Type 2 diabetes in migrants: register-based analyses

(WORK-IN-PROGRESS DRAFT OF!) PhD dissertation

Anders Aasted Isaksen

Faculty of Health
Aarhus University
March 2023

AARHUS UNIVERSITY
HEALTH

Website

Supervisors and assessment committee

PhD student:

Anders Aasted Isaksen, MD
Department of Public Health, Aarhus University, Denmark

Supervisors:

Professor Annelli Sandbæk, PhD (main supervisor)
Department of Public Health, Aarhus University, Denmark
Steno Diabetes Center Aarhus, Denmark

Associate Professor Mette Vinther Skriver, PhD
Department of Public Health, Aarhus University, Denmark

Lasse Bjerg, PhD
Department of Public Health, Aarhus University, Denmark
Steno Diabetes Center Aarhus, Denmark

Assessment committee:

Professor Reimar Wernich Thomsen, PhD (chairman)
Department of Clinical Epidemiology, Department of Clinical Medicine, Aarhus University Hospital, Denmark

Professor Per Wändell, PhD
Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet

Senior researcher Nana Folmann Hempler, PhD
Diabetes Management Research, Steno Diabetes Center Copenhagen, Denmark

Financial disclosure

Financial disclosures

This PhD project was supported by a research training supplement grant from **Aarhus University**, and grants from the **Public Health in Central Denmark Region - a collaboration between municipalities and the region** foundation (Danish: *Folkesundhed i Midten*) and **Steno Diabetes Center Aarhus**, which is partially funded by an unrestricted donation from the *Novo Nordisk Foundation*.

Acknowledgements

Thanks yous and nostalgia go here….

Not least Gregers and the SDCC epi group. Also thanks to the R-foundation, RStudio/Posit, open-source package developers and the R biostatistics group at University of Bergen for providing a Quarto template for this dissertation.

Also thanks to family, friends & colleagues at the Department of Public Health and the Research Unit for General Practice being nice. Thanks to supervisors for discovering when I didn't know what I was doing and for being patient, supportive and insightful throughout.

Thank you to fellow researchers at Steno Diabetes Center Aarhus for great discussions about diabetes, tools and methodology.

Also a special thanks to Luke Johnston for inspiring me to strive for openness and better programming practices, and having a lot of fun learning along the way. Thank you to fellow teachers and contributors to the *R for Reproducible Research* courses. I hope my work reflects the principles and spirit of open science we strive for.

Preface

Motivation

During my work as a junior doctor in a general practice in Gellerup, a deprived suburb of Aarhus with a large migrant population, I was immediately struck by the high frequency of type 2 diabetes (T2D) and other chronic diseases among day-to-day consultations. I would quickly learn that T2D held more surprises for me than just a high prevalence, and after just a few weeks it seemed that much of what I had learned about the disease in medical school and studied in clinical guidelines did not apply in my new setting. Patients were not only more prevalent, they were also much younger, and hemoglobin-A1c (HbA1c) levels rarely came close to guideline targets despite our best efforts to intensify treatment. Sadly, these experiences would repeat themselves in my subsequent work with migrants as a general practitioner (GP) in health clinics of the Danish Red Cross at asylum residence centers and pre-removal detention centers.

As I reviewed the literature, I understood why medical school and guidelines had failed to prepare me for the challenges I faced as a clinician with T2D in migrants: the existing literature was inadequate. While increased prevalence of T2D appeared well-established, and migrants with T2D were also ascribed a higher mortality than their native counterparts, this provided little guidance for my day-to-day work. Contrarily, there was hardly any evidence on disparities in the time between diagnosis and death - the time when care from the GP is needed the most. So I was left to wonder what might cause this discordance between what I knew about T2D from books and guidelines, and what I encountered in migrants in the clinic.

*Are migrants with T2D more prone to under-treatment, and in which areas of care? Are some migrant groups more prone to under-treatment than others?

Knowledge on these clinical questions would enable GPs and healthcare planners to address disparities and improve care in migrants by prioritizing and focusing care accordingly. Migrants are currently a younger demographic than the rest of the population, but as time causes the migrant demographic to age, it is likely that migrants will constitute an increasing proportion of the T2D population in the coming years. Therefore, I was excited to explore these clinical questions as a researcher, hoping I could provide

answers to myself and fellow GPs that could lead to better care - and, ultimately, better health - in a vulnerable and challenging group of patients.

During my time as a PhD student, I quickly encountered the first of many challenges on the way to studying migrants with T2D in the Danish registers: there was no validated definition of T2D, nor a common consensus among researchers. This led to the PhD project expanding its scope to develop and validate a tool to define T2D in the Danish registers, which allowed me to answer my research questions based on robust findings. This PhD project benefited greatly from open-source tools, and in the spirit of open science, the source code of the validated diabetes classifier was made available to other researchers in the *osdc* package for the *R* statistical programming language. As a final commitment to openness, this dissertation was made available to a global audience in website format at aastedet.github.io/dissertation/.

Outline of the dissertation

Chapter 1 introduces type 2 diabetes (T2D) and the special risk it poses to migrants. It then describes the context of T2D care in Denmark and summarizes clinical guideline recommendations and how they relate migrants with T2D. Finally, it introduces the reader to identification of diabetes patients in Danish healthcare registers.

Chapter 2 states the overall aims of this PhD project and each individual study.

Chapter 3 describes the setting, data sources, methods and study designs used in the studies of this PhD project.

Chapter 4 presents the main results of the studies.

Chapter 5 contains a discussion of the methods used and their potential impact on results.

Chapter 6 discusses the findings in light of the methods used.

Chapter 7 presents the main conclusions and their clinical implications.

Chapter 8 draws up future perspectives of the PhD study and research field.

Papers associated with the dissertation

Coming soon

Table of contents

List of Figures

List of Tables

Abbreviations

ACEI: Angiotensin-converting enzyme-inhibitors
ARB: Angiotensin receptor blockers
ATC: Anatomical Therapeutic Chemical (classification)
CI: Confidence interval
CVD: Cardiovascular disease
DKD: Diabetic kidney disease
GP: General practitioner
GDM: Gestational diabetes mellitus
GLD: Glucose-lowering drugs
HbA1c: Hemoglobin-A1c
LDL-C: low-density lipoprotein cholesterol
LLD: Lipid-lowering drugs
OSDC: Open-Source Diabetes Classifier
RSCD: Register of Selected Chronic Diseases
T1D: Type 1 diabetes mellitus
T2D: Type 2 diabetes mellitus
UACR: Urine albumin-to-creatinine ratio

1. Introduction

This chapter describes type 2 diabetes (T2D) and how migrants are at particular risk. It then describes the context of migration and health care in Denmark and summarizes clinical guideline recommendations and how they relate to the needs of migrants with T2D. Finally, it introduces the reader to identification of diabetes patients in Danish healthcare registers as a means to provide much-needed knowledge on potential disparities in T2D care between migrants and native Danes.

1.1. Migrants and type 2 diabetes

T2D constitutes the vast majority of all diabetes cases, and disproportionally affects the socially deprived demographics in societies across the globe.[1] The prevalence of type 2 diabetes (T2D) today exceeds half a *billion* people globally and is on the rise due to several factors, including population aging, urbanization, adoption of energy-dense diets, and sedentary lifestyles. These transitions are occurring at a rapid pace in developing countries, where they affect the largest number of people. Furthermore, many developing countries have populations with a genetic predisposition to developing T2D, which exacerbates the impact of these transitions.[2]

After being granted residence permit in Denmark and other developed nations, migrants are often housed in densely-populated urban areas and exposed to drastic lifestyle transitions. Migrants may also suffer from previous mental and physical trauma, which - in addition to cultural and language barriers - can limit their ability to gain higher-income employment and escape their disadvantageous socioeconomic position.[3] This leaves migrants particularly vulnerable to T2D, as they are positioned at the point of convergence of these exposures: genetic predisposition, rapid lifestyle transitions, and the disparate social impact of T2D.

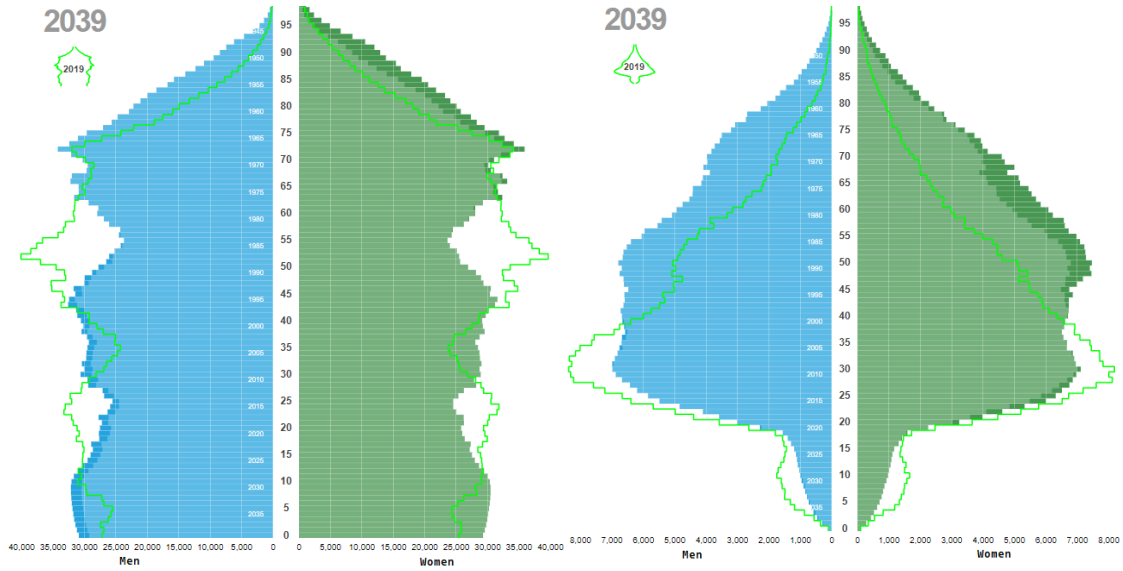
In Denmark and many other developed nations, healthcare systems are facing a challenge to sustain the expected high quality of care in an aging population with a growing burden of chronic diseases and comorbidities.[4,5] Prioritization of limited resources is crucial to ensure that those with the highest need for (and potential benefit from) healthcare receive adequate care. Migrant origin is a well-established risk factor of T2D, and migrants have

a higher prevalence of T2D compared to native populations today, but the increased risk of developing T2D is partially mitigated by their younger age.[6,7] However, this only delays the impact of their increased T2D risk, and as these migrant populations age in the coming years, they will make up a growing proportion of T2D patients in the country - 9% of immigrants in Denmark were aged 66 years and above in 2019, a proportion projected to double by 2039[8]. Therefore, understanding potential disparities in care and the specific healthcare needs of migrant populations with T2D can inform the prioritization of healthcare resources, ultimately ensuring that all individuals with T2D can continue to receive appropriate care.

Currently, there is a lack of evidence on T2D care in migrants, and research in this area faces several challenges that may explain the scarcity of evidence. Not only are migrants a minority, which limits the size of potential study populations, they are often a hard-to-reach group in studies that require patient engagement for inclusion, which exacerbates sample size problems and can induce selection bias in such study designs.[9] Furthermore, migrants have historically been defined by their racial appearance and ethnic minorities categorized accordingly, particularly in Europe and the United States. While these categories allow researchers to identify racial minorities, they obscure the role of migrants in the modern era, and provide only indirect evidence on migrant minority groups[10].

1.2. Migrants in Denmark

In 2019, Denmark had a population of approximately 5.8 million residents, of whom 600.000 (10%) were migrants and 1 million (18%) were aged 66 years and older. By 2039, the population is expected to have grown to 6.2 million, with a migrant population of 750.000 (12%) and 1.5 million individuals (23%) aged 66 years and older.[8] Figure 1.1 shows the projected demographics of native Danes and migrants.



(a) Native Danes 2019-2039

(b) Migrants 2019-2039

Figure 1.1.: Demographic projections of native Danes (a) and migrants (b) in 2039. Demographics in 2019 are outlined in green. Adapted from Statistics Denmark[8]

1.2.1. Migration

Individuals migrate between countries for various reasons, such as the pursuit of work, education and family reunification, or to seek protection. Due to different international rights, a clear legal distinction is made between 'refugees' and 'migrants'. Refugees are defined by the United Nations High Commissioner for Refugees, and protected in accordance with international law:

> Refugees are persons who are outside their country of origin for reasons of feared persecution, conflict, generalized violence, or other circumstances that have seriously disturbed public order and, as a result, require international protection.[11]

By contrast, there is no formal legal definition of an international migrant, although anyone changing country of usual residence, irrespective of the reason for migration or legal status, is generally considered a migrant.[11]

In Denmark, migrants seeking protection are processed by authorities to determine if they quality for refugee status, and whether they have previously applied for asylum elsewhere in the European Union per the Dublin Regulation.[12] Individuals qualifying for

refugee status are granted asylum and receive residence permits. Conversely, individuals with immediate family members holding Danish citizenship (or with residence permit granted for reasons other than refugee status) can apply for a residence permit on grounds of family reunification before entering the country,[13] and residence permits can also be granted to citizens of the European Union, as well as individuals with work or study obligations in Denmark.[14,15] Citizens of Finland, Iceland, Norway and Sweden are free to enter, reside and work in Denmark.[16]

Once residence has been granted, regardless of the reason, these individuals have access to the same public healthcare services as Danish citizens. In addition, any patient in the public healthcare system who is deemed in need of interpreting assistance (by the treating physician) has a right to a professional interpreter. The public healthcare system covers the initial cost of interpreter services, but since February 2018, they have been obliged by law to charge a fee to patients residing in Denmark for more than three years at the time of interpreting assistance.[17]

1.2.2. Migrants in the Danish data infrastructure

In the context of Danish registers and this PhD project, both refugees and immigrants are referred to as migrants after obtaining residence. In contrast to the automated collection of data on events occurring in Denmark, migrants and their past pose a challenge to register data collection.

