

# Type 2 diabetes in migrants: a register-based analysis

PhD dissertation

Anders Aasted Isaksen

Health  
Aarhus University  
2023





# Supervisors and assessment committee

## PhD student:

**Anders Aasted Isaksen, MD**

Department of Public Health, Aarhus University, Denmark

## Supervisors:

**Professor Anneli 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

# Notes and disclosures

## 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** foundation (Danish: *Folkesundhed i Midten*) and **Steno Diabetes Center Aarhus**, which is partially funded by an unrestricted donation from the *Novo Nordisk Foundation*.



## Notes on website version

This dissertation is available for viewing online at [aastedet.github.io/dissertation/](https://aastedet.github.io/dissertation/) (or by scanning the QR code in the lower right-hand corner of the title page). While the core contents are identical, the website may offer a more comfortable reading experience compared to the physical booklet for some readers. In particular, the website is recommended for readers with impaired vision and others who may be uncomfortable reading the small physical book.

# Acknowledgements

Looking back on the last three years against a backdrop of pandemics and full-scale invasions, my time as a PhD student appears minuscule and seems to have gone by in an instant. As it draws to an end, I feel truly grateful for the opportunity to engage in a research topic that was close to me, and for the creative freedom to pursue it in a way that was interesting to me as a researcher and person. This was in large part due to the great people around me during this time, who I owe my thanks.

First of all, I would like to thank my supervisor group. I could not have asked for a more insightful and supportive group of researchers to guide me along. Especially to Anneli, your sound advice, positive spirit and confidence in me has been vital throughout the project. To Lasse, for always looking out for me and patiently guiding me to make sense. To Mette, for guidance, particularly in the early days of data exploration (and often with one of my children in attendance at meetings). To Gregers, for bearing with me in the initial stages until my project found its footing, and for excellent advice and interesting discussions later.

I have also benefited greatly from the support and company provided by colleagues at the Department of Public Health and the Research Unit for General Practice - particularly my former and fellow PhD students, junior researchers and office mates. Thank you all for the coffee breaks, the discussions, the advice given, and all the laughs shared across the years. Too many to mention, I'm grateful to you all!

I would like to thank Marianne Pedersen of the Department of Public Health for assistance navigating through the dark jungle that is applying for data to Statistics Denmark. Thanks also to Else-Marie Dalsgaard, Kasper Norman and Eva Therkildsen for assistance in all things practical and impractical along the way, and to Lone Niedziella for assistance on linguistic issues.

A special thanks to the Clinical Epidemiology Research Group of Steno Diabetes Center Copenhagen for being great hosts and for the many interesting and fruitful discussions during my visit there. I'm also grateful to members of the Epidemiology and Cross-sectoral research groups at Steno Diabetes Center Aarhus for inviting me to be a part of their communities. In particular, I owe a great deal of thanks to Luke Johnston for inspiring me to strive for openness and better programming practices, and for having a lot of fun while learning immensely 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.

This dissertation was built with free, open-source tools, and I would like to extend my gratitude to the many unsung heroes of open-source software development. In particular, to the *R Core Team* for the *R* scripting language and to *Posit* for the *RStudio IDE* and *Quarto* tools.

I'm grateful to the participants in the *Health in Central Denmark* survey, without whom this project would not be what it is today, and to the funding sources for believing in me and my project - and providing the means to realize it.

Finally, I owe a debt of gratitude to my family. To my wife, Thea, for everything you are. To Svend, Frode, Ejnar and Aase for filling my life with wonder and laughter (and mess and noise).

To my mother for always supporting and inspiring me, to my father, who I would have loved to share this journey with, and to my sister, my life's first supervisor.

Thank you.

# 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 soon learn that T2D held more surprises for me than just a high prevalence. After 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. Diabetes was not only more prevalent; the patients were also much younger, and their haemoglobin-A1c 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 centres and pre-removal detention centres.

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 could enable GPs and healthcare planners to address disparities and improve care in migrants by prioritising and focussing care accordingly. While migrants are currently a younger demographic group than the rest of the population, it is likely that migrants will constitute an increasing proportion of the T2D population in the coming years as the group ages. 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 an expansion of the scope of the PhD project 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. In the spirit of open-source, 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 and accessibility, this dissertation was made available to a global public audience in website format.

## Outline of the dissertation

**Chapter 1** describes type 2 diabetes (T2D) and how migrants are at particular risk. It then outlines the context of migration and health care in Denmark and clinical guideline recommendations for 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 main findings of each study and compares them to existing literature. Finally, the clinical implications of the results are discussed.

**Chapter 7** presents the main conclusions.

**Chapter 8** addresses the future perspectives raised by the dissertation and discusses the opportunities for research in the evolving research field.

The rest of the dissertation includes a list of **references**, English and Danish **summaries**, a **supplementary** validation analysis, and the three **papers** of the dissertation and their supplementary materials.



# Papers in the dissertation

## Study I

Isaksen AA, Sandbæk A, Bjerg L. Validation of Register-Based Diabetes Classifiers in Danish Data. *Clinical Epidemiology*. 2023;15:569-581. <https://doi.org/10.2147/CLEP.S407019>

## Study II

Isaksen AA, Sandbæk A, Skriver MV, Andersen GS, Bjerg L. Guideline-level monitoring, biomarker levels and pharmacological treatment in migrants and native Danes with type 2 diabetes: population-wide analyses. *PLOS Global Public Health*. In review 2023.

## Study III

Isaksen, AA, Sandbæk, A, Skriver, MV, Bjerg, L. Glucose-lowering drug use in migrants and native Danes with type 2 diabetes: Disparities in combination therapy and drug types. *Diabetes Obesity and Metabolism*. 2023; 1-10. <https://doi.org/10.1111/dom.15230>

# Table of contents

<b>Supervisors and assessment committee</b>	<b>1</b>
<b>Notes and disclosures</b>	<b>2</b>
Financial disclosures . . . . .	2
<b>Acknowledgements</b>	<b>3</b>
<b>Preface</b>	<b>5</b>
Motivation . . . . .	5
Outline of the dissertation . . . . .	6
<b>Papers in the dissertation</b>	<b>7</b>
Study I . . . . .	7
Study II . . . . .	7
Study III . . . . .	7
<b>Table of contents</b>	<b>8</b>
<b>List of Figures</b>	<b>13</b>
<b>List of Tables</b>	<b>15</b>
<b>Abbreviations</b>	<b>16</b>
<b>1. Introduction</b>	<b>17</b>
1.1. Migrants and type 2 diabetes . . . . .	17

## TABLE OF CONTENTS

---

1.2. Migrants in Denmark . . . . .	18
1.2.1. Migration . . . . .	19
1.2.2. Migrants in the Danish data infrastructure . . . . .	20
1.2.3. Migrant groups in Denmark . . . . .	20
1.2.4. Descendants . . . . .	23
1.3. Type 2 diabetes care . . . . .	23
1.3.1. Complications and risk factors . . . . .	23
1.3.2. Pharmacological treatment . . . . .	24
1.3.3. Monitoring . . . . .	25
1.4. Clinical guidelines for type 2 diabetes care . . . . .	25
1.4.1. Context . . . . .	25
1.4.2. Diagnosis . . . . .	27
1.4.3. Monitoring . . . . .	28
1.4.4. Biomarker levels . . . . .	28
1.4.5. Pharmacological treatment . . . . .	28
1.5. Identification of type 2 diabetes cases in healthcare registers . . . . .	29
1.5.1. Danish register data infrastructure . . . . .	29
1.5.2. Benefits . . . . .	30
1.5.3. Challenges . . . . .	30
1.5.4. Register-based diabetes classifiers . . . . .	31
1.6. Introduction at a glance . . . . .	31
<b>2. Aims</b>	<b>32</b>
2.1. Overall aims . . . . .	32
2.2. Specific aims . . . . .	32
2.3. Outputs . . . . .	32

## TABLE OF CONTENTS

---

<b>3. Material and methods</b>	<b>33</b>
3.1. Setting: The Danish healthcare system . . . . .	33
3.2. Register data sources . . . . .	33
3.2.1. The Danish Civil Registration System . . . . .	34
3.2.2. The Danish National Patient Register . . . . .	34
3.2.3. The Danish National Health Service Register . . . . .	34
3.2.4. The Danish National Prescription Registry . . . . .	35
3.2.5. The Danish Register of Laboratory Results for Research . . . . .	35
3.3. Diabetes classification algorithms . . . . .	35
3.3.1. The Open-Source Diabetes Classifier . . . . .	37
3.3.2. The Register of Selected Chronic Diseases . . . . .	38
3.4. Questionnaire data used in validation . . . . .	39
3.4.1. The Health In Central Denmark survey . . . . .	39
3.4.2. The National Health Survey . . . . .	40
3.5. Study design and population . . . . .	40
3.5.1. Study populations . . . . .	40
3.5.2. Outcomes . . . . .	45
3.5.3. Primary variable of interest: migrant status . . . . .	47
3.5.4. Other explanatory variables . . . . .	48
3.6. Statistical analysis . . . . .	48
3.7. Ethics and approvals . . . . .	52
<b>4. Results in summary</b>	<b>53</b>
4.1. Register-based type 2 diabetes population . . . . .	53
4.1.1. Flow of individuals in the Open-Source Diabetes Classifier . . . . .	53
4.2. Study I . . . . .	56
4.3. Study II . . . . .	61
4.3.1. Prevalence . . . . .	61

## TABLE OF CONTENTS

---

4.3.2. Monitoring . . . . .	63
4.3.3. Biomarker control . . . . .	65
4.3.4. Pharmacological treatment . . . . .	67
4.4. Study III . . . . .	69
<b>5. Discussion of methods</b>	<b>73</b>
5.1. Design of diabetes classifiers . . . . .	73
5.1.1. Approaches to censoring gestational diabetes mellitus . . . . .	73
5.1.2. Approaches to censoring polycystic ovary syndrome . . . . .	75
5.1.3. Approaches to censoring other sources of false inclusions . . . . .	75
5.1.4. Approaches to classification of diabetes type . . . . .	76
5.1.5. Incidence and demographic characteristics of register-based diabetes cohorts . . . . .	76
5.2. Study designs . . . . .	83
5.3. Internal validity . . . . .	83
5.3.1. Selection bias . . . . .	83
5.3.2. Information bias . . . . .	85
5.3.3. Confounding . . . . .	87
5.4. External validity . . . . .	88
5.4.1. Generalisability . . . . .	88
5.5. Analyses . . . . .	89
<b>6. Discussion of results</b>	<b>91</b>
6.1. Key main findings . . . . .	91
6.2. Study I . . . . .	91
6.3. Study II . . . . .	92
6.4. Study III . . . . .	94
6.5. Clinical implications . . . . .	96

## TABLE OF CONTENTS

---

<b>7. Conclusions</b>	<b>98</b>
7.1. Main conclusions . . . . .	98
<b>8. Perspectives and future research</b>	<b>100</b>
8.1. Perspectives . . . . .	100
8.2. Future research . . . . .	100
<b>References</b>	<b>102</b>
<b>English summary</b>	<b>116</b>
<b>Dansk resumé</b>	<b>118</b>
<b>Appendices</b>	<b>121</b>
<b>A. Supplementary validation analyses</b>	<b>121</b>
A.0.1. Validation of unspecified diabetes in the Health In Central Den- mark survey and in the National Health Survey. . . . .	121
<b>B. Full papers</b>	<b>124</b>
B.1. Paper I . . . . .	124
B.2. Paper II . . . . .	125
B.3. Paper III . . . . .	126
<b>C. Addenda and errata</b>	<b>127</b>
C.1. Addenda . . . . .	127
C.2. Errata . . . . .	127

## List of Figures

1.1. Demographic projections of native Danes and migrants . . . . .	19
1.2. Primary reason for migration among the largest migrant groups in Denmark . . . . .	22
3.1. Overall diabetes classifier design differences and population flows	36
3.2. Flow of individuals into the <i>Health In Central Denmark</i> survey and the study population of study I . . . . .	42
3.3. Relationship between odds ratio and relative risk depending on baseline risk . . . . .	49
4.1. Inclusion and censoring in the Open-Source Diabetes Classifier .	54
4.2. Logic and flow of classification of diabetes type in the Open- Source Diabetes Classifier . . . . .	55
4.3. Prevalence of type 2 diabetes . . . . .	62
4.4. Risk of not receiving guideline-recommended monitoring . . . . .	64
4.5. Risk of exceeding guideline-recommended biomarker targets . . .	66
4.6. Risk of not receiving guideline-recommended pharmacological treatment . . . . .	68
4.7. Glucose-lowering drug combination therapy . . . . .	70
4.8. Oral glucose-lowering drug types . . . . .	71
4.9. Injection-based glucose-lowering drug types . . . . .	72
5.1. Annually registered births . . . . .	74
5.2. Calendar year of inclusion among prevalent diabetes cases on 1 January 2019 . . . . .	78

## LIST OF FIGURES

---

5.3. Age at inclusion for each sex among prevalent diabetes cases on 1 January 2019 . . . . .	80
5.4. Age at inclusion for each sex among prevalent diabetes cases on 1 January 2019, stratified by calendar year of inclusion . . . . .	82



## List of Tables

3.1. Overview of studies . . . . .	41
3.3. Concordance table and validation metrics . . . . .	51
4.1. Validation of register-based classification of type 1 diabetes . . .	57
4.2. Validation of register-based classification of type 2 diabetes . . .	58
4.3. Validation of classification of type 1 diabetes stratified by age at onset . . . . .	59
4.4. Validation of classification of type 2 diabetes stratified by age at onset . . . . .	60
7.1. Highlights of disparities within each migrant group . . . . .	99
A.1. Supplementary: Validation of unspecified diabetes in the HICD study population . . . . .	122
A.2. Supplementary: Validation of unspecified diabetes in the Na- tional Health Survey . . . . .	123

## Abbreviations

**ACEI:** Angiotensin-converting enzyme-inhibitors  
**APT:** Antiplatelet therapy  
**ARB:** Angiotensin receptor blockers  
**ATC:** Anatomical Therapeutic Chemical (classification)  
**CI:** Confidence interval  
**CVD:** Cardiovascular disease  
**DKD:** Diabetic kidney disease  
**DPP4i:** Dipeptidyl peptidase-4 inhibitors  
**GP:** General practitioner  
**GDM:** Gestational diabetes mellitus  
**GLD:** Glucose-lowering drugs  
**GLP1RA:** Glucagon-like peptide-1 receptor agonists  
**HbA1c:** Haemoglobin-A1c  
**LDL-C:** Low-density lipoprotein cholesterol  
**LLD:** Lipid-lowering drugs  
**NPV:** Negative predictive value  
**OR:** Odds ratio  
**OSDC:** Open-Source Diabetes Classifier  
**PCOS:** Polycystic ovary syndrome  
**PPV:** Positive predictive value  
**RR:** Relative risk  
**RSCD:** Register of Selected Chronic Diseases  
**SGLT2i:** Sodium glucose co-transporter type 2 inhibitors  
**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 outlines 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 providing 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 disproportionately affects the socially deprived demographics in societies across the globe.<sup>1</sup> Today, the prevalence of type 2 diabetes (T2D) exceeds half a *billion* people globally and is on the rise due to several factors, including population ageing, 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.<sup>2</sup>

After being granted residence permit in Denmark and other developed nations, migrants are often housed in densely-populated urban areas and are 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.<sup>3</sup> 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 ageing population with a growing burden of chronic diseases and comorbidities.<sup>4,5</sup> Prioritisation 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.<sup>6,7</sup> However, this only delays the impact of their increased T2D risk. As these migrant populations age in the coming years, they will make up a growing proportion of T2D patients in the country. In 2019, 9% of immigrants in Denmark were aged 66 years and above, a proportion projected to have doubled by 2039.<sup>8</sup> Therefore, understanding potential disparities in care and the specific healthcare needs of migrant populations with T2D can inform the prioritisation 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.<sup>9</sup> Furthermore, migrants have historically been defined by their racial appearance and ethnic minorities categorised 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.<sup>10</sup>

### 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. Figure 1.1 shows the projected demographics of native Danes and migrants in 2039 (2019 demographics outlined in light green (adapted from Statistics Denmark)).<sup>8</sup>



Figure 1.1.: Demographic projections of native Danes and migrants

### 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 under 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.<sup>11</sup>

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.<sup>11</sup>

In Denmark, applications from migrants seeking protection are processed by authorities to determine if they qualify for refugee status, and whether they have previously applied for asylum elsewhere in the European Union under the Dublin Regulation.<sup>12</sup> Individuals eligible for refugee status are granted asylum and a residence permit. Conversely,

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

Once residence has been granted, regardless of the reason, an individual has access to the same public healthcare services as a Danish citizen. 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 the interpreter services, but since 2018, the regional public healthcare administrations have been obliged by law to charge a fee from patients with more than three years of residence.<sup>17</sup>

### 1.2.2. Migrants in the Danish data infrastructure

In this PhD project, refugees and immigrants are referred to collectively as migrants after obtaining residence. The automated, prospective collection of register data in Denmark is challenged by the missing data on migrants and their past.

