabetes, although the latter is likely to be the predominant form among our study population. While the potential for inaccurate cancer diagnosis codes exists, this bias is likely to affect both the control and diabetes groups equally. However, the exclusion of individuals with preexisting cancer diagnoses prior to 2010 may have introduced some bias, as preexisting malignancies from childhood or early adulthood may not have been consistently recorded in the available data. The inclusion of patients with untreated diabetes within the control group potentially bias the results. This is due to the fact that we only considered those as patients with diabetes who had filled an antidiabetic drug prescription. In addition, the database on antidiabetic drug dispensing included only medication prescribed with reimbursement, therefore people with diabetes taking other antidiabetic drugs could also be included in the control group. According to Jermendy et al. this is not a significant number (69). However, biasing factors that place people with diabetes in the control group do not weaken but strengthen our findings, meaning that the association between diabetes and cancer that our research suggests is underestimated and so probably even stronger in reality. It should be also noted that the exclusion of

individuals with pregnancy-obstetric or PCOS codes from both groups limits the generalizability of our findings to these specific populations.

5.3. The applicability of the NHIFA database in BDA

The NHIFA database undoubtedly qualifies as big data. It encompasses decades of healthcare data for the entire Hungarian population utilizing publicly funded services (volume), including diverse information on diseases, individuals, providers, and interventions (variety), which can updated and expanded on a monthly basis (velocity) and can be easily augmented (variability). A significant portion of the input data is standardized (e.g., OECD, ICD codes) (veracity), and can be linked to other databases through various identifiers (e.g. SSN, provider ID) after depersonalization (valence). Its value is evidenced by previous domestic research and is further reinforced by our current study, which expands its application to a new area. Visualization currently depends on the specific application and statistical methods used.

Based on our test analyses, the six dimensions of data quality (152) regarding the NHIFA database can be described as follows. The data relevance was only partial as we could reliably study only laboratory tests among the diabetes care process indicators. Due to the lack of test results, only the fact that tests were performed could be evaluated. Data accuracy also required caution, as, for example, we could not use the most obvious method of defining people with diabetes, the diagnosis code. The timeliness of the data was adequate, as we could work with a current dataset at the start of the research. However, only one member of our research team had access to the database, which limited the smoothness of the analyses but complied with security regulations. Data selection for analysis was sometimes a lengthy process, lasting several days, due to the specific characteristics of the available infrastructure. The knowledge required for interpretation was widely available regarding the origin of the database's data content, which we were able to refine and specify through interviews. Analytical algorithms were preserved during data analyses. Data from different sources within the NHIFA database were merged based on SSN numbers, but merging with other databases did not occur. Based on all these factors, it can be said that the data quality in the database is adequate if used for analyzing appropriate research questions after sufficient familiarization. Thus,

data quality can be evaluated depending on the focus of each analysis, but our example also shows that the database can be successfully applied to previously unexplored areas.

The NHIFA database has already been used for a variety of descriptive research, but there is still room for further exploration of its potential. In the realm of big data analytics, one of the most frequently studied areas, diabetes, has also revealed new application areas, which have been successfully explored. By studying process indicators in healthcare, we can identify systemic intervention points in the healthcare system, providing a deeper understanding of the strengths and weaknesses within the system that influence outcomes and performance. For example, an analysis of laboratory tests in diabetes care has highlighted issues such as unclear responsibilities, weaknesses in multidisciplinary care, the need to reconsider certain eligibility criteria, and gaps in professional evaluation. Additionally, a deeper national-level analysis of the relationship between diabetes and cancer could serve as a foundation for preventive public health initiatives. For example, our analysis has revealed vulnerable age groups that are currently overlooked by age-based screening programs, and by studying the temporal relationships, we can optimize screening schedules.

As the NHIFA database contains real-world data on both medical procedures and drug therapies, population-based retrospective cohort studies open the door to the world of real-world evidence, which is often used to examine for example the real-world effectiveness or associations of drugs. These studies complement randomized controlled trials (RCTs) and can be used to corroborate or question scientific evidence, given that the data available are far more representative than in RCTs (15,18). Our study for example, has reinforced several pieces of evidence regarding the association between diabetes and cancer, however, it did not support the inverse relationship between diabetes and prostate cancer.

Although the scope of our research did not allow for it, based on the content and characteristics of the database, it can be concluded that it has the potential to support both predictive and prescriptive analyses. Given that the database encompasses all healthcare areas and levels of care, and includes the entire domestic population, it could be used to discover and explore previously unknown or poorly understood relationships, especially with the help of artificial intelligence. However, to fully leverage its potential, it is recommended to link the database with nonadministrative healthcare data. As the diabetes

example shows, the NHIFA database lacks sufficient quality information on prognostic, contributing and influencing factors, and the detailed results of outcomes of interventions are not available in numerical, textual, or visual formats. Future prospectively integrated data systems including other databases can offer improved representativeness, consistency, accuracy, comprehensiveness, and timeliness. These data systems can then provide substantial input for prescriptive analyses in the future (18).

Based on our findings, the following recommendations can be formulated for addressing the challenges related to the domestic application of BDA. Given the national scope of the database, it is advisable to address these challenges also at a national, systemic level. BDA and BDA-based interventions in the healthcare system must be accepted by the stakeholders. Although healthcare professionals are open to decision-support tools and any solution that reduces their administrative burden, they may question the reliability of artificial intelligence due to a lack of understanding of its operation, security concerns, and the vulnerability of direct patient relationships, making them skeptical and resistant to BDA. In addition to persuading and educating them, it is crucial to train specialists who can conduct BDA, as there is currently no doubt a shortage of professionals with the necessary skills in our country. Besides investing in human capital, well-considered planning and implementation of infrastructural and technological investments are also critical. The development of complex, prospective data systems that leverage current capabilities requires the establishment of cloud-based or at least hybrid systems capable of supporting the dynamically evolving and increasingly sophisticated artificial intelligence technologies, and the real-time evolving database. To sustain continuous development, it is essential to share not only data but also the results of data analyses. Naturally, this requires the establishment and ongoing maintenance of numerous regulations, covering data ownership, privacy, data security, metadata management, authorization of wearable devices and applications, as well as related ethical and equality issues. It must not be forgotten that, especially in the case of patient-generated data, ethnic minorities and socioeconomically disadvantaged groups may disproportionately benefit from big data and BDA (3,5,17,153).

Big data and BDA offer immense opportunities for new research, leading to new discoveries and the development of new initiatives and innovations. However, it cannot be overemphasized that the results of analyses are fundamentally determined by the data

used. We cannot measure everything, we must avoid the quantitative fallacy, i.e. only measuring what is easy, forgetting about unmeasured factors and believng that what cannot be measured does not exist. Current healthcare databases inevitably fail to capture factors influencing health and care, such as empathy compassion, understanding, previous experiences, and unconscious bias (13). Therefore, conclusions should be interpreted with care. So when we think about how much knowledge we can gain from BDA, let us not forget how much we cannot.

6. CONCLUSIONS

The findings of our study can be summarized as follows in response to the research questions.

1. What are the diabetes research areas, if any, where NHIFA database analysis could be a valuable tool but have not been investigated in Hungary so far?

Based on a review of the diabetes BDA literature, evaluation of the NHIFA database data, and interviews with experts on coding practices and diabetes care, we identified two areas of research that would be worth exploring as test analyses using the NHIFA database in Hungary: the quality of diabetes care focusing on process indicators, and the association between diabetes and cancer among analyses for diabetes complications.

2. What specific analytical questions can be substantively addressed within the delineated diabetes research domains?

Within the context of diabetes care quality, our research aimed to investigate the extent to which laboratory tests recommended by relevant professional guidelines are performed in diabetes care and to identify the factors influencing their completion. Regarding the association between diabetes and cancer, we had two main research questions. First, we aimed to determine the risk of developing cancer in patients with diabetes compared to the population without diabetes. Secondly, we focused on examining how the time of cancer diagnosis is related to the time of diabetes diagnosis.

3. What novel discoveries and recommendations can be revealed regarding domestic diabetes care based on the test analysis/analyses?