Basic civil registration data on migrants is collected during processing of residence applications. Generally, registrations pertaining to events that occurred in another country prior to immigration require technical evidence (e.g. clinical imaging to determine age and verify date of birth) or legal proof from the issuing authority of the country where the event occurred. However, in cases where such documentation is unavailable, registrations may be supported by undocumented information provided by the migrant.[18] In cases where information was not systematically collected at the time of migration, e.g. an individual's level of education attained before migration, surveys have been used to collect this data retrospectively, or it may be imputed based on available data.[19]

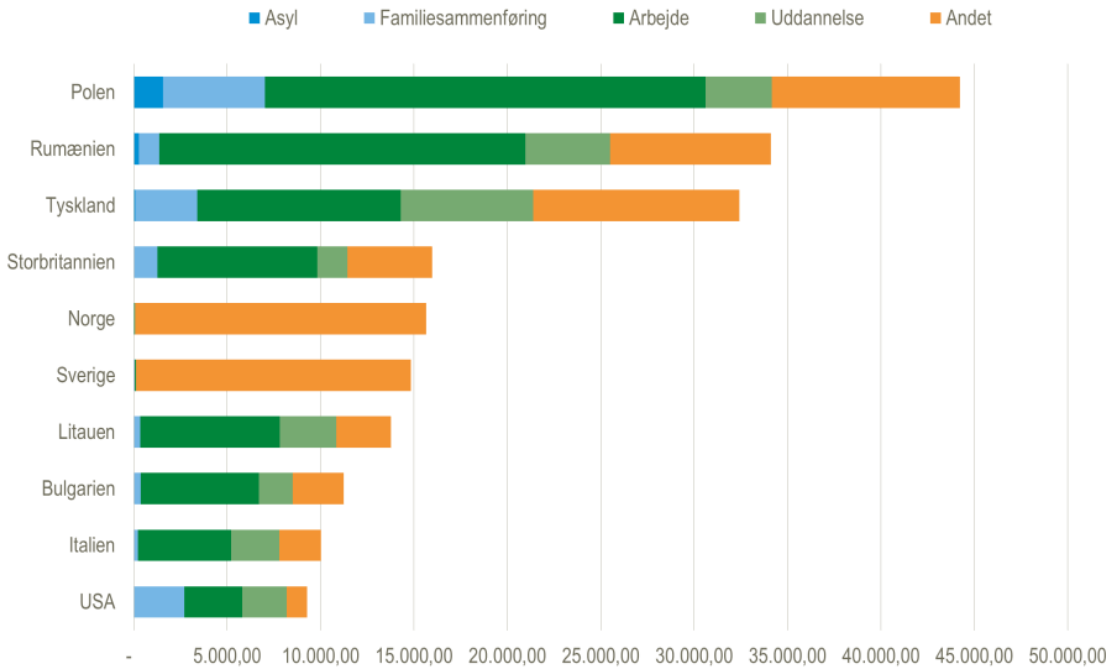
1.2.3. Migrant groups in Denmark

In Danish civil register data, a migrant's country of origin is defined by Statistics Denmark according to the following rules (paraphrased):

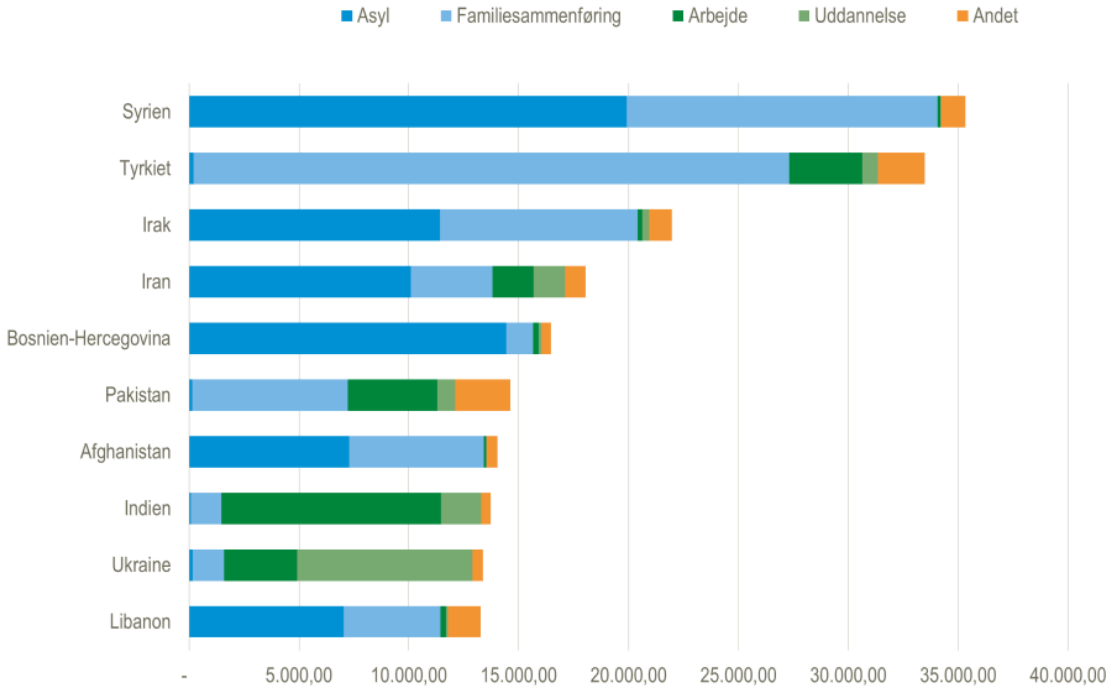
> An immigrant is a person born abroad to non-Danish citizens. If Danish civil records contain data on a migrant's mother, country of origin is defined

> by the mother's country of birth. Otherwise the father's country of birth is used, or - if no data is available - country of origin is defined by the country of birth of the immigrant themselves.[20]

Migrants from more than 200 different countries of origin reside in Denmark, making the migrant group very heterogenous with respect to geographic and cultural background.[20] Furthermore, the context of migration differs greatly between migrants of different origins in terms of when and why a particular minority migrated to Denmark. For example, the vast majority of migrants from Turkey were granted residence due to work and family reunification in the 1960s and 1970s, while almost all migrants from Bosnia-Herzegovina were refugees from the Yugoslav wars of the 1990s. By contrast, migrants from other European countries came to Denmark mainly for work, education or other reasons at various points in time. Figure 1.2 shows the primary reason for migration among the largest migrant origin groups residing in Denmark, while Figure 1.3 shows the calendar year of immigration in different migrant groups.



(a) Reason for migration: Western countries



(b) Reason for migration: Non-western countries

Figure 1.2.: Primary reason for migration among the 10 largest migrant groups in Denmark from Western (a) and Non-western countries (b). Adapted from Statistics Denmark.[21]

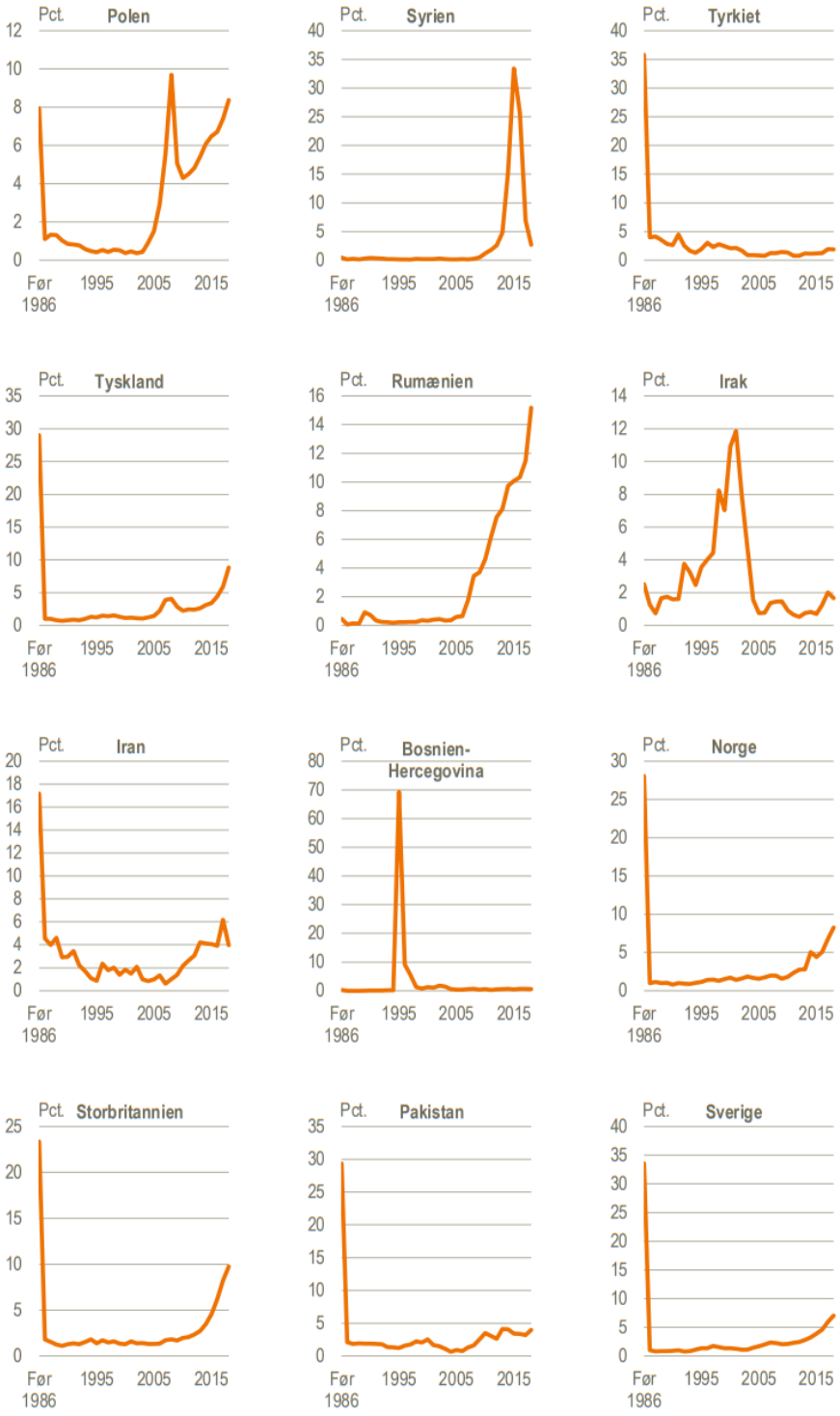


Figure 1.3.: Calendar year of first immigration among the 12 largest migrant groups residing in Denmark in 2019. Viewer discretion is advised: Y-axes vary between plots, and the line graphs should be interpreted as histograms with a bin for every calendar year. Adapted from Statistics Denmark.[20]

The limited size of each ethnic minority and the strongly contrasting characteristics between them poses a challenge to statistical processing. Researchers must strike a balance between aggregation to achieve adequate sample sizes while preserving clinically meaningful minority groups, and avoid introducing biases due to aggregation of dissimilar groups.[22] For statistical purposes, different geoschemes have been used by different institutions depending on the context and purpose. The Statistical Division of the United Nations developed the *M49 Standard Country or Area Codes for Statistical Use,*[23] while Statistics Denmark group migrants into Western and Non-western countries, respectively, a grouping that resembles the more widely-used classification of developed and developing countries.[24] See Chapter 3 for a description of grouping of countries of origin in the context of this PhD project.

1.2.4. Descendants

Descendants of migrants pose an additional challenge to classification for research purposes, as they can be considered part of the migrant group, the native population, or their own distinct demographic. In the context of T2D today, descendants of migrants in Denmark are a very young demographic unlikely to constitute a substantial proportion of the patient population (less than 1% of descendants are aged 66 years or older).[8] There is also some evidence from Sweden suggesting that descendants are a distinct group from first-generation immigrants in terms of T2D and mortality.[25] In the context of this PhD project, descendants were not a focus of investigation, and the term migrants specifically refers to first-generation migrants.

1.3. Type 2 diabetes care

1.3.1. Complications and risk factors

1.3.1.1. Overall

Complications of diabetes are traditionally divided into macrovascular (e.g. *cardiovascular disease* (CVD)) and microvascular complications (e.g. kidney disease, retinopathy, neuropathy), all of which contribute to morbidity and mortality in T2D. Several lifestyle and physiological factors influence the risk of developing complications, and clinical care revolves around management of these. In addition to being key risk factors of complications, blood pressure and levels of *hemoglobin-A1C* (HbA1c) and *low-density lipoprotein cholesterol* (LDL-C) drive indications for treatment and clinical decision-making in T2D.[26]

1.3.1.2. In migrants

In Denmark and several other European countries, HbA1c-levels are higher in migrants than in native populations[27–32], but LDL-C levels have only been sparsely described.[33] The evidence on complication risk is inconsistent, but indicates a higher risk compared to native populations, although risks differ between complication types, migrant groups and countries. Similarly, higher mortality among migrants with diabetes has been reported in some studies,[7,34] while other studies have found similar or lower risk [manglerrefs]

1.3.2. Pharmacological treatment

1.3.2.1. Overall

Due to their effects on HbA1c, LDL-C, and blood pressure, *glucose-lowering drugs* (GLD), *lipid-lowering drugs* (LLD), and antihypertensive drugs are critical parts of T2D care. In T2D patients with particularly high risk of complications, pharmacological treatment also includes *antiplatelet therapy* (APT), as the complication risk-lowering effect outweighs the risk of adverse events in these patients[35,36]. In addition to their effect on biomarker levels, certain drug types within the above classes of drugs have other positive effects. In particular, the GLD types *sodium glucose co-transporter type 2 inhibitors* (SGLT2i) and *glucagon-like peptide-1 receptor agonists* (GLP1RA) substantially reduce the risk of adverse cardiovascular and renal events beyond their effect on HbA1c. Similarly, the antihypertensive drugs types *angiotensin-converting enzyme-inhibitors* (ACEI) and *angiotensin receptor blockers* (ARB) reduce the risk and improve the prognosis of *diabetic kidney disease* (DKD) beyond their effect on blood pressure. [manglerref]

Combination therapy - the use of multiple drug types with differing mechanisms of action - is a way to increase treatment intensity and achieve treatment goals with fewer side-effects. Combination therapy is particularly important in T2D, as GLD monotherapy fails to achieve target levels of HbA1c in many patients, even at maximal dosage. [manglerref]

1.3.2.2. In migrants

Evidence from several European countries shows that GLDs are prescribed earlier and/or to a larger proportion of migrants compared to native populations[manglerref]. However, disparities in the quality of the prescribed GLD treatment, such as the use of combination therapy or the types of GLD prescribed, has not been studied. Indirect evidence suggests that such disparities are likely, as non-white minorities in the UK were slower and less likely to intensify to combination therapy than white groups[manglerref9], and

migrants in Italy with T2D purchased fewer packages of non-insulin GLD per person-year than their native counterparts[manglerref10], and socioeconomic[manglerref21-25] and racial[manglerref22, 24, 27] disparities in prescribing of SGLT2i and GLP1RA have been reported in several countries.

In contrast, studies from European countries have reported lower rates of prescribing of LLD in migrants with T2D *[care_qual_migrants refs 40, 5, Eastwood]*. Treatment with LLD is not indicated in all patients with T2D, however, and only a few of these studies accounted for treatment indication.

Evidence on ACEI/ARB use and APT in migrants with T2D is very limited, but a study from Italy found lower odds of treatment in most migrant groups [42], indicating a potential for disparities in these areas of care.

1.3.3. Monitoring

1.3.3.1. Overall

In T2D, timely monitoring of risk factors and complications allows faster adjustment of treatment, and monitoring has previously been used as a process indicator of care quality, although the direct evidence on its influence on outcomes is sparse. *[see refs* [7-12, e.g. https://doi.org/10.1093/intqhc/mzl023, , evt. DVDD rapport]*

1.3.3.2. In migrants

The evidence on monitoring of T2D in migrants is inconsistent between countries and different aspects of monitoring. In Norway, migrants with T2D received similar or more monitoring than native Norwegians[manglerrefTran2010], while migrants with T2D in Italy received fewer referrals for consultations than native Italians[manglerrefMarchesini2014].

1.4. Clinical guidelines for type 2 diabetes care

1.4.1. Context

In Denmark, T2D care is most often provided by *general practitioners* (GPs) and other actors in the primary care sector, although some patients are treated in the outpatient hospital setting. The Danish College of General Practitioners publishes national clinical guidelines for T2D in cooperation with the Danish Endocrine Society, and as well as

separate guidelines for pharmacological treatment. Although these guidelines advise that monitoring intervals, biomarker goals and treatment intensity are adapted to fit the individual patient, they do provide specific recommendations that may be used as indicators of care quality.