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.<sup>18</sup> 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.<sup>19</sup>

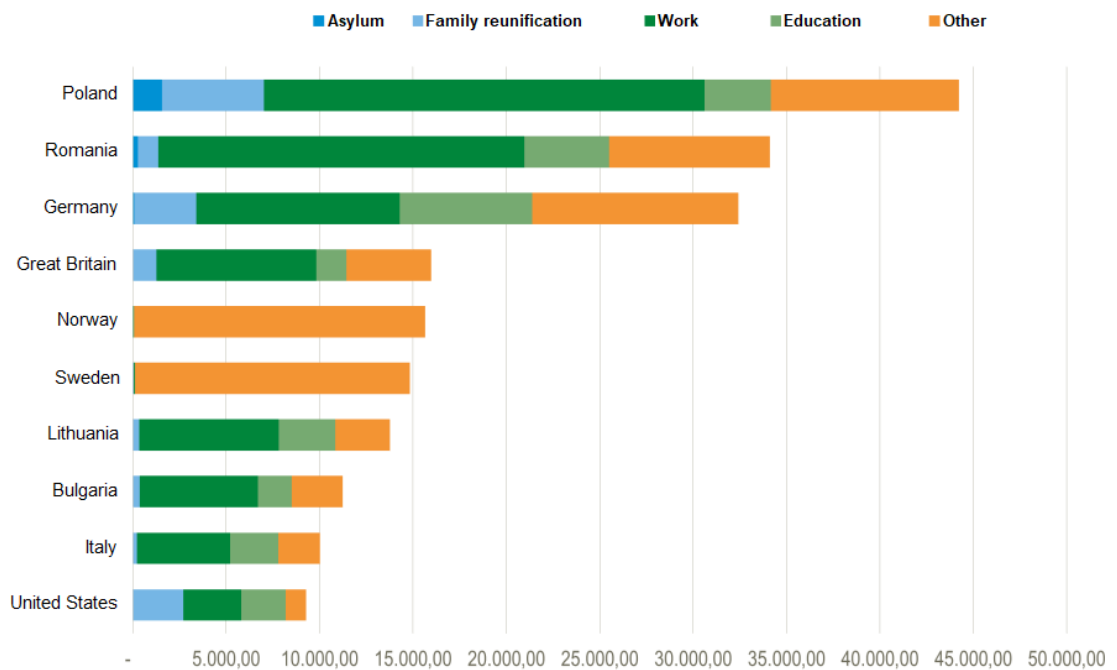
### 1.2.3. Migrant groups in Denmark

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.<sup>20</sup>

Migrants from more than 200 different countries of origin reside in Denmark, which makes the migrant group very heterogenous with respect to geographic and cultural background.<sup>20</sup> 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.<sup>20</sup> Figure 1.2 shows the primary reason for migration among the largest migrant groups in Denmark (adapted from Statistics Denmark).<sup>21</sup>

## 1. Introduction



(a) Reason for migration: Western countries



(b) Reason for migration: Non-western countries

Figure 1.2.: **Primary reason for migration among the largest migrant groups in Denmark**



The limited size of each ethnic minority and the strongly contrasting characteristics between them poses a challenge to statistical processing. Minority categories should be clinically meaningful, and researchers must balance the need for aggregation to achieve adequate sample sizes against the risk of introducing biases due to aggregation of dissimilar groups.<sup>22</sup> 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*,<sup>23</sup> whereas Statistics Denmark group migrants into Western and Non-western countries, respectively, using a grouping scheme that resembles the more widely-used classification of developed and developing countries.<sup>24</sup> 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 category. In the context of T2D today, descendants of migrants in Denmark are a very young group unlikely to constitute a substantial proportion of the patient population (less than 1% of descendants are aged 66 years or older).<sup>8</sup> Some evidence from Sweden suggests that descendants are a distinct group from first-generation immigrants in terms of T2D and mortality.<sup>25</sup> 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 *haemoglobin-A1C* (HbA1c) and *low-density lipoprotein cholesterol* (LDL-C) drive indications for treatment and clinical decision-making in T2D.<sup>26</sup>

### 1.3.1.2. In migrants

In Denmark and several other European countries, HbA1c levels are higher in migrants than in native populations,<sup>27–32</sup> but LDL-C levels have only been sparsely described.<sup>33</sup> The evidence on complication risk is inconsistent, but there is some indication that the risk is higher in migrants compared to native populations, although the risks differ between complication types, migrant groups and countries. Similarly, some studies have reported higher mortality among migrants with diabetes,<sup>7,34</sup> whereas other studies have found similar or lower risk.<sup>35</sup>

### 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).<sup>36,37</sup> In addition to their effect on biomarker levels, certain types of the above classes of drugs have other positive effects. In particular, the GLDs *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 *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.<sup>38</sup>

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 because GLD monotherapy fails to achieve target levels of HbA1c in many patients, even at maximal dosage.<sup>39,40</sup>

#### 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.<sup>29,31</sup> However, disparities in the quality of the prescribed GLD treatment, such as the use of combination therapy or the types of GLD prescribed, have not been studied. Indirect evidence suggests that such disparities are likely, as non-white minorities in the UK were found to be slower and less likely to intensify to combination therapy than white groups.<sup>41</sup>

Likewise, migrants in Italy with T2D have been reported to purchase fewer packages of non-insulin GLDs per person-year than their native counterparts.<sup>42</sup> Moreover, socioeconomic and racial disparities in prescribing of SGLT2i and GLP1RA have been reported in several countries.<sup>32,43–47</sup>

In contrast, studies from European countries have reported lower prescription rates of LLDs in migrants with T2D.<sup>29,33,48</sup> However, treatment with LLD is not indicated in all patients with T2D, 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. A study from Italy found lower odds of treatment in most migrant groups,<sup>49</sup> 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.<sup>50,51</sup>

#### 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,<sup>29</sup> while migrants with T2D in Italy received fewer referrals for consultations than native Italians.<sup>42</sup>

## 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 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

The Danish guidelines for T2D contain a chapter on vulnerable populations. This chapter includes 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. The guidelines recommend the use of a professional interpreter 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 accommodation of insulin-therapy during religious fasting. In the guidelines on pharmacological treatment of T2D, migrants are not addressed in the main contents, nor in the appendix on treatment of special groups.<sup>52</sup>

The sparse recommendations on clinical management of migrants with T2D are 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, the Swedish guidelines do not mention migrants at all.<sup>53</sup>
- Germany: Ethnicity is noted as one of several contextual factors that may influence morbidity in T2D, but migrants or ethnic minorities are not mentioned otherwise.<sup>54</sup>
- The Netherlands: The Dutch guidelines recommend increased screening for T2D in individuals of Hindustani, Turkish, Moroccan or Surinamese origin. The guidelines state that provision of diabetes care in these groups should be similar to that in the native Dutch population. No specific recommendations on care in migrants are provided, although it is noted that some treatment options appear to be less effective in these groups (lifestyle interventions, bariatric surgery).<sup>55</sup>
- The United Kingdom: The UK guidelines advise that BMI thresholds for obesity be adjusted in ethnic minorities, and an increased risk of complications in some ethnic groups is noted. No specific recommendations for clinical management of migrants are made. Citing a lack of evidence on ethnicity-specific effects of pharmacological treatment, the guidelines recommend research into the effectiveness of SGLT2i in different ethnic groups.<sup>56</sup>

- Norway: The special needs of migrants are recognized in the 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 not to do so among all the 13 guideline topics), and the recommendations specific to migrants are limited to more liberal screening for diabetes, use of a professional interpreter, and attention to diet and religious fasting.<sup>57</sup>

### 1.4.1.2. Historical revisions

The 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, which recommended the use of a professional interpreter when encountering language barriers and noted a 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.<sup>58</sup>

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<sup>59</sup> and 2019<sup>60</sup> revisions. These revisions brought minimal changes to the recommendations relevant to this PhD project, as they were largely unchanged from 2004 until the time period of the studies of T2D care in this dissertation. These recommendations are described below as they appeared in the prevailing guidelines at the time, with highlights of any mentioning of migrants.

Some readers will note that parts of the 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).

### 1.4.2. Diagnosis

HbA1c is the main diagnostic standard, with values  $\geq 48\text{mmol/mol}$  being diagnostic of type 2 diabetes, although 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 haemoglobinopathy. Once diagnosed, patients with T2D should be considered to be permanently affected by the disease, and the associated risk factors for complications should be treated regardless of subsequent normalisation of HbA1c.

The 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 age and other risk factors.

### 1.4.3. Monitoring

The risk biomarkers HbA1c and LDL-C should be measured at least once a year along with urine albumin-to-creatinine ratio (UACR) to screen for DKD (repeated samples of  $\text{UACR} \geq 300\text{mg/g}$  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.<sup>61</sup> Similarly, screening for diabetic foot disease should be performed by a podiatrist every year or, alternatively, by the GP if podiatrist service is not feasible for the patient. At the initial diagnosis of T2D, a baseline-screening of all five types of monitoring is recommended.

The 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 hypoglycaemia or inappropriate polypharmacy. Specific targets at  $48\text{mmol/mol}$ ,  $53\text{mmol/mol}$ ,  $58\text{mmol/mol}$ , and  $70\text{mmol/mol}$  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  $2.6\text{mmol/L}$  is provided, while the target in patients with complications should be below  $1.8\text{mmol/L}$ .

The 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  $\geq 48\text{mmol/mol}$ , 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:** The guidelines do not present a fixed hierarchy between GLD drug type, but recommend factoring in all effects, positive and negative, as well as cost when choosing between individual drug types.

### 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.

The 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, which are stored in the same structure, unless altered by the researcher (or their data manager). The identical raw data and the common virtual working environment potentially enables reproducible research. This means that any data processing workflow should be transferable between research projects if the underlying code is designed with reproducibility in mind and the code is shared (“open-sourced”).<sup>62</sup> 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.<sup>63</sup>

### 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 possibility 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 they 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.<sup>64,65</sup> Considerable efforts have been made towards establishing such a tool for diabetes research in several countries, including Denmark.<sup>66–68</sup>

Internationally, some algorithms classifying T1D and T2D have been validated in cohorts of individuals with diabetes,<sup>69–73</sup> 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 it must also identify diabetes while accounting for events that might lead to inclusion of non-cases, such as the 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 through register data was the National Diabetes Register, which was established in 2006.<sup>74</sup> 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.<sup>75</sup>

Launched in 2014, the Register of Selected Chronic Diseases (RSCD) is currently the only publicly available resource to identify diabetes cases through Danish register data (by application to the Danish Health Data Authority), but the RSCD has not been publicly validated, and the source code behind the algorithm has not been made publicly available. Notably, the algorithm lacks inclusion based on elevated HbA1c levels.<sup>76</sup>

Since the launch of the RSCD, nationwide laboratory data on HbA1c testing has become available in the Danish register ecosystem,<sup>77</sup> but this data is 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 to reduce the risk of complications and death. In an ageing population, the burden of T2D on healthcare services is rising, and resources must be prioritised to those with higher needs to ensure adequate care.

Migrants are particularly vulnerable to T2D, as they have an increased prevalence of the disease compared to native Danes. Despite having higher HbA1c levels than their native counterparts, the evidence on complication risk and mortality in migrants is inconsistent, but some evidence suggests a higher risk. Evidence on care disparities that may contribute to this excess risk in migrants is limited. Additionally, 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 no validated method currently exists to identify T2D cases in a general population, which leaves register-based studies in this area vulnerable to biases.

In Denmark, clinical guidelines are available to direct clinicians and patients towards the best possible outcomes. Despite the special challenges faced by migrants, the current guidelines do not contain specific recommendations for T2D care in migrants, and the evidence on care disparities between migrants and native Danes is limited.

## 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. To develop a register-based classification of type 1 and type 2 diabetes that incorporates previous recommendations on data sources, and to make it available to other researchers.
- II. To validate register-based diabetes classifiers available to researchers in Denmark.
- III. To assess the prevalence of T2D in migrant minorities of Denmark, and to explore care disparities between migrant groups and native Danes in terms of monitoring, biomarker control and pharmacological treatment as recommended in the guidelines.
- IV. To 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

Aim I was pursued in the creation of the *Open-Source Diabetes Classifier* (OSDC), made available in the open-source *R* package *osdc*.<sup>78</sup> Aims II, III, and IV were pursued in the respective papers I, II, and III of this dissertation.

## 3. Material and methods

This chapter describes the healthcare setting and the register data sources used in the studies of this PhD project. The diabetes classification algorithms used in the project are described along with the sources of survey data used to validate them. Finally, the study designs, populations, statistical methods and ethical approvals of the three papers in the dissertation are described.

### 3.1. Setting: The Danish healthcare system

Denmark is divided into five 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, other specialist practices (e.g. ophthalmologists) and 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 programmes on managing T2D through diet and physical exercise.

Although referral to outpatient hospital care is possible, most patients with T2D are treated by the GP. All native Danes and migrants with residence permit have free access to general practice. In the tax-funded Danish public healthcare system, Danish citizens and migrants with residence permit have equal access to services. Nearly all services are provided free of charge, although T2D patients must pay some out-of-pocket fees for podiatrist care and prescription drugs at pharmacies. Yearly medication expenditures are partially reimbursed from €132, and expenditures beyond €553 are fully covered (2019 limits).<sup>79</sup> In contrast, 50% of all expenditures related to diabetes-specific podiatrist care are covered by the public health insurance regardless of the patient's total expenditures.<sup>80</sup>

### 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 resi-

dents at immigration or birth. This covers the entire population and allows complete linkage between data sources; this forms the foundation of several nationwide Danish healthcare registers.<sup>[81]</sup> The following section briefly describes the data extracted from these registers and from other data sources used in the PhD project. A detailed list of variables and code values used in each study is available in the supplementary material attached to each paper.

#### 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 together with data on age, sex, equivalised disposable household income,<sup>82</sup> 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 along with data on procedures relating to retinal screening and treatment of cardiovascular complications was extracted for 1994-2018. At the time of this PhD project, the clinical coding system in use was the International Classification of Diseases, 10<sup>th</sup> revision (*ICD-10*), which was implemented in 1994.<sup>83</sup>

#### 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 healthcare billing information, and serves as the basis for payment from the public health insurance to the healthcare providers. 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 together with point-of-care HbA1c testing performed in general practice.

#### 3.2.4. The Danish National Prescription Registry

The Danish National Prescription Registry contains detailed information on all prescription drugs purchased at Danish pharmacies. These registrations serve to determine reimbursement awarded to each purchase from the public health insurance. Each purchase is classified according to the Anatomical Therapeutic Chemical (ATC) classification. Information on purchases of GLD, LLD, ACEI, ARB, and APT was extracted for 1995-2018.

#### 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 Danish hospitals. Except for point-of-care tests, all biomarker samples taken in the primary healthcare sector are analysed in hospital laboratories. Compared to the other registers, this is a relatively new data source; it was initially launched in 2013 and only recently achieved nationwide coverage.<sup>77</sup>

At the time of data extraction, the Central Denmark Region, one of five administrative regions in Denmark, which covers a population of 1.3 million inhabitants, corresponding to 22% of the entire Danish population, was not yet covered in the register. To address this issue, laboratory results from this part of 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 for 2011-2018.

### 3.3. Diabetes classification algorithms

As described previously in Section 1.5.4, the currently available register-based diabetes classifiers have yet to incorporate the emerging register data on routine HbA1c testing. 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.<sup>76</sup>

The following sections describe the algorithms behind the RSCD and OSDC diabetes classifiers. The overall design and the flow of populations are summarised in Figure 3.1. The advantages and disadvantages of their designs are discussed in Chapter 5.

### 3. Material and methods

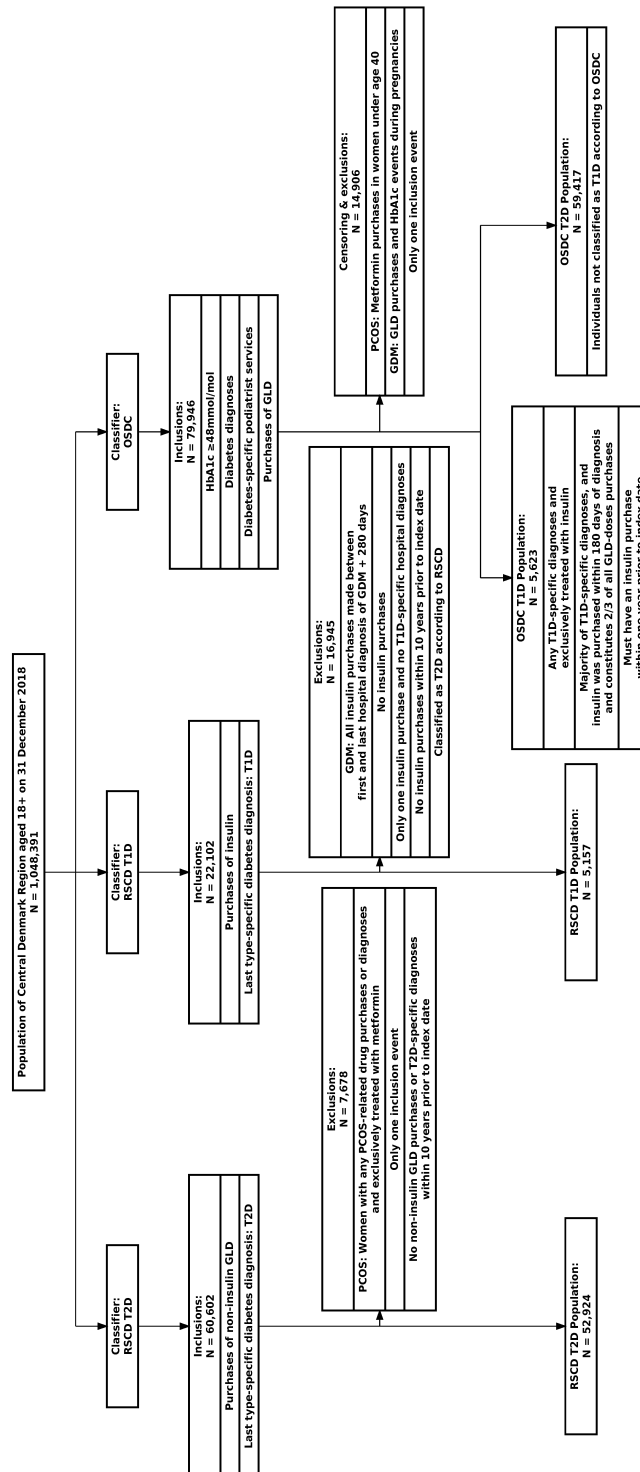


Figure 3.1.: Overall diabetes classifier design differences and population flows

#### 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 the 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 fulfil 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.<sup>84</sup>

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

Individuals not classified as T1D cases are classified as T2D cases.

#### 3.3.2. The Register of Selected Chronic Diseases

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

In the 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 inclusion criteria of the algorithm without fulfilling any of its exclusion criteria.

The RSCD was implemented as described below, based on available documentation,<sup>76</sup> which may not completely reflect the source code use by the Danish Health Data Authority to generate the register population. Although the implementation is generally ad verbatim to the documentation, a few areas of the documentation did not address all outlier cases, and in these cases we expanded the algorithm slightly from the documentation. For example, the documentation did not present a method to classify individuals with both a diagnosis of T1D and T2D on the day of their latest diagnosis, if these individuals also had an equal number of T1D and T2D diagnoses in their records. Following documentation to the letter would omit these patients completely from the diabetes population, and we assessed that this was likely a limitation in the documentation rather than the algorithm. Our implementation included these individuals as T2D cases, which seems most likely a priori.

##### 3.3.2.1. Classification of type 2 diabetes

For classification of T2D, the inclusion criteria are:

- 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 purchased only metformin and have any diagnoses of PCOS, purchases of clomifene or combination drugs containing antiandrogens and oestrogens.
- 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, the 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 purchased GLD only in the period from 280 days prior to their first diagnosis of GDM until 280 days after their last diagnosis of GDM.
- b. Individuals classified as T2D.
- c. Individuals without any purchases of GLD, or only one purchase and 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 are described below.

### 3.4.1. The Health In Central Denmark survey

*Health In Central Denmark* is a survey conducted in the Central Denmark Region using digital and postal questionnaires. The first wave of the survey was performed in 2020 and included all inhabitants aged 18-74 years with prevalent diabetes according to the OSDC on 31 December 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 a special focus on items related to diabetes mellitus, such as current disease, diabetes type, and age at onset. A total of 44,659 OSDC diabetes cases and 46,195 matched OSDC non-diabetes cases were invited to the survey, and 51,854 (57%) responded.<sup>85</sup>

#### 3.4.2. The National Health Survey

The National Health Survey is a nationwide repeated digital and postal questionnaire survey, which is conducted in 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.<sup>86</sup> This project utilised the subset of data from participants in the 2017 survey, who resided in Central Denmark Region. In this regional subset of the survey, a total of 52,000 individuals were invited and 32,400 (62%) responded.<sup>87</sup> Data from this survey was used only for 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 the register-based definitions of T1D and T2D. After validating the two candidates, the most accurate classifier, the OSDC, was chosen to identify the study populations of T2D for use in studies II and III.

##### 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. A total of 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 be evaluated as 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 survey respondents remained. However, as survey invites were conditioned on OSDC diabetes status, the survey population was biased towards the OSDC classification, and the 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

### 3.5. Study design and population

Table 3.1.: Overview of studies

	Study I	Study II	Study III
<b>Study population</b>	Participants in the Health In Central Denmark survey	Native Danes and first-generation migrants with T2D on 1 Jan 2018.	Native Danes and first-generation migrants with T2D on 1 Jan 2019
<b>Sample size</b>	29,391	254,097	253,364
<b>Calendar period</b>	Register-classified diabetes type on 31 Dec 2018	<i>Monitoring:</i> Oct 2015/2016 to Dec 2017. <i>Biomarker levels:</i> 1 Jan 2018. <i>Pharmacological treatment:</i> during 2018.	GLD use during 2019.
<b>Design</b>	Validation study	Cross-sectional study	Cross-sectional study
<b>Data sources</b>	Questionnaire and register data	Register data	Register data
<b>Age</b>	18-74 years	25-99 years	25 years and older
<b>Variables</b>	<i>Validated standard:</i> Register-classified diabetes type <i>Gold standard:</i> Self-reported diabetes type	<i>Primary independent variable:</i> Migrant origin <i>Dependent variables:</i> Timely T2D monitoring. HbA1c and LDL-C control. Use of recommended pharmacological treatment.	<i>Primary independent variable:</i> Migrant origin <i>Dependent variables:</i> Use of GLD combination therapy. Use of oral GLD types. Use of injection-based GLD types
<b>Statistical analyses</b>	<i>Validation metrics:</i> Sensitivity, specificity, positive and negative predictive value	<i>Robust Poisson regression:</i> Relative risk of non-fulfilment of guideline recommendations.	<i>Robust Poisson regression:</i> Relative risk of GLD use

Notes

- Some details were left out for the sake of brevity, i.e. the analysis of T2D prevalence in study II and exclusion criteria.
- Abbreviations: GLD: Glucose-lowering drugs. T2D: Type 2 diabetes.