The main findings of the analysis of laboratory tests among adult people with diabetes:

- Only testing rates could be examined using the NHIFA database, test results were not available.
- Over 70% of people with diabetes underwent blood glucose and serum creatinine testing in at least three different years over the four-year period of follow-up, while HbA1c, urine glucose, and urine ketone testing were performed in 50-60% of patients.

- The low performance rates of LDL cholesterol and urine albumin tests (less than 30%) are striking, although literature data from other countries also show them to be the lowest.
- The impact of the COVID-19 pandemic was evident in a decrease in testing rates during 2020-2021.
- Testing was less likely in males, the youngest age group studied (40–49-year-olds), and those taking only metformin and/or sulfonylureas, i.e. probably those treated in primary care.
- Variability of the county-level results of laboratory test rates is also worth mentioning.
- Recommendations for developing domestic diabetes care based on the analysis of laboratory tests are as follows:
    - Strengthening the role of primary care in diabetes care and the review of the range of tests and the range of antidiabetic drugs that can be prescribed by general practitioners, at least in cases where the general practitioner has other relevant specialist qualifications or licenses.
    - Development of integrated care or its individual elements with the clarification of roles and responsibilities.
    - Through automated solutions integrated into the general practitioner's information system, notifications can be sent about when the necessity of regular tests of a person with diabetes becomes timely or when test results exceed the treshold.
    - Active involvement of people with diabetes and their relatives with reconsideration and renewal of patient education activities.
    - Building a national diabetes register and audit system as a prospective data system by linking different databases (including EESZT) would allow not only the analysis of the performance of individual tests, but also the assessment of the effectiveness of care activities, with the possibility of evaluating therapeutic target values without additional administrative burden.

The main findings of the analysis of association between diabetes and cancer:

- For each cancer site, a significantly higher risk of cancer was demostrated in the group with diabetes compared to the control group. This association was most pronounced for pancreatic and liver cancer.

- Our findings revealed a significant positive association between diabetes and prostate cancer, despite most studies have shown an inverse association.
- No significant sex-specific interactions were observed for most cancer sites. Significant difference was found (only at 5% significance level) for kidney cancer, where the presence of diabetes conferred a significantly higher risk increase in females, whereas for liver cancer, the diabetes-associated risk reduction was less pronounced in females.
- Significant interaction was revealed between age and diabetes in influencing cancer risk, showing that mainly for kidney, prostate, and colon cancers, the presence of diabetes increases the risk of cancer diagnosis less and less as age increases. Diabetes conferred the highest cancer risk in the youngest age group (40-54 years), and the lowest risk in the oldest age group (≥70 years).
- As for the temporal relationship, the cancer incidence starts to increase before the diagnosis of diabetes and peaks in the year after. Following this, the incidence drops close to the control group.
- Recommendations for developing domestic diabetes care based on the analysis of association between diabetes and cancer are as follows:
    - Refining cancer screening strategies for people with diabetes considering the vulnerability of the youngest age group studied, the potential role of hyperinsulinemia and the temporal association between the two diseases.
    - Education of health care professionals and people with diabetes about the association between the two diseases and the importance of cancer screening.
    - Diabetes guidelines could be complemented with cancer prevention aspects.
    - Further studies on the association between diabetes and sex-related cancers (prostate and breast cancer) could help clarify the conflicting results.

4. What general conclusions can be inferred about the NHIFA database for its application in BDA, based on the outcomes of the test analysis/analyses?

General conclusions about the NHIFA database can be summarized as follows:

- Data within the NHIFA database undoubtedly qualifies as big data.
- Due to the lack of test results, mainly elements linked to perfomance, appaerance, occurance can be tested.

- Standardized coding systems supports BDA, however have their own limitation related to the specific research question. NHIFA database collects data primarly for financing purposes which may also limit its data accuracy.
- The database can be successfully applied to previously unexplored areas such as process evaluation or deeper analysis of relationships among different prognostic factors, diseases, outcomes, interventions or medications.
- NHIFA database contains real world data, the analysis of which can provide real world evidence. It is suitable for population-based retrospective cohort studies which studies can complement RCTs and can be used to corroborate or question scientific evidence.

5. What prospective recommendations can be formulated for the utilization of the database?

To summarize, the recommendations for future utilization are:

- It is advisable to address challenges of BDA at a national, systemic level.
- Education of every stakeholder regarding BDA and BDA-based interventions can support their acceptance.
- Substantial future BDA reasearch requires adequate human resources which can be provided by programmed specialist training. Similarly, well-considered planning and implementation of infrastructural and technological investments are also critical.
- Development of complex, prospective data systems seem to be the most forward-looking, however, it requires the establishment of cloud-based or at least hybrid systems capable of supporting the dynamically evolving and increasingly sophisticated artificial intelligence technologies, the real-time evolving database and the sharing of data and the results of data analyses
- Establishing and ongoing maintenance of related regulations can support privacy, data security, and handle ethical and equity issues.
- The interpretation of the results of BDA requires special attention to avoid making erroneous decisions or taking inappropriate actions.

7. SUMMARY

Big data analysis (BDA) in healthcare holds significant potential to transform healthcare by acquiring, storing, processing, and analyzing diverse health data to generate valuable insights for users. It can support diabetes care in many ways, such as facilitating patient self-management, image analysis, clinical decision support, or improving quality of care. Our goal was to evaluate the National Health Insurance Fund Administration (NHIFA) database from an analyzability perspective, and by testing new analytical possibilities in the field of diabetes, to identify and perform them, thus enriching the body of available results on domestic diabetes care.

We performed a retrospective cohort study to analyze the completion rates of laboratory tests, as recommended by clinical guidelines, among people with diabetes in 4 cohorts over a 4-year follow-up period and to analyse the role of diabetes in the development of cancer. Multivariate logistic regression analysis was used to identify factors associated with test performance, while univariate and multivariate Cox regression were used to examine how diabetes status, in relationship with age and sex are related to the time to cancer diagnosis.

Perfomance rates of laboratory tests were low when compared to international standards, and testing was less likely in males, the youngest age group studied, and those taking only metformin and/or sulfonylureas. For each cancer site, a significantly higher risk of cancer was demostrated in the group with diabetes and the youngest age group revealed to be the most vulnerable. Cancer incidence started to increase before the diagnosis of diabetes and peaked in the year after. Following that, the incidence dropped close to the control group. Therefore, the most important recommendation is to establish and reinforce the components of integrated care, as well as to revise cancer screening protocols.

NHIFA database found to be successfully applicable to previously unexplored areas. Containing real word data, it can facilitate the generation of real word evidence, particularly by means of retrospective cohort studies. To overcome future challanges, national-level interventions are necessary by programmed specialist training, well-considered implementation of infrastructural and technological investments, education of stakeholders, developing complex, prospective data systems and continuously maintening the related regulations. The interpretation of the results of BDA requires special attention to avoid making erroneous decisions or taking inappropriate actions.

8. REFERENCES

1. Batko K, Ślęzak A. The use of Big Data Analytics in healthcare. J Big Data. 2022 Dec;9(1):3.

2. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data. 2019 Dec;6(1):54.

3. Imran S, Mahmood T, Morshed A, Sellis T. Big data analytics in healthcare A systematic literature review and roadmap for practical implementation. IEEECAA J Autom Sin. 2021 Jan;8(1):1–22.

4. Khanra S, Dhir A, Islam AKMN, Mäntymäki M. Big data analytics in healthcare: a systematic literature review. Enterp Inf Syst. 2020 Aug 8;14(7):878–912.

5. Mohammad Yousef M. Big Data Analytics in Health Care: A Review Paper. Int J Comput Sci Inf Technol. 2021 Apr 30;13(2):17–28.

6. Prableen Kaur. Big Data Analytics in Healthcare: A Review. Int J Eng Res Technol. 2021 Jun;10(6).

7. Kuchipudi Sravanthi, Tatireddy Subba Reddy. Applications of Big data in Various Fields. Int J Comput Sci Inf Technol. 2015;6(5):4629–4632.

8. Wang Y, Hajli N. Exploring the path to big data analytics success in healthcare. J Bus Res. 2017 Jan;70:287–299.

9. Islam MS, Hasan MM, Wang X, Germack HD, Noor-E-Alam M. A Systematic Review on Healthcare Analytics: Application and Theoretical Perspective of Data Mining. Healthcare. 2018 May 23;6(2):54.