1.4.1.1. Migrants in guidelines

Danish guidelines contain a chapter on vulnerable populations, which features a section on cultural and language barriers. Here, the special challenges facing migrants in T2D are noted (high prevalence and early onset of disease, poor metabolic control and early onset of complications), and guideline recommend the use of a professional interpreted when facing language barriers, but specific recommendations on care are not provided. An appendix on ethnic minorities further notes how diabetes-friendly diet and physical exercise can be particularly challenging in these groups, but does not provide recommendations for specific care, except on accomodation of insulin-therapy during religious fasting. In guidelines on pharmacological treatment, migrants are not addressed in the main contents, nor in the appendix on treatment of special groups.[37]

The sparse recommendations on clinical management of migrants with T2D is not unique to Denmark, however, as guidelines in nearby countries are similarly lacking:

- Sweden: Besides stating the need for providing care regardless of ethnicity or other personal and sociodemographic characteristics, Swedish guidelines contain no mentioning of migrants at all.[38]
- Germany: Ethnicity is noted as one of several contextual factors influencing morbidity in T2D, but migrants or ethnic minorities are not mentioned otherwise.[39]
- The Netherlands: Dutch guidelines recommend increased screening for T2D in individuals of Hindustani, Turkish, Moroccan or Surinamese origin. Guidelines state that provision of diabetes care in these groups should be similar to that in the native Dutch population, and no specific recommendations on care in migrants are provided, although it is noted that some treatment options appears to be less effective in these groups (lifestyle interventions, bariatric surgery).[40]
- UK: In UK guidelines, BMI-thresholds for obesity are adjusted in ethnic minorities, and an increased risk of complications in some ethnic groups is noted. No specific recommendations for clinical management of migrants is made, however, citing a lack of evidence on ethnicity-specific effects of pharmacological treatment - instead, a recommendation for research into the effectiveness of SGLT2i in different ethnic groups is made.[41]

- Norway: The special needs of migrants were recognized in Norwegian guidelines, and a reference group on migrants and diabetes was established among the authors. However, the topic did not receive a separate guideline chapter (the only topic among 13 author group topics without a separate chapter), and recommendations specific to migrants are limited to more liberal screening for diabetes, use of a professional interpreter, and attention to diet and religious fasting.[42]

1.4.1.2. Historical revisions

Danish guidelines for type 2 diabetes care have existed for several decades, but a strictly evidence-based format was first introduced in the 2004 revision. This edition included a chapter on T2D in migrants that recommended the use of a professional interpreter when encountering language barriers, and noted the high prevalence and early incidence of T2D in migrants from third-world countries. Notably, this separate chapter on migrants was not included in revisions after 2004.[43]

The guidelines are continuously updated as new evidence emerges. In this PhD project, T2D care was studied in the time period covered by the 2012[44] and 2019[45] revisions. These revision brought minimal changes to the recommendations relevant to this PhD project, which were largely unchanged from 2004 until the time period of the studies associated with this dissertation. These recommendations are described below as they appeared in guidelines in effect at the time - with highlights of any mentioning of migrants.

Current and future readers will note that some recommendations on pharmacological treatment have been rendered obsolete by newer evidence (e.g. recommendations for SGLT2i and GLP1RA were expanded to include all T2D patients with CVD, multiple risk factors for CVD, or chronic kidney disease, irrespective of HbA1c level in the 2022 revision, which also expanded the recommendation for APT to include individuals with kidney failure and concurrent microalbuminuria).

1.4.2. Diagnosis

HbA1c values $\geq 48 mmol/mol$ are diagnostic of type 2 diabetes, but diagnosis must be confirmed with a repeated sample on a different day, unless symptoms of diabetes are present. HbA1c is not suitable for diagnosis of T2D in certain groups, e.g. pregnant women and individuals with hemoglobinopathy. Once diagnosed, patients with T2D should be considered permanently affected by the disease, and the associated risk factors for complications should be treated regardless of subsequent normalization of HbA1c.

Guidelines mention ethnicity as a risk factor for T2D, and recommend considering HbA1c screening in migrants from Asia, Africa and the Middle East, depending on other risk factors and age.

1.4.3. Monitoring

The risk biomarkers HbA1c and LDL-C should be measured at least once a year, as well as urine albumin-to-creatinine ratio (UACR) to screen for DKD. Repeated samples of UACR $\geq 300 mg/g$ is considering diagnostic of diabetic kidney disease (DKD). Diabetic retinopathy should be screened by an ophthalmologist every second year, although ophthalmologists can adapt flexible intervals depending on individual retinopathy status and other risk factors.[46] Similarly, screening for diabetic foot disease should be performed by a podiatrist every year - or, alternatively, by the GP if podiatrist service is unfeasible for the patient. At the initial diagnosis of T2D, a baseline-screening of all five types of monitoring is recommended.

Guidelines provide no specific recommendations on monitoring in migrants.

1.4.4. Biomarker levels

For HbA1c, the goal should be the lowest possible level without hypoglycemia or inappropriate polypharmacy. Specific targets at $48 mmol/mol$, $53 mmol/mol$, $58 mmol/mol$, and $70 mmol/mol$ are provided depending on diabetes duration and severity, complications, age and comorbidities.

For LDL-C, the level should be as low as possible. A general target below $2.6 mmol/L$ is provided, while the target in patients with complications should be below $1.8 mmol/L$.

Guidelines provide no specific recommendations on HbA1c or LDL-C targets in migrants.

1.4.5. Pharmacological treatment

1.4.5.1. Glucose-lowering drugs

Combination therapy: GLD therapy is recommended in patients with HbA1c 48 mmol/mol, and metformin is the recommended first-line treatment. There is no overall recommendation of second- or third-line drugs, although SGLT2i or GLP1RA should be considered in patients with CVD when metformin or other treatment is insufficient to reach the patient's HbA1c target.

Individual drug types: Guidelines recommend factoring in all effects - positive and negative - as well as cost when choosing between individual drug types, and do not present a fixed hierarchy between GLD drug type. [*maybe summarise guideline considerations on second- and third-line drugs, similar to* https://vejledninger.dsam.dk/media/238.png]

1.4.5.2. Lipid-lowering drugs

Treatment with LLD is recommended to all individuals with prevalent CVD, diabetic nephropathy, or LDL-C above 2.5 mmol/L.

1.4.5.3. Antihypertensive drugs

In addition to treatment of hypertension, use of either ACEI or ARB is recommended in individuals with prevalent CVD, microalbuminuria, or diabetic kidney disease.

1.4.5.4. Antiplatelet therapy

APT is recommended in individuals with prevalent CVD, diabetic kidney disease or very high risk of CVD due to other risk factors.

Guidelines provide no specific recommendations on any aspect of pharmacological treatment in migrants.

1.5. Identification of type 2 diabetes cases in healthcare registers

1.5.1. Danish register data infrastructure

Several types of individual-level data, e.g. civil registration, public healthcare contacts, and drug prescriptions, are automatically collected on all residents in Denmark and stored in nationwide Danish registers by Statistics Denmark and the Danish Health Data Authority (see Chapter 3 for details on contents of registers). These agencies are legally allowed to share the register data for research purposes, thus creating a set of common data sources shared by researchers with access to the data. Any researcher associated with an approved Danish research institute (mainly Danish universities) can apply for access, but fees and conditions apply.

Register data is generally accessed and processed by approved researchers on remote servers operated by Statistics Denmark and the Danish Health Data Authority, and all researchers can potentially access the same raw data - stored in the same structure, unless altered by the researcher (or their data manager). The identical raw data and virtual working environment potentially enables reproducible research. That is, any data processing workflow should be transferable between research projects if the underlying code is designed with reproducibility in mind - if the code is shared ("open-sourced"). While reproducibility in research relates to transparent reporting of methods to enable others to reproduce analyses and experiments, this also applies to a diabetes classification program, which - if reproducible - could be reused by any researcher with access to the necessary register data to dynamically identify a study population of individuals with diabetes for their research needs.

1.5.2. Benefits

Healthcare registers provide a powerful data source for population-wide studies without the need to engage the patient for inclusion. The ability to include all individuals in the population makes them particularly suited for studies of migrants with T2D, where sample size is a limiting factor and other data sources may be vulnerable to selection bias in these groups.

1.5.3. Challenges

General-purpose registers and other administrative databases often provide the basis of diabetes epidemiology, but rarely contain validated diabetes-specific data. If the diabetes-specific data is not accurate, bias may be induced into studies. Thus, it is important to have an accurate tool to identify individuals with diabetes in the registers, as findings may differ with various diabetes definitions. Considerable efforts have been made towards establishing such a tool for diabetes research in several countries, including Denmark.

Internationally, some algorithms classifying T1D and T2D have been validated in cohorts of individuals with diabetes, but none have been validated in a general population. In a general population, classification algorithms (classifiers) must not only discern type 1 diabetes (T1D) from T2D, but also identify diabetes while accounting for events that might lead to inclusion of non-cases, such as use of GLD in the treatment of other conditions. Currently, no type-specific diabetes classifier has been validated in a general population, which leaves register-based studies in this area vulnerable to biases.

1.5.4. Register-based diabetes classifiers

In Denmark, the first resource readily available to researchers to identify diabetes cases using register data was the National Diabetes Register, established in 2006.[47] The National Diabetes Register was discontinued in 2012, and a later validation study questioned its validity and called for future registers to adopt inclusion based on elevated HbA1c levels.[48]

Launched in 2014, the Register of Selected Chronic Diseases (RSCD) is currently the only publicly available resource to identify diabetes cases in Danish register data (by application to the Danish Health Data Authority), but it has not been publicly validated nor is the source code behind the algorithm publicly available. Notably, the algorithm lacks inclusion based on elevated HbA1c levels.[49]

Since the launch of RSCD, nationwide laboratory data on HbA1c testing has become available in the Danish register ecosystem,[50] but have yet to be incorporated into available diabetes classifiers.

1.6. Introduction at a glance

T2D is a disease that carries serious risks, but several treatment options are available to control the risk factors in T2D and reduce the risk of complications and death. In an aging population, the burden of T2D on healthcare services is rising, and resources must be prioritized to those with higher needs to ensure adequate care.

Migrants are particularly vulnerable to T2D, having an increased prevalence of the disease compared to native Danes. Despite having higher HbA1c levels than their native counterparts, evidence on complication risk and mortality in migrants is inconsistent, but some evidence suggests a higher risk. Evidence on disparities in care that may contribute to this excess risk in migrants is limited, and research in this area faces several challenges, such as limited sample sizes and migrants being hard-to-reach for intervention and survey studies. Register-based studies are well-suited to handle these limitations, but currently there is no validated method to identify T2D cases in a general population, which leaves register-based studies in this area vulnerable to biases.

In Denmark, guidelines are available to direct clinicians and patients towards the best possible outcomes while prioritizing resources. Despite the special challenges facing migrants, current guidelines do not contain specific recommendations for T2D care in migrants, perhaps due to the limited evidence on care disparities between migrants and native Danes.

2. Aims

2.1. Overall aims

The overall aim of this PhD dissertation is to provide robust knowledge on specific areas of T2D care where disparities between migrant groups and native Danes may be present.

2.2. Specific aims

I. Develop a register-based classification of type 1 and type 2 diabetes incorporating previous recommendations on data sources and make it available to other researchers.

II. Validate register-based diabetes classifiers available to researchers in Denmark.

III. Assess prevalence of T2D in migrant minorities of Denmark, and explore disparities between migrant groups and native Danes in indicators of guideline-recommendations on monitoring, biomarker control, and pharmacological treatment.

IV. Evaluate disparities in the quality of glucose-lowering pharmacological treatment between migrants and native Danes in terms of combination therapy and drug types.

2.3. Outputs

Aims II, III, and IV were pursued in the respective papers I, II, and III associated with this dissertation, while aim I was pursued in the creation of the *Open-Source Diabetes Classifier* (OSDC), available in the open-source R package *osdc*.[51]

3. Material and methods

This chapter describes the healthcare setting and the register data sources underlying the studies of this PhD project. Secondly, the diabetes classification algorithms used in the project are described, as well as the sources of survey data used to validate them. The last half of the chapter goes over the study designs, populations and statistical methods of the three papers associated with the dissertation.

3.1. Setting: The Danish healthcare system

Denmark is divided into 5 administrative regions that are responsible for providing public healthcare services to their residents through public hospitals and contract agreements with private actors in the primary care sector. Private actors include GPs and other specialist practices (e.g. ophthalmologists), as well as other healthcare professionals (e.g. podiatrists). In addition, the 98 Danish municipalities are obliged to offer resources to prevent and/or manage chronic disease, such as programs on managing T2D through diet and physical exercise.

Although referral to outpatient hospital care is possible, patients with T2D are primarily treated at the GP, which native Danes and migrants with residence permit have free access to. In the tax-funded Danish public healthcare system, Danish citizens and migrants with residence permit have equal access to services, and nearly all services are provided free of charge, although T2D patients incur some out-of-pocket co-payments for podiatrist care and prescription drug purchases at pharmacies. Yearly co-payments towards medication expenditures are partially covered from €132 and expenditures beyond €553 are fully covered (2019 limits),[52] while 50% of all expenditures towards diabetes-specific podiatrist care are covered by the the public health insurance.[53]

3.2. Register data sources

In Denmark, civil registration, public healthcare contacts, and drug prescriptions are recorded with a unique personal identifier (the *CPR number*) given to all Danish residents at immigration or birth. This enables coverage of the entire population and complete

linkage between data sources, and forms the foundation of several nationwide Danish healthcare registers.[54] The following section briefly describes the data extracted from these registers, as well as other data sources used in the PhD project. A detailed list of variables and code values used is available in Supplementary material B.

3.2.1. The Danish Civil Registration System

The Danish Civil Registration System contains civil registration data on all residents in Denmark. Information on migrations and country of origin was extracted, in addition to data on age, sex, equivalised disposable household income,[55] employment status, and region of residence.

3.2.2. The Danish National Patient Register

The Danish National Patient Register contains information on all hospital admissions and outpatient contacts. Data on diagnoses of diabetes and cardiovascular complications, and procedures relating to retinal screening and treatment of cardiovascular complications was extracted from 1994 through 2018. At the time of this PhD project, the clinical coding system in use was the International Classification of Diseases, 10*th* revision (*ICD-10*), implemented in 1994.[56]

3.2.3. The Danish National Health Service Register

The Danish National Health Service Register contains information on all public healthcare services provided in the primary care sector. The data is essentially billing information, and serves as the basis for payment to the healthcare provider from the Danish regions. Each type of service has a unique *SPECIALE*-code (e.g. all diabetes-specific podiatrist care services are contained as sub-codes under *SPECIALE*-code 54) Information on diabetes-specific services performed at podiatrist and ophthalmologist practices from 1990 through 2018 was extracted, as well as point-of-care HbA1c testing performed at GPs.

3.2.4. The Danish National Prescription Registry

The Danish National Prescription Registry contains information on all prescription drug purchases at Danish pharmacies, classified according to the Anatomical Therapeutic Chemical (ATC) Classification. These registrations serve to determine reimbursement

awarded to each purchase from the public health insurance. Information on purchases of GLD, LLD, ACEI, ARB, and APT from 1995 through 2018 was extracted.