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.

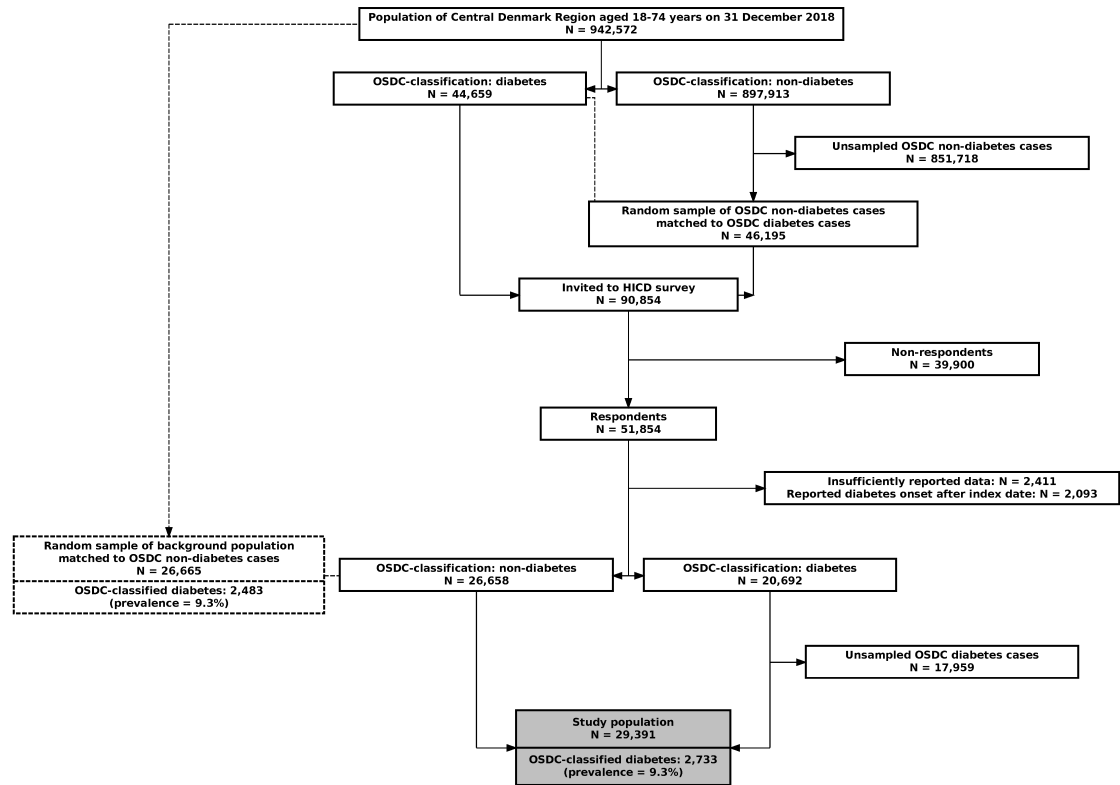


Figure 3.2.: Flow of individuals into the *Health In Central Denmark* survey and the study population of study I

#### 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 in Denmark for at least three years on the index date.

::: .content-hidden unless-format="pdf"}

T2D prevalence was studied in the above population. To study T2D care where clinically relevant, the eleven analyses of T2D care were performed in distinct subsets of the population with T2D, as outlined below and in [?@fig-study-2-flow-pdf](#):

:::

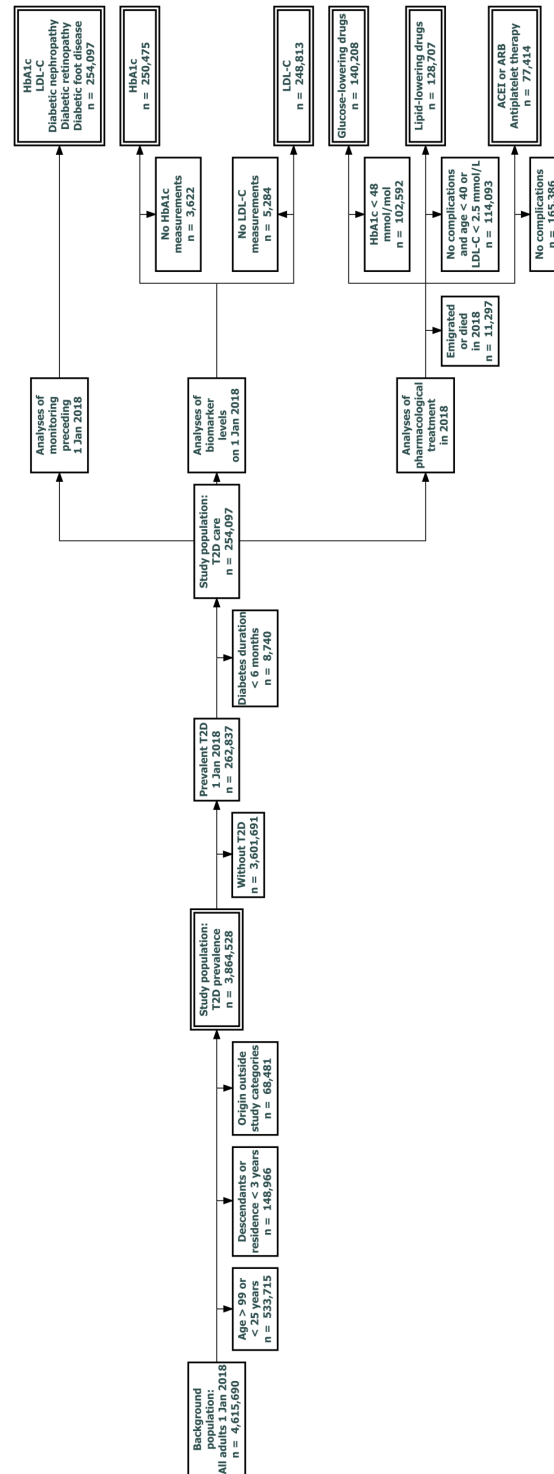
::: .content-hidden when-format="pdf"}

T2D prevalence was studied in the above population. To study T2D care where clinically relevant, the eleven analyses of T2D care were performed in distinct subsets of the population with T2D, as outlined below and in [?@fig-study-2-flow-nonpdf](#):

:::

- All analyses of T2D care were performed in a subset of 254,097 individuals with T2D onset at least six months prior to the index date 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 residing in Denmark in the year following the index date, 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. Material and methods



## Study population flow in analyses in study II

### 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 years. Among individuals with T2D, 269,011 native Danes or first-generation migrants from the selected origin categories residing for at least three years in Denmark on the index date were included. After exclusion of 15,647 individuals without records of HbA1c or LDL-cholesterol within three years prior to the index date, the study population contained 253,364 individuals.

### 3.5.2. Outcomes

#### 3.5.2.1. Study I

Self-reported diabetes type, the validation gold standard, was categorised as either T1D (self-reported T1D), T2D (all other types of diabetes) or no diabetes, corresponding to the diabetes types discernible from the register-based classifiers.

#### 3.5.2.2. Study II

Twelve outcomes were defined as follows:

1. 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, biomarker levels and pharmacological treatment.

T2D monitoring was evaluated in time-frames prior to index date that matched guideline-recommendations, with three months of leeway added (e.g. monitoring within 15 months was considered to fulfil the guideline recommendation, even if the specific type of monitoring is recommended yearly in the guidelines):

2. Monitoring of HbA1c was identified by records of biomarker samples between 1 October 2016 and 31 December 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 *Danish National Health Service Register*.

3. Monitoring of LDL-C: was identified by records of biomarker samples between 1 October 2016 and 31 December 2017 in the *Register of Laboratory Results for Research* or the *Clinical Laboratory Information System*.
4. UACR-screening for diabetic nephropathy was identified by records of biomarker samples between 1 October 2016 and 31 December 2017 in the *Register of Laboratory Results for Research* or the *Clinical Laboratory Information System*.
5. Screening for diabetic retinopathy was identified by records of retinal photo screening for diabetic retinopathy between 1 October 2015 and 31 December 2017, either at hospital departments in the *Danish National Patient Register*, or diabetes-specific examinations at ophthalmologist practice recorded in the *Danish National Health Service Register*.
6. Screening for diabetic foot disease was identified by records of diabetes-specific podiatrist services received between 1 October 2016 and 31 December 2017 in the *Danish National Health Service Register*.

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

7. The outcome measure for HbA1c control was 53 mmol/mol.
8. The outcome measure for LDL-C control was 2.6 mmol/L.

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

9. GLD: ATC codes A10.
10. LLD: ATC codes C10.
11. ACEI/ARB: ATC codes C09A-C09D.
12. APT: ATC codes B01AC.

#### 3.5.2.3. Study III

Eight outcomes were defined across combination therapy and individual GLD types used.

GLD combination therapy was analysed using the following two outcome definitions:

1. Any combination therapy: two or more GLD types used.
2. Triple combination therapy: three or more GLD types used.



For each GLD type, individuals were defined as users if at least two purchases were made during 2019, and at least one of these was made in the last six months of the year. The following types of GLDs (ATC codes including subcodes) counted towards the number of GLD types used by an individual in a year (fixed-dose combination drugs counted towards both of their component drug types), and they also defined the six outcomes for analysis of individual GLD type usage:

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

In addition, thiazolidinediones (ATC codes A10BG) and repaglinide (ATC codes A10BX02) counted towards GLD combination therapy, but they were not analysed as individual drug types due to the very low prevalence of use of these drugs.

#### 3.5.3. Primary variable of interest: migrant status

Migrant status and country of origin were defined according to the definitions by Statistics Denmark. To accommodate the expected heterogeneity between different countries of origin, aggregation of countries was kept to a minimum and limited to countries with similar reasons for migration and with racial and cultural similarities. Attempting to preserve the individual origin countries with the most T2D cases in Denmark, and aggregating some countries based on the United Nations M49 geoscheme, the population was grouped into nine origin categories:

- Native Danes
- Europe (M49: Europe, excluding countries of the former Yugoslavia)
- Middle East (M49: Northern Africa, Western Asia, excluding 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 the index dates were treated as continuous variables. Household income was based on the three-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 categorised in the maximally acculturated group. Employment status was categorised 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, ischaemic 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 and LLD use within a year prior to the index date. Finally, the analyses of the use of individual GLD drug types in study III adjusted for combination therapy categorised as *Monotherapy or no GLD*, *Dual combination* and *Triple combination or more*.

#### 3.6. Statistical analysis

Statistical analyses were performed in RStudio<sup>88</sup> using R version 4<sup>89</sup> and a range of open-source code packages for computations<sup>90–94</sup> and visualisation.<sup>95,96</sup>

This section describes our overall statistical analysis strategy and 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). However, ORs approximate RR **only** when the outcome is *rare*. If this is not the case, interpreting ORs as RRs will exaggerate risk estimates,<sup>97</sup> as illustrated in Figure 3.3 (adapted from Wikipedia Commons).<sup>98</sup>

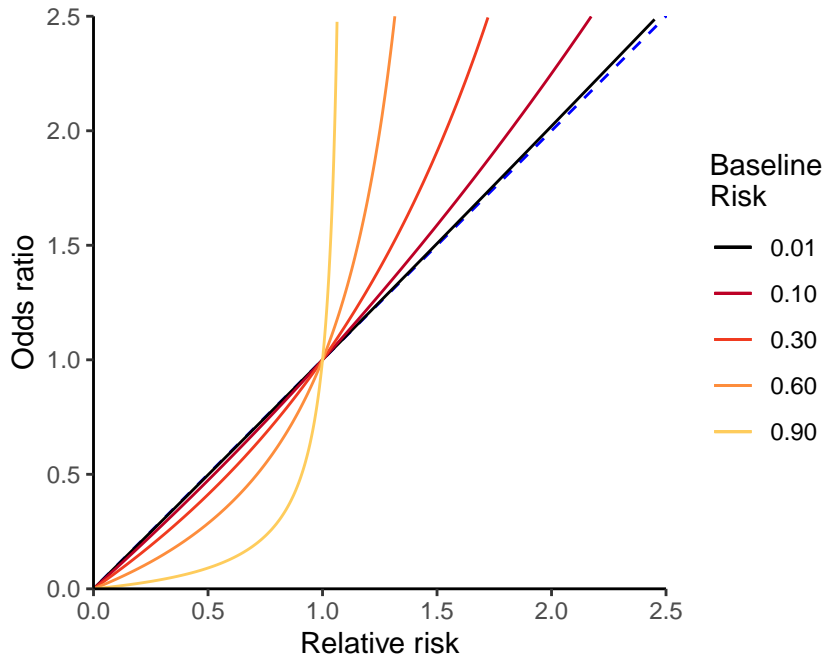


Figure 3.3.: Relationship between odds ratio and relative risk depending on baseline risk

None of the outcomes in studies II & III were expected to be rare. Thus, logistic regression was unsuitable. Although log-binomial regression models are the most natural choice for computing relative risk in a mathematical sense, they are prone to convergence failure.<sup>99</sup> 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)<sup>100</sup> presented an appealing alternative for modelling RR. A comprehensive description of the technique is beyond the scope of this dissertation, but a brief explanation is required, since ordinary Poisson regression is not the right tool for modelling binary outcomes as such data violate the assumptions behind the model.<sup>101</sup> 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 generalised estimating equations (so-called *sandwich*) variance estimates, which provide accurate standard errors in the presence of over-dispersion and excess zeros,<sup>102</sup> such as in a binary distribution. Ultimately, this allows robust Poisson regression to estimate RR with valid confidence intervals in data with binary outcomes.<sup>103</sup> It is worth noting that RR is merely a measure of relative probability, despite the semantics around *risk* implying that an adverse outcome is under study, which is not necessarily the case.

All regression models were modelled with native Danes as the reference group. To account for potential non-linearity, continuous variables were modelled as natural splines with knots placed at each quintile of the distribution in the respective study population. Supplementary analyses, where continuous variables were modelled as categorical variables (quintiles), were conducted for easier interpretation of these variables. In addition, the supplementary analyses included estimates of absolute risk differences corresponding to the above analyses. In these analyses, a generalised 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. Continuous variables were presented with means and standard differences, and categorical variables were represented by absolute numbers and percentage proportions. In study I, the characteristics were tabulated according to self-reported, OSDC-classified and RSCD-classified diabetes type, whereas studies II and III tabulated the distribution of covariates and outcomes by origin group.

The following sections describe the statistical analyses of each study.

#### 3.6.0.1. Study I

Validation analyses were performed separately for T1D and T2D, and each register-based diabetes classifier was validated against self-reported diabetes type. In the analyses of each diabetes type, diabetes type was modelled as a distinct binary variable; does the individual have this diabetes type or not (e.g. in the analyses of T1D, diabetes type was modelled 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 Table 3.3 and 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 1,000 random subsamples in order to assess robustness.

Table 3.3.: Concordance table and validation metrics

	Survey positives	Survey negatives	
Classifier positives	True positives	False positives	<i>PPV</i>
Classifier negatives	False negatives	True negatives	<i>NPV</i>
	<i>Sensitivity</i>	<i>Specificity</i>	

$$Sensitivity = \frac{N_{truepositives}}{N_{truepositives} + N_{falsenegatives}} \quad (3.1)$$

$$Specificity = \frac{N_{truenegatives}}{N_{truenegatives} + N_{falsepositives}} \quad (3.2)$$

$$PPV = \frac{N_{truepositives}}{N_{truepositives} + N_{falsepositives}} \quad (3.3)$$

$$NPV = \frac{N_{truenegatives}}{N_{truenegatives} + N_{falsenegatives}} \quad (3.4)$$

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. This was done 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 modelled 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 visualise 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 behaviour and healthcare service usage (employment status, household income, duration of residence and region of residence).

#### 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 study III 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 included 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 in this PhD project was registered in the internal register of research projects (Danish: *fortegnelsen*) of Aarhus University, under file number 2016-051-000001, serial number 1339. Study II was approved by the *Health In Central Denmark* steering committee. The *Health In Central Denmark* research project is registered in the internal register of research projects (file number 1-16-02-165-20) in the Central Denmark Region.

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

This chapter presents the nationwide register-based T2D cohort identified by the OSDC and details the flow of individuals through the algorithm. Next, the main results of each study associated with the dissertation are presented.

Additional details are available in the respective papers and their supplementary material, attached in Appendix B.

### 4.1. Register-based type 2 diabetes population

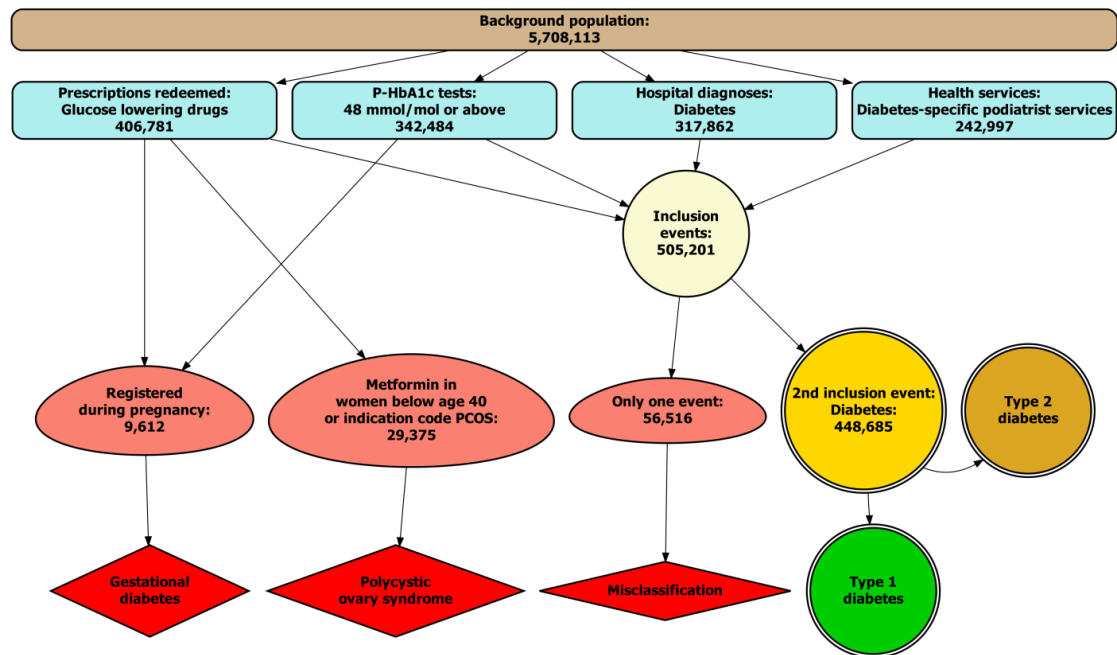
#### 4.1.1. Flow of individuals in the Open-Source Diabetes Classifier

This section shows the OSDC diabetes classification algorithm applied on the register data and details how it identifies individuals with T2D (i.e. the study populations in subsequent studies).