10. Cozzoli N, Salvatore FP, Faccilongo N, Milone M. How can big data analytics be used for healthcare organization management? Literary framework and future research from a systematic review. BMC Health Serv Res. 2022 Dec;22(1):809.

11. Wang Y, Kung L, Wang WYC, Cegielski CG. An integrated big data analytics-enabled transformation model: Application to health care. Inf Manage. 2018

Jan;55(1):64–79.

12. Magliano D, Boyko EJ. IDF diabetes atlas. 10th edition. Brussels: International Diabetes Federation; 2021.

13. Kerr D, Klonoff DC. Digital Diabetes Data and Artificial Intelligence: A Time for Humility Not Hubris. J Diabetes Sci Technol. 2019 Jan;13(1):123–127.

14. Rumbold JMM, O'Kane M, Philip N, Pierscionek BK. Big Data and diabetes: the applications of Big Data for diabetes care now and in the future. Diabet Med. 2020 Feb;37(2):187–193.

15. Zaccardi F, Davies MJ, Khunti K. The present and future scope of real-world evidence research in diabetes: What questions can and cannot be answered and what might be possible in the future? Diabetes Obes Metab. 2020 Apr;22(S3):21–34.

16. Begg A. Diabetes care: is big data the future? Pract Diabetes. 2022 May;39(3):7–9.

17. Mackenzie SC, Sainsbury CAR, Wake DJ. Diabetes and artificial intelligence beyond the closed loop: a review of the landscape, promise and challenges. Diabetologia. 2024 Feb;67(2):223–235.

18. Riddle MC, Blonde L, Gerstein HC, Gregg EW, Holman RR, Lachin JM, Nichols GA, Turchin A, Cefalu WT. *Diabetes Care* Editors' Expert Forum 2018: Managing Big Data for Diabetes Research and Care. Diabetes Care. 2019 Jun 1;42(6):1136–1146.

19. Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner AD, Migh E, Schweizer L, Santos A, Bzorek M, Naimy S, Rahbek-Gjerdrum LM, Dyring-Andersen B, Bulkescher J, Lukas C, Eckert MA, Lengyel E, Gnann C, Lundberg E, Horvath P, Mann M. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol. 2022 Aug;40(8):1231–1240.

20. Rácz A, Bajusz D, Miranda-Quintana RA, Héberger K. Machine learning models for classification tasks related to drug safety. Mol Divers. 2021 Aug;25(3):1409–1424.

21. Szocska M, Pollner P, Schiszler I, Joo T, Palicz T, McKee M, Asztalos A, Bencze L, Kapronczay M, Petrecz P, Toth B, Szabo A, Weninger A, Ader K, Bacskai P, Karaszi P, Terplan G, Tuboly G, Sohonyai A, Szoke J, Toth A, Gaal P. Countrywide population movement monitoring using mobile devices generated (big) data during the COVID-19 crisis. Sci Rep. 2021 Mar 15;11(1):5943.

22. Nagy B, Galata DL, Farkas A, Nagy ZK. Application of Artificial Neural Networks in the Process Analytical Technology of Pharmaceutical Manufacturing—a Review. AAPS J. 2022 Jul;24(4):74.

23. Nagyné Elek R, Haidegger T. Next in Surgical Data Science: Autonomous Non-Technical Skill Assessment in Minimally Invasive Surgery Training. J Clin Med. 2022 Dec 19;11(24):7533.

24. Emri M. [Methods of artificial intelligence and their application in imaging diagnostics]. Magy Onkol. 2020 Jun 10;64(2):145–152.

25. Balogh DB, Hudelist G, Bļizņuks D, Raghothama J, Becker CM, Horace R, KrentelI H, Horne AW, Bourdel N, Marki G, Tomassetti C, Kirk UB, Acs N, Bokor A. FEMaLe: The use of machine learning for early diagnosis of endometriosis based on patient self-reported data-Study protocol of a multicenter trial. PloS One. 2024;19(5):e0300186.

26. Jambor T, Juhasz G, Eszlari N. Towards personalised antidepressive medicine based on ‘big data’: an up-to-date review on robust factors affecting treatment response. Neuropsychopharmacol Hung Magy Pszichofarmakologiai Egyesulet Lapja Off J Hung Assoc Psychopharmacol. 2022 Mar 1;24(1):17–28.

27. Gajdács M, Bátori Z, Burián K. Interplay between Phenotypic Resistance to Relevant Antibiotics in Gram-Negative Urinary Pathogens: A Data-Driven Analysis of 10 Years’ Worth of Antibiogram Data. Life. 2021 Oct 8;11(10):1059.

28. Béres E, Nagy J, Tóth J, Árkosy P, Sipka S. [Standardized data on the incidence and mortality of female and male breast cancers in Hungary between 2000 and 2016]. Orv Hetil. 2022;163(5):181–186.

29. Gálffy G, Szabó GT, Tamási L, Müller V, Moldvay J, Sárosi V, Kerpel-Fronius A, Kardos T, Csada E, Pápai-Székely Z, Szász Z, Király Z, Hódi G, Kovács Z, Balogh É, Kovács KA, Darida M, Buga V, Rokszin G, Abonyi-Tóth Z, Kiss Z, Vokó Z, Bogos K. Decreasing incidence and mortality of lung cancer in Hungary between 2011 and 2021 revealed by robust estimates reconciling multiple data sources. Pathol Oncol Res POR. 2024;30:1611754.

30. Jánosi A, Pach FP, Erdős G, Csató G, Pápai G, Andréka P. Incidence, pre-hospital delay and prognosis of acute myocardial infarction in big regions of Hungary: Population data from the Hungarian myocardial infarction registry. Int J Clin Pract. 2021;75(11):e14831.

31. Kenessey I, Szőke G, Dobozi M, Szatmári I, Wéber A, Fogarassy G, Nagy P, Kásler M, Polgár C, Vathy-Fogarassy Á. Comparison of Cancer Survival Trends in Hungary in the Periods 2001-2005 and 2011-2015 According to a Population-Based Cancer Registry. Pathol Oncol Res POR. 2022;28:1610668.

32. Zsirai L, Kun A, Visolyi GÁ, Svébis MM, Domján BA, Tabák Á. Birthweight trends and their explanatory factors in Hungary between 1999 and 2018: an analysis of the Hungarian Tauffer registry. Reprod Health. 2024 Apr 12;21(1):52.

33. Gálffy G, Vastag A, Bogos K, Kiss Z, Ostoros G, Müller V, Urbán L, Bittner N, Sárosi V, Polányi Z, Nagy-Erdei Z, Daniel A, Knollmajer K, Várnai M, Szegner P, Vokó Z, Nagy B, Horváth K, Rokszin G, Abonyi-Tóth Z, Pozsgai É, Barcza Z, Moldvay J, Tamási L. Significant Regional Differences in Lung Cancer Incidence in Hungary: Epidemiological Study Between 2011 and 2016. Pathol Oncol Res POR. 2021;27:1609916.

34. Sánta B, Tomisa G, Horváth A, Balázs T, Németh L, Gálffy G. Severe exacerbations and mortality in COPD patients: A retrospective analysis of the database of the Hungarian National Health Insurance Fund. Pulmonology. 2023;29(4):284–291.

35. Tomisa G, Sánta B, Horváth A, Németh L, Tamás B, Gálffy G, Tamási L, Eszes N. Risk of exacerbation and mortality in asthma: a 10-year retrospective financial database analysis of the Hungarian Health Insurance Fund. BMJ Open Respir Res.

2024 Feb 27;11(1):e002006.

36. Vincze K, Bíró M, Kamocsai M, Andriska P, Sánta B, Lázár Z. [Epidemiological and treatment characteristics of patients with chronic obstructive pulmonary disease on maintenance inhaled therapy between 2011 and 2019 in Hungary]. Orv Hetil. 2024 Mar 3;165(9):338–345.

37. Jánosi A, Ferenci T, Bársony G, Szabóné Póth A, Vajer P, Andréka P. [Practice of antiplatelet therapy in acute myocardial infarction]. Orv Hetil. 2022 May 8;163(19):743–749.