3.2.5. The Danish Register of Laboratory Results for Research

The Danish Register of Laboratory Results for Research contains information on clinical biomarker samples analysed at clinical biochemistry departments in hospitals around Denmark. With the exception of point-of-care tests, all biomarker samples taken in the primary healthcare sector are sent for analysis in hospital laboratories. Compared to the other registers, this is a relatively new data source, initially launched in 2013, and only recently achieved nationwide coverage.[50]

At the time of data extraction, the Central Denmark Region, one of five administrative regions in Denmark (population of 1.3 million inhabitants, 22% of the entire Danish population) was not yet covered in the register. To address this issue, laboratory results from this part of the the population were extracted directly from the *Clinical Laboratory Information System*, a clinical database that stores all laboratory results analysed in Central Denmark Region. Data on HbA1c, LDL-C, and UACR analyses were extracted from 2011 through 2018.

3.3. Diabetes classification algorithms

As described previously in Section 1.5.4, currently available register-based diabetes classifiers have yet to incorporate the emerging register data on routine HbA1c testing, despite recommendations to do so. Wishing to take advantage of this data, we developed a new classifier, the OSDC. As a point of reference, we developed an implementation of the RSCD diabetes classifier (version 1.0, August 2016) according to official documentation.[49]

The following sections describe the algorithms behind these diabetes classifiers. Pros and cons of their design are discussed in Chapter 5.

3.3.1. The Open-Source Diabetes Classifier

This classifier first identifies a population of individuals with any type of diabetes mellitus, and then splits this population into T1D and T2D by identifying individuals with T1D and classifying the remainder of the diabetes population as T2D.

In OSDC, diabetes is defined at the second occurrence of any event across the four types of inclusion events listed below. All available data is used, except purchases of GLD, which are restricted to data from 1997 onwards.

i. HbA1c measurements of 48 mmol/mol, after censoring:
    - Results of samples taken during pregnancies - potential gestational diabetes mellitus (GDM).

ii. Hospital diagnoses of diabetes.
iii. Diabetes-specific services received at podiatrist.
iv. Purchases of GLD after censoring:
    1. Brand drugs for weight loss, e.g. *Saxenda.*
    2. Purchases during pregnancies - potential treatment of GDM.
    3. Metformin purchases in women below age 40 - potential treatment of polycystic ovary syndrome (PCOS).

Diabetes type is classified as either T1D or T2D based on patterns of purchases of insulins (including analogues) and hospital primary diagnoses of T1D and T2D. Classification as T1D requires an individual to fulfill either of the following criteria:

a. Must have purchased only insulins and never any other type of GLD, and have at least one diagnosis of T1D

b. Must have a majority of T1D diagnoses from endocrinological departments (or from other medical departments, in the absence of contacts to endocrinological departments), and a purchase of insulin within 180 days after onset of diabetes, with insulin contributing at least two thirds of all defined daily doses of GLD purchased.[57]

In populations generated on a fixed index date (such as the cross-sectional studies associated with this dissertation), individuals classified as T1D must have purchased insulins in the last year prior to the index date.

Individuals not classified as T1D are classified as T2D.

3.3.2. The Register of Selected Chronic Diseases

Formally, the RSCD includes two separate algorithms, RSCD-T1D and RSCD-T2D, that identify T1D and T2D, respectively. In practice, however, they are inter-dependent, as individuals identified by the T2D classifier are excluded from the T1D classifier.

In RSCD, diabetes is defined by type-specific diagnoses of T1D or T2D, and purchases of GLD. Individuals are defined as having T1D or T2D if they meet any of the algorithm's inclusion criteria without fulfilling any of its exclusion criteria.

3.3.2.1. Classification of type 2 diabetes

For classification of T2D, inclusion criteria include:

i. Any purchases of non-insulin GLD.
ii. A hospital diagnosis of T2D as the most recent type-specific diabetes diagnosis.

Exclusions from the T2D population include:

a. Women that have only purchased metformin and have any diagnoses of PCOS, purchases of clomifene or combination drugs containing antiandrogens and estrogens.
b. Individuals with only one recorded inclusion event.
c. Individuals with no recorded inclusion events in the last 10 years prior to the index date.

3.3.2.2. Classification of type 1 diabetes

For classification of T1D, inclusion criteria include:

i. Any purchases of insulins.
ii. A hospital diagnosis of T1D as the most recent type-specific diabetes diagnosis.

Exclusions from the T1D population include:

a. Women with any diagnoses of GDM, who have made purchases of GLD only in the period between 280 days prior to their first diagnosis of GDM and 280 days after their last diagnosis of GDM.
b. Individuals classified as T2D.
c. Individuals without any purchases of GLD, or have made only one purchase and have no hospital records of T1D.
d. Individuals with no insulin purchases in the last 10 years prior to the index date.

3.4. Questionnaire data used in validation

To validate the above diabetes classifiers, self-reported data from two surveys was retrieved. The design and contents of these surveys is described below.

3.4.1. The Health In Central Denmark survey

Health in Central Denmark is a digital and postal questionnaire survey conducted in the Central Denmark Region. The survey was conducted in 2020 on all inhabitants aged 18 to 74 years identified as prevalent diabetes cases by OSDC on 31 Dec 2018, plus an equally-sized group of OSDC non-diabetes cases (matched to diabetes cases by sex, age, and municipality). The survey collected self-reported data related to health in general, with an additional focus on items related to diabetes mellitus, such as current disease, diabetes type, and age at onset.In total, 44,659 OSDC diabetes cases and 46,195 matched OSDC non-diabetes cases were invited to the survey, and 51,854 (57%) responded.[58]

3.4.2. The National Health Survey

The National Health Survey is a repeated digital and postal questionnaire survey conducted among a representative sample of Danish residents aged 16 or above. It collects self-reported data related to mental and physical health in general, such as current or former chronic diseases, and includes a question on diabetes of any type.[59] In the 2017 survey of Central Denmark Region, 52,000 individuals were invited and 32,400 (62%) responded. Data from this survey was used in supplementary validation analyses.

3.5. Study design and population

In addition to developing a code package for the statistical programming language *R* containing the source-code behind the diabetes classifiers, three studies were conducted. Study I was designed to examine the validity of register-based definitions of T1D and T2D. After validating the two candidate classifiers, the most accurate classifier - OSDC - was chosen to identify study populations of T2D for use in studies II and III.

[Overview table of the three studies]

3.5.1. Study populations

3.5.1.1. Study I

Validation analyses were performed in a subset of respondents in the Health in Central Denmark survey.

Among survey respondents, 2,411 were excluded due to missing data on diabetes items. 2,093 individuals with self-reported onset of diabetes after 31 December 2018 were excluded, as diabetes cases onset after the index date would erroneously evaluate to false-negative cases in the register-based classifiers, due to the delay from register-classification on the index date (31 December 2018) until questionnaire responses (November 2020).

After these exclusions, 47,350 individuals remained, but due to survey invites being conditioned on OSDC diabetes status, the survey population was biased towards the OSDC classification and OSDC diabetes prevalence was 43.7% (20,692 individuals). To account for this, we first estimated the OSDC diabetes prevalence to be 9.3% (2,483 individuals) in a random sample of 26,665 individuals from the background population with the same age, sex and municipality distributions as the OSDC non-diabetes cases of respondents. To offset the oversampling of OSDC diabetes cases in the survey, OSDC diabetes cases were randomly subsampled to 2,733 individuals to achieve an unbiased OSDC diabetes prevalence of 9.3% in the final study population of 29,391 individuals.

3.5.1.2. Study II

Register data was used to identify 262,837 individuals with prevalent T2D on 1 January 2018 (index date) from a background population of 3,864,528 native Danes or first-generation migrants from the selected origin categories aged 25 - 99 years, and residing at least 3 years in Denmark on the index date.

T2D prevalence was studied in the above population. In order to study T2D care where it was clinically relevant, the analyses of T2D care were performed in a distinct subsets of the population with T2D, as outlined below:

- All analyses of T2D care were performed in a subset of 254,097 individuals with T2D onset at least 6 months prior to the index date in order to allow time for baseline-screening, initiation of pharmacological treatment and subsequent biomarker control to occur in the newly-diagnosed.
  - Monitoring was evaluated among all these individuals:
    - Monitoring of HbA1c, LDL-C, nephropathy, retinopathy and foot disease: 254,097 individuals.
  - Biomarker levels were evaluated in those with at least one measurement prior to the index date:
    - HbA1c levels: 250,075 individuals.
    - LDL-C levels: 248,813 individuals.

- Each analysis of pharmacological treatment was limited to individuals with a clear guideline-recommendation for that particular treatment according to complication status or biomarker levels on the index date (excluding individuals not alive and resident in Denmark in the year following the index date and thus unable to purchase medication):
  - GLD. HbA1c 48 mmol/mol at the most recent measurement: 140,208 individuals.
  - LLD. Prevalent macrovascular complications or DKD, or age above 40 with LDL-C above 2.5 mmol/L: 128,707 individuals.
  - ACEI/ARB & APT. Prevalent macrovascular complications or DKD: 77,414 individuals.

3.5.1.3. Study III

Register data was used to identify 275,525 individuals with prevalent T2D on 1 January 2019 (index date) from a background population of 4,084,564 individuals above age 25. Among individuals with T2D, 269,011 native Danes or first-generation migrants from the selected origin categories residing at least 3 years in Denmark on the index date were included. After exclusion of 5,518 individuals without records of HbA1c or LDL-cholesterol, the study population contained 263,393 individuals. A subsequent analysis of GLD use in the year 2021 was performed on 241,140 individuals from the above study population still alive and residing in Denmark on 1 January 2021.

3.5.2. Outcomes

3.5.2.1. Study I

The validation golden standard, self-reported diabetes type, was categorized as either T1D (self-reported T1D), T2D (all other types of diabetes) or no diabetes, corresponding to the diabetes types discernable by the register-based classifiers.

3.5.2.2. Study II

Twelve outcomes were defined as follows

- Prevalent T2D was defined as OSDC-classified T2D prevalent on the index date.

T2D care quality was evaluated in the scope of failure to meet guideline recommendations in terms of monitoring (allowing for three months of leeway), biomarker levels, and pharmacological treatment.

T2D monitoring was evaluated in time-frames prior to index date, according to guideline-recommendations:

- Monitoring of HbA1c and LDL-C, as well as UACR-screening for diabetic nephropathy were identified by records of biomarker samples between October 1, 2016 and December 31, 2017 in the *Register of Laboratory Results for Research* or the *Clinical Laboratory Information System*. Additional monitoring of HbA1c using point-of-care tests was identified by records in the *National Health Service Register*.
- Screening for diabetic retinopathy was identified by records of retinal photoscreening for diabetic retinopathy between October 1, 2015 and December 31, 2017 - either at hospital departments in the *National Patient Register*, or diabetes-specific examinations at ophthalmologist practice recorded in the *National Health Service Register* .
- Screening for diabetic foot disease was identified by records of diabetes-specific podiatrist services received between October 1, 2016 and December 31, 2017 in the *National Health Service Register*.

Biomarker levels on the index date was evaluated using the most recent measurement of HbA1c and LDL-C before the index date:

- The outcome measure for HbA1c control was a level 53 mmol/mol.
- For LDL-C, the outcome measure was a level 2.6 mmol/L.

Pharmacological treatment was assessed in the year following the index date, and treatment was defined by at least one purchase of the respective type of drug based on the following ATC codes (including subcodes):

- GLD: ATC A10.
- LLD: ATC C10.
- ACEI/ARB: ATC C09A-C09D.
- APT: ATC B01AC.

3.5.2.3. Study III

Use of individual GLD types was defined by any purchases made during the respective year (2019 and 2021). Purchase of the following types of GLD counted towards the total

number of drugs used by an individual in a year (fixed-dose combination drugs counting towards both of their component drug types):

1. Metformin: A10BA02, A10BD0, A10BD10-16, A10BD20, A10BD23.
2. Sulfonylureas: A10BB.
3. DPP4i: A10BH, A10BD0, A10BD10-11, A10BD13, A10BD19, A10BD21, A10BD24.
4. SGLT2i: A10BK, A10BD15-24, A10BX09, A10BX11-12.
5. Insulins (including analogues): A10A.
6. GLP1RA: A10BJ, A10AE5, A10BX04, A10BX07, A10BX10, A10BX13-14.
7. Thiazolidinediones: A10BG.
8. Repaglinide: A10BX02.

Separate outcomes for the use of each individual GLD type were defined for metformin, sulfonylurea, DPP4i, SGLT2i, insulins, and GLP1RA. In addition, combination therapy was analysed using the following two outcome definitions:

- Any combination therapy: two or more drug types purchased in the year
- Triple combination therapy: three or more drug types purchased in the year.

3.5.3. Primary variable of interest: migrant status

Migrant status and country of origin was defined according to the definitions by Statistics Denmark. Based on the national origins with the most T2D cases and the United Nations M49 geoscheme, the population was grouped into nine origin categories:

- Native Danes
- Europe (M49: Europe, except countries of the former Yugoslavia)
- Middle East (M49: Northern Africa, Western Asia except Turkey, plus Afghanistan and Iran)

And the six remaining individual countries with the most T2D cases (in order of number of cases):

- Turkey
- Former Yugoslavia
- Pakistan
- Sri Lanka
- Somalia
- Vietnam

3.5.4. Other explanatory variables

3.5.4.1. Socioeconomic variables

Age and diabetes-duration on index dates were treated as continuous variables. Household income was based on the 3-year average of equivalised disposable household income prior to index dates, converted to the corresponding percentile in the background population and treated as a continuous variable. As a measure of acculturation, duration of residence was calculated from the date of first migration into Denmark, converted to a categorical variable with native Danes categorized in the maximally acculturated group. Employment status was categorized as *Employed* (including students and others), *Retired* or *Unemployed* based on the year prior to index dates.

3.5.4.2. Clinical variables

Prevalent macrovascular complications were assessed by primary diagnosis codes of stroke, ischemic heart disease and peripheral arterial disease (and associated procedure codes) prior to index dates. Diabetic kidney disease was defined as a UACR 300 mg/g measured within three years prior to index dates (study II: within 27 months). Other types of microvascular complications were not included due to insufficient data.

Study III used HbA1c levels and LDL-C levels at baseline as continuous covariates for adjustment, as well as LLD use within a year prior to the index date. Finally, analyses of the use of individual GLD drug types in study III adjusted for combination therapy categorized as *Monotherapy or no GLD*, *Dual combination* and *Triple combination or more*.

3.6. Statistical analysis

Statistical analyses were performed in RStudio[60] using R version 4[61] and a range of open-source code packages for computations[62–66] and visualisation.[67,68]

This section describes our overall statistical analysis strategy, followed by the specific statistical analyses of each study.

Logistic regression - a statistical model that computes odds ratios (OR) between groups - is often applied in cross-sectional studies to estimate relative risk (RR, or risk ratio). However, odds ratios approximate relative risk **only** when the outcome is *rare*. If this is not the case, interpreting ORs as RRs will exaggerate risk estimates,[69] as illustrated in Figure 3.1 (adapted from D Wells - CC-BY-SA).