When applied to nationwide register data (described in Chapter 3) on all individuals aged 18 years and above, the OSDC identifies a nationwide cohort of individuals with prevalent diabetes at any point in time of 448,685 individuals, as shown in Figure 4.1. Of the four inclusion criteria, GLD purchase was most the common, being present in roughly 90% of diabetes cases, followed by records of elevated HbA1c levels in 75%, hospital diagnoses of diabetes in 70%, and diabetes-specific podiatrist services in 55% of cases. Censoring of potential GDM affected fewer individuals relative to censoring of potential PCOS and exclusion of cases with only a single inclusion event.

In this diabetes cohort, 34,079 were classified as T1D and 425,300 as T2D, of whom 24,026 and 276,081 were alive and residing in Denmark with prevalent T1D and T2D on 1 January 2019 after adjusting for insulin use in the previous year, as shown in Figure 4.2. Among individuals exclusively treated with insulin (left half of Figure 4.2), around 75% had at least one primary hospital diagnosis of T1D from a medical department. These constituted the vast majority of T1D cases, as the classification branch containing individuals with any non-insulin GLD purchase (right half of Figure 4.2) contributed only 2,748 (8%) of all T1D cases in the cohort. Correcting the cross-sectional classifier populations for recent insulin purchases prior to a particular index date had little impact

(e.g. on 1 January 2019, only 413 (1.7%) of the T1D classifications in the cohort had not made a purchase of insulin in the previous year). The classification of T2D cases in the cohort was more homogeneous, as the vast majority (382,116 (90%) of 425,300 cases) were individuals with purchases of non-insulin GLD and no primary hospital diagnoses of T1D from a medical department (the rightmost path of Figure 4.2).



Counts indicate the number of individuals with any records of each inclusion criteria, and with any censored inclusion events (thus, an individual may contribute to several counts, and individuals with censored events may have uncensored events recorded at other points in time).

Figure 4.1.: Inclusion and censoring in the Open-Source Diabetes Classifier



## 4.1. Register-based type 2 diabetes population

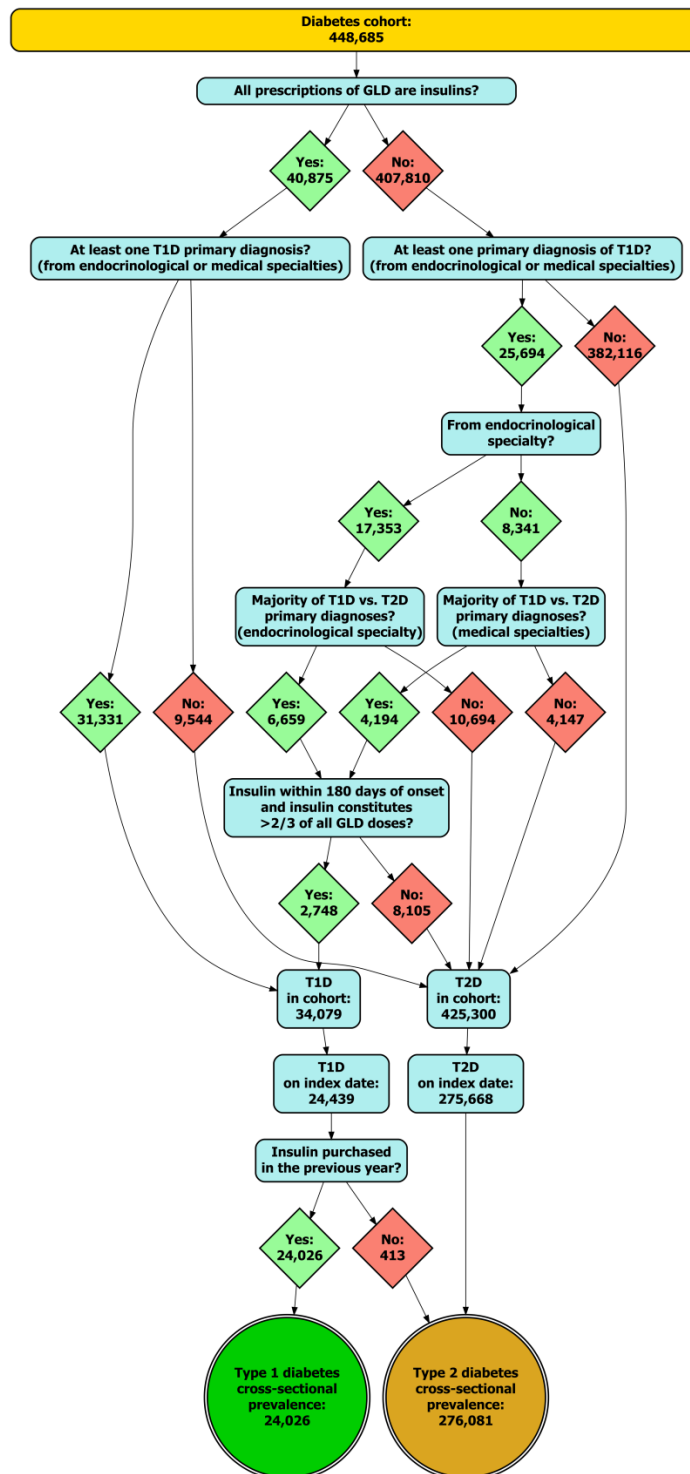


Figure 4.2.: Logic and flow of classification of diabetes type in the Open-Source Diabetes Classifier

### 4.2. Study I

This study validated the OSDC and the RSCD, two predefined register-based classifiers of diabetes type, against self-reported diabetes type in a general survey population. Among 29,391 individuals aged 18-74 years, the prevalence of self-reported diabetes was 2,633 (9.0%); 410 (1.4%) reported T1D and 2,223 (7.6%) reported T2D.

In the classification of T1D and T2D, the sensitivity of the OSDC was 0.773 (95% CI [0.730; 0.813]) and 0.944 (95% CI [0.933; 0.953]), respectively. Compared to the RSCD, sensitivity of the OSDC was substantially higher in the classification of T1D (7% higher) and in the classification of T2D (4% higher). In the classification of T1D, the specificity was practically identical in the two algorithms at 0.999 (95% CI [0.999; 1.000]), while the specificity of the OSDC in the classification of T2D was 0.989 (95% CI [0.988; 0.990]), which was 0.3% lower than for the RSCD. This corresponded to a PPV of 0.875 (95% CI [0.861; 0.888]), which was 2.3% lower than for the RSCD. See Table 4.1 and Table 4.2 for details.

In the analyses stratified by age at onset of diabetes, both sensitivity and PPV in the classification of T1D were much higher in individuals with diabetes onset before age 40 years than in individuals with diabetes onset later in life. Similarly, both sensitivity and PPV in the classification of T2D were much higher in individuals with diabetes onset after age 40 than in individuals with earlier onset of diabetes. See Table 4.3 and Table 4.4 for details.

Table 4.1.: Validation of register-based classification of type 1 diabetes

(a) The Open-Source Diabetes Classifier (OSDC)

OSDC	Survey: +T1D	Survey: -T1D	Total N
<b>OSDC: +T1D</b>	317	19	336
<b>OSDC: -T1D</b>	93	28,962	29,055
<b>Total N</b>	410	28,981	29,391
<b>Sensitivity:</b>	0.773 (0.730, 0.813)		
<b>Specificity:</b>	0.999 (0.999, 1.000)		
<b>PPV:</b>	0.943 (0.913, 0.966)		
<b>NPV:</b>	0.997 (0.996, 0.997)		

(b) The Register for Selected Chronic Diseases (RSCD)

RSCD	Survey: +T1D	Survey: -T1D	Total N
<b>RSCD: +T1D</b>	287	17	304
<b>RSCD: -T1D</b>	123	28,964	29,087
<b>Total N</b>	410	28,981	29,391
<b>Sensitivity:</b>	0.700 (0.653, 0.744)		
<b>Specificity:</b>	0.999 (0.999, 1.000)		
<b>PPV:</b>	0.944 (0.912, 0.967)		
<b>NPV:</b>	0.996 (0.995, 0.996)		

Notes

- “-T1D” designates individuals with type 2 diabetes or no diabetes according to the source (classifier or survey).
- Abbreviations: T1D: Type 1 diabetes. OSDC: Open-Source Diabetes Classifier. RSCD: Register for Selected Chronic Diseases. PPV: Positive predictive value. NPV: Negative predictive value.

Table 4.2.: Validation of register-based classification of type 2 diabetes

(a) The Open-Source Diabetes Classifier (OSDC)

OSDC	Survey: +T2D	Survey: -T2D	Total N
<b>OSDC: +T2D</b>	2,098	299	2,397
<b>OSDC: -T2D</b>	125	26,869	26,994
<b>Total N</b>	2,223	27,168	29,391
<b>Sensitivity:</b>	0.944 (0.933, 0.953)		
<b>Specificity:</b>	0.989 (0.988, 0.990)		
<b>PPV:</b>	0.875 (0.861, 0.888)		
<b>NPV:</b>	0.995 (0.994, 0.996)		

(b) The Register for Selected Chronic Diseases (RSCD)

RSCD	Survey: +T2D	Survey: -T2D	Total N
<b>RSCD: +T2D</b>	2,011	229	2,24
<b>RSCD: -T2D</b>	212	26,939	27,151
<b>Total N</b>	2,223	27,168	29,391
<b>Sensitivity:</b>	0.905 (0.892, 0.917)		
<b>Specificity:</b>	0.992 (0.990, 0.993)		
<b>PPV:</b>	0.898 (0.884, 0.910)		
<b>NPV:</b>	0.992 (0.991, 0.993)		

Notes:

- “-T2D” designates individuals with type 1 diabetes or no diabetes according to the source (classifier or survey).
- Abbreviations: T2D: Type 2 diabetes. OSDC: Open-Source Diabetes Classifier. RSCD: Register for Selected Chronic Diseases. PPV: Positive predictive value. NPV: Negative predictive value.

Table 4.3.: Validation of classification of type 1 diabetes stratified by age at onset

(a) Type 1 diabetes onset before age 40 years

	Survey: +T1D	Survey: -T1D	Sensitivity	PPV
<b>OSDC</b>				
+T1D	283	13	0.884 (0.844, 0.917)	0.956 (0.926, 0.976)
-T1D	37	27,000		
<b>RSCD</b>				
+T1D	262	11	0.819 (0.772, 0.859)	0.960 (0.929, 0.980)
-T1D	58	27,002		

(b) Type 1 diabetes onset after age 40 years

	Survey: +T1D	Survey: -T1D	Sensitivity	PPV
<b>OSDC</b>				
+T1D	34	14	0.378 (0.278, 0.486)	0.708 (0.559, 0.830)
-T1D	56	28,712		
<b>RSCD</b>				
+T1D	25	13	0.278 (0.189, 0.382)	0.658 (0.486, 0.804)
-T1D	65	28,713		

Notes

- “-T1D” designates individuals with type 2 diabetes or no diabetes according to the source (classifier or survey).
- Self-reported non-diabetes cases were included in both strata of age at onset.
- Abbreviations: T1D: Type 1 diabetes. OSDC: Open-Source Diabetes Classifier. RSCD: Register for Selected Chronic Diseases. PPV: Positive predictive value.

Table 4.4.: Validation of classification of type 2 diabetes stratified by age at onset

(a) Type 2 diabetes onset before age 40 years

	Survey: +T2D	Survey: -T2D	Sensitivity	PPV
<b>OSDC</b>				
+T2D	220	247	0.863 (0.814, 0.902)	0.471 (0.425, 0.517)
-T2D	35	26,831		
<b>RSCD</b>				
+T2D	218	169	0.855 (0.806, 0.896)	0.563 (0.512, 0.613)
-T2D	37	26,909		

(b) Type 2 diabetes onset after age 40 years

	Survey: +T2D	Survey: -T2D	Sensitivity	PPV
<b>OSDC</b>				
+T2D	1,878	265	0.954 (0.944, 0.963)	0.876 (0.862, 0.890)
-T2D	90	26,583		
<b>RSCD</b>				
+T2D	1,793	175	0.911 (0.898, 0.923)	0.911 (0.898, 0.923)
-T2D	175	26,673		

Notes

- “-T2D” designates individuals with type 1 diabetes or no diabetes according to the source (classifier or survey).
- Self-reported non-diabetes cases were included in both strata of age at onset.
- Abbreviations: T2D: Type 2 diabetes. OSDC: Open-Source Diabetes Classifier. RSCD: Register for Selected Chronic Diseases. PPV: Positive predictive value.

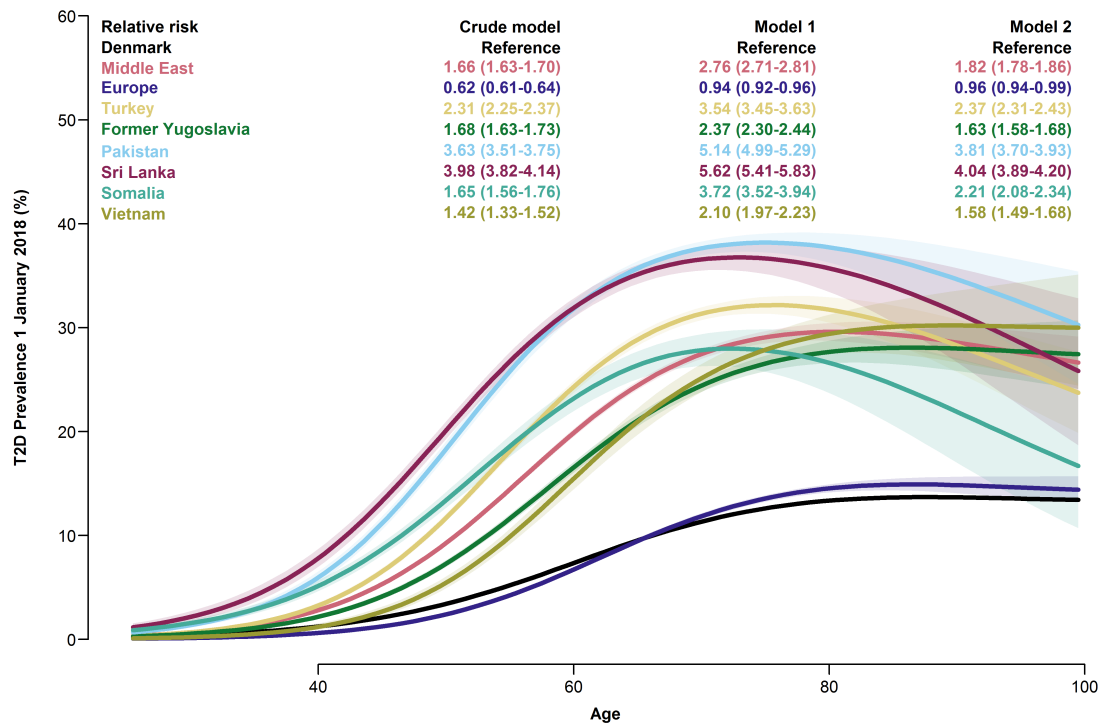
### 4.3. Study II

This study contained a series of cross-sectional analyses on T2D prevalence and eleven indicators of guideline-level care across monitoring, biomarker control and pharmacological treatment.

#### 4.3.1. Prevalence

Compared to native Danes, the risk of prevalent T2D was elevated in all migrants, except the Europe group, regardless of model, and it was highest in the Sri Lanka and Pakistan groups (crude RR 3.98 [3.82-4.14] and 3.63 [3.51-3.75], resp.). In the age-specific model, the increase in prevalence was discernible in all age groups, including the youngest. See Figure 4.3 for details.

## 4. Results in summary



Overall relative risks and age-specific prevalences of type 2 diabetes with 95% CIs for each migrant group.

- Model 1 adjusted for age and sex.
- Model 2 further adjusted for employment status, household income, duration of residence and region of residence.

Figure 4.3.: Prevalence of type 2 diabetes



### 4.3.2. Monitoring

Among individuals with T2D, the proportion without diabetes monitoring within the guideline-recommended intervals varied between the five types of monitoring assessed. In native Danes, the proportions without monitoring of HbA1c, LDL-C, diabetic nephropathy, diabetic retinopathy and diabetic foot disease were 6.8%, 13.1%, 43.6%, 43.2% and 57.0%, respectively. Compared to native Danes, migrant groups had similar or higher crude RR in these indicators of diabetes monitoring. The only exception was the Sri Lanka group, which had lower risk in all analyses, except screening for diabetic foot disease (RRs from 0.64 [0.51-0.80] for HbA1c to 0.79, [0.74-0.84] for diabetic nephropathy). Risk estimates were stable in the Europe-group, but they attenuated with increasing adjustment in other migrant groups. See Figure 4.4 for details.

## 4. Results in summary



Baseline risk in native Danes and relative risk in each migrant group with 95% CIs.

- Model 1 adjusted for age, sex, diabetes duration and prevalent complications.
- Model 2 further adjusted for employment status, household income, duration of residence and region of residence.
- Abbreviations: HbA1c: hemoglobin-A1c. LDL-C: low-density lipoprotein cholesterol.

Figure 4.4.: Risk of not receiving guideline-recommended monitoring

### 4.3.3. Biomarker control

Among native Danes, 37.2% had an HbA1c level  $\geq 53$  mmol/mol and 28.3% had an LDL-C level  $\geq 2.6$  mmol/L. In both biomarkers, the risk of having an elevated level was higher in most migrant groups compared to native Danes. The crude risk of dysglycaemia was increased in all migrant groups, except the Europe and Vietnam groups (RRs from 1.27 [1.24-1.30] in the Middle East group to 1.46 [1.41-1.52] in the Pakistan group), while dyslipidaemia risk was increased in all migrant groups, except the Sri Lanka and Vietnam groups (RRs from 1.08 [1.03-1.14] in the Former Yugoslavia group to 1.78 [1.67-1.90] in the Somalia group). Adjustment for clinical factors (model 1) roughly halved the size of the increased risks compared to crude estimates, while further adjustment for socioeconomic factors (model 2) only had minor effect on the estimates. See Figure 4.5 for details.



Baseline risk in native Danes and relative risk in each migrant group with 95% CIs.

- Model 1 adjusted for age, sex, diabetes duration and prevalent complications.
- Model 2 further adjusted for employment status, household income, duration of residence and region of residence.
- Abbreviations: HbA1c: haemoglobin-A1c. LDL-C: low-density lipoprotein cholesterol.

Figure 4.5.: Risk of exceeding guideline-recommended biomarker targets

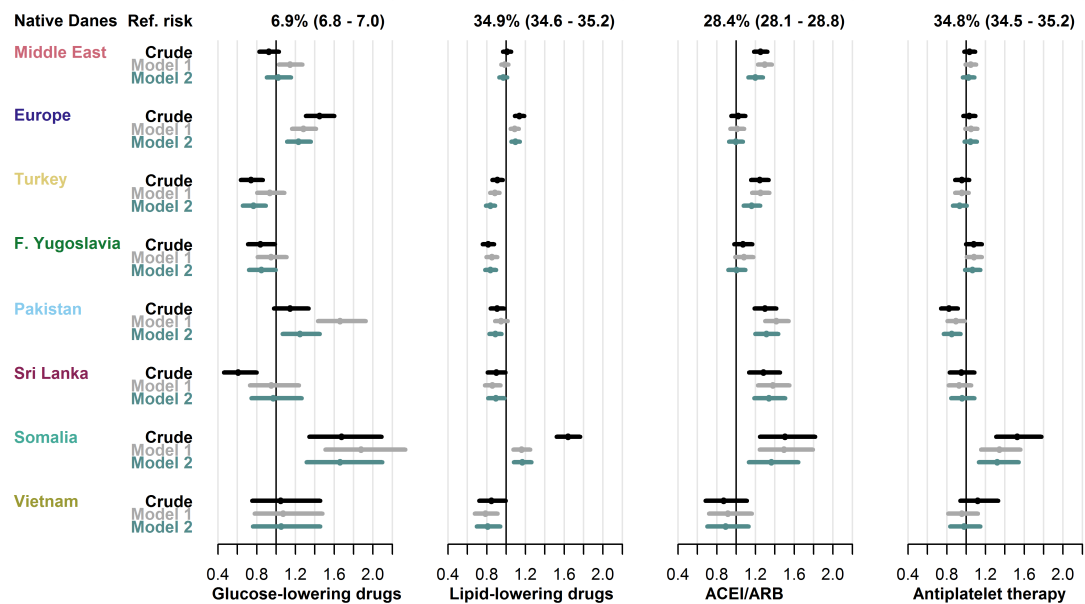
### 4.3.4. Pharmacological treatment

Among native Danes with a treatment indication for GLD, LLD, ACEI/ARB and APT, 6.9%, 34.9%, 28.4% and 34.8%, respectively, did not receive treatment. In pharmacological treatment, the overall risk patterns in migrants were less clear than found in previous outcomes, but some groups had notably high risks.