38. Papp T, Kiss Z, Rokszin G, Fábián I, Márk L, Bagoly Z, Becker D, Merkely B, Aradi D, Dézsi CA, Járai Z, Csanádi Z. Mortality on DOACs Versus on Vitamin K Antagonists in Atrial Fibrillation: Analysis of the Hungarian Health Insurance Fund Database. Clin Ther. 2023 Apr;45(4):333–346.

39. Simonyi G, Paksy A, Várnai R, Medvegy M. [Real-world adherence to oral anticoagulants in atrial fibrillation]. Orv Hetil. 2020 May 1;161(20):839–845.

40. Batar P, Alizadeh H, Rokszin G, Abonyi-Toth Z, Demeter J. Comorbidities and outcomes of patients with chronic myeloid leukemia treated with tyrosine kinase inhibitors: a real-world, nationwide, retrospective study from Hungary. Pathol Oncol Res POR. 2024;30:1611497.

41. Kiss Z, Gálffy G, Müller V, Moldvay J, Sárosi V, Pápai-Székely Z, Csada E, Kerpel-Fronius A, Király Z, Szász Z, Hódi G, Polányi Z, Kovács K, Karamousouli E, Knollmajer K, Szabó TG, Berta A, Vokó Z, Rokszin G, Abonyi-Tóth Z, Barcza Z, Tamási L, Bogos K. Significant changes in advanced lung cancer survival during the past decade in Hungary: impact of modern immunotherapy and the COVID-19 pandemic. Front Oncol. 2023;13:1207295.

42. Kiss Z, Szabó TG, Polgár C, Horváth Z, Nagy P, Fábián I, Kovács V, Surján G, Barcza Z, Kenessey I, Wéber A, Wittmann I, Molnár GA, Gyöngyösi E, Benedek A, Karamousouli E, Abonyi-Tóth Z, Bertókné Tamás R, Fürtős DV, Bogos K, Moldvay J, Gálffy G, Tamási L, Müller V, Krasznai ZT, Ostoros G, Pápai-Székely

Z, Maráz A, Branyiczkiné Géczy G, Hilbert L, Tamás Berki L, Rokszin G, Vokó Z. Revising cancer incidence in a Central European country: a Hungarian nationwide study between 2011-2019 based on a health insurance fund database. Front Oncol. 2024;14:1393132.

43. Kiss ZN, Bogos K, Tamási L, Ostoros G, Müller V, Bittner N, Sárosi V, Vastag A, Knollmajer K, Várnai M, Kovács K, Berta A, Köveskuti I, Karamousouli E, Rokszin G, Abonyi-Tóth Z, Barcza Z, Kenessey I, Weber A, Nagy P, Freyler-Fadgyas P, Szócska M, Szegner P, Hilbert L, Géczy GB, Surján G, Moldvay J, Vokó Z, Gálffy G, Polányi Z. Underlying reasons for post-mortem diagnosed lung cancer cases - A robust retrospective comparative study from Hungary (HULC study). Front Oncol. 2022;12:batar.

44. Kunovszki P, Milassin Á, Gimesi-Országh J, Takács P, Szántó K, Bálint A, Farkas K, Borsi A, Lakatos PL, Szamosi T, Molnár T. Epidemiology, mortality and prevalence of colorectal cancer in ulcerative colitis patients between 2010-2016 in Hungary - a population-based study. PloS One. 2020;15(5):e0233238.

45. Liszkay G, Benedek A, Polgár C, Oláh J, Holló P, Emri G, Csejtei A, Kenessey I, Polányi Z, Knollmajer K, Várnai M, Vokó Z, Nagy B, Rokszin G, Fábián I, Barcza Z, Gyulai R, Kiss Z. Significant improvement in melanoma survival over the last decade: A Hungarian nationwide study between 2011 and 2019. J Eur Acad Dermatol Venereol JEADV. 2023 May;37(5):932–940.

46. Molnár A, Pálfi E, Belák B, Blasszauer C, Reibl D, Lövey J. Positive correlation between persistence of medical nutrition therapy and overall survival in patients with head and neck cancer. Pathol Oncol Res POR. 2024;30:1611664.

47. Tamási L, Horváth K, Kiss Z, Bogos K, Ostoros G, Müller V, Urbán L, Bittner N, Sárosi V, Vastag A, Polányi Z, Nagy-Erdei Z, Daniel A, Nagy B, Rokszin G, Abonyi-Tóth Z, Moldvay J, Vokó Z, Gálffy G. Age and Gender Specific Lung Cancer Incidence and Mortality in Hungary: Trends from 2011 Through 2016. Pathol Oncol Res POR. 2021;27:598862.

48. Bereczki D, Bálint M, Ajtay A, Oberfrank F, Vastagh I. Pregestational neurological

disorders among women of childbearing age-Nationwide data from a 13-year period in Hungary. PloS One. 2022;17(9):e0274873.

49. Iljicsov A, Bereczki D, Dobi B, Oberfrank F, Bálint M, Ajtay A, Milanovich D, Simó M. [Age and gender characteristics of patients affected with multiple sclerosis in Hungary between 2004 and 2016]. Orv Hetil. 2021 May 9;162(19):746–753.

50. Mátyás K, Bobál T, Abonyi Z. [Assessment of compliance and patient pathway among multiple sclerosis patients on disease modifying treatment]. Ideggyogyaszati Szle. 2023 Sep 30;76(9–10):309–317.

51. Bitter I, Szekeres G, Cai Q, Feher L, Gimesi-Orszagh J, Kunovszki P, El Khoury AC, Dome P, Rihmer Z. Mortality in patients with major depressive disorder: A nationwide population-based cohort study with 11-year follow-up. Eur Psychiatry J Assoc Eur Psychiatr. 2024 Sep 30;67(1):e63.

52. Makkos Z, Kovats T, Gaal P. [Analysis of antipsychotic prescription data and attendance of patients with schizophrenia in the Hungarian health system]. Neuropsychopharmacol Hung Magy Pszichofarmakologiai Egyesulet Lapja Off J Hung Assoc Psychopharmacol. 2019;21(4):164–169.

53. Rihmer Z, Dome P, Szekeres G, Feher L, Kunovszki P, Gimesi-Orszagh J, Cai Q, El Khoury AC, Bitter I. Comparison of overall survival and healthcare resource utilization among patients with major depressive disorder with or without psychiatric emergency admission: A real-world study from Hungary. J Affect Disord. 2024;372:184–190.

54. Boncz I, Kovács LG. A nemzeti egészségbiztosítási adatvagyon használata betegségteher-elemzésekhez Magyarországon. Orv Hetil. 2021 Mar 28;162(Supplement-1):2–5.

55. Kívés Z, Bódis J, Hunyady B, Pónusz-Kovács D, Elmer D, Kajos FL, Csákvári T, Vajda R, Boncz I. [Participation indicators of colorectal screenings in Hungary between 2008-2021]. Magy Onkol. 2022 Oct 5;66(3):209–217.

56. Laczó A, Bódis J, Bogner P, Molnár K, Vajda R, Pónusz-Kovács D, Elmer D, Kajos

FL, Csákvári T, Kívés Z, Boncz I. [Participation indicators of organized mammography screening in Hungary between 2012-2021]. Magy Onkol. 2022 Oct 5;66(3):195–200.

57. Vajda R, Bódis J, Pónusz-Kovács D, Elmer D, Kajos FL, Csákvári T, Kívés Z, Boncz I. [Participation indicators of organized cervical cancer screening in Hungary]. Magy Onkol. 2022 Oct 5;66(3):186–193.

58. Barkai L, Kiss Z, Rokszin G, Abonyi-Tóth Z, Jermendy G, Wittmann I, Kempler P. Changes in the incidence and prevalence of type 1 and type 2 diabetes among 2 million children and adolescents in Hungary between 2001 and 2016 - a nationwide population-based study. Arch Med Sci AMS. 2020;16(1):34–41.