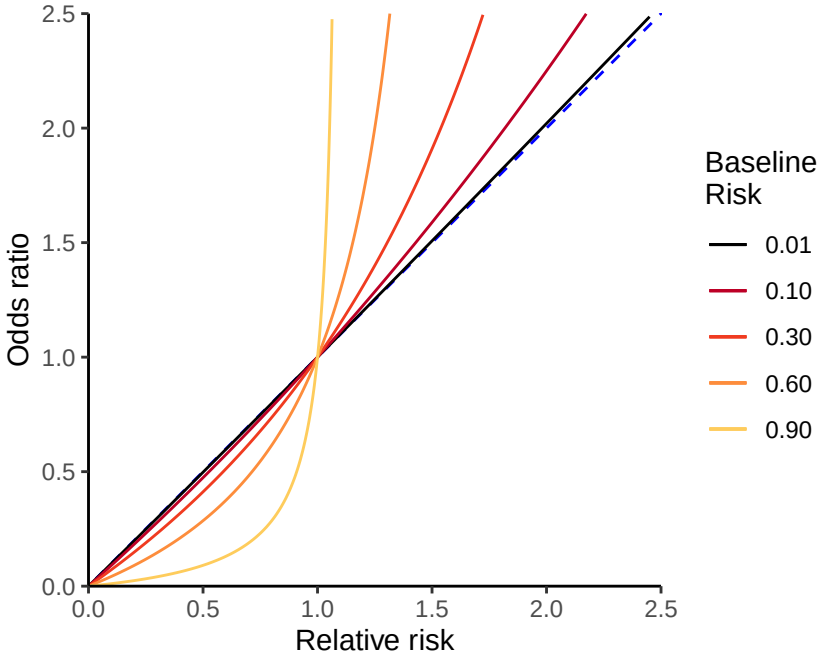


Figure 3.1.: Relationship between odds ratio and relative risk depending on baseline risk.

None of the outcomes in studies II & III were expected to be rare, and thus logistic regression was unsuitable. While log-binomial regression models are the most natural choice for computing relative risk in a mathematical sense, they are prone to convergence failure.[70] Given the large number of outcomes and models that needed fitting, this presented a serious (although not insurmountable) inconvenience, and log-binomial regression was also discarded.

Robust Poisson regression (sometimes termed modified Poisson regression)[71] presented an appealing alternative for modeling RR. A comprehensive description of the technique is beyond the scope of this dissertation, but a brief explanation is in order, since ordinary Poisson regression is not the right tool for modeling binary outcomes as such data violate the assumptions behind the model.[72] When applied to binomial data, the error for the estimated relative risk will be overestimated in an ordinary Poisson model. However, this problem can be addressed by using robust generalized estimating equations (so-called *sandwich*) variance estimates, which provide accurate standard errors in the presence of over-dispersion and excess zeros,[73] such as in a binary distribution. Ultimately, this allows robust Poisson regression to estimate RR with valid confidence intervals in data with binary outcomes.[74]

All regression models were modeled with native Danes as the reference group. To account

for potential non-linearity, continuous variables were modeled as natural splines with knots placed at each quintile of the distribution in the respective study population. Supplementary analyses where continuous variable were modeled as categorical variables (quintiles) were conducted for easier interpretation of these variables. Supplementary analyses also included estimates of absolute risk differences corresponding to the above analyses. In these analyses, a generalized linear model using a Gaussian distribution and identity link with robust *sandwich* variance estimates was used to compute absolute risk difference with 95% confidence intervals.

In all three studies, the characteristics of study populations were presented, describing continuous variables with means and standard differences, and categorical variables by absolute numbers and percentage proportions. In study I, characteristics were tabulated according to self-reported, OSDC-classified, and RSCD-classified diabetes type, while study II and III tabulated the distribution of covariates and outcomes by origin group.

The following sections describes the statistical analyses of each study.

3.6.0.1. Study I

Validation analyses were performed separately for T1D and T2D, where each register-based diabetes classifier was validated against self-reported diabetes type. Within analyses of each diabetes type, diabetes type was modeled as a distinct binary variable: does the individual have this diabetes type or not (e.g. in the analyses of T1D, diabetes type was modeled as a binary variable of T1D vs. no T1D, the latter category including both T2D and non-diabetes cases).

Concordance tables and associated validation metrics were computed (with 95% confidence intervals) as defined in equations 3.1, 3.2, 3.3, and 3.4: sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). These validation analyses were bootstrapped in 1000 random subsamples in order to assess robustness.

$$Sensitivity = \frac{N_{truepositives}}{N_{truepositives} + N_{falsenegatives}} \tag{3.1}$$

$$Specificity = \frac{N_{truenegatives}}{N_{truenegatives} + N_{falsepositives}} \tag{3.2}$$

$$PPV = \frac{N_{truepositives}}{N_{truepositives} + N_{falsepositives}} \tag{3.3}$$

$$NPV = \frac{N_{truenegatives}}{N_{truenegatives} + N_{falsenegatives}} \tag{3.4}$$

Furthermore, analyses stratified by self-reported age at onset of diabetes (including all self-reported non-diabetes cases in both strata of age at onset) were performed to assess the influence of age at onset of diabetes on sensitivity and PPV.

3.6.0.2. Study II

Robust Poisson regression was used to compute RR and 95% confidence intervals in a series of cross-sectional analyses of T2D prevalence and eleven outcomes within T2D care.

In the analysis of T2D prevalence, RR of prevalent T2D was computed in a crude model and in two models with different levels of adjustment to explore the effects of clinical (model 1) and socioeconomic factors (model 2) on migrant risks. The first model adjusted for sex and age (model 1); the second model further adjusted for employment status, household income, duration of residence, and region of residence (model 2). Additionally, age-specific T2D prevalence in each migrant group was modeled using a binomial model with log-link and restricted cubic splines as a function of age.

In the eleven analyses of T2D care, the RR of not receiving each type of guideline-level T2D care was computed in a crude model and in two adjusted models to visualize risk disparities in a clinical context separately from the fully-adjusted model. The first model (model 1) adjusted for clinical risk factors in diabetes that may influence decision-making when planning diabetes care with the patient (sex, age, diabetes duration, prevalent macrovascular complications and DKD). The fully-adjusted model (model 2) further adjusted for socioeconomic factors that may influence a patient's health behavior and healthcare service usage (employment status, household income, duration of residence, and region of residence).

In order to assess potential disparities at other cut-off values of HbA1c and LDL-C, cumulative empirical distributions were plotted with 95% confidence intervals computed with Kolmogorov-Smirnov's D.

3.6.0.3. Study III

Similar to study II, robust Poisson regression was used to compute RR with 95% confidence intervals in three models for each outcome: an unadjusted model (crude), a model

adjusted for clinical risk factors (model 1), and a model adjusted for clinical and socioeconomic factors (model 2). Compared to the models in study II, all adjusted models in this study contained adjustment for additional clinical risk factors (HbA1c level, LDL-C level, and lipid-lowering drug usage), and the analyses of individual GLD type also including adjustment for the degree of combination therapy.

3.7. Ethics and approvals

In compliance with the General Data Protection Regulation, the handling of personal data associated with this PhD project was registered in the internal register of research projects (Danish: *fortegnelsen*) of Aarhus University, journal number 2016-051-000001, serial number 1339. Study II was approved by the Health in Central Denmark steering committee. The Health in Central Denmark project is registered in the Central Denmark Region internal register of research projects (reg. no. 1-16-02-165-20).

Access to register data was provided and approved by the Danish Health Data Authority and Statistics Denmark. In Denmark, studies based entirely on survey and register data do not require further ethical approval.

4. Results in summary

[Align structure with each individual study? Or split into themes? Kind of doing both for now]

4.1. Exploratory results during development of diabetes classifiers

Insulin non-users among T1D -> Women with diabetes onset before 1997.

Self-reported no diabetes (or previous diabetes)

Insufficient obstetric coding practices before 1997.

Arguments for and against design choices goes into discussion chapter

[detailed discussion of choices in OSDC, e.g. on obstetric coding, and pros/cons of RSCD goes in chapter 5]

4.2. Study I

In this study, we identified… Number of participants, diabetes cases etc.

4.2.1. Register-based classification of type 1 diabetes

metrics

4.2.2. Register-based classification of type 2 diabetes

metrics

4.3. Study II

[Maybe for each outcome?] In this study, we identified… Number of participants, diabetes cases etc.

4.3.1. Prevalence

4.3.2. Monitoring

4.3.3. Biomarker control

4.3.4. Pharmacological treatment

4.3.5.

4.4. Study III

In this study, we identified… Number of participants, diabetes cases etc.

4.4.1. Combination therapy

4.4.2. Drug types

5. Discussion of methods

Before discussing the methodological design of the studies associated with the dissertation, this chapter first discusses algorithm design choices in OSDC and RSCD in light of discoveries made during development, implementation and validation of the classifiers.

5.1. Design of diabetes classifiers

Using data from general-purpose healthcare registers to classify diabetes cases and their specific diabetes type faces several limitations. This section discusses issues and challenges identified during the development and implementation of diabetes classifiers in this PhD project prior to the validation performed in study I, while the validation study design is discussed in Section 5.2 and its results are discussed in Chapter 6. Future perspectives of register-based diabetes classification are discussed in Chapter 8.

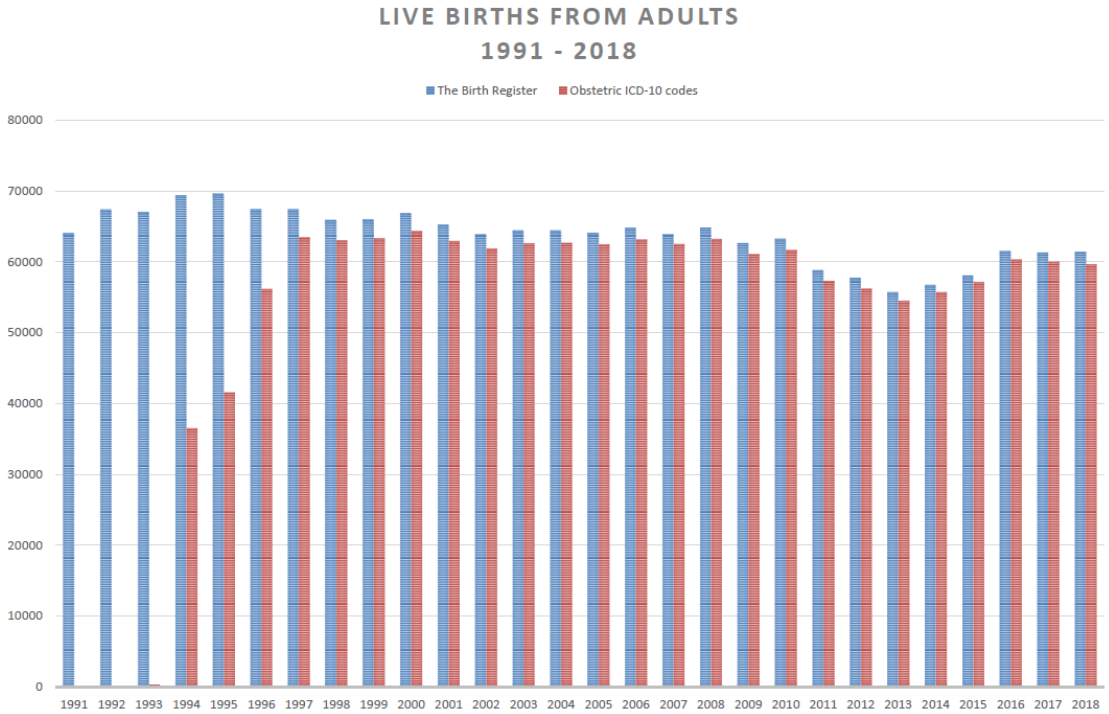
5.1.1. Approaches to censoring gestational diabetes mellitus

In an early version of the OSDC, we discovered that when using all the available data on GLD purchases, including the earliest years of the National Prescription Register, a substantial proportion of individuals classified as having T1D had no purchases of insulin during the previous year (around 8% of the T1D population in most years). This seemed unlikely from a clinical perspective, and exploring these cases among participants in the National Health Survey, we found that none of them reported having diabetes in the data, although a few reported having previously had diabetes (data not shown). This led to the decision to correct diabetes type on the index-date of cross-sectional studies.

However, pursuing this issue further showed that most of these cases were younger women with diabetes onset before 1997, which might indicate that censoring of GDM was insufficient during this period of time. Censoring of GDM relied on clinical obstetric coding in the National Patient Register to identify individual pregnancy windows, and a rudimentary validation of the number of births registered by codes in the National Patient Register against the number of annual births recorded in the National Birth Register was attempted. This showed that clinical obstetric coding was indeed lacking

in the first few years after the adoption of ICD-10 coding in 1994, but appeared sufficient from 1997 onward, as shown in Figure 5.1. This led to the decision to restrict inclusion from prescription data to 1997 onward.

Any algorithm censoring GDM using obstetrical code from data in the National Patient Register is likely to include false-positive diabetes cases among women prior to 1997. The RSCD censors GDM based on hospital diagnoses of GDM, but ultimate manages these old cases (and other potentially misclassified cases) by retrospectively excluding individuals without any inclusion events 10 years prior to a given index date. While such a mechanism solves this issue in a cross-sectional study, it creates new problems, as study designs with dynamic index dates (e.g. longitudinal studies with case-specific index dates at the time of diabetes onset) are no longer possible with the algorithm.



Births identified in the National Patient Register with obstetrical ICD-10 codes vs. registrations in the National Birth Register. Note the discrepancy in births per year during 1994-1996 after adoption of the ICD-10 coding scheme in 1994.

Figure 5.1.: **Births registered by year**

5.1.2. Approaches to censoring polycystic ovary syndrome

In Denmark, women with PCOS are mainly diagnosed and treated in the primary care sector without any registration in the National Patient Register to assist researchers identify these patients. Metformin, a GLD, is used in the treatment of PCOS, and leads to inclusion of false-positive diabetes cases among women treated with metformin. As PCOS mainly affects younger women, censoring metformin purchases among women below an arbitrary age-cutoff is a possible solution.

In OSDC, we adopted the 40-year age-cutoff previously used in the Danish National Diabetes Register,[47] and also censored metformin purchases in women where the indication code of the prescription was for PCOS. This approach over-censors purchases among younger women, but we expected this to be a minor issue, as most of these patients should be included based on HbA1c data instead. Some women may be treated for PCOS with metformin after the age of 40, and unless this was extremely rare, these were unlikely to be censored correctly, as data on indication codes only contributed very few cases of censoring.

The RSCD applies a very different approach to identify PCOS cases and censor metformin purchases, using data on purchases of contraceptives used in the treatment of PCOS, as well as any diagnoses of PCOS in the National Patient Register. This approach provides an age-agnostic censoring criteria, and as contraceptives are the first-line treatment drugs of PCOS guidelines, metformin is unlikely to be prescribed for PCOS without these. However, these contraceptives are also used in women without PCOS, which risks overcensoring women with T2D without PCOS.