Compared to native Danes, the crude risk of not receiving GLD was increased in the Europe (RR 1.45 [1.31-1.60]) and the Somalia (RR 1.67 [1.34-2.09]) groups. Both groups also had increased risk of not receiving LLD (RR 1.14 [1.09-1.19] in the Europe group and 1.64 [1.52-1.77] in the Somalia group). In the remaining migrant groups, the risk of not receiving GLD and LLD was similar or lower than the risk among native Danes. The risk of not receiving ACEI/ARB was higher in five of the eight migrant groups (RRs from 1.24 [1.15-1.34] in the Turkey group to 1.50 [1.25-1.82] in the Somalia-group), but only the Somalia group was at higher risk of not receiving APT than native Danes (RR 1.53 [1.31-1.78]).

Adjustment for clinical and socioeconomic factors had varying effect on risk estimates in the analysis of GLD and had little effect in the analyses of LLD, ACEI/ARB and APT. In particular, the high RR of not receiving LLD in the Somalia-group attenuated substantially following adjustment, but the risk remained higher than in native Danes. See Figure 4.6 for details.

## 4. Results in summary



Baseline risk in native Danes and relative risk in each migrant group with 95% CIs.

- Model 1 adjusted for age, sex, diabetes duration and prevalent complications.
- Model 2 further adjusted for employment status, household income, duration of residence and region of residence.
- Abbreviations: ACEI/ARB: angiotensin-converting enzyme-inhibitors or angiotensin receptor blockers.

Figure 4.6.: Risk of not receiving guideline-recommended pharmacological treatment

## 4.4. Study III

This nationwide cross-sectional study examined user-prevalence of combination therapy and individual GLD types within T2D in 2019, focusing on disparities in migrants.

During 2019, 34.7% of native Danes received any combination therapy, and 12.9% received triple combination therapy. In the adjusted models, the RR was lower in most migrant groups for any combination therapy, as the fully adjusted RR (model 2) ranged from 0.77 [0.71-0.85] (Somalia group) to 1.00 [0.97-1.04] (Former Yugoslavia group). Similarly, the fully adjusted RR of using triple combination therapy ranged from 0.56 [0.46-0.68] (Somalia group) to 1.03 [0.96-1.10] (Former Yugoslavia group). See Figure 4.7 for details.

Among orally administered GLDs, the most widely used drug type during 2019 in native Danes was metformin (used by 62.1%), followed by DPP4i (13.3%), SGLT2i (11.9%) and sulfonylureas (5.2%). In these GLD types, the crude and adjusted RR was higher in most migrant groups. For each drug type, the range of the fully adjusted RRs (model 2) were: **metformin**: 0.99 [0.97-1.01] (Europe group) to 1.10 [1.06-1.15] (Somalia group), **sulfonylureas**: 1.05 [0.99-1.11] (Europe group) to 1.95 [1.59-2.39] (Vietnam group), **DPP4i**: 1.05 [0.99-1.10] (Europe group) to 1.66 [1.46-1.90] (Somalia group) and **SGLT2i**: 1.06 [0.92-1.21] (Vietnam group) to 1.14 [1.09-1.19] (Turkey group). See Figure 4.8 for details.

Among injection-based GLD types, 18.7% of native Danes received insulins during 2019, and 13.3% received GLP1RA. In these GLD types, the fully adjusted RR (model 2) was lower in all migrant groups, as the RRs ranged from 0.66 [0.62-0.71] (Sri Lanka group) to 0.94 [0.89-0.99] (Europe group) for insulin use, and from 0.29 [0.22-0.39] (Somalia group) to 0.95 [0.89-1.01] (Europe group) for GLP1RA use. See Figure 4.9 for details.

## 4. Results in summary



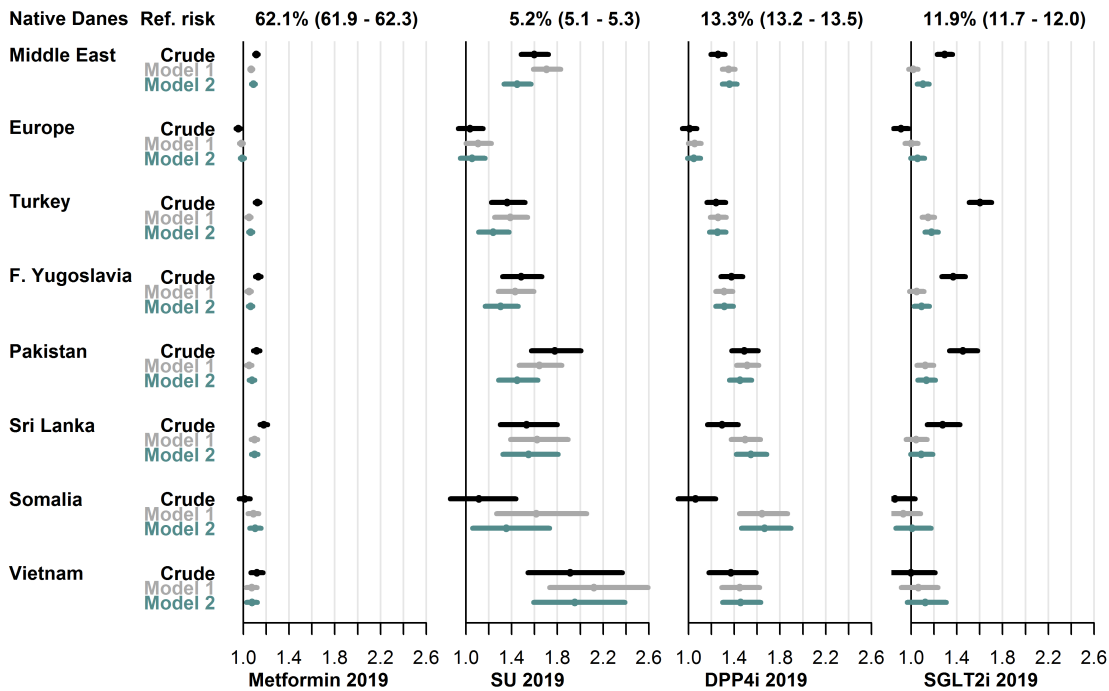
User-prevalence in native Danes and relative risk in each migrant group with 95% CIs.

- Model 1 adjusted for sex, age, diabetes duration, hemoglobin-A1c level, low-density lipoprotein cholesterol level, lipid-lowering drug use, prevalent complications.
- Model 2 further adjusted for employment status, household income, duration of residence, region of residence.

Figure 4.7.: Glucose-lowering drug combination therapy



#### 4.4. Study III



User-prevalence in native Danes and relative risk in each migrant group with 95% CIs.

- Model 1 adjusted for sex, age, diabetes duration, hemoglobin-A1c level, low-density lipoprotein cholesterol level, lipid-lowering drug use, prevalent complications, glucose-lowering drug combination therapy.
- Model 2 further adjusted for employment status, household income, duration of residence, region of residence.
- Abbreviations: SU: sulfonylureas. DPP4i: dipeptidyl peptidase-4 inhibitors. SGLT2i: sodium glucose cotransporter-2 inhibitors.

Figure 4.8.: Oral glucose-lowering drug types

## 4. Results in summary



User-prevalence in native Danes and relative risk in each migrant group with 95% CIs.

- Model 1 adjusted for sex, age, diabetes duration, hemoglobin-A1c level, low-density lipoprotein cholesterol level, lipid-lowering drug use, prevalent complications, glucose-lowering drug combination therapy.
- Model 2 further adjusted for employment status, household income, duration of residence, region of residence.
- Abbreviations: GLP1RA: glucagon-like peptide-1 receptor agonists.

Figure 4.9.: Injection-based glucose-lowering drug types

## 5. Discussion of methods

Before discussing the methodological design of the studies in the dissertation, this chapter first discusses the algorithm design choices made in the OSDC and the 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 has several limitations. This section discusses the challenges identified during the development and implementation of diabetes classifiers in this PhD project prior to the validation performed in study I. 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 addressed 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 the 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. When 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 in cross-sectional studies.

However, pursuing this issue further showed that most of these cases were young 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 the clinical obstetric coding recorded in the National Patient Register to identify individual pregnancy windows. Therefore, 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

Danish Medical Birth Register<sup>104</sup> was attempted. This showed that the clinical obstetric diagnosis codes use to capture births in the National Patient Register was indeed lacking in the first few years after the adoption of the ICD-10 coding system in 1994. However, it appeared to be sufficient from 1997 onwards, as shown in Figure 5.1 (note the gap in annual births between the two sources during 1993-1996 highlighted in grey). This led to the decision to restrict inclusion from prescription data to 1997 onward.

Any algorithm censoring GDM by using obstetrical codes from data prior to 1997 in the National Patient Register is likely to include false-positive diabetes cases among women. The RSCD censors GDM based on hospital diagnoses of GDM, but it ultimately manages these old cases (and other potentially misclassified cases) by retrospectively excluding individuals without any inclusion events in the 10 years preceding a given index date. While such a mechanism solves this issue in a cross-sectional study, it may be problematic for study designs with dynamic index dates (e.g. longitudinal studies with index dates at the time of diabetes onset are not possible).

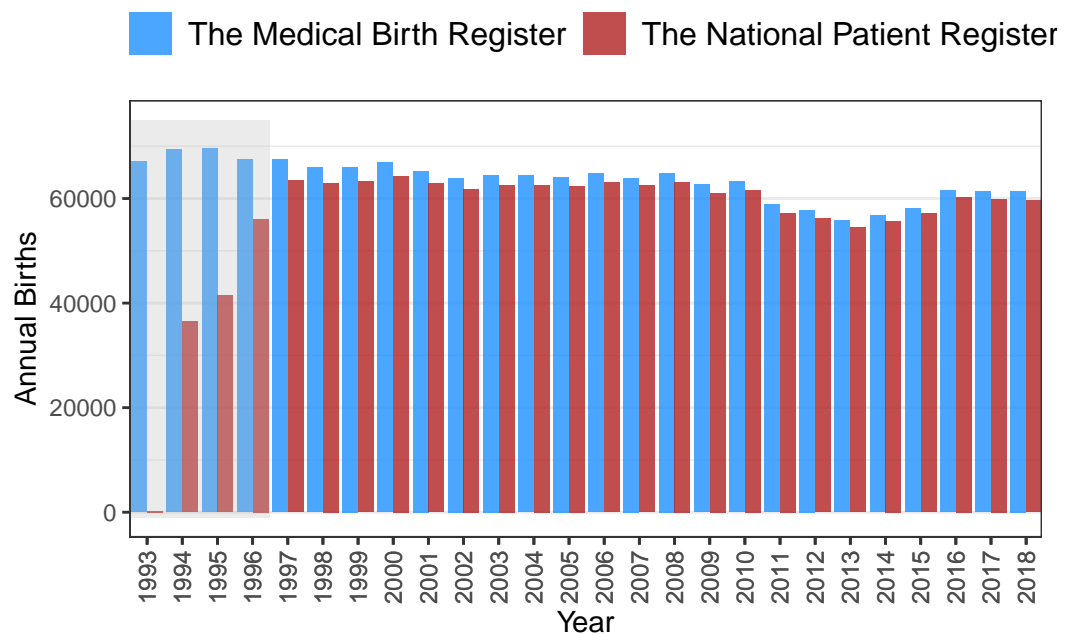


Figure 5.1.: Annually registered births

### 5.1.2. Approaches to censoring polycystic ovary syndrome

In Denmark, women with PCOS are diagnosed and treated mainly in the primary care sector, and therefore these cases are not registered in the *Danish National Patient Register*. Metformin, a GLD, is used in the treatment of PCOS, and this may lead to inclusion of false-positive diabetes cases among women treated with metformin. As PCOS affects mainly young women, a possible solution can be to censor metformin purchases among women below an arbitrary age cut-off.

In OSDC, we adopted a cut-off of age 40 years, which has previously been used in the *Danish National Diabetes Register*.<sup>74</sup> We also censored metformin purchases in women when the indication code of the prescription suggested that it was for PCOS. This approach may over-censor 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 years. Unless this was extremely rare, these were unlikely to be censored correctly, as the data on the indication codes contributed only very few cases of censoring.

The RSCD took a very different approach to identifying PCOS cases and censoring metformin purchases; it used data on purchases of contraceptives for the treatment of PCOS and any diagnoses of PCOS in the *Danish National Patient Register*. This approach provides an age-agnostic censoring criterion. As contraceptives are the first-line treatment drugs in the guidelines for PCOS, metformin is unlikely to be prescribed for PCOS without these. However, these contraceptives are also used in women without PCOS, which holds a risk of over-censoring women with T2D taking these contraceptives without having 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. Thus, entries into the data may be 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 the OSDC, diabetes diagnosis was defined at the second occurrence of an event across any of the four types of inclusion events. This approach was taken to account for errors in the register data and to comply with the HbA1c-testing recommendations from the World Health Organization and clinical guidelines, which recommend repeated samples before diagnosis.<sup>60,105</sup> 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 in the 10 years preceding a given index date. This approach avoids unnecessary delay, but it accumulates 10 years of falsely included cases on any given index date.

### 5.1.4. Approaches to classification of diabetes type

With no accurate, population-wide marker of diabetes type available, researchers are forced to using indirect markers to classify diabetes type from register data. Thus, algorithm design decisions are based on assumptions, which may qualify as so-called expert clinical knowledge, and these come with benefits and drawbacks. For example, for type-classification in the OSDC, we chose to omit diagnoses from surgical departments, as we assumed that diagnoses from medical departments, particularly endocrinology departments, would be more accurate as these departments provide care specifically for T1D and T2D.

In the OSDC, we deliberately chose to make classification of T1D cases more restrictive compared to T2D. This choice was made because 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 (see Figure 4.2).

As previously mentioned, around 10% of T1D cases in the OSDC had purchased non-insulin GLD at one or more points in time. In the RSCD version 1.0, which was implemented and validated in the first study of this PhD project, these cases would all be classified as T2D and omitted from the T1D population, thereby lowering the sensitivity of T1D classification. Recent revisions of the RSCD addressed this by allowing cases with non-insulin GLD purchases in the first year after onset of diabetes to be classified as T1D.<sup>106</sup>

The concordance was high between the two algorithms in the classification of diabetes of any type vs. no diabetes. 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 demographic characteristics 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 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 spike is not surprising as this was the year after HbA1c data was made 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.



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



The different approaches to censoring between the classifiers affect the inclusions, which is evident in the distribution of the age at inclusion of each sex, as shown in Figure 5.3. For 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 age 40 years old. In the RSCD, the distribution is left-skewed by women included before age 40 years.



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

As discussed previously, we expected the addition of HbA1c data in the OSDC classifier to compensate for the over-censoring of metformin prescription data. In density plots of age at inclusion stratified by calendar year period, the observed spike in T2D among women at age 40 years appeared to be due to cases with onset during calendar 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, a spike of inclusions is seen for women with T2D included at age 40 years. This spike is absent in the subsequent years, but the distribution is slightly skewed with a plateau in the age range 40-45 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 visual sign of cross-contamination, which may indicate that the RSCD is more prone to misclassifying cases during the first years after onset/inclusion.

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



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

## 5.2. Study designs

All three studies in this dissertation were designed as observational, cross-sectional studies based on prospectively collected register data covering 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.

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

The use of nationwide register data enabled the largest possible sample size and provided sufficient statistical power for analyses on minority groups. The drawback was that the data was not collected nor validated for this specific research purpose.

Factors affecting the generalisability of the findings of this dissertation are discussed in the following sections.

## 5.3. Internal validity

### 5.3.1. Selection bias

In study I, validation was performed in a population of respondents to a survey; a design that is prone to selection bias. Survey invitations were conditioned on being aged 18-74 years and on being classified as a diabetes case by the OSDC or being sampled into 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 this did not change the other characteristics in the survey population. As a result, the study population contained fewer women and a skewed age distribution compared to a random sample of the general population and also a higher diabetes prevalence.

The accuracy of the classifiers may be different among individuals older than age 75 years, 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 obtained in the *Health in Central Denmark* survey. This does not rule out

selection bias, however, as the data in both surveys covered only the Central Denmark Region, and the supplementary analyses were limited to classification of diabetes of any type vs. no diabetes (see Appendix A).

In addition to the distribution of survey invitations, non-response among invitees may also be a source of selection bias as the analyses were restricted to respondents. However, several factors suggest that this bias was limited; the response rates were high (>50%) in the T1D, T2D and no-diabetes groups, and the three groups shared similar non-response patterns.<sup>85</sup>

The unique strength of studies II and III was the large study population with data on socio-demographic characteristics, healthcare services, biomarker levels and pharmacological treatment for an entire nation, which minimises the risk of selection bias. However, the statistical analyses were performed as complete-case analyses, which omits individuals with missing data from the analyses and may induce selection bias. To avoid this, individuals with very short residence durations on the index dates of the studies were excluded to allow time for the data to be captured correctly. In addition, the variables included in the analyses were chosen to minimize missing data, e.g. we omitted education level from the analyses due to a substantial proportion of migrants missing this data.

Laboratories from different regions of the country did not contribute data for the same length of time prior to the index dates of the studies. Migrant groups were distributed unevenly across these regions. This could potentially have induced a differential bias on biomarker levels between migrant groups if biomarker levels differed between individuals with a recent sample and those with only an older sample. In study II, this potential bias is likely to be minor, however, since most of the study population underwent yearly biomarker monitoring prior to the index date. Analyses of biomarker levels and pharmacological treatment in study II - and all analyses in study III - were conditioned on individuals having an HbA1c or/and LDL-C level recorded, and this selection could potentially induce a bias. However, even in study III, which excluded individuals without a samples in the three years prior to the index date to minimise information bias, this excluded no more than a few percent of the study population, an indication that substantial selection bias was unlikely.

In study II, the analyses on pharmacological treatment were conditioned on individuals still being alive and residing in Denmark at the end of the following year, which could have induced selection bias. However, the alternative, i.e. leaving these individuals in the analyses, would have the potential to induce substantial bias, as any individual dying or emigrating before their first drug purchase in the year would have had their outcome misclassified if being in pharmacological treatment prior to the index date, which was most likely (the majority of individuals with an indication for treatment did receive treatment in the year after the index date). In study III, the analyses were not

performed in a population with a strict indication for any given treatment. Therefore, we expected outcomes to be more rare, and we did not condition on being alive and residing in Denmark for the entire year after the index date. This may have lead to underestimated user-prevalences due to deaths and emigrations in the first months after the index date, but this is unlikely to be a major source of bias.