59. Vesztergom D, Téglás G, Bahrehmand K, Török A, Balla L, Forgács V, Konc J, Tándor Z, Várnagy Á, Boga P, Zádori J, Sipos M, Mánfai Z, Novák Z. Reducing radicality in fertility-sparing surgery is associated with improved in vitro fertilization outcome in early-stage cervical cancer: A national retrospective study. Gynecol Oncol. 2024 Jul;186:35–41.

60. Bogos K, Kiss Z, Tamási L, Ostoros G, Müller V, Urbán L, Bittner N, Sárosi V, Vastag A, Polányi Z, Nagy-Erdei Z, Daniel A, Vokó Z, Nagy B, Horváth K, Rokszin G, Abonyi-Tóth Z, Barcza Z, Gálffy G and Moldvay J. Improvement in Lung Cancer Survival: 6-Year Trends of Overall Survival at Hungarian Patients Diagnosed in 2011-2016. Pathol Oncol Res POR. 2021;27:603937.

61. Csatordai M, Benkő R, Matuz M, Bor A, Lengyel C, Doró P. Use of glucose-lowering drugs in Hungary between 2008 and 2017: the increasing use of novel glucose-lowering drug groups. Diabet Med J Br Diabet Assoc. 2019;36(12):1612–1620.

62. Hankó BZ, Reszegi CA, Kumli P, Vincze Z. [Practice of antidiabetic therapy in Hungary]. Acta Pharm Hung. 2005;75(2):77–86.

63. Jermendy G, Wittmann I, Nagy L, Kiss Z, Rokszin G, Abonyi-Tóth Z, Katona L, Paragh G, Karádi I, Mekely B. Persistence of initial oral antidiabetic treatment in

patients with type 2 diabetes mellitus. Med Sci Monit Int Med J Exp Clin Res. 2012 Feb;18(2):CR72-77.

64. Jermendy G, Kiss Z, Rokszin G, Abonyi-Tóth Z, Wittmann I, Kempler P. [Trends in antidiabetic treatment prescribed for patients with type 2 diabetes in Hungary between 2001 and 2014 - results from the database analysis of the National Health Insurance Fund]. Orv Hetil. 2017 May;158(20):770–778.

65. Jermendy G, Kiss Z, Rokszin G, Abonyi-Tóth Z, Lengyel C, Kempler P, Wittmann I. Changing Patterns of Antihyperglycaemic Treatment among Patients with Type 2 Diabetes in Hungary between 2015 and 2020—Nationwide Data from a Register-Based Analysis. Medicina (Mex). 2022 Oct 1;58(10):1382.

66. Rokszin G, Kiss Z, Sütő G, Kempler P, Jermendy G, Fábián I, Szekanecz Z, Poór G,Wittmann I, Molnár GA. Sodium-Glucose Co-Transporter 2 Inhibitors May Change the Development of Urinary Tract and Hematological Malignancies as Compared With Dipeptidyl Peptidase-4 Inhibitors: Data of the Post-Hoc Analysis of a Nationwide Study. Front Oncol. 2021;11:725465.

67. Jermendy G, Kiss Z, Rokszin G, Abonyi-Tóth Z, Wittmann I, Kempler P. Decreasing incidence of pharmacologically treated Type 2 diabetes in Hungary from 2001 to 2016: A nationwide cohort study. Diabetes Res Clin Pract. 2019 Sep;155:107788.

68. Jermendy G, Kiss Z, Rokszin G, Abonyi-Tóth Z, Wittmann I, Kempler P. Antidiabetikummal kezelt 2-es típusú cukorbetegek epidemiológiai adatai Magyarországon 2016-ban – központi regiszter adatbázisának elemzése. Diabetol Hung. 2019;27(4):205–211.

69. Jermendy G, Kiss Z, Rokszin G, Fábián I, Wittmann I, Kempler P. Changes in mortality rates and ratios in people with pharmacologically treated type 2 diabetes mellitus between 2001 and 2016 in Hungary. Diabetes Res Clin Pract. 2020 May;163:108134.

70. Kiss Z, Jermendy G, Rokszin G, Abonyi-Tóth Z, Fábián I, Kempler P, Wittmann I.

A 2-es típusú hazai cukorbetegek többlethalálozásának nem és korcsoport szerinti értékelése. Diabetol Hung. 2021;29(3):161–173.

71. Kiss Z, Rokszin G, Abonyi-Tóth Z, Jermendy G, Kempler P, Barkai L, Wittmann I. Young adult patients with type 1 diabetes have a higher risk of mortality than those of similar age with type 2 diabetes: A nationwide analysis in Hungary. Diabetes Metab Res Rev. 2019;35(8):e3190.

72. Csákvári T, Elmer D, Németh N, Kívés Z, Wittmann I, Janszky J, Boncz I. [Assessing the epidemiological disease burden and health insurance treatment cost of diabetic polyneuropathy based on health insurance claims data]. Orv Hetil. 2021 Mar 28;162(162 Suppl 1):38–45.

73. Iski G, Rurik SE, Rurik I. Expenditures Of Metabolic Diseases - An Estimation on National Health Care Expenditures of Diabetes and Obesity, Hungary 2013. Exp Clin Endocrinol Diabetes Off J Ger Soc Endocrinol Ger Diabetes Assoc. 2019;127(1):62–67.

74. Kiss Z, Rokszin G, Abonyi-Tóth Z, Jermendy G, Kempler P, Aradi D, Wittmann I. Dissimilar impact of type 2 diabetes on cardiovascular outcomes according to age categories: a nationwide population study from Hungary. Cardiovasc Diabetol. 2018 Dec;17(1):107.

75. Kiss Z, Rokszin G, Abonyi-Tóth Z, Jermendy G, Kempler P, Aradi D, Wittmann I. A 2-es típusú cukorbetegség kardiovaszkuláris megbetegedésre és összmortalitásra kifejtett korfüggő hatása: országos elemzés a NEAK adatbázisa alapján. Diabetol Hung. 2019;27(1):19–27.

76. Kiss Z, Rokszin G, Abonyi-Tóth Z, Jermendy G, Kempler P, Wittmann I. Different Changes of Risks for Stroke and Myocardial Infarction in Patients With Type 2 Diabetes in Hungary Between the Two Periods of 2001-2004 and 2010-2013. Front Endocrinol. 2019;10:170.

77. Klebovich A, Hankó B, Orbán K, Zelkó R. Antipsychotic treatment and the prevalence of diabetes among elderly patients in psychiatric rehabilitation. Arch

Gerontol Geriatr. 2009;48(1):19–21.

78. Molnár GA, Vokó Z, Sütő G, Rokszin G, Nagy D, Surján G, Surján O, Nagy P, Kenessey I, Wéber A, Pálosi M, Müller C, Kásler M, Wittmann I, Kiss Z. Effectiveness of SARS-CoV-2 primary vaccines and boosters in patients with type 2 diabetes mellitus in Hungary (HUN-VE 4 Study). BMJ Open Diabetes Res Care. 2024;12(1):e003777.

79. Safadi H, Balogh Á, Lám J, Nagy A, Belicza É. Associations between diabetes and cancer: A 10-year national population-based retrospective cohort study. Diabetes Res Clin Pract. 2024 May;211:111665.

80. Lester WT, Zai AH, Chueh HC, Grant RW. Diabetes Information Technology: Designing Informatics Systems to Catalyze Change in Clinical Care. J Diabetes Sci Technol. 2008 Mar;2(2):275–283.

81. Agnello D, Tello J. Registries and information systems for diabetes care in the WHO European Region: preliminary findings for consultation [Internet]. World Health Organization; 2021. Available from: https://cdn.who.int/media/docs/librariesprovider2/euro-health-topics/diabetes/registries-information-systems-diabetes-consultation-eng.pdf?sfvrsn=d9f3551f_2&download=true

82. Carinci F, Štotl I, Cunningham SG, Poljicanin T, Pristas I, Traynor V, Olympios G, Scoutellas V, Azzopardi J, Doggen K, Sandor J, Adany R, Løvaas KF, Jarosz-Chobot P, Polanska J, Pruna S, de Lusignan S, Monesi M, Di Bartolo P, Scheidt-Nave C, Heidemann C, Zucker I, Maurina A, Lepiksone J, Rossing P, Arffman M, Keskimäki I, Gudbjornsdottir S, Di Iorio CT, Dupont E, de Sabata S, Klazinga S, Benedetti MM. Making Use of Comparable Health Data to Improve Quality of Care and Outcomes in Diabetes: The EUBIROD Review of Diabetes Registries and Data Sources in Europe. Front Clin Diabetes Healthc. 2021 Oct 11;2:744516.