5.1.3. Approaches to censoring other sources of false inclusions

Despite much of the data collection being automated, the health registers are ultimately subject to human error, and entries into the data may be the erroneous and lead to false inclusions by the classifier, e.g. a prescription for GLD intended for another patient may be written and purchased by an individual without diabetes or a wrong diagnosis may have been recorded by a hospital physician. In OSDC, diabetes diagnosis was defined at the second occurrence of an event across any of the four types of inclusion events in order to account for errors in the register data and comply with HbA1c-testing recommendations from the World Health Organization and clinical guidelines, which recommend a repeated samples before diagnosis.[45,75] The drawback of this approach was the potential delay of inclusion after diabetes onset, particularly in the years before HbA1c data was available.

As previously mentioned, the RSCD retrospectively excludes individuals without any inclusion events 10 years prior to a given index date, which avoids unnecessary delay, but accumulates 10 years of falsely included cases on any given index date.

5.1.4. Approaches to classification of diabetes type

Without an accurate, population-wide marker of diabetes type in the register data, researchers are forced to classify diabetes type using indirect markers in the data instead. Algorithm design decision in this regard are based on assumptions - which may qualify as so-called expert clinical knowledge - and come with benefits and drawbacks. For example, for type-classification in OSDC, we chose to omit diagnoses from surgical departments, as we assumed that diagnoses from medical departments, particularly endocrinology departments, would be more accurate due to these departments providing specific care for T1D and T2D.

In OSDC, we deliberately chose to make classification of T1D cases more restrictive compared to T2D, as the T1D population is much smaller than the T2D population and therefore more vulnerable to bias from misclassification, while the T2D population is more robust due to its larger size. The algorithm specifying diabetes type reflects this, as all cases of T1D must pass separate checks on medication-purchases and hospital diagnoses, and failure in either check results in classification as T2D (refer to **?@fig-osdc-type-flow**).

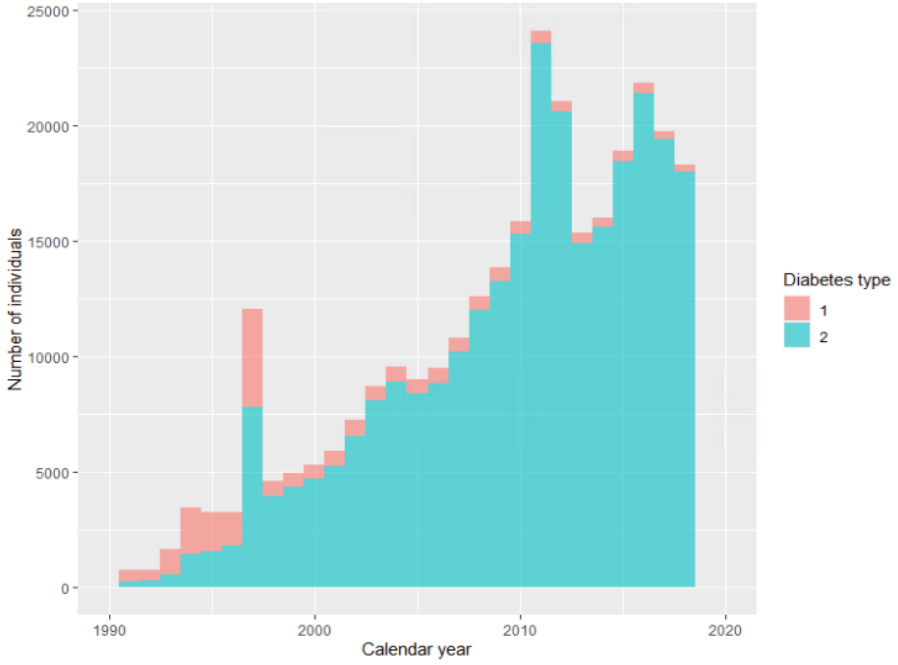
As previously presented, around 10% of T1D cases in OSDC had purchased non-insulin GLD at one or more points in time. In RSCD version 1.0, which was implemented and validated in the first study of this PhD project, these would all be classified as T2D and omitted from the T1D population, lowering the sensitivity of T1D-classification. Recent revisions of RSCD addressed this by allowing cases with non-insulin GLD purchases in the first year after onset of diabetes to be classified as T1D.[76]

Concordance was high between the two algorithms in classification of diabetes of any type vs. no diabetes, and the different approaches to classifying diabetes type is likely the main cause of the performance differences observed between the two algorithms.

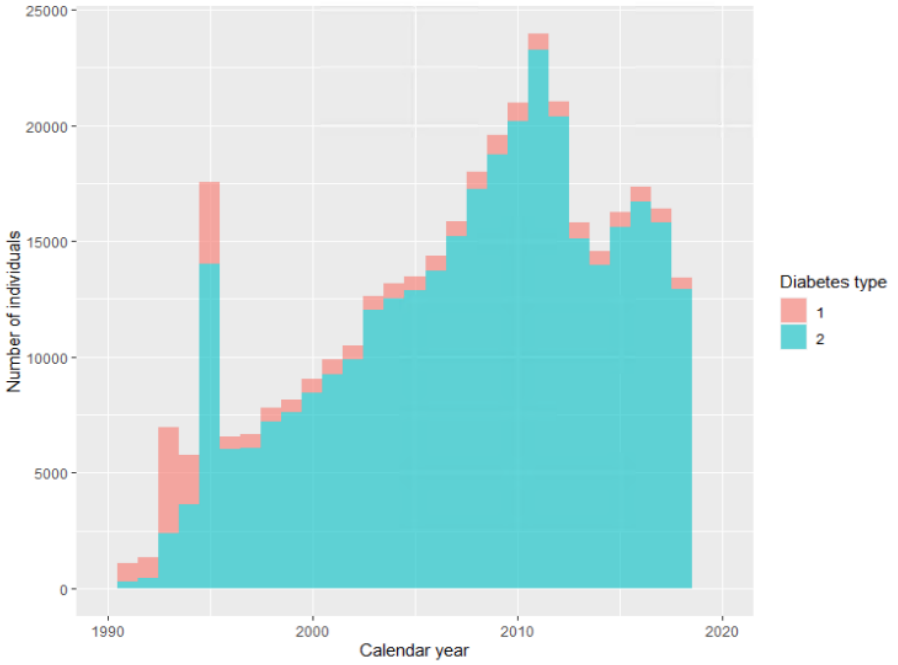
5.1.5. Incidence and demography of register-based diabetes cohorts

Both the OSDC and the RSCD depend on data sources with varying duration of data coverage. In addition, methods to diagnose diabetes have changed in the period of time covered by the data, which leads to spikes in the incidence of diabetes inclusions, as shown in Figure 5.2. Both classifiers show three incidence spikes:

- 1995/1997: The year after incorporation of prescription data.
- 2011: The year after HbA1c becoming the diagnostic standard.
- 2016: In the OSDC, this spikes is not surprising, given that this was the year after HbA1c data becoming available from the last of the five Danish regions. The reason for the increased incidence in the RSCD is less clear, but it might indicate that this is not a spike, but rather the end of a plateau that followed after the previous spike in 2011.



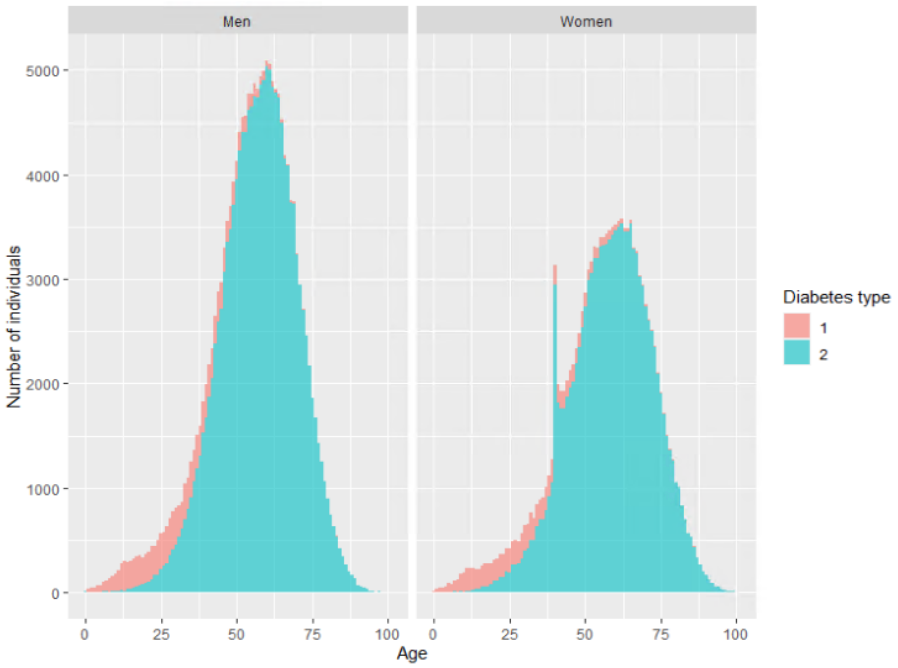
Open-Source Diabetes Classifier



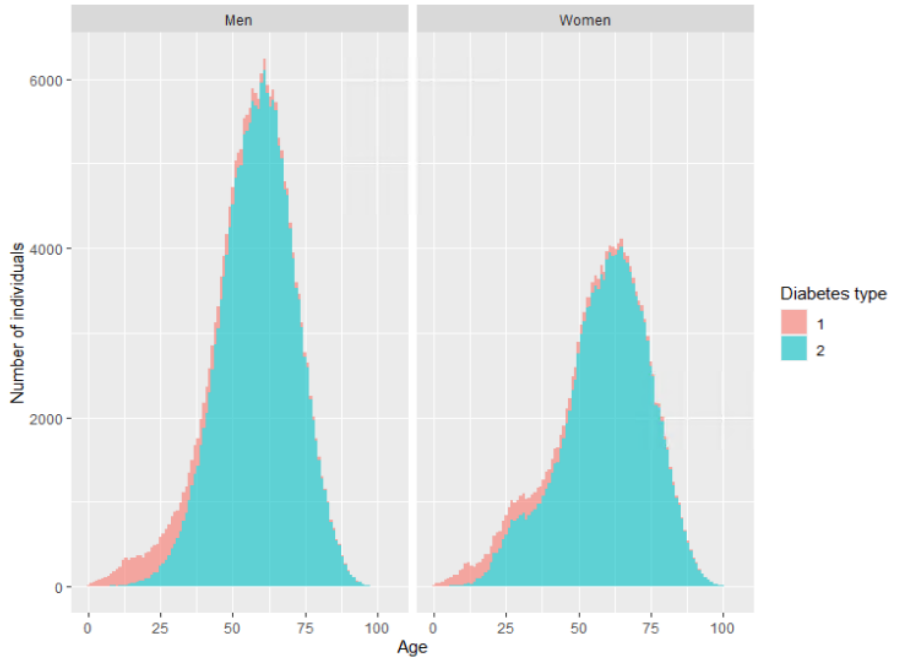
Register for Selected Chronic Diseases

Figure 5.2.: Calendar year of inclusion among prevalent diabetes cases on January 1 2019.

The different approaches to censoring between the classifiers affect inclusions, which is evident in the distribution of age at inclusion of each sex, as shown in Figure 5.3. In men, where no censoring of GDM nor PCOS is necessary, age at inclusion of T2D is normally distributed and similar between the two classifiers. For women, however, this distribution differs between the classifiers. In the OSDC, the age-based censoring of PCOS results in a spike of T2D cases included at 40 years old, while in the RSCD, the distribution is left-skewed by women included before age 40.



Open-Source Diabetes Classifier

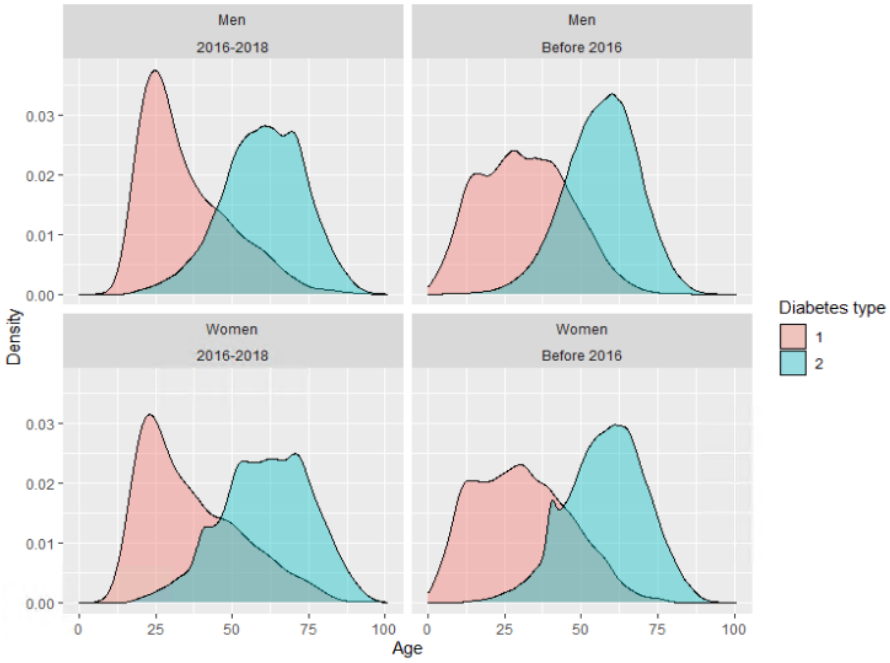


Register for Selected Chronic Diseases

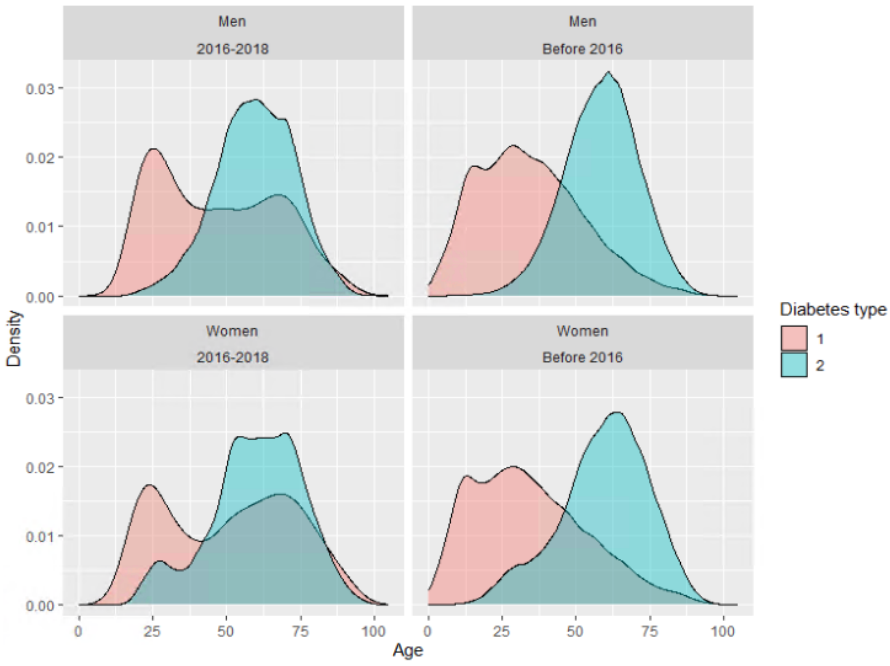
Figure 5.3.: Age at inclusion for each sex among prevalent diabetes cases on January 1 2019.