#### 5.3.2. Information bias

In study I, the gold standard for validation was self-reported diabetes type, which is an imperfect gold standard that may be prone to recall bias. As age at onset in the true T1D population was much lower compared to age at onset in the T2D population (which was an order of magnitude larger in size) we would expect the mean age at onset of diabetes in the self-reported T1D population to increase substantially if the self-reported diabetes type was inaccurate. However, the distribution of self-reported age at onset of T1D was similar to previously reported estimates.<sup>107</sup> This indicates that the self-reported data on T1D was accurate overall. Although inaccuracies in the self-reported diabetes data cannot be ruled out, especially in groups with low health literacy (e.g. migrants),<sup>108</sup> we expect these to be minor and non-differential between the two register-based classifiers. Thus, any inaccuracies would result in bias towards the mean: an underestimation of performance in both classifiers and an attenuation of differences between them.<sup>109</sup> Indeed, a substantial proportion of self-reported T1D cases among migrants in our study may have been true T2D cases, as the proportion of self-reported T1D was higher in migrants than in native Danes, which contradicts previous findings.<sup>110,111</sup> This does not imply that the classifiers perform worse in migrants, only that the ability to validate the classifiers in these groups is limited in this study design.

In light of the results from study I, the OSDC classification of T2D used to define study populations in studies II and III was found to be highly valid. The quality of the register data is likely to be very high overall, as it is captured automatically and used in civil registration, billing of public health care services, reimbursement of prescription drug purchases, and for clinical use and administrative purposes in hospitals. The biomarker data is captured automatically by the clinical databases and this data is likely to be accurate, as clinical laboratories are compelled to uphold quality assurance standards.<sup>112</sup> Thus, the outcomes defined in studies II and III through billing records from primary care (monitoring), biomarker samples (monitoring and biomarker levels) and prescription drug purchases (pharmacological treatment) were likely to be valid. However, migrants tend to have more healthcare contacts than native Danes<sup>113</sup> and may be more likely to receive testing for HbA1c, LDL-C and UACR that does not represent routine diabetes monitoring. While we expect this effect to be minor, it might have overestimated the

quality of monitoring in migrants compared to native Danes. In addition, while guidelines recommend that screening for diabetic foot to be performed by a podiatrist, GPs may also perform it. This would not be captured in the data, which may have overestimated the risk of non-fulfilment of this aspect of monitoring in groups where screening for diabetic foot is more often performed by the GP. Migrants, being poorer than native Danes, may be more likely to have screening for diabetic foot performed at the GP if they are deterred from using podiatrist services due to the associated financial barriers. This would lead to exaggerated risks in migrants if screening for diabetic foot by GPs is considered equal to podiatrist screening in terms of care quality.

In study III, the lack of data on active ordinations of GLD meant that an indirect definition based on prescription purchases had to be used. The greatest challenge in this regard is managing patients switching from one type of GLD to another, which - in a fixed window of prescription purchase data - is hard to discern from patients adding an additional type of GLD to their treatment. If a drug switch is misclassified as the use of two GLD types rather than one, user-prevalences will be overestimated. In contrast, patients purchasing large amounts of GLD at once may not need to make another purchase during the data window despite being users, which will lead to underestimation of user-prevalences. This is a particular problem since metformin, the most commonly used GLD type in Denmark, is available for purchase in package of 400 tablets, which will last the user for several months - even at maximum daily dosage and complete adherence to medication.

Our definition of GLD use - requiring two purchases in a year, and at least one in the last six months of the year - attempts to limit the potential of both types of errors. By requiring two purchases of each GLD, we expect to have accounted for most cases of switching/discontinuation due to side-effects, which is likely to occur during use of the first purchase. In addition, by requiring a purchase to be made within a six month window, the potential for GLD switching is minimised, due to the low likelihood of switching from one GLD to another (and making two purchases of the new drug) in such a short window of time.<sup>114,115</sup> Furthermore, this user-definition is likely to correctly identify users of metformin, even users of sub-maximal dosages purchasing in bulk. Naturally, we may have omitted individuals purchasing very large amounts of a GLD type at once, and may incorrectly have identified patients discontinuing/switching from one type of GLD to another as being users of both types if the switch happened during the last six months of the year. However, we do not expect this to have influenced our findings substantially, neither in any of the individual GLD types nor in combination therapy.

Alternative definitions of use could have pursued a more restrictive identification of drug discontinuation/switching by using distinct rules for metformin-use, or by requiring repeated purchases of each GLD during subsequent time-windows. The drawback of the latter is the inability to allow for GLD types to be added during subsequent time



windows, while the former is unlikely to provide enough improvement to justify the added complexity (unless using a definition with a shorter time-frame than ours, e.g. requiring purchases within a three-month window).

The exposure definition of migrant origin is based on systematic registrations by the immigration office, which are likely to be highly valid. While this classification based on country of origin is able to identify distinct ethnic minorities better than definitions based on race,<sup>10</sup> it does not fully capture ethnic groups within each country (e.g. ethnic Turks vs. ethnic Kurds in the Turkey group) or pooled category (e.g. Moroccan vs. Afghan migrants in the Middle East group). This may lead to underestimation of disparities due to *bias towards the mean* if an origin group contains multiple ethnic groups with opposite risk patterns.

Among variables used for adjustment, some are particularly vulnerable to inaccuracies. For example, data on hospital diagnoses may be vulnerable to inaccuracies and errors, as these data potentially rely on registrations entered by a single doctor (possibly sleep-deprived and pressured for time), who may not prioritise accuracy of the registration over more pressing clinical concerns. In studies II and III, hospital diagnoses of macrovascular complications were used as adjustment variables, and these primary diagnoses had previously been validated for peripheral arterial disease, stroke and cardiovascular disease with PPVs between 75% and 99%, which was sufficient for their purpose in our studies.<sup>116–118</sup>

In contrast, to the best of my knowledge, register-defined duration of diabetes has never been the subject of validation in any previous studies. Supplementary analyses in study I indicated that register-classified diabetes duration may be biased among individuals with T2D. However, this bias seems unlikely to be differential between migrant groups and the outcomes investigated in studies II and III, although it might have led to an underestimation of disparities due to *bias towards the mean*.

#### 5.3.3. Confounding

None of the studies in this dissertation attempted to establish causation. Nevertheless, potential confounding of the observed risk disparities between migrants and native Danes deserves to be discussed since migrant origin, socioeconomic position, lifestyle and healthcare utilisation are intertwined beyond the information captured in the data. Studies II and III mapped the risks in migrants (not causal associations). However, in order to visualise and interpret disparities in a clinical and social context, the analyses were adjusted for some potential clinical and socioeconomic risk factors. As the focus was placed on risk rather than causality, we did not adjust for possible differences in GP behaviour, which is obviously an important factor in care provision. Additionally,

the factors adjusted for may not fully capture the clinical status of a patient, and the distributions of hypertension, smoking and diabetic complications beyond macrovascular complications and kidney disease are likely to differ between migrant groups and the outcomes studied. However, any residual confounding of the associations is assessed to be minor, as we adjusted for several key clinical risk factors.

### 5.4. External validity

#### 5.4.1. Generalisability

In study I, both classifiers achieved excellent overall accuracy, despite their different approaches to classifying diabetes type, handling potentially erroneous data and censoring PCOS and GDM cases. This might indicate that the quality of the underlying data is more important for accurate classification of diabetes type than the specific technical design. If this is the case, the overall performance estimates are likely to be generalisable to other diabetes definitions based on Danish register data from similar time periods and with comparable inclusion and censoring criteria. Register-based studies set in healthcare systems of other countries may use the algorithms as a starting point. However, a careful analysis of whether a given register data infrastructure captures valid information on type-specific diagnoses, inclusion and censoring criteria is required before similar classification accuracy can be assumed.

Migrant populations differ between nations, and T2D care quality differs between nations with different healthcare systems. In addition, diabetes risks in migrants appear to be sensitive not only to the country of origin and the country of residence, but also to the socioeconomic gradient between the two.<sup>34</sup> Thus, generalising the findings on care quality to migrants in other countries depends on several context-specific factors. Despite these limitations, it seems likely that the disparities found in migrant groups in Denmark may also be present in migrants from the same origin groups in other European countries with universal public healthcare systems. In study II, reference risks in most indicators of T2D care were similar to reports from other Scandinavian countries and the UK. This suggests that these settings are comparable in terms of diabetes care, but the levels of HbA1c and LDL-C were lower in the Danish T2D population which may limit the generalisability of our results within these outcomes.<sup>32,119</sup>

The vast majority of migrant groups were covered in the origin categories studied in this dissertation. The remaining migrants are a very heterogeneous group in terms of origin, but the overall risk patterns may generalise to these as well. The results on diabetes care from study II may not generalise to migrants with a short duration of diabetes or residence in Denmark, who were excluded from the analyses. Still, the ramifications of

this are limited, since it only concerns relatively few individuals, who transition out of these states after a limited period of time.

### 5.5. Analyses

In study I, the great strength of the survey data was that it provided information on individuals regardless of diabetes status, which enabled analyses of all four standard validation metrics (sensitivity, specificity, PPV and NPV). This contrasts clinical audit of electronic patient records, a commonly used validation study design, which provides only an estimate of PPV. Subsampling and excluding most OSDC-classified diabetes cases from the study population may seem counter-intuitive, as it reduces the statistical power, but this was necessary to provide unbiased validation estimates. Without subsampling, the study population would have been far from representative of the target population in terms of diabetes prevalence, and it would also induce circular bias towards the OSDC classifier; among individuals with diabetes, true positive OSDC diabetes cases would be much more likely to be included in the study population than false-negative OSDC diabetes cases. Another undesirable effect of subsampling is that it may induce a random bias, but this was examined by drawing multiple samples and bootstrapping repeated analyses. The bootstrapped analyses showed that the study population sample was robust and that sampling had not induced any major random bias.

In studies II and III, the use of robust Poisson regression provided intuitive estimates of relative risk, despite most outcomes being common. While relative risk is probably the most relevant risk measure for the individual patient and the clinician providing treatment, public health planners may be more interested in disparities measured as absolute risk differences between groups.

In studies II and III, the choice of statistical models and variables for adjustment merits some discussion. Both studies presented risk estimates from three different models with different contexts: crude, adjusted for clinical variables (model 1) and adjusted for clinical and socioeconomic variables (model 2). In study II, the outcomes were closely aligned with guideline recommendations, most of which apply to all T2D patients regardless of patient characteristics. Therefore, the emphasis in study II was placed on the crude risk estimates. In study III, the outcomes were not tied to absolute recommendations, and indication for treatment depended more on the risk profile and characteristics of each individual patient. For this reason, the models were expanded to capture more clinical characteristics in study III, where the emphasis was placed on the fully adjusted estimates (model 2). Due to the very different distributions of some characteristics between migrants and native Danes (e.g. most migrant groups with T2D were much younger and had much lower household incomes), the adjusted estimates should be interpreted with

caution, as they compare the risk in migrants with the risk in the most vulnerable minority of native Danes with T2D. Indeed, researchers focussing on causal inference could argue that being young and poor is a causal effect of being a migrant with T2D. Furthermore, if receiving poor T2D care quality is considered a causal effect of being young and poor, then these variables would (in a causal framework) be considered intermediary variables on the causal pathway (and adjusting for them would lead to underestimation of the association). Ultimately, the three models contribute to understanding and interpreting the risk in migrants from different perspectives, and the most relevant model will depend on the aims of the reader.

## 6. Discussion of results

### 6.1. Key main findings

This dissertation identified several important findings in the three studies. First, study I determined whether register-based classifiers can identify valid populations of T1D and T2D in a general Danish population. Second, disparities in T2D care between migrants and native Danes were investigated in two cross-sectional studies. Study II analysed T2D prevalence and disparity in care quality as outlined in guideline recommendations for monitoring, biomarker levels and pharmacological treatment, while study III explored disparities in GLD treatment in terms of combination therapy and types of GLD used.

In this chapter, the main findings of each study are discussed separately and compared to existing literature before clinical implications of the results of studies II and III are discussed. Risk factors in T2D care beyond migrant origin are of great importance in a broader scope of equity in T2D care provision. However, they are not the focus of this dissertation and are only briefly discussed as they relate to migrants.

### 6.2. Study I

This study validated two predefined algorithms (the OSDC and the RSCD), which classified diabetes type in Danish register data against self-reported diabetes type in a Danish survey population. Overall, both classifiers performed excellent in terms of PPV in both T1D and T2D as well as sensitivity in T2D classification (all estimates 0.875 and above), and they had near-perfect accuracy in terms of specificity and NPV in both T1D and T2D (all estimates 0.989 and above). Both classifiers were unable to accurately classify diabetes type in individuals with T1D onset after age 40 years and T2D onset before age 40 years.

Sensitivity and NPV were higher in the OSDC than in the RSCD for both diabetes types. In T1D classification, these differences are attributable to differences in the algorithms, as both algorithms relied on the same data sources for identifying T1D (GLD use and type-specific diabetes diagnoses).<sup>76</sup> In T2D, this difference may be explained by the use

of HbA1c data in the OSDC, which enables inclusion of diabetes cases at the time of diagnosis rather than requiring subsequent initiation of GLD treatment or hospitalisation. Specificity and PPV in T2D classification were higher in the RSCD than in the OSDC. This could possibly be explained by inclusion of milder cases in the latter, as this population may be less likely to correctly report having diabetes (e.g. if an individual had never purchased GLD or been hospitalised for the disease). Notably, the demographic characteristics of the register-classified diabetes populations differed, particularly in T1D, where the higher sensitivity of the OSDC in women (0.846 vs. 0.725 in supplementary analyses) resulted in a higher prevalence of women in the OSDC population compared to the RSCD population. This is in line with a previous study comparing the Danish National Diabetes Register (which did not use HbA1c) against a local database containing HbA1c as an alternative inclusion criterion, which found higher prevalences of diabetes in the HbA1c-augmented definition, especially among women.<sup>64</sup>

To the best of my knowledge, this study was the first to validate the performance of a type-specific diabetes classifier in a general population of individuals with and without diabetes. Other studies have validated type-specific diabetes classifier performance in populations consisting only of individuals with diabetes, which fails to test the ability of the algorithm to distinguish individuals with diabetes from those without. This overestimates the accuracy of the algorithm compared to its performance in a general population. In addition, both algorithms were pre-specified, as opposed to the data-driven, exploratory approaches used in other validation studies to design algorithms optimized for a particular dataset, which holds a risk of overfitting the algorithm to the dataset and reducing the external validity of findings.<sup>69,72,120</sup> Despite these differences, the performance of both our classifiers was superior to the performance of T1D classification reported in studies in the United States and Hong Kong,<sup>69,73</sup> and comparable to that reported in the United Kingdom,<sup>121</sup> but inferior to that reported in a study in Canada.<sup>72</sup>

The poor performance of the register-based classifiers in subgroups with diabetes onset at atypical age is likely to reflect clinical uncertainty in these cases. This may result in inaccurate type-specific diabetes diagnoses and uncharacteristic GLD purchase patterns,<sup>122</sup> which may lead the algorithms to misclassify these individuals. Sensitivity in T1D classification has previously been reported to be highly dependent on age at onset,<sup>120</sup> and our findings indicate that this issue also extends to PPV and to T2D classification.

### 6.3. Study II

This study analysed the prevalence of T2D and disparities between migrant groups and native Danes across eleven indicators of guideline recommendations on monitoring,

biomarker levels and pharmacological treatment.

The prevalence of T2D was found to be higher in migrants compared to native Danes, especially in migrants from Sri Lanka and Pakistan, who were four to five times more likely to have T2D compared to native Danes. While similar findings have previously been reported in Denmark<sup>7,28</sup> and across Europe,<sup>6</sup> our study shows that this risk persists after adjusting for differences in socioeconomic position.

Failure to fulfil guideline recommendations and targets for T2D care was common. The proportion of native Danes unable to fulfil recommendations was below 10% in only two indicators (HbA1c-monitoring and GLD-treatment) and above 25% in eight (all remaining indicators excluding LDL-C monitoring). Non-fulfilment of guideline recommendations was more common in migrants, particularly in terms of monitoring and controlling HbA1c and LDL-C levels. In most of these indicators, we observed increased crude risks at a significance level of  $\alpha = 0.05$  in most migrant groups, whereas no clear overall pattern was observed across migrant groups in the indicators of pharmacological treatment.

The disparity between migrants and native Danes was largest in the indicators of glycaemic control and screening for diabetic foot disease (crude risk was increased by more than one-third in most migrant groups). Migrants from Somalia stood out across all aspects of T2D care. The group had increased crude risk in all eleven indicators, and the highest risk of all migrants in nine. In addition, the largest single disparity was observed in migrants from Somalia, who had almost twice the crude risk of native Danes in the indicator of lipid control.

Our findings of lower likelihood of T2D monitoring in most migrant groups contrast the findings in prior studies, where migrants received similar or more monitoring compared to native Danes,<sup>28</sup> Norwegians<sup>29</sup> and Swedes.<sup>31</sup> Still, it is in line with recent findings of less timely monitoring in ethnic minority groups in England.<sup>32</sup> The proportion of the T2D population without timely monitoring varied substantially between each type of monitoring, but socioeconomic factors such as household income and region of residence were risk factors in all types of monitoring. Screening for diabetic foot disease is the only service that is not fully covered by the public health insurance in Denmark (requiring the patient to pay a fee), which may explain why this indicator had the highest proportion of individuals without timely monitoring. In addition, the large disparity between migrants and native Danes in this indicator, which persisted after adjusting for socioeconomic variables, suggests that economic barriers may disproportionately limit the access to care in migrant groups.

In line with previous studies, most migrant groups were less likely than native Danes to achieve glycaemic control.<sup>31</sup> The magnitude of the risk was similar to what has previously been reported, although it may seem lower compared to estimates reported as ORs in

other studies, since ORs may lead to exaggerated risk estimates if interpreted as RR when the outcome is common.<sup>123</sup> For example, a study in Scotland reported a crude OR of dysglycaemia of 2.2 in the Pakistani group and a dysglycaemia-prevalence of roughly 50% in the reference group;<sup>30</sup> this corresponds to an RR of 1.4,<sup>97</sup> which is similar to our estimate.

To the best of my knowledge, increased LDL-C levels among migrants have not previously been reported,<sup>33</sup> although a small US study noted a minor trend towards higher lipid levels in migrants from Somalia.<sup>124</sup> Three studies have examined obesity in migrants from Somalia and found high prevalences of obesity and increased waist-to-hip ratio in women, but not in men.<sup>125–127</sup> If this pattern between sexes extends to migrants from Somalia in Denmark, it seems unlikely that anthropometric factors alone can explain the high levels of LDL-C in this population, as the risk remained unchanged in supplementary analyses stratified by sex (not included in this dissertation).

Most studies of pharmacological treatment in T2D have analysed the whole population (rather than analysing groups with a clinical indication for treatment) and have evaluated incident medication use in cohorts of newly diagnosed, making direct comparisons to our findings difficult. The finding of lower or similar risk of GLD non-use in most migrant groups is in line with previous studies comparing migrants to native populations.<sup>29,31</sup> However, the lower risk of LLD non-use in most migrant groups contrasts findings from prior studies.<sup>29,33,48</sup> The increased risk of dysglycaemia in the Somalia group coincided with an increased risk of not receiving GLD. This pattern was also observed for dyslipidaemia and LLD non-use in the Somalia and Europe groups, indicating a particular need to the increase uptake of pharmacological treatment in these groups. Similar to a study from Italy, we found a higher risk of not receiving ACEI/ARB treatment in most migrant groups. In contrast, the risk of not receiving APT was similar in most groups in our study.<sup>49</sup> As the indications for treatment with ACEI/ARB and APT were identical in our study, one might expect similar risk patterns in these outcomes. Yet, this was not the case, which might be due to unmeasured differences between the groups in terms of indications, counter-indications, awareness of and attitude to these drugs.