83. Cunningham SG, Carinci F, Brillante M, Leese GP, McAlpine RR, Azzopardi J, Beck P, Bratina N, Bocquet V, Doggen K, Jarosz-Chobot PK, Jecht M, Lindblad U, Moulton T, Metelko Ž, Nagy A, Olympios G, Pruna S, Skeie S, Storms F, Di Iorio

CT, Massi Benedettiet M. Core Standards of the EUBIROD Project: Defining a European Diabetes Data Dictionary for Clinical Audit and Healthcare Delivery. Methods Inf Med. 2016;55(02):166–176.

84. Chan JCN, Lim LL, Wareham NJ, Shaw JE, Orchard TJ, Zhang P, Lau ESH, Eliasson B, Kong APS, Ezzati M, Aguilar-Salinas CA, McGill M, Levitt NS, Ning G, So WY, Adams J, Bracco P, Forouhi NG, Gregory GA, Guo J, Hua X, Klatman EL, Magliano DJ, Ng BP, Ogilvie D, Panter J, Pavkov M, Shao H, Unwin N, White M, Wou C, Ma RCW, Schmidt MI, Ramachandran A, Seino Y, Bennett PH, Oldenburg B, Gagliardino JJ, Luk AOY, Clarke PM, Ogle GD, Davies MJ, Holman RR, Gregg EW. The Lancet Commission on diabetes: using data to transform diabetes care and patient lives. The Lancet. 2020 Dec;396(10267):2019–2082.

85. Monash University, School of Public Health and Preventive Medicine. Australian National Diabetes Audit Annual Report 2022. Australian National Diabetes Audit; 2023.

86. Holman N, Knighton P, Wild SH, Sattar N, Dew C, Gregg EW, Khunti K, Valabhji J, Young B. Cohort profile: National Diabetes Audit for England and Wales. Diabet Med. 2021 Sep;38(9):e14616.

87. Belügyminisztérium – Egészségügyi Államtitkárság Egészségügyi Szakmai Kollégium. Egészségügyi szakmai irányelv A diabetes mellitus kórismézéséről, a cukorbetegek antihyperglykaemiás kezeléséről és gondozásáról felnőttkorban [Internet]. Belügyminisztérium; 2023. Available from: https://kollegium.aeek.hu

88. American Diabetes Association Professional Practice Committee, ElSayed NA, Aleppo G, Bannuru RR, Bruemmer D, Collins BS, et al. 4. Comprehensive Medical Evaluation and Assessment of Comorbidities: *Standards of Care in Diabetes—2024*. Diabetes Care. 2024 Jan 1;47(Supplement_1):S52–76.

89. Emberi Erőforrások Minisztériuma – Egészségügyért Felelős Államtitkárság EGÉSZSÉGÜGYI SZAKMAI KOLLÉGIUM. Egészségügyi szakmai irányelv – A diabetes mellitus kórismézéséről, a cukorbetegek antihyperglykaemiás kezeléséről és gondozásáról felnőttkorban. 2017.

90. Emberi Erőforrások Minisztériuma – Egészségügyért Felelős Államtitkárság EGÉSZSÉGÜGYI SZAKMAI KOLLÉGIUM. Egészségügyi szakmai irányelv – A diabetes mellitus kórismézéséről, a cukorbetegek antihyperglykaemiás kezeléséről és gondozásáról felnőttkorban. 2020.

91. György Jermendy. A diabetes mellitus kórismézése, a cukorbetegek kezelése és gondozása a felnőttkorban. A Magyar Diabetes Társaság szakmai irányelve, 2009. 2009.

92. Pearson-Stuttard J, Zhou B, Kontis V, Bentham J, Gunter MJ, Ezzati M. Worldwide burden of cancer attributable to diabetes and high body-mass index: a comparative risk assessment. Lancet Diabetes Endocrinol. 2018 Jun;6(6):e6–15.

93. Pearson-Stuttard J, Papadimitriou N, Markozannes G, Cividini S, Kakourou A, Gill D, Rizos EC Monori G, Ward HA, Kyrgiou M, Gunter MJ, Tsilidis KK. Type 2 Diabetes and Cancer: An Umbrella Review of Observational and Mendelian Randomization Studies. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2021 Jun;30(6):1218–1228.

94. Wang M, Yang Y, Liao Z. Diabetes and cancer: Epidemiological and biological links. World J Diabetes [Internet]. 2020 Jun 15 [cited 2023 Mar 27];11(6):227–38. Available from: https://www.wjgnet.com/1948-9358/full/v11/i6/227.htm

95. Ling S, Brown K, Miksza JK, Howells L, Morrison A, Issa E, Yates T, Khunti K, Davies MJ, Zaccardi F. Association of Type 2 Diabetes With Cancer: A Meta-analysis With Bias Analysis for Unmeasured Confounding in 151 Cohorts Comprising 32 Million People. Diabetes Care. 2020 Sep;43(9):2313–2322.

96. Zhu B, Qu S. The Relationship Between Diabetes Mellitus and Cancers and Its Underlying Mechanisms. Front Endocrinol. 2022;13:800995.

97. Ohkuma T, Peters SAE, Woodward M. Sex differences in the association between diabetes and cancer: a systematic review and meta-analysis of 121 cohorts including 20 million individuals and one million events. Diabetologia. 2018 Oct;61(10):2140–2154.

98. Yuan S, Kar S, Carter P, Vithayathil M, Mason AM, Burgess S, Larsson SC. Is Type 2 Diabetes Causally Associated With Cancer Risk? Evidence From a Two-Sample Mendelian Randomization Study. Diabetes. 2020 Jul;69(7):1588–1596.

99. Quoc Lam B, Shrivastava SK, Shrivastava A, Shankar S, Srivastava RK. The Impact of obesity and diabetes mellitus on pancreatic cancer: Molecular mechanisms and clinical perspectives. J Cell Mol Med. 2020 Jul;24(14):7706–7716.

100. Zhang AMY, Wellberg EA, Kopp JL, Johnson JD. Hyperinsulinemia in Obesity, Inflammation, and Cancer. Diabetes Metab J. 2021 May;45(3):285–311.

101. Kim DS, Scherer PE. Obesity, Diabetes, and Increased Cancer Progression. Diabetes Metab J. 2021 Nov;45(6):799–812.

102. Yang J, Nishihara R, Zhang X, Ogino S, Qian ZR. Energy sensing pathways: Bridging type 2 diabetes and colorectal cancer? J Diabetes Complications. 2017 Jul;31(7):1228–1236.

103. Klil-Drori AJ, Azoulay L, Pollak MN. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? Nat Rev Clin Oncol. 2017 Feb;14(2):85–99.

104. Cignarelli A, Genchi VA, Caruso I, Natalicchio A, Perrini S, Laviola L, Giorgino F. Diabetes and cancer: Pathophysiological fundamentals of a ‘dangerous affair’. Diabetes Res Clin Pract. 2018 Sep;143:378–388.

105. Lega IC, Wilton AS, Austin PC, Fischer HD, Johnson JA, Lipscombe LL. The temporal relationship between diabetes and cancer: A population-based study: Population-Based Study of Diabetes and Cancer. Cancer. 2016 Sep 1;122(17):2731–2738.

106. Bowker SL, Lin M, Eurich DT, Johnson JA. Time-Varying Risk for Breast Cancer Following Initiation of Glucose-Lowering Therapy in Women with Type 2 Diabetes: Exploring Detection Bias. Can J Diabetes. 2017 Apr;41(2):204–210.

107. Carstensen B, Witte DR, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. Diabetologia. 2012 Apr;55(4):948–958.

108. Hu Y, Zhang X, Ma Y, Yuan C, Wang M, Wu K, Tabung FK, Tobias D, Hu FB, Giovannucci E, Song M. Incident Type 2 Diabetes Duration and Cancer Risk: A Prospective Study in Two US Cohorts. J Natl Cancer Inst. 2021 Apr 6;113(4):381–389.