As discussed previously, we expected the addition of HbA1c data in the OSDC classifier to compensate for the overcensoring of metformin prescription data. In density plots of age at inclusion stratified by calendar year period, the observed spike in T2D among women on age 40 appeared to be due to cases with onset during years with partial or no coverage of HbA1c data, as shown in Figure 5.4. Before 2016, when HbA1c data was not available from all regions of Denmark, there is a clear spike in women with T2D at age 40. This spike is absent in subsequent years, but the distribution is slightly skewed with a plateau in the age-range from 40 to 45 years. This is likely a residual effect of women in this age-range having been prematurely included in the preceding years. In plots of the RSCD classification, the distributions of T2D and T1D are more skewed among individuals included after 2016 than before 2016. Indeed, among cases included during the most recent years, the RSCD T1D and T2D distributions appear bimodal, a visible indication of misclassification and cross-contamination of true T1D and T2D cases in the RSCD classifications.

While the different design approaches between the two classifiers resulted in differences in incidence and demography between their cohorts, their performance in classification of prevalent diabetes was comparable - as described in Chapter 4 - and the differences were not specific to T2D in women, where the design of the algorithms differed the most.



Open-Source Diabetes Classifier



Register for Selected Chronic Diseases

Figure 5.4.: Age at inclusion for each sex among prevalent diabetes cases on January 1 2019, stratified by calendar year of inclusion.

5.2. Study designs

All three studies associated with this dissertation were designed as observational, cross-sectional studies based on prospectively collected register data covered the entire adult population in Denmark. The first study also incorporated questionnaire data from the *Health in Central Denmark* survey to serve as the gold standard for validation.

This cross-sectional design is useful for assessing prevalence of an outcome or exposure at a particular point in time, and any associations between them. Compared to a longitudinal design, the major limitation of this design is the inability to establish temporal or causal relationships between exposures and outcomes. These limitations were not a concern in the context of this dissertation, as the aim was to map risks in migrants, rather than infer causality, and since the exposure (migrant origin) was defined at birth, long before any of the outcomes were assessed.

The use of nationwide register data enables the largest possible sample size and statistical power, with the drawback that the data is not collected nor validated for a specific research purpose.

The generalizability of the findings of this dissertation are discussed in the following sections.

5.3. Internal validity

5.3.1. Selection bias

> Selection bias is a type of bias that may arise when the selection of participants into a study is not representative of the target population, which can lead to inaccurate or biased estimates of the association between exposure and outcome that do not apply to the target population.

In study I, validation was performed in a population of respondents to a survey, a design prone to selection bias. Invites to the survey were conditioned on being aged 18 to 74 years and classified as a diabetes case according to OSDC, or being sampled in a group matched by sex, age and municipality. The choice to subsample OSDC diabetes cases when deriving the study population compensated for the oversampling of OSDC cases compared to non-cases in the survey population, but did not change the other characteristics in the survey population. As a results, the study population contained fewer women and a skewed age distribution compared to a random sample of the general population - and a higher diabetes prevalence.

The accuracy of the classifiers may be different among individuals older than 75, and the predictive values will vary depending on diabetes prevalence. However, supplementary validation analyses performed in the *National Health Survey* - which contains a representative sample of the general adult population - provided results similar to those in the *Health in Central Denmark* survey. This does not rule out selection bias, however, as the data in both surveys only covered the Central Denmark Region, and the supplementary analyses were limited to classification of diabetes of any type vs. no diabetes - see Appendix C for details.

In addition to survey invites, non-response among those invited may also be a source of selection bias, as analyses were restricted to respondents. However, several factors suggest that this bias was limited, as response rates were high (>50%) in both the T1D, T2D and no-diabetes groups, and the three groups shared similar non-response patterns.[58]

The register-based design of studies II and III meant that data was available on the entire population, minimizing selection bias. Statistical analyses were performed as complete-case analyses, which omits individuals with missing data from analyses and may induce selection bias. In order avoid this, individuals with very short durations of residence on the index dates of the studies were excluded in order to allow time for data to be captured correctly. In addition, the variables included in analyses were chosen to minimize missing data, e.g. we omitted education level from analyses due to a substantial proportion of migrants missing this data.

Analyses of pharmacological treatment in study II and all analyses in study III were conditioned on individuals having an HbA1c or/and LDL-C level recorded, but this excluded no more than 2% of the study population and substantial bias due to this selection was unlikely. In study II, the analyses on pharmacological treatment were also conditioned on individuals still being alive and residing in Denmark at the end of the following year, which could induce selection bias. However, the alternative - leaving these individuals in the analyses - would likely have had a greater biasing effect, as any individual dying or emigrating before their first drug purchase in next the year would have resulted in their outcome being misclassified if they were in pharmacological treatment prior to the index date (as most individuals with an indication for treatment were, at least in the following year).

5.3.2. Information bias

Information bias In study I, T2D was validated! Diabetes duration Migrant ethnicity based on country of origin is more accurate than race or continent, but does not capture ethnic differences within the countries of origin.

Employment status

Pros and cons of algorithm design choices.

RLRR heterogenous, coverage issue, LDL-C in 2018

self-reported gold standard of diabetes type

5.3.3. Confounding

Studies II and III mapped risks, not causal associations

5.4. External validity

5.4.1. Generalizability

Population-wide: representative of the Danish healthcare system and migrant populations in their current state. Risks are likely to be representative of other, similar healthcare systems and migrant populations found in other Nordic countries - but not in other contexts.

Migrant group characteristics, such as distribution of other covariables may provide different results between countries, as well as difference outcome prevalence in the native population.

Results generalize to migrants with short durations of stay, but not to descendants?

5.5. Analyses

Robust poisson regression provides more intuitive estimates of relative risk than logistic regression in studies where the outcome of interest is not rare.

Which model to emphasize? Adjusting for age, age at onset/duration and income/employment status compares the risk in migrants with a segment of native danes with T2D that is very unrepresentative of the overall population of native Danes with T2D.

6. Discussion of results

6.1. Key main findings

Register-based classifiers identified valid populations of T1D and T2D in a general population, but diabetes type in cases with atypical age at onset of diabetes should be interpreted with caution.

Disparities between migrant groups and native Danes, yes, but much can be attributed to differences in clinical characteristics. Largest disparities in podiatrist monitoring, control of HbA1c and LDL-cholesterol and GLP1RA use. Migrants from Somalia received "poorer" care in all outcomes, and poorest care of all groups in most.

In the following sections, the main findings out each study will be discussed separately in light of existing literature.

6.1.1. Study I

[probably mostly CP from paper?]

Comparisons to NDR?

RSCD is implemented ad verbatim from documentation, which may not complete reflect the source code use by the Danish Health Data Authority to generate the register population. We deviated from documentation where we assessed it to be inaccurate, e.g. having no tie-breaker for classifying individuals with equal numbers of T1D-diagnoses and T2D diagnoses who also had a diagnosis of T1D and T2D on the same day as their latest diagnoses. Following documentation to the letter would omit these patients completely from the diabetes population, while our implementation included them as T2D cases.

6.1.2. Study II

[probably mostly CP from paper?]

6.1.3. Study III

[probably mostly CP from paper?]

7. Conclusions and implications

7.1. Main conclusions

[probably mostly CP from papers]

7.2. Implications

Validation: Diabetes researchers can trust OSDC and RSCD, and very likely also classifiers with similar algorithms used on similar source data and populations.

osdc-package

Atypical age at onset of diabetes

Migrants:

[Structure similar to discussion in paper III]

8. Perspectives and future research

8.1. Perspectives

The findings call for bla bla bla updated guidelines, increased awareness….

Action card for physicians: which areas of care to pay particular attention in each migrant group, when delivering care to a migrant with T2D.

8.2. Future research

Validated diabetes classifiers, or similar, should be used. Validation studies on diabetes duration/age at onset of diabetes should be performed.

Qualitative studies? The reasons for the observed disparities should be studied, and intervention studies are needed to address these reasons in migrants in an evidence-based way.

Other vulnerable groups?

Discerning T1D from T2D using insulin vs. non-insulin GLD purchases is likely to become a greater challenge in the future, as SGLT2i and GLP1RA may establish themselves as treatment options in T1D due to their effects on cardiovascular and kidney risk. With novel GLD types becoming indicated in heart and kidney patients without diabetes, inclusion criteria in register-based algorithms will have to adapt accordingly.

References

1. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. International Journal of Epidemiology [Internet] 2011;40(3):804–18. Available from: https://doi.org/10.1093/ije/dyr029
2. Sun H, Saeedi P, Karuranga S, et al. IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Research and Clinical Practice [Internet] 2022;183:109119. Available from: https://www.sciencedirect.com/science/article/pii/S0168822721004782
3. Eurostat. Migrant integration statistics [Internet]. {2022, last accessed 7 Feb 2023};Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migrant_integration_statistics
4. Eurostat. Population structure and ageing [Internet]. {2022, last accessed 7 Feb 2023};Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing
5. Eurostat. Ageing Europe - statistics on health and disability [Internet]. {2022, last accessed 7 Feb 2023};Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Ageing_Europe_-_statistics_on_health_and_disability
6. Meeks KA, Freitas-Da-Silva D, Adeyemo A, et al. Disparities in type 2 diabetes prevalence among ethnic minority groups resident in Europe: a systematic review and meta-analysis. Intern Emerg Med 2016;Apr; 11(3):327-40.
7. Andersen GS, Kamper-Jørgensen Z, Carstensen B, Norredam M, Bygbjerg IC, Jørgensen ME. Diabetes among migrants in Denmark: Incidence, mortality, and prevalence based on a longitudinal register study of the entire Danish population. Diabetes Res Clin Pract 2016;Dec:122:9–16.
8. Danmarks Statistik [Statistics Denmark]. Befolkningsfremskrivning: Indvandrere og efterkommere [English: Population projection: Immigrants and descendants] [Internet]. {2019-2039; last accessed on 7 Feb 2023};Available from: https://extranet.dst.dk/pyramide/pyramide.htm#!y=2039&o=2019&c=4

9. Ahlmark N, Algren MH, Holmberg T, et al. Survey nonresponse among ethnic minorities in a national health survey – a mixed-method study of participation, barriers, and potentials. Ethnicity & Health 2015;20(6):611–32.

10. Erayil SE, Smith MK, Gebreslasse T, et al. The value and interpretation of race and ethnicity data in the era of global migration: A change is in order. The American Journal of Tropical Medicine and Hygiene [Internet] 2021;105(6):1453–5. Available from: https://www.ajtmh.org/view/journals/tpmd/105/6/article-p1453.xml

11. UNHCR. "Refugees" and "Migrants" – Frequently Asked Questions (FAQs) [Internet]. {2016; last accessed on 7 Feb 2023};Available from: https://www.unhcr.org/en-us/news/latest/2016/3/56e95c676/refugees-migrants-frequently-asked-questions-faqs.html

12. European Commission, Migration and Home Affairs. Country responsible for asylum application (Dublin Regulation) [Internet]. {last accessed on 7 Feb 2023};Available from: https://home-affairs.ec.europa.eu/policies/migration-and-asylum/common-european-asylum-system/country-responsible-asylum-application-dublin-regulation_en

13. The Danish Immigration Service. New to Denmark: You want to apply as a family member [Internet]. {last accessed on 7 Feb 2023};Available from: https://nyidanmark.dk/uk-UA/You-want-to-apply/Family

14. The Danish Immigration Service. New to Denmark: You want to apply for a work permit [Internet]. {last accessed on 7 Feb 2023};Available from: https://nyidanmark.dk/uk-UA/You-want-to-apply/Work

15. The Danish Immigration Service. New to Denmark: You want to apply for a residence and work permit as a student [Internet]. {last accessed on 7 Feb 2023};Available from: https://nyidanmark.dk/uk-UA/You-want-to-apply/Study

16. The Danish Immigration Service. New to Denmark: Nordic citizens [Internet]. {last accessed on 7 Feb 2023};Available from: https://nyidanmark.dk/uk-UA/You-want-to-apply/Residence-as-a-Nordic-citizen-or-EU-or-EEA-citizen/Nordic-citizens

17. Indenrigs- og Boligministeriet [English: Ministry of the Interior and Health]. Bekendtgørelse om tolkebistand efter sundhedsloven [English: Executive Order on Interpretation Services under the Health Care Act] [Internet]. {2018; last accessed on 7 Feb 2023};Available from: https://www.retsinformation.dk/eli/lta/2018/855

18. Indenrigs- og Boligministeriet [English: Ministry of the Interior and Housing]. Vejledning om folkeregistrering, bilag 1: Registrering af hændelser i udlandet [English: Civil registration guidelines, appendix 1: Registration of events abroad] [Internet]. {2013; last accessed on 7 Feb 2023};Available from: https://www.retsinformation.dk/eli/retsinfo/2013/9273#Bil1

19. Danmarks Statistik [English: Statistics Denmark]. Indvandrernes medbragte uddannelse [English: Education carried by immigrants] [Internet]. {2020; last accessed on 7 Feb 2023};Available from: https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=formid&fo=medbragt-udd--pdf

20. Danmarks Statistik [English: Statistics Denmark]. Indvandrere i Danmark 2019 [English: Immigrants in Denmark 2019] [Internet]. {2019; last accessed on 7 Feb 2023};Available from: https://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=29446&sid=indv2019

21. Danmarks Statistik [English: Statistics Denmark]. Fakta om indvandrere og efterkommere i Danmark [English: Facts on immigrants and descendants in Denmark] [Internet]. {2022; last accessed on 7 Feb 2023};Available from: https://www.dst.dk/da/Statistik/nyheder-analyser-publ/bagtal/2022/2022-08-18-fakta-om-indvandrere

22. Kaneshiro B, Geling O, Gellert K, Millar L. The challenges of collecting data on race and ethnicity in a diverse, multiethnic state. Hawaii Med J 2011;70(8):168–71.

23. United Nations Statistics Division. Standard country or area codes for statistical use (M49) [Internet]. {2022; last accessed on 7 Feb 2023};Available from: https://unstats.un.org/unsd/methodology/m49/

24. Statistics Denmark. Documentation of statistics for Population Projections 2019 [Internet]. {2019; last accessed on 7 Feb 2023};Available from: https://www.dst.dk/Site/Dst/SingleFiles/GetArchiveFile.aspx?fi=49459100303&fo=0&ext=kvaldel

25. Bennet L, Udumyan R, Östgren CJ, Rolandsson O, Jansson SPO, Wändell P. Mortality in first- and second-generation immigrants to sweden diagnosed with type 2 diabetes: A 10 year nationwide cohort study. Diabetologia 2021;64(1):95–108.

26. American Diabetes Association Professional Practice Committee. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2022. Diabetes Care [Internet] 2021;45(Supplement_1):S17–38. Available from: https://doi.org/10.2337/dc22-S002

27. Holm AL, Andersen GS, Diderichsen F, Jørgensen ME. Is the Rule of Halves framework relevant for diabetes care in Copenhagen today? A register-based cross-sectional study. BMJ Open 2018;8:e023211.

28. Kristensen J, Bak J, Wittrup I, Lauritzen T. Diabetes prevalence and quality of diabetes care among Lebanese or Turkish immigrants compared to a native Danish population. Prim Care Diabetes 2007;Sep:1(3):159–65.