Some of the disparities between migrants and native Danes could be attributed to a higher prevalence of risk factors of poor T2D care in migrants; specifically young age, low household income and residence in the Capital Region of Denmark.

### 6.4. Study III

This study explored disparities in GLD treatment in terms of combination therapy and types of GLD used, adding to the analysis of the indicator of GLD treatment performed in study II. The analysis of GLD in study II found no overall pattern of disparity across



migrant groups, but it did not capture nuances in the treatment provided, because the analysis was a simple evaluation of any GLD treatment being present or not. In this study, most migrant groups were found to be less likely to receive combination GLD therapy than native Danes. We also found that, compared to native Danes, most migrant groups were more likely to receive oral GLD types (metformin, sulfonylureas, DPP4i, and SGLT2i) and less likely to receive injection-based GLD types (insulins and GLP1RA). Disparity was largest in GLP1RA use, where the RR ranged from 0.30 to 0.60 in most migrant groups.

As in study II, migrants from Somalia stood out; they had the lowest likelihood of receiving combination therapy and GLP1RA of all groups.

To the best of my knowledge, combination therapy usage has not previously been studied in migrants. Still, disparities between racial groups have been examined, and these proxies of ethnicity provide a basis for comparisons. With the notable exception of migrants in the four smallest migrant groups (Pakistan, Sri Lanka, Somalia and Vietnam), the migrants in our study received combination therapy on a level close to native Danes after adjustments (RRs around 0.95 to 1.05). Since combination therapy is not a hard clinical endpoint, we argue that the modest disparities observed in the largest groups of migrants (covering more than three-quarters of the total migrant T2D population) were insignificant from a clinical perspective. At a glance, this may contrast another UK study reporting a high odds ratio of experiencing inertia in initiation of combination therapy in non-white groups compared to Whites (e.g. odds ratio in Blacks: 1.43) and reporting their findings as persuasive evidence of disparities.<sup>41</sup> However, the high prevalence of inertia in the reference group (67.1%) of the study exaggerates this risk estimate if interpreted as RR, and the corresponding RR (around 1.11 in Blacks)<sup>97</sup> is more in line with our findings.

The higher prevalence of metformin, sulfonylurea and DPP4i in migrant groups is in line with previous reports in non-white minorities.<sup>32,45</sup> The particularly high prevalence of sulfonylureas in migrants from Vietnam may reflect care being influenced by practice in the origin country, as sulfonylureas are used in roughly three quarters of all T2D patients in Vietnam.<sup>128</sup> In contrast, the higher prevalence of SGLT2 found in our study contradicts previous studies on racial minorities.<sup>45,129,130</sup> The very low prevalence of GLP1RA use in migrants compared to the native population has not previously been reported, but is in line with previous reports of disparities in non-white minorities.<sup>32,45,129</sup> Similarly, the low prevalence of insulin treatment in migrants with T2D is consistent with an Italian study<sup>42</sup> and studies on non-white minorities elsewhere.<sup>32,131</sup>

The main analyses were limited to GLD use in 2019, as we only had access to data on co-variables in the time before 2019. GLD treatment practice changed considerably in the subsequent years (even before the broadening of GLP1RA/SGLT2i-recommendations in

2022), and our findings may not persist in the new clinical context. However, supplementary analyses of GLD use in 2021 found that, despite substantial changes to GLD use compared to 2019, including >50% increases in GLP1RA and SGLT2i user-prevalences, disparities between migrants and native Danes remained unchanged. This may indicate that our findings are robust rather than limited to a clinical setting of narrow indication or low prevalence of use of these drugs.

### 6.5. Clinical implications

In study II, the evaluation of quality of care using binary process and outcome indicators does not capture qualitative differences in the provided care, nor does it account for valid patient-specific reasons for not providing care or achieving biomarker control. In some individuals, failure to achieve biomarker control may be unrelated to the quality of provided care (e.g. due to genetic factors), and a lack of monitoring may not affect risk factors and the risk of adverse outcomes. In contrast, there is less leeway for the timely provision of pharmacological treatment to patients with a guideline indication for treatment. Drug side-effects or contra-indications may be valid clinical reasons for a patient to not receive a particular drug. However, this is likely to only explain a very small fraction of non-users, as it is unlikely that these reasons prohibit all treatment options offered by guidelines (e.g. all types of GLD or LLD). Thus, relative to monitoring and biomarker levels, the clinical implications of disparities in pharmacological treatment are clearer, and they require further discussion below.

Barriers to pharmacological treatment may originate from the healthcare system, the clinician, and the patient. Therefore, efforts to increase the use should target these.

Healthcare policies affecting patient-incurred costs are major determinants of pharmacological treatment accessibility.<sup>132,133</sup> In the Danish universal healthcare system, the annual drug expenditures are partially covered from €132 and fully covered from €553 (2019 limits). While patients with higher medication expenses will incur similar costs, regardless of the number or types of GLD used, patients treated with fewer, less expensive drugs will receive only limited coverage, and they may perceive costs as a barrier to using the newer, more expensive GLD types. In study III there were indications of financial barriers to using the more expensive GLD regimens, as lower household income was associated with less use of triple combination therapy, SGLT2i and GLP1RA (see supplementary material). This could indicate that broader coverage of GLD expenditures may serve to increase the use of combination therapy, SGLT2i and GLP1RA in the poorest patient groups, in which migrants constitute a disproportionately large share.

Clinicians may be reluctant to prescribe insulin in migrants if a language barrier is present, given the complex nature of administrations and the risk of hypoglycaemia.

However, the use of GLP1RA does not carry these risks and should not take up more resources of the prescriber (e.g. physician or staff consultation time) over oral GLD, even if language barriers are present. Increasing GP awareness of disparities in pharmacological treatment of migrants with T2D (e.g. by increasing focus on migrants in clinical guidelines) may reduce the risk of suboptimal pharmacological treatment in these groups.

Patient preference of oral administration over injection-based drugs is also likely to influence prescribing. Migrant origin may influence this preference, which could explain the divergent findings between oral and injection drugs, particularly between GLP1RA and SGLT2i. Other studies have reported several reasons for insulin-aversion among migrants, including fears of weight gain, loss of independence, risk of hypoglycemia and death.<sup>134–136</sup> Similar fears may be present for GLP1RA use in migrants due to the administration similarities to insulin injections, and patient education may be able to address these.

The massive disparities faced by migrants from Somalia in all areas of care studied in this PhD project hold particular implications for clinicians (and perhaps also for health planners and local communities). Overall, they indicate that clinicians treating patients in this group should consider close attention to care quality in this group, particularly in relation to pharmacological treatment with GLDs and LLDs. Specifically, LLDs are available in relatively inexpensive oral form, and expenditures and patient preference seem less likely to be major barriers to use of these drugs. Therefore, increasing awareness among GPs to the high lipid levels and need for LLD treatment in patients from Somalia with T2D may be a potential approach to increase treatment quality and reduce complication risk in this group.

## 7. Conclusions

### 7.1. Main conclusions

Five main conclusions can be derived from this dissertation.

First, both of the register-based diabetes classifiers identified valid populations of T1D and T2D in a general Danish population, with substantially higher sensitivity in the OSDC compared to the RSCD. Neither of the two algorithms was able to accurately classify diabetes type in individuals with T1D onset after age 40 years, nor T2D onset before age 40 years, and results from register-based studies of these groups should be interpreted with caution.

Second, compared to native Danes, the prevalence of T2D was higher in most migrant groups, the quality of monitoring was inferior, and most migrant groups had higher risks of not achieving glycaemic and lipid control, despite similar prevalence of pharmacological treatment.

Third, most migrant groups were less likely than native Danes to use combination GLD therapy, and migrants were more likely to use oral GLDs and less likely to use injection-based GLDs, particularly GLP1RA.

Fourth, T2D care was poorest in migrants from Somalia, who had higher risk than native Danes in all eleven guideline indicators assessed, and they also had the lowest likelihood of using combination therapy and GLP1RA of all groups. Their almost double risk of dyslipidaemia combined with a higher risk of LLD non-use compared to native Danes appeared to be an area of care with a particular potential for improvement to reduce complication risk.

Finally, clinicians treating migrants with T2D should be aware of the the most substantial disparities in T2D care within each migrant group. These are summarised in the table below, which highlights the areas of T2D care, where disparities relative to native Danes are greater than 10%.

Table 7.1.: Highlights of disparities within each migrant group

Migrant group	Monitoring	Biomarker control	Pharmacological treatment
Middle East	Nephropathy Retinopathy <b>Podiatrist</b>	HbA1c	GLD <b>ACEI/ARB</b> <b>GLP1RA</b>
Europe	<b>HbA1c</b> LDL-C Nephropathy Retinopathy Podiatrist	LDL-C	<b>GLD</b>
Turkey	<b>Podiatrist</b>	HbA1c	<b>ACEI/ARB</b> <b>GLP1RA</b>
F. Yugoslavia	Retinopathy <b>Podiatrist</b>	HbA1c	<b>GLP1RA</b>
Pakistan	Nephropathy <b>Retinopathy</b> <b>Podiatrist</b>	HbA1c	<b>GLD</b> <b>ACEI/ARB</b> Combination therapy <b>GLP1RA</b>
Sri Lanka	None >1.10	None >1.10	<b>ACEI/ARB</b> <b>Combination therapy</b> <b>GLP1RA</b>
Somalia	<b>Nephropathy</b> Retinopathy <b>Podiatrist</b>	HbA1c <b>LDL-C</b>	<b>GLD</b> LLD <b>ACEI/ARB</b> <b>APT</b> <b>Combination therapy</b> <b>GLP1RA</b>
Vietnam	<b>HbA1c</b> <b>Podiatrist</b>	None >1.10	<b>Combination therapy</b> <b>GLP1RA</b>

Notes

- Based on estimates of relative risk in studies II and III adjusted for differences in clinical risk factors (model 1).
- Disparities less than 10% compared to native Danes are not shown.
- Disparities greater than 20% are highlighted in **bold**.
- Abbreviations: HbA1c: haemoglobin-A1C. LDL-C: Low-density lipoprotein cholesterol. ACEI/ARB: Angiotensin-converting enzyme-inhibitors or angiotensin receptor blockers. GLP1RA: Glucagon-like peptide-1 receptor agonists.

## 8. Perspectives and future research

### 8.1. Perspectives

Based on the findings of this dissertation, diabetes epidemiologists can rely on OSDC and RSCD to provide valid study populations of T1D and T2D for Danish register-based research, and the validated, open-source classifiers provide robust and transparent tools for researchers using these data. In addition, the findings indicate that other register-based diabetes definitions based on Danish register data from similar time-periods with comparable algorithm designs are likely to provide valid T1D and T2D populations as well. Some caveats remain, however, as register-classified diabetes type in cases with atypical age at onset of diabetes should be interpreted with caution, and the onset of diabetes may not be accurately captured by the classifiers.

Migrants are an increasingly important part of the T2D population due to their high T2D prevalence, which will continue to rise in the future as their demography ages. Accordingly, addressing the particular care needs of migrants with T2D is going to become an increasingly pressing matter to clinicians and healthcare planners, if all individuals with T2D are to receive appropriate care. This dissertation answered an easy part of this challenge, the *what*: what are the disparities in care quality faced by migrants? The hard part is to answer the *why*: why is it harder to deliver adequate care in migrants? - and ultimately, the *how*: how can healthcare resources be prioritized to provide adequate T2D care to those with the greatest needs?

Migrant origin is a complex entity, T2D is a complex disease with several care modalities, and in combination they are a formidable challenge to clinicians and researchers. Without more supporting knowledge on the causes of care disparities or how to effectively address them in migrants with T2D, our findings alone are insufficient to guide healthcare planners or sway clinical practice.

### 8.2. Future research

In register-based classification of diabetes, 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. In addition, as these GLD types are increasingly used in heart and kidney patients without diabetes, and inclusion criteria in register-based algorithms will have to adapt accordingly. In light of these changes, further validation studies are needed in order to ensure robust classifications in the future.

Disparities in diabetes care in migrants should be studied further, particularly the causes of disparities, how they affect complication risk, and how interventions can be designed to effectively reduce complication risk in migrant patients. Due to the differences in health-care delivery and migrant group composition between countries, studies from several countries are needed to provide knowledge on as many migrant groups and healthcare system contexts as possible. Disparities in monitoring, biomarker levels, and pharmacological treatment of T2D are likely the result of underlying cultural and lifestyle factors in migrants that are poorly understood. A better understanding of these factors is needed to design interventions that can effectively address migrants and their challenges in a culturally appropriate and patient-centered way. Thus, future research should not be limited to traditional, clinical aspects of T2D care surrounding these disparities, but some questions are particularly relevant to this dissertation. Beyond combination therapy and drug type, several factors contribute to optimal pharmacological treatment, but have yet to be examined. Factors such as daily dosage of each GLD and adherence to medication may hold potential for improving risk profiles in patients with T2D, and sub-optimal care in these factors could also be more common in migrants than native Danes, if the observed disparities in other aspects of care are a general indication. Furthermore, while there is no evidence to the contrary, it is not clear if HbA1c, LDL-C, and GLP1RA have the same impact on complication risk in migrants as in the (predominantly) white, native European populations in which their effects have mainly been studied. In order to prioritize future interventions and best inform intervention design, future studies are needed to assess the risk-modifying effects of these disparities, as well as their causes. Finally, the poor level of care in migrants from Somalia in Denmark is striking, and merits further studies in other countries with large Somali minorities to assess if these findings are applicable in a broader, international context.

## 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](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](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](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](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#Bill>
19. Danmarks Statistik [English: Statistics Denmark]. Indvandrerne 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. Agyemang C, Linden EL van der, Bennet L. Type 2 diabetes burden among migrants in europe: Unravelling the causal pathways. *Diabetologia* [Internet] 2021;64(12):2665–75. Available from: <https://doi.org/10.1007/s00125-021-05586-1>
36. 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>
37. 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>

38. Stephens JW, Brown KE, Min T. Chronic kidney disease in type 2 diabetes: Implications for managing glycaemic control, cardiovascular and renal risk. *Diabetes, Obesity and Metabolism* [Internet] 2020;22(S1):32–45. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.13942>
39. Kim K, Unni S, Brixner DI, et al. Longitudinal changes in glycated haemoglobin following treatment intensification after inadequate response to two oral antidiabetic agents in patients with type 2 diabetes. *Diabetes, Obesity and Metabolism* [Internet] 2019;21(7):1725–33. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.13694>
40. Bonnet F, Chen H, Cooper A, et al. What are the factors associated with long-term glycaemic control in patients with type 2 diabetes and elevated glycated haemoglobin (7.0). *Diabetes, Obesity and Metabolism* [Internet] 2021;23(10):2336–43. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.14476>
41. Mathur R, Farmer RE, Eastwood SV, Chaturvedi N, Douglas I, Smeeth L. Ethnic disparities in initiation and intensification of diabetes treatment in adults with type 2 diabetes in the UK, 1990-2017: A cohort study. *PLoS Med* 2020;17(5):e1003106.
42. Marchesini G, Bernardi D, Miccoli R, et al. Under-treatment of migrants with diabetes in a universalistic health care system: The ARNO Observatory. *Nutr Metab Cardiovasc Dis* 2014;24.
43. Vasti EC, Basina M, Calma J, Maron DJ, Rodriguez F, Sandhu AT. Disparities in adoption of new diabetic therapies with cardiovascular benefits. *Diabetes Research and Clinical Practice* [Internet] 2022;110233. Available from: <https://www.sciencedirect.com/science/article/pii/S0168822722010476>
44. Morton JI, Ilomaki J, Magliano DJ, Shaw JE. The association of socioeconomic disadvantage and remoteness with receipt of type 2 diabetes medications in Australia: A nationwide registry study. *Diabetologia* [Internet] 2021;64(2):349–60. Available from: <https://doi.org/10.1007/s00125-020-05304-3>
45. Elhussein A, Anderson A, Bancks MP, et al. Racial/ethnic and socioeconomic disparities in the use of newer diabetes medications in the look AHEAD study. *Lancet Reg Health Am* 2022;6.
46. Falkentoft AC, Andersen J, Malik ME, et al. Impact of socioeconomic position on initiation of SGLT-2 inhibitors or GLP-1 receptor agonists in patients with type 2 diabetes – a danish nationwide observational study. *The Lancet Regional Health - Europe* [Internet] 2022;14:100308. Available from: <https://www.sciencedirect.com/science/article/pii/S2666776222000011>

47. Nicolucci A, Charbonnel B, Gomes MB, et al. Treatment patterns and associated factors in 14 668 people with type 2 diabetes initiating a second-line therapy: Results from the global DISCOVER study programme. *Diabetes, Obesity and Metabolism* [Internet] 2019;21(11):2474–85. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.13830>
48. Ramirez-Sanchez DC, Krasnik A, Kildemoes HW. Do immigrants from Turkey, Pakistan and Ex-Yugoslavia with newly diagnosed type 2 diabetes initiate recommended statin therapy to the same extent as Danish-born residents? A nationwide register study. *Eur J Clin Pharmacol* 2013;69:87–95.
49. Marzona I, Avanzini F, Tettamanti M, et al. Prevalence and management of diabetes in immigrants resident in the Lombardy Region: the importance of ethnicity and duration of stay. *Acta Diabetol* 2018;55:355–62.
50. Nicolucci A, Greenfield S, Mattke S. Selecting indicators for the quality of diabetes care at the health systems level in OECD countries. *International Journal for Quality in Health Care* [Internet] 2006;18(suppl\_1):26–30. Available from: <https://doi.org/10.1093/intqhc/mzl023>
51. Jørgensen ME, Kristensen JK, Reventlov Husted G, Cerqueira C, Rossing P. The danish adult diabetes registry. *Clinical epidemiology* [Internet] 2016;8:429–34. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/27843339>  
<https://www.ncbi.nlm.nih.gov/pmc/PMC5098513/>
52. 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/>
53. 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/>
54. Nationale VersorgungsLeitlinie [English: National Care Guidelines]. NVL Typ-2-Diabetes [Internet]. 2021; Available from: <https://www.leitlinien.de/themen/diabetes/2-auflage/>
55. Nederlands Huisartsen 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>

56. 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>
57. 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>
58. 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](https://vejledninger.dsam.dk/media/files/20/type_2_diabetes_2004_rettet.pdf)
59. 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>
60. 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/>
61. 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>
62. Marszalek RT, Flintoft L. Being open: Our policy on source code. *Genome Biology* [Internet] 2016;17(1):172. Available from: <https://doi.org/10.1186/s13059-016-1040-y>
63. Dima AL, Dediu D. Computation of adherence to medication and visualization of medication histories in r with AdhereR: Towards transparent and reproducible use of electronic healthcare data. *PLoS One* 2017;12(4):e0174426.
64. Nielsen AA, Christensen H, Lund ED, Christensen C, Brandslund I, Green A. Diabetes mortality differs between registers due to various disease definitions. *Dan Med J* 2014;61(5):A4840.
65. Rawshani A, Landin-Olsson M, Svensson AM, et al. The incidence of diabetes among 0-34 year olds in sweden: New data and better methods. *Diabetologia* 2014;57(7):1375–81.