109. Safadi H, Lam J, Tislér A, Keresztes K, Belicza E. Analysis of laboratory tests in adults with diabetes in Hungary A retrospective cohort study based on administrative data from 2014 to 2021. Orv Hetil. 2024;165(52):2047–2060.

110. Svensson AM, Eliasson B, Linder E, Almskog I, Eeg-Olofsson K, Miftaraj M, Gudbjörnsdóttir S, Franzén S, Hermansson-Carter V. The Swedish National Diabetes Register (NDR) - Nationwide results 1996-2020 [Internet]. The Swedish National Diabetes Register; 2021 [cited 2024 Jun 8] p. 16473344 byte. Available from: http://refdocs.registercentrum.se/10.18158/ryNUNVPiu

111. Bonora E, Cataudella S, Marchesini G, Miccoli R, Vaccaro O, Fadini GP, Martini N, Rossi E. A view on the quality of diabetes care in Italy and the role of Diabetes Clinics from the 2018 ARNO Diabetes Observatory. Nutr Metab Cardiovasc Dis. 2020 Oct;30(11):1945–1953.

112. Háziorvosok által kérhető laborvizsgálatok [Internet]. Available from: https://haziorvosinfo.hu/wp-content/uploads/2023/01/Kerheto-laborvizsgalatok.pdf

113. Nemzeti Egészségbiztosítási Alapkezelő. A HÁZIORVOSI SZOLGÁLATOK INDIKÁTOR ALAPÚ TELJESÍTMÉNYÉRTÉKELÉSE 2023. JÚLIUS [Internet]. Available from: http://www.neak.gov.hu/pfile/file?path=/letoltheto/altfin_dok/altfin_virt_dok2/Alapellatas/hsz_indikator/Haziorvosi_indikator_utmutato_202307.pdf&inline=true

114. Palmer K, Marengoni A, Forjaz MJ, Jureviciene E, Laatikainen T, Mammarella F, Muth C, Navickas R, Prados-Torres A, Rijken M, Rothe U, Souchet L, Valderas J, Vontetsianos T, Zaletel J, Onder G. Multimorbidity care model: Recommendations from the consensus meeting of the Joint Action on Chronic Diseases and Promoting Healthy Ageing across the Life Cycle (JA-CHRODIS). Health Policy. 2018 Jan;122(1):4–11.

115. European Comission Best Practices Portal [Internet]. Available from: https://webgate.ec.europa.eu/dyna/bp-portal/submission/search

116. Ali Sarah N, Alicea Sarah, Avery Lorraine, Beba Hannah, Kanumilli Naresh, Milne Nicola. Best Practice in the Delivery of Diabetes Care in the Primary Care Network [Internet]. 2021. Available from: https://www.diabetes.org.uk/for-professionals/improving-care/good-practice/primary-and-community-care/pcn-delivery-of-care

117. Diabetes UK. IMPROVING THE DELIVERY OF ADULT DIABETES CARE THROUGH INTEGRATION [Internet]. Available from: https://www.diabetes.org.uk/for-professionals/improving-care/clinical-recommendations-for-professionals/integrated-diabetes-care

118. Australian Government, Department of Health. Australian National Diabetes Audit (ANDA) - COVID-19 Service Survey Report 2020. 2021.

119. Pranata R, Henrina J, Raffaello WM, Lawrensia S, Huang I. Diabetes and COVID-19: The past, the present, and the future. Metabolism. 2021 Aug;121:154814.

120. Mohseni M, Ahmadi S, Azami-Aghdash S, Mousavi Isfahani H, Moosavi A, Fardid M, Etemadi M, Ghazanfari F. Challenges of routine diabetes care during COVID-19 era: A systematic search and narrative review. Prim Care Diabetes. 2021 Dec;15(6):918–922.

121. Harding JL, Wang Y, McDonald B, Gander JC, Oviedo SA, Ali MK. Disruption to routine diabetes care processes during the pandemic: evidence from a large integrated health system in the Southeast United States. BMJ Open Diabetes Res Care. 2023 Jul;11(4):e003466.

122. Ikesu R, Miyawaki A, Sugiyama T, Nakamura M, Ninomiya H, Kobayashi Y. Trends in Diabetes Care during the COVID-19 Outbreak in Japan: an Observational Study. J Gen Intern Med. 2021 May;36(5):1460–1462.

123. Abdel-Rahman N, Manor O, Cohen A, Elran E, Cohen AG, Krieger M, Paltiel O, Valinsky L, Ben-Yehuda A, Calderon-Margalit R, Ang YG. Adherence to diabetes

quality indicators in primary care and all-cause mortality: A nationwide population-based historical cohort study. Ang YG, editor. PLOS ONE. 2024 May 9;19(5):e0302422.

124. Corrao G, Rea F, Di Martino M, Lallo A, Davoli M, De Palma R, Belotti L, Merlino L, Pisanti P, Lispi L, Skrami E, Carle F. Effectiveness of adherence to recommended clinical examinations of diabetic patients in preventing diabetes-related hospitalizations. Int J Qual Health Care. 2019 Jul 1;31(6):464–472.

125. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ [Internet]. 2015 Jan 2 [cited 2023 Apr 23];350(jan02 1):g7607–g7607. Available from: https://www.bmj.com/lookup/doi/10.1136/bmj.g7607

126. Ling S, Brown K, Miksza JK, Howells LM, Morrison A, Issa E, Yates T, Khunti K, Davies MJ, Zaccardi F. Risk of cancer incidence and mortality associated with diabetes: A systematic review with trend analysis of 203 cohorts. Nutr Metab Cardiovasc Dis NMCD. 2021 0;31(1):14–22.

127. Peila R, Rohan TE. Diabetes, Glycated Hemoglobin, and Risk of Cancer in the UK Biobank Study. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2020 Jun;29(6):1107–1119.

128. Gurney J, Stanley J, Teng A, Krebs J, Koea J, Lao C, Lawrenson R, Meredith I, Sika-Paotonu D, Sarfati D. Cancer and diabetes co-occurrence: A national study with 44 million person-years of follow-up. PloS One. 2022;17(11):e0276913.

129. Bjornsdottir HH, Rawshani A, Rawshani A, Franzén S, Svensson AM, Sattar N, Gudbjörnsdottir S. A national observation study of cancer incidence and mortality risks in type 2 diabetes compared to the background population over time. Sci Rep. 2020 Oct 15;10(1):17376.

130. Maskarinec G, Jacobs S, Park SY, Haiman CA, Setiawan VW, Wilkens LR, Le Marchand L. Type II Diabetes, Obesity, and Breast Cancer Risk: The Multiethnic Cohort. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored

Am Soc Prev Oncol. 2017 Jun;26(6):854–861.

131. Zhang JJ, Jia JP, Shao Q, Wang YK. Diabetes mellitus and risk of pancreatic cancer in China: A meta-analysis based on 26 case-control studies. Prim Care Diabetes. 2019 Jun;13(3):276–282.

132. Xu H, Jiang H wen, Ding G xiong, Zhang H, Zhang L min, Mao S hua, Ding Q. Diabetes mellitus and prostate cancer risk of different grade or stage: A systematic review and meta-analysis. Diabetes Res Clin Pract. 2013 Mar;99(3):241–249.

133. Bansal D, Bhansali A, Kapil G, Undela K, Tiwari P. Type 2 diabetes and risk of prostate cancer: a meta-analysis of observational studies. Prostate Cancer Prostatic Dis. 2013 Jun;16(2):151–158.

134. Hong SK, Oh JJ, Byun SS, Hwang SI, Lee HJ, Choe G, Lee SE. Impact of diabetes mellitus on the detection of prostate cancer via contemporary multi (≥12)-core prostate biopsy: Diabetes Mellitus and Prostate Biopsy Outcome. The Prostate. 2012 Jan;72(1):51–57.

135. Li Q, Kuriyama S, Kakizaki M, Yan H, Sone T, Nagai M, Sugawara Y, Ohmori-Matsuda K, Hozawa A, Nishino Y, Tsuji I. History of diabetes mellitus and the risk of prostate cancer: the Ohsaki Cohort Study. Cancer Causes Control. 2010 Jul;21(7):1025–1032.