29. Tran AT, Diep LM, Cooper JG, et al. Quality of care for patients with type 2 diabetes in general practice according to patients' ethnic background: a cross-sectional study from Oslo, Norway. BMC Health Serv Res 2010;10, 145.

30. Negandhi PH, Ghouri N, Colhoun HM, et al. Ethnic Differences in Glycaemic Control in People with Type 2 Diabetes Mellitus Living in Scotland. PLoS ONE 2013;8.

31. Rawshani A, Svensson A, Rosengren A, Zethelius B, Eliasson B, Gudbjörnsdottir S. Impact of ethnicity on progress of glycaemic control in 131935 newly diagnosed patients with type 2 diabetes: a nationwide observational study from the Swedish National Diabetes Register. BMJ Open 2015;5;e007599.

32. Whyte MB, Hinton W, McGovern A, et al. Disparities in glycaemic control, monitoring, and treatment of type 2 diabetes in England: A retrospective cohort analysis. PLoS Med 2019;Oct 7; 16(10):e1002942.

33. Eastwood SV, Mathur R, Sattar N, Smeeth L, Bhaskaran K, Chaturvedi N. Ethnic differences in guideline-indicated statin initiation for people with type 2 diabetes in UK primary care, 2006-2019: A cohort study. PLoS Med 2021;18.

34. Vandenheede H, Deboosere P, Stirbu I, et al. Migrant mortality from diabetes mellitus across Europe: the importance of socio-economic change. Eur J Epidemiol 2012;Feb; 27(2):109-17.

35. Haller H, Ito S, Izzo JL, et al. Olmesartan for the delay or prevention of microalbuminuria in type 2 diabetes. New England Journal of Medicine [Internet] 2011;364(10):907–17. Available from: https://doi.org/10.1056/NEJMoa1007994

36. Zoungas S, Galan BE de, Ninomiya T, et al. Combined Effects of Routine Blood Pressure Lowering and Intensive Glucose Control on Macrovascular and Microvascular Outcomes in Patients With Type 2 Diabetes: New results from the ADVANCE trial. Diabetes Care [Internet] 2009;32(11):2068–74. Available from: https://doi.org/10.2337/dc09-0959

37. Dansk Selskab for Almen Medicin & Dansk Endokrinologisk Selskab [English: The Danish College of General Practitioners & The Danish Endocrine Society. Farmakologisk behandling af type 2-diabetes [English: Pharmacological treatment of type 2 diabetes] [Internet]. {2022; 2023 revision};Available from: https://vejledninger.dsam.dk/fbv-t2dm/

38. Socialstyrelsen [English: The National Board of Health and Welfare]. Nationella riktlinjer för diabetesvård [English: National guidelines for diabetes care] [Internet]. 2018;Available from: https://www.socialstyrelsen.se/kunskapsstod-och-regler/regler-och-riktlinjer/nationella-riktlinjer/riktlinjer-och-utvarderingar/diabetes/

39. Nationale VersorgungsLeitlinie [English: National Care Guidelines]. NVL Typ-2-Diabetes [Internet]. 2021;Available from: https://www.leitlinien.de/themen/diabetes/2-auflage/

40. Nederlands Huisarten Genootschap [English: Dutch College of General Practitioners]. NHG STANDARD: Diabetes mellitus type 2 [Internet]. {2018, latest revision January 2023};Available from: https://richtlijnen.nhg.org/standaarden/diabetes-mellitus-type-2

41. National Institute for Health and Care Excellence. NICE guideline [NG28]: Type 2 diabetes in adults: management [Internet]. {2015, latest revision June 2022};Available from: https://www.nice.org.uk/guidance/ng28

42. Helsedirektoratet [English: Norwegian Directorate of Health]. Nasjonal faglig retningslinje for diabetes [English: National professional guideline for diabetes] [Internet]. {2016, latest revision March 2021};Available from: https://www.helsedirektoratet.no/retningslinjer/diabetes

43. Dansk Selskab for Almen Medicin [English: The Danish College of General Practitioners]. Type 2-diabetes i almen praksis - en evidensbaseret vejledning [English: Type 2 diabetes in general practice - an evidence-based guideline] [Internet]. {2004; last accessed on 7 Feb 2023};Available from: https://vejledninger.dsam.dk/media/files/20/type_2_diabetes_2004_rettet.pdf

44. Dansk Selskab for Almen Medicin [English: The Danish College of General Practitioners]. Type 2-diabetes - et metabolisk syndrom [English: Type 2 diabetes - a metabolic syndrome] [Internet]. {2012; last accessed on 7 Feb 2023};Available from: https://vejledninger.dsam.dk/media/files/20/type-2-diabetes-2012-med-links.pdf

45. Dansk Selskab for Almen Medicin [English: The Danish College of General Practitioners]. Type 2-diabetes - opfølgning og behandling [English: Type 2 diabetes - follow-up and treatment] [Internet]. {2019; last accessed on 7 Feb 2023};Available from: https://vejledninger.dsam.dk/type2/

46. Danish Ophthalmological Society. National retningslinje for screening for diabetisk retinopati [English: National guideline on screening for diabetic retinopathy] [Internet]. {2018; last accessed on 7 Feb 2023};Available from: https://dansk-oftalmologisk-selskab.dk/wp-content/uploads/2021/05/National-retningslinje-for-screening-af-diabetisk-retinopati.pdf

47. Carstensen B, Kristensen JK, Marcussen MM, Borch-Johnsen K. The national diabetes register. Scand J Public Health 2011;39(7 Suppl):58–61.

48. Green A, Sortsø C, Jensen PB, Emneus M. Validation of the danish national diabetes register. Clinical epidemiology [Internet] 2014;7:5–15. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25565889 https://www.ncbi.nlm.nih.gov/pmc/PMC4274151/

49. Sundhedsdatastyrelsen [English: Danish Health Data Authority]. Algoritmer for udvalgte kroniske sygdomme og svære psykiske lidelser [English: Algorithms for selected chronic diseases and serious mental illnesses] [Internet]. {2016, last accessed 7 Feb 2023};Available from: https://www.esundhed.dk/-/media/Files/Dokumentation/Register-for-Udvalgte-Kroniske-Sygdomme-og-Svaere-Psykiske-Lidelser/29_Algoritmer_for_RUKS.ashx

50. Sundhedsdatastyrelsen [English: Danish Health Data Authority]. Dokumentation af Laboratoriedatabasens Forskertabel – Version 1.3 [English: Documentation on Register of Laboratory Results for Research - version 1.3] [Internet]. {2018, last accessed 7 Feb 2023};Available from: https://sundhedsdatastyrelsen.dk/-/media/sds/filer/registre-og-services/nationale-sundhedsregistre/doedsaarsager-og-biologisk-materiale/laboratoriedatabasen/dokumentation-af-labdatabasens-forskertabel.pdf

51. Anders Aasted Isaksen. The Open-Source Diabetes Classifier [Internet]. {last accessed 7 Feb 2023};Available from: https://github.com/Aastedet/osdc

52. Danish Medicines Agency. Reimbursement thresholds [Internet]. {2023; last accessed on 7 Feb 2023};Available from: https://laegemiddelstyrelsen.dk/en/reimbursement/calculate-reimbursement/reimbursement-thresholds/

53. Danish Regions. OKPortal: Takstmapper [English: Contract agreement Portal: Rate folders] [Internet]. {2023; last accessed on 7 Feb 2023};Available from: https://okportal.dk/Default.aspx?area=Praksis&type=Takstmapper

54. Schmidt M, Schmidt SAJ, Adelborg K, et al. The danish health care system and epidemiological research: From health care contacts to database records. Clin Epidemiol 2019;11:563–91.

55. Eurostat. Glossary: Equivalised disposable income [Internet]. {last accessed 7 Feb 2023};Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Equivalised_disposable_income

56. World Health Organisation. International statistical classification of diseases and related health problems, 10th revision: Instruction manual, fifth edition [Internet]. {2016, last accessed 7 Feb 2023};Available from: https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf

57. Wertheimer AI. The defined daily dose system (DDD) for drug utilization review. Hosp Pharm 1986;21(3):233–4, 239–41, 258.

58. Bjerg L, Dalsgaard E-M, Norman K, Isaksen AA, Sandbæk A. Cohort profile: Health in central denmark (HICD) cohort - a register-based questionnaire survey on diabetes and related complications in the central denmark region. BMJ Open [Internet] 2022;12(7):e060410. Available from: http://bmjopen.bmj.com/content/12/7/e060410.abstract

59. Christensen AI, Ekholm O, Glümer C, et al. The danish national health survey 2010. Study design and respondent characteristics. Scandinavian Journal of Public Health [Internet] 2012;40(4):391–7. Available from: https://doi.org/10.1177/1403494812451412

60. RStudio Team. RStudio: Integrated development environment for r [Internet]. Boston, MA: RStudio, PBC; 2021. Available from: http://www.rstudio.com/

61. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

62. Dowle M, Srinivasan A. Data.table: Extension of 'data.frame' [Internet]. 2021. Available from: https://CRAN.R-project.org/package=data.table

63. Carstensen B, Plummer M, Laara E, Hills M. Epi: A package for statistical analysis in epidemiology [Internet]. 2021. Available from: https://CRAN.R-project.org/package=Epi

64. Harrison E, Drake T, Ots R. Finalfit: Quickly create elegant regression results tables and plots when modelling [Internet]. 2021. Available from: https://CRAN.R-project.org/package=finalfit

65. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. Journal of Open Source Software 2019;4(43):1686.

66. Halekoh U, Højsgaard S, Yan J. The r package geepack for generalized estimating equations. Journal of Statistical Software 2006;15/2:1–11.

67. Iannone R. DiagrammeR: Graph/network visualization [Internet]. 2020. Available from: https://CRAN.R-project.org/package=DiagrammeR

68. Frerebeau N. Khroma: Colour schemes for scientific data visualization [Internet]. Pessac, France: Université Bordeaux Montaigne; 2021. Available from: https://doi.org/10.5281/zenodo.1472077

69. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. BMJ 2014;348.

70. Williamson T, Eliasziw M, Fick GH. Log-binomial models: Exploring failed convergence. Emerging Themes in Epidemiology [Internet] 2013;10(1):14. Available from: https://doi.org/10.1186/1742-7622-10-14

71. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. American Journal of Epidemiology 2004;159:702–6.

72. Zeileis, Achim and Kleiber, Christian and Jackman, Simon. Regression Models for Count Data in R [Internet]. {last accessed 7 Feb 2023};Available from: https://cran.r-project.org/web/packages/pscl/vignettes/countreg.pdf

73. Greenland, Sander. Model-based Estimation of Relative Risks and Other Epidemiologic Measures in Studies of Common Outcomes and in Case-Control Studies. American Journal of Epidemiology 2004;160(4):301–5.

74. Chen W, Qian L, Shi J, Franklin M. Comparing performance between log-binomial and robust poisson regression models for estimating risk ratios under model misspecification. BMC Medical Research Methodology [Internet] 2018;18(1):63. Available from: https://doi.org/10.1186/s12874-018-0519-5

75. World Health Organization. Use of glycated haemoglobin (HbA1c) in diagnosis of diabetes mellitus: abbreviated report of a WHO consultation [Internet]. {2011; last accessed on 12 Jan 2022};WHO/NMH/CHP/CPM/11.1. Available from: https://apps.who.int/iris/handle/10665/70523

76. Sundhedsdatastyrelsen [English: Danish Health Data Authority]. Algoritmer for udvalgte kroniske sygdomme og svære psykiske lidelser [English: Algorithms for selected chronic diseases and serious mental illnesses] [Internet]. {2022, last accessed 7 Feb 2023};Available from: https://www.esundhed.dk/-/media/Files/Publikationer/Emner/Operationer-og-diagnoser/Udvalgte-kroniske-sygdomme-svaere-psykiske-lidelser/Algoritmer-for-Udvalgte-Kroniske-Sygdomme-og-svre-psykiske-lidelser-RUKS-2022.ashx

English summary

This is what I spent three years on...

Dansk resumé

Det her brugte jeg tre år på...

A. Appendix A

A.1. Papers associated with the dissertation

B. Appendix B

B.1. Supplementary material

B.1.1. Study I

B.1.1.1. Coding scheme

Link to blob

B.1.1.2. Supplementary analyses

Link to blob

B.1.2. Study II

B.1.2.1. Coding scheme

Link to blob

B.1.2.2. Supplementary analyses

Link to blob

B.1.3. Study III

B.1.3.1. Coding scheme

Link to blob

B.1.3.2. Supplementary analyses

Link to blob

B.2. To do/Questions:

- Revisit the prefacing section of each chapter after revisions
- Figures of OSDC logic flow need nationwide counts
  - Update main text in results chapter accordingly.
    - And each inclusion criteria should perhaps also show counts of unique contributions?
- Consider revising classifierdiffs figures for improve legibility/nationwide counts.
- Translate/revise figures from DST
- Tilføj NHS analyser fra methods selection bias discussion i (separat?) supplementary?
- Statistical analyses section describes only main analyses. Should a summary of supplementary/sensitivity analyses be included (elsewhere?)?

B.2.1. Introduction:

B.2.1.1. Clinical guidelines:

Social ulighed: Beskriv evt. kort begreb og relation/forskelle ift. projektet. Henvis til supplementary analyser i results?

Fig. 1.3: Indvandringsår: Skal den med, og hvor meget skal der skrives for at redde den.

B.2.2. Methods

Table: Overview of 3 studies

B.2.3. Results

Ret tal

Fortløbende figurer 1 - Inf, eller samlet optælling af figurer + tabeller i TOC?

Hvor meget skal med? Alle figurerne fra artiklerne, så man kan få et overblik over alle hoved-resultaterne ét sted (minus ECD-grafer for HbA1c og LDL-C).

Opbryd teksten med figurerne for hvert outcome.

B.2.3.1. Study I:

4.1: Flyttes til diskussion. Erstattes af den endelige algoritme, pastelflowchart osv.: Sådan har vi defineret T2D, og sådan påvirker designet den endelige population undervejs.

Alle resultater som i artiklen, eller mere overfladisk som i det øverste og nederste afsnit?

B.2.4. Discussion

Diskutér det forskellige fokus fra studie 2 til 3: Guideline-anbefalinger er mere absolutte, derfor primært fokus på ujusterede estimater i studie II. Der er ikke sort-hvide mål for kombinationsbehandling/præparatvalg, så behandlingen hviler på et samlet klinisk billede, hvorfor det primære fokus på justerede estimater i studie III.

OSDC: Censoring of PCOS: incidens-spike ved 40-års alder, evt. histogram for incidente cases 2012/16 og frem.

Migrants: Some of the risks in migrants are due to other clinical and socioeconomic risk factors that are more prevalent in migrants compared to native Danes. These risk factors are not the focus of this dissertation and are not presented in the main results. However, these risk factors are of key importance in a greater scope of equity in T2D care provision, and readers are urged to explore the online appendix, where they are presented. The following section provides a brief summary of overall patterns found among covariates of analyses:

BLOB:

Kap 5.1.2: PCOS:

The age-based censoring constituted the vast majority of censored events, as data on indication codes was available in only a minority of prescriptions (contrary to the 72% of all prescriptions described by official documentation).[@lmdbdoc]