66. Bak JCG, Serné EH, Kramer MHH, Nieuwdorp M, Verheugt CL. National diabetes registries: Do they make a difference? *Acta Diabetologica* [Internet] 2021;58(3):267–78. Available from: <https://doi.org/10.1007/s00592-020-01576-8>
67. Hallgren Elfgren I-M, Grodzinsky E, Törnvall E. The swedish national diabetes register in clinical practice and evaluation in primary health care. *Primary Health Care Research & Development* [Internet] 2016;17(6):549–58. Available from: <https://www.cambridge.org/core/article/swedish-national-diabetes-register-in-clinical-practice-and-evaluation-in-primary-health-care/F3A9B74FED1E7B36147FF63382939B9A>
68. Cooper JG, Thue G, Claudi T, Løvaas K, Carlsen S, Sandberg S. The norwegian diabetes register for adults – an overview of the first years. *Norsk Epidemiologi* [Internet] 2013;23(1). Available from: <https://www.ntnu.no/ojs/index.php/norepid/article/view/1599>
69. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. *Diabetes Care* 2013;36(4):914–21.
70. Schroeder EB, Donahoo WT, Goodrich GK, Raebel MA. Validation of an algorithm for identifying type 1 diabetes in adults based on electronic health record data. *Pharmacoepidemiol Drug Saf* 2018;27(10):1053–9.
71. Lo-Ciganic W, Zgibor JC, Ruppert K, Arena VC, Stone RA. Identifying type 1 and type 2 diabetic cases using administrative data: A tree-structured model. *J Diabetes Sci Technol* 2011;5(3):486–93.
72. Weisman A, Tu K, Young J, et al. Validation of a type 1 diabetes algorithm using electronic medical records and administrative healthcare data to study the population incidence and prevalence of type 1 diabetes in ontario, canada. *BMJ Open Diabetes Res Care* 2020;8(1).
73. Lynam A, McDonald T, Hill A, et al. Development and validation of multivariable clinical diagnostic models to identify type 1 diabetes requiring rapid insulin therapy in adults aged 18-50 years. *BMJ Open* 2019;9(9):e031586.
74. Carstensen B, Kristensen JK, Marcussen MM, Borch-Johnsen K. The national diabetes register. *Scand J Public Health* 2011;39(7 Suppl):58–61.
75. 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/>



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]. {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](https://www.esundhed.dk/-/media/Files/Dokumentation/Register-for-Udvalgte-Kroniske-Sygdomme-og-Svaere-Psykiske-Lidelser/29_Algoritmer_for_RUKS.ashx)
77. Sundhedsdatastyrelsen [English: Danish Health Data Authority]. Dokumentation af Laboratedatabasens 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/doesaarsager-og-biologisk-materiale/laboratedatabasen/dokumentation-af-labdatasens-forskertabel.pdf>
78. Anders Aasted Isaksen. The Open-Source Diabetes Classifier [Internet]. {last accessed 7 Feb 2023}; Available from: <https://github.com/Aastedet/osdc>
79. Danish Medicines Agency. Reimbursement thresholds [Internet]. {2023; last accessed on 7 Feb 2023}; Available from: <https://laegemiddelstyrelsen.dk/en/reimbursement/calculate-reimbursement/reimbursement-thresholds/>
80. 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>
81. 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.
82. 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](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Equivalised_disposable_income)
83. 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](https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf)
84. Wertheimer AI. The defined daily dose system (DDD) for drug utilization review. *Hosp Pharm* 1986;21(3):233–4, 239–41, 258.
85. 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>



86. 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>
87. DEFACTUM: Social & Health Services and Labour Market. DEFACTUM: Population Health Surveys [Internet]. {2022, last accessed 7 Feb 2023}; Available from: <https://www.defactum.net/research-areas/bif/>
88. RStudio Team. RStudio: Integrated development environment for r [Internet]. Boston, MA: RStudio, PBC; 2021. Available from: <http://www.rstudio.com/>
89. 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/>
90. Dowle M, Srinivasan A. Data.table: Extension of ‘data.frame’ [Internet]. 2021. Available from: <https://CRAN.R-project.org/package=data.table>
91. 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>
92. 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>
93. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software* 2019;4(43):1686.
94. Halekoh U, Højsgaard S, Yan J. The r package geepack for generalized estimating equations. *Journal of Statistical Software* 2006;15/2:1–11.
95. Iannone R. DiagrammeR: Graph/network visualization [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=DiagrammeR>
96. 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>
97. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ* 2014;348.
98. Wells, D. Wikipedia Commons: Risk Ratio vs Odds Ratio.svg [Internet]. {2022, last accessed 7 Feb 2023}; Available from: <https://commons.wikimedia.org/w/index.php?curid=100175653>
99. 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>

100. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology* 2004;159:702–6.
101. 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>
102. 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.
103. 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>
104. Bliddal M, Broe A, Pottegård A, Olsen J, Langhoff-Roos J. The danish medical birth register. *European Journal of Epidemiology* [Internet] 2018;33(1):27–36. Available from: <https://doi.org/10.1007/s10654-018-0356-1>
105. 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>
106. 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>
107. Green A, Hede SM, Patterson CC, et al. Type 1 diabetes in 2017: Global estimates of incident and prevalent cases in children and adults. *Diabetologia* 2021;64(12):2741–50.
108. Pettersson S, Hadziabdic E, Marklund H, Hjelm K. Lower knowledge about diabetes among foreign-born compared to swedish-born persons with diabetes-a descriptive study. *Nurs Open* 2019;6(2):367–76.
109. Fox MP, Lash TL, Bodnar LM. Common misconceptions about validation studies. *International Journal of Epidemiology* [Internet] 2020;49(4):1392–6. Available from: <https://doi.org/10.1093/ije/dyaa090>

110. Hussen HI, Yang D, Cnattingius S, Moradi T. Type i diabetes among children and young adults: The role of country of birth, socioeconomic position and sex. *Pediatric Diabetes* [Internet] 2013;14(2):138–48. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1399-5448.2012.00904.x>
111. Neu A, Willasch A, Ehehalt S, Kehrner M, Hub R, Ranke MB. Diabetes incidence in children of different nationalities: An epidemiological approach to the pathogenesis of diabetes. *Diabetologia* 2001;44 Suppl 3:B21–6.
112. The International Organization for Standardization. ISO 15189: Medical laboratories — Requirements for quality and competence [Internet]. {2022, last accessed 7 Feb 2023}; Available from: <https://www.iso.org/standard/76677.html>
113. Nielsen SS, Hempler NF, Waldorff FB, Kreiner S, Krasnik A. Is there equity in use of healthcare services among immigrants, their descendants, and ethnic Danes? *Scandinavian Journal of Public Health* 2012;40:260–70.
114. Tan X, Yang L, Khunti K, et al. Factors associated with switching from sulphonylureas to dipeptidyl peptidase 4 inhibitors among patients with type 2 diabetes in the united states. *Diabetes, Obesity and Metabolism* [Internet] 2021;23(10):2251–60. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.14466>
115. Carney G, Kim JD, O’Sullivan C, et al. Treatment pattern trends of medications for type 2 diabetes in british columbia, canada. *BMJ Open Diabetes Research and Care* [Internet] 2022;10(6). Available from: <https://drc.bmj.com/content/10/6/e002995>
116. Lasota AN, Overvad K, Eriksen HH, Tjonneland A, Schmidt EB, Gronholdt MM. Validity of peripheral arterial disease diagnoses in the danish national patient registry. *Eur J Vasc Endovasc Surg* 2017;53(5):679–85.
117. Lühndorf P, Overvad K, Schmidt EB, Johnsen SP, Bach FW. Predictive value of stroke discharge diagnoses in the danish national patient register. *Scandinavian Journal of Public Health* [Internet] 2017;45(6):630–6. Available from: <https://doi.org/10.1177/1403494817716582>
118. Sundboll J, Adelborg K, Munch T, et al. Positive predictive value of cardiovascular diagnoses in the danish national patient registry: A validation study. *BMJ Open* 2016;6(11):e012832.
119. Knudsen ST, Bodegård J, Birkeland KI, et al. Risk factor management of type 2 diabetic patients in primary care in the Scandinavian countries between 2003 and 2015. *Primary Care Diabetes* 2021;15:262–8.

120. Ke C, Stukel TA, Luk A, et al. Development and validation of algorithms to classify type 1 and 2 diabetes according to age at diagnosis using electronic health records. *BMC Medical Research Methodology* [Internet] 2020;20(1):35. Available from: <https://doi.org/10.1186/s12874-020-00921-3>
121. Thomas NJ, McGovern A, Young KG, et al. Identifying type 1 and 2 diabetes in research datasets where classification biomarkers are unavailable: Assessing the accuracy of published approaches. *J Clin Epidemiol* [Internet] 2022; Available from: <https://www.ncbi.nlm.nih.gov/pubmed/36368478>
122. Thomas NJ, Lynam AL, Hill AV, et al. Type 1 diabetes defined by severe insulin deficiency occurs after 30 years of age and is commonly treated as type 2 diabetes. *Diabetologia* 2019;62(7):1167–72.
123. Holcomb WL, Chaiworapongsa T, Luke DA, Burgdorf KD. An odd measure of risk: use and misuse of the odds ratio. *Obstetrics & Gynecology* 2001;98(4):685–8.
124. Wieland ML, Morrison TB, Cha SS, Rahman AS, Chaudhry R. Diabetes Care Among Somali Immigrants and Refugees. *J Community Health* 2012;37:680–4.
125. Gele AA, Mbalilaki AJ. Overweight and obesity among African immigrants in Oslo. *BMC Res Notes* 2013;6.
126. Skogberg N, Laatikainen T, Lundqvist A, Lilja E, Hrk nen T, Koponen P. Which anthropometric measures best indicate type 2 diabetes among Russian, Somali and Kurdish migrants in Finland? A cross-sectional study. *BMJ Open* 2018;8.
127. Ahmed S, Meyer HE, Kj llesdal MK, Madar MA. Prevalence and Predictors of Overweight and Obesity among Somalis in Norway and Somaliland: A Comparative Study. *Journal of Obesity* 2018;
128. Lim L-L, Lau ESH, Cheung JTK, et al. Real-world usage of sulphonylureas in asian patients with type 2 diabetes using the joint asia diabetes evaluation (JADE) register. *Diabetes, Obesity and Metabolism* [Internet] 2023;25(1):208–21. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.14865>
129. Lamprea-Montealegre JA, Madden E, Tummalapalli SL, et al. Association of race and ethnicity with prescription of SGLT2 inhibitors and GLP1 receptor agonists among patients with type 2 diabetes in the veterans health administration system. *Jama* 2022;328(9):861–71.
130. Bidulka P, Mathur R, Lugo-Palacios DG, et al. Ethnic and socioeconomic disparities in initiation of second-line antidiabetic treatment for people with type 2 diabetes in england: A cross-sectional study. *Diabetes, Obesity and Metabolism* [Internet] 2023;25(1):282–92. Available from: <https://dom-pubs.onlinelibrary.wiley.com/doi/abs/10.1111/dom.14874>

131. Li P, Zhang P, Guan D, et al. Changes in racial and ethnic disparities in glucose-lowering drug utilization and glycated haemoglobin A1c in US adults with diabetes: 2005-2018. *Diabetes Obes Metab* 2023;25(2):516–25.
132. Dor A, Encinosa W. How does cost-sharing affect drug purchases? Insurance regimes in the private market for prescription drugs. National Bureau of Economic Research Working Paper Series [Internet] 2004;No. 10738. Available from: <http://www.nber.org/papers/w10738>
133. McManus P, Donnelly N, Henry D, Hall W, Primrose J, Lindner J. Prescription drug utilization following patient co-payment changes in australia. *Pharmacoepidemiol Drug Saf* 1996;5(6):385–92.
134. Khan H, Lasker SS, Chowdhury TA. Prevalence and reasons for insulin refusal in bangladeshi patients with poorly controlled type 2 diabetes in east london. *Diabet Med* 2008;25(9):1108–11.
135. Hu J, Amirehsani KA, Wallace DC, Letvak S. The meaning of insulin to hispanic immigrants with type 2 diabetes and their families. *The Diabetes Educator* [Internet] 2012;38(2):263–70. Available from: <https://journals.sagepub.com/doi/abs/10.1177/0145721712437559>
136. Mull DS, Nguyen N, Mull JD. Vietnamese diabetic patients and their physicians: What ethnography can teach us. *West J Med* 2001;175(5):307–11.

## English summary

**Background:** Type 2 diabetes (T2D) is an increasing burden on healthcare services due to changes in demography and lifestyle. Migrants are particularly vulnerable to T2D, having a higher prevalence of the disease compared to native populations in western Europe, with some evidence suggesting a higher risk of complications and mortality. Evidence on disparities in care that may contribute to this excess risk in migrants is sparse, but register-based studies are well-suited to explore this in a Danish context. However, these studies are vulnerable to biases due to the lack of a validated method to identify T2D cases in the general population.

**Aims:** This dissertation aimed to develop and validate a register-based classification of type 1 diabetes (T1D) and T2D, and utilise it to explore specific areas of monitoring, biomarker levels, and pharmacological treatment, where disparities in T2D care between migrant groups and native Danes may contribute to excess complication risk in migrants.

**Methods:** Three cross-sectional register-based studies were performed. Study I validated two register-based classifications of diabetes - the *Open-Source Diabetes Classifier* (OSDC), and the *Register of Selected Chronic Diseases* (RSCD) - in a population of survey respondents from Central Denmark Region (*Health In Central Denmark*), while the two subsequent studies examined T2D prevalence and care in nationwide populations, using native Danes as the reference group. Study II computed relative risk (RR) of prevalent T2D and non-fulfilment of eleven indicators corresponding to guideline-recommendations. Study III computed RR of using glucose-lowering drug (GLD) combination therapy, and RR of using each type of GLD.

**Results:** In study I, both register-based diabetes classifiers identified valid populations of T1D and T2D in a general population (positive predictive values from 87.5% to 94.4%, negative predictive values above 98.9%), although sensitivity in the OSDC was substantially higher compared to the RSCD in both T1D (77.3% vs. 70.0%) and in T2D (94.4% vs. 90.5%). Neither algorithm was able to accurately classify diabetes type in individuals with T1D onset after age 40, nor T2D onset before age 40. In study II, prevalence of T2D was higher in all migrants excluding the Europe-group (fully-adjusted RR from 0.96 [95% confidence interval: 0.94-0.99] (Europe group) to 4.04 [3.89-4.20] (Sri Lanka group)). Non-fulfilment of recommendations for care was

common regardless of origin: in eight of the eleven indicators of care studied >25% of native Danes did not meet guideline-recommendations. Apart from monitoring in the Sri Lanka group, migrants were at similar or higher risk of non-fulfilment than native Danes across all indicators of monitoring (crude RR from 0.64 [0.51-0.80] (hemoglobin-A1c (HbA1c) monitoring, Sri Lanka group) to 1.62 [1.36-1.95] (HbA1c monitoring, Somalia group)), HbA1c control (crude RR from 1.27 [1.24-1.30] (Middle East group) to 1.46 [1.41-1.52] (Pakistan group)), and low-density lipoprotein cholesterol (LDL-C) control (crude RR from 1.08 [1.03-1.14] (Former Yugoslavia group) to 1.78 [1.67-1.90] (Somalia group)), while no overall disparities were observed for the presence of pharmacological treatment (crude RR from 0.61 [0.46-0.80] (GLD, Sri Lanka group) to 1.67 [1.34-2.09] (GLD, Somalia group)). In study III, migrants were less likely to use GLD combination therapy than native Danes (fully adjusted RR from 0.77 [0.71-0.85] (Somalia group) to 1.00 [0.97-1.04] (Former Yugoslavia group)). Migrants were more likely to use oral GLD types (fully adjusted RR (use of any oral GLD) from 0.99 [0.97-1.01] (Europe group) to 1.09 [1.06-1.11] (Sri Lanka group)), but less likely to use injection-based GLD types such as insulins (fully adjusted RR from 0.66 [0.62-0.71] (Sri Lanka group) to 0.94 [0.89-0.99] (Europe group)) and, especially, glucagon-like peptide-1 receptor agonists (GLP1RA) (fully adjusted RR from 0.29 [0.22-0.39] (Somalia group) to 0.95 [0.89-1.01] (Europe group)).

**Conclusion and perspectives** This dissertation was able to accurately identify populations of T1D and T2D in register data from the general Danish population. Diabetes epidemiologists can rely on the OSDC and the RSCD to provide valid study populations for Danish register-based research, but researchers should be aware of limitations in cases with atypical age at onset and potential challenges due to evolving GLD indications in the future. This dissertation found room for improvement in T2D care in Denmark, and presented several disparities in T2D care between migrants and native Danes. These may contribute excess risk of complication and mortality in migrants, especially among migrants from Somalia, who received the poorest care of all groups overall. Overall, the large disparity in use of GLP1RA between migrants and native Danes may provide a target for future interventions to improve care and reduce complication risk among migrants with T2D.

## Dansk resumé

**Baggrund:** Type 2 diabetes (T2D) er en stigende belastning på sundhedsvæsenet på grund af ændringer i demografi og livsstil. Indvandrere er særligt sårbare overfor T2D, idet de har en højere forekomst af sygdommen i forhold til indfødte befolkninger i Vesteuropa, ligesom nogle undersøgelser peger på en højere risiko for diabetiske komplikationer og død. Potentielle uligheder i behandlingen, der kan bidrage til den øgede risiko hos indvandrere med T2D, er dårligt belyst, men registerbaserede studier er velegnede til at undersøge dette i en dansk sammenhæng. Disse undersøgelser er imidlertid sårbare, idet der ikke foreligger nogen validerede metoder til at identificere T2D-tilfælde i den generelle danske befolkning.

**Formål:** Denne afhandling havde til formål at udvikle og validere en registerbaseret klassifikation af type 1 diabetes (T1D) og T2D, og anvende den til at undersøge specifikke områder i behandlingen af T2D - kontroller, biomarkør-niveauer og medicinsk behandling - hvor uligheder mellem indvandrergupper og indfødte danskere kan bidrage til øget risiko for komplikationer hos indvandrere.

**Metoder:** Som led i afhandlingen blev tre tværnsnittstudier baseret på register-data gennemført. Studie I validerede to registerbaserede klassifikationer af diabetes - *Open-Source Diabetes Classifier* (OSDC) og *Register for Udvalgte Kroniske Sygdomme* (RUKS) - blandt spørgeskema-besvarelser fra personer i Region Midtjylland (*Helbred i Midt*), mens de to efterfølgende studier undersøgte forekomst og behandling af T2D i hele landet og sammenlignede indvandrergupper med indfødte danskere. Studie II beregnede relativ risiko (RR) for forekomst af T2D og for manglende opfyldelse af elleve indikatorer for behandlingskvalitet svarende til anbefalinger i gældende kliniske retningslinjer. Studie III beregnede RR for brug af kombinationsbehandling med blodsukker-nedsættende lægemidler (GLD) og RR for brug af hver type GLD.

**Resultater:** I studie I identificerede begge registerbaserede diabetesklassifikationer gyldige populationer af T1D og T2D i en bred befolkning (positive prædiktive værdier fra 87,5% til 94,4%, negative prædiktive værdier over 98,9%), dog med en væsentlig højere sensitivitet i OSDC sammenlignet med RUKS både i T1D (77,3% vs. 70,0%) og i T2D (94,4% vs. 90,5%). Ingen af algoritmerne var i stand til nøjagtigt at klassificere diabetestype hos personer med T1D debut efter 40-årsalderen eller T2D debut før 40-årsalderen. I studie II var forekomsten af T2D højere i alle indvandrergupper



undtagen Europa-gruppen (fuldt justeret RR fra 0,96 [95% konfidensinterval: 0,94-0,99] (Europa-gruppen) til 4,04 [3,89-4,20] (Sri Lanka-gruppen)). Manglende opfyldelse af de vejledende behandlingsanbefalinger var hyppig - uanset oprindelse: i otte ud af elleve indikatorer opfyldte >25% af indfødte danskere ikke anbefalingerne. Fraset rutinemæssig kontrolundersøgelser i Sri