136. Kasper JS, Giovannucci E. A Meta-analysis of Diabetes Mellitus and the Risk of Prostate Cancer. Cancer Epidemiol Biomarkers Prev. 2006 Nov 1;15(11):2056–2062.

137. Jian Gang P, Mo L, Lu Y, Runqi L, Xing Z. Diabetes mellitus and the risk of prostate cancer: an update and cumulative meta-analysis. Endocr Res. 2015 Jan 2;40(1):54–61.

138. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, Pollak M, Regensteiner JG, Yee D. Diabetes and Cancer. Diabetes Care. 2010 Jul 1;33(7):1674–1685.

139. Qiang JK, Lipscombe LL, Lega IC. Association between diabetes, obesity, aging, and cancer: review of recent literature. Transl Cancer Res. 2020 Sep;9(9):5743–5759.

140. Kim HI, Lim H, Moon A. Sex Differences in Cancer: Epidemiology, Genetics and Therapy. Biomol Ther. 2018 Jul 1;26(4):335–342.

141. Dorak MT, Karpuzoglu E. Gender Differences in Cancer Susceptibility: An Inadequately Addressed Issue. Front Genet. 2012;3.

142. Jackson SS, Marks MA, Katki HA, Cook MB, Hyun N, Freedman ND, Kahle LL, Castle PE, Graubard BI, Chaturvedi AK. Sex disparities in the incidence of 21 cancer types: Quantification of the contribution of risk factors. Cancer. 2022 Oct;128(19):3531–3540.

143. Fang HJ, Shan SB, Zhou YH, Zhong LY. Diabetes mellitus and the risk of gastrointestinal cancer in women compared with men: a meta-analysis of cohort studies. BMC Cancer. 2018 Apr 16;18(1):422.

144. Tseng CH. Diabetes and Risk of Prostate Cancer. Diabetes Care. 2011 Mar 1;34(3):616–621.

145. Yang Z, Wu Y, Xu L, Zhu Z, Li T, Yu L, Gao K, Zhang X, Shen P, Lin H, Shui L, Tang M, Jin M, Sun Y, Wang J, Chen K. Age at diagnosis modifies associations of type 2 diabetes with cancer incidence and mortality: a retrospective matched-cohort study. Diabetologia . 2023 Aug;66(8):1450–1459.

146. Johnson JA, Bowker SL, Richardson K, Marra CA. Time-varying incidence of cancer after the onset of type 2 diabetes: evidence of potential detection bias. Diabetologia. 2011 Sep;54(9):2263–2271.

147. Ballotari P, Vicentini M, Manicardi V, Gallo M, Chiatamone Ranieri S, Greci M, Giorgi Rossi P. Diabetes and risk of cancer incidence: results from a population-based cohort study in northern Italy. BMC Cancer. 2017 Oct 25;17(1):703.

148. Chari S, Leibson C, Rabe K, Ransom J, Deandrade M, Petersen G. Probability of

Pancreatic Cancer Following Diabetes: A Population-Based Study. Gastroenterology. 2005 Aug;129(2):504–511.

149. Bhatia D, Lega IC, Wu W, Lipscombe LL. Breast, cervical and colorectal cancer screening in adults with diabetes: a systematic review and meta-analysis. Diabetologia. 2020 Jan;63(1):34–48.

150. Sharma A, Kandlakunta H, Nagpal SJS, Feng Z, Hoos W, Petersen GM, Chari ST. Model to Determine Risk of Pancreatic Cancer in Patients With New-Onset Diabetes. Gastroenterology. 2018 Sep;155(3):730-739.e3.

151. Gregg EW, Sattar N, Ali MK. The changing face of diabetes complications. Lancet Diabetes Endocrinol. 2016 Jun;4(6):537–547.

152. United Nations Economic Commission for Europe. Conference of European Statisticians Recommendations of the 2010 Censuses of Population and Housing. Prepared in coopera- tion with the Statistical Office of the European Communities (EUROSTAT). New York and Geneva: United Nations Economic Commission for Europe; 2006.

153. Karatas M, Eriskin L, Deveci M, Pamucar D, Garg H. Big Data for Healthcare Industry 4.0: Applications, challenges and future perspectives. Expert Syst Appl. 2022 Aug;200:116912.

9. BIBLIOGRAPHY OF PUBLICATIONS

Publications related to the thesis:

Safadi, Heléna; Balogh, Ágnes; Lám, Judit; Nagy, Attila; Belicza, Éva
Associations between diabetes and cancer: A 10-year national population-based retrospective cohort study
DIABETES RESEARCH AND CLINICAL PRACTICE 211 Paper: 111665, 12 p. (2024)

Safadi, Heléna; Lám, Judit; Tislér, András; Keresztes, Katalin; Belicza, Éva
A felnőtt korú cukorbetegek laboratóriumi vizsgálatainak elemzése hazánkban: Retrospektív kohorszvizsgálat a 2014–2021 közötti igénybevételi adatok alapján
ORVOSI HETILAP 165: 52 pp. 2047-2060., 14 p. (2024)

Publications not related to the thesis:

Varga, Beatrix; Kovács, Aranka Katalin; Safadi, Heléna
Audit ellenőrzőlista kidolgozása a gasztroenterológiai endoszkópos vizsgálóhelyek higiénés értékelésére
IME 23: 4 pp. 38-46., 9 p. (2024)

Safadi, Heléna; Lám, Judit; Baranyi, Ivett; Belicza, Éva
An intermediate step in bridging the gap between evidence and practice: developing and applying a methodology for “general good practices”
IME 23: 4 pp. 25-37., 13 p. (2024)

Báldy, Barbara; Safadi, Heléna; Lám, Judit
Betegek a betegbiztonságról
IME 22: 3 pp. 44-51., 8 p. (2023)

Sziráki, Márton; Váradi, Zsolt; Nagy, Tünde; Pénzes, Melinda; Bedros, J. Róbert; Safadi, Heléna
Az újraélesztés folyamatának és oktatásának újraszervezése a Dél-budai Centrumkórház Szent Imre Egyetemi Oktatókórházban
IME 22: 2 pp. 23-31., 9 p. (2023)

Bálity, Csaba; Cserháti, Zoltán; Lám, Judit; Palicz, Tamás; Safadi, Heléna
Humán szolgáltatások képzési programjainak értékelési rendszere
IME 19: 3 pp. 9-14., 6 p. (2020)

Lám, Judit; Safadi, Heléna; Pitás, Eszter; Belicza, Éva
Ki mit tud? - A betegátadásról: Egy magyar és lengyel mintán végzett tudásfelmérés tapasztalatai
ORVOSI HETILAP 160: 44 pp. 1735-1743., 9 p. (2019)

Mikos, Borbála; Molnár, Eszter; Hantos, Mónika; Safadi, Heléna; Velkey, György János
Az antibiotikus terápia múltja, jelene és jövőképe
GYERMEKGYÓGYÁSZAT 70: 1 pp. 32-37., 6 p. (2019)

10. ACKNOWLEDGEMENTS

“And what is it to work with love? (...)
It is to charge all things you fashion with a breath of your own spirit,
And to know that all the blessed dead are standing about you and watching.” – writes Kahlil Gibran in his book The Prophet. And for this, I first and foremost thank my many loved ones who have passed away, for standing about me and watching. I am nourished by your love every day.

I am so grateful to my family for, instead of standing about me and watching, stepping up and doing things for me, bearing my absence – I know it was really a great sacrifice.

Thanks to my supervisor, Éva Belicza, for the trust she has placed in me, for introducing me to the world of quality of care and for many years guiding me through it, as did Judit Lám and my colleagues in the Patient Safety Department – thanks for your support and encouragement. I also like to thank my colleagues in the Health Services Management Training Centre as well as Marianna Török and the Workgroup for Science Management, Doctoral School for their practical advice, which helped me to complete my dissertation.

Finally, I thank my friends who so many times could not believe that I had not finished this work yet – this dissertation is nothing but a gift to you for appreciating me so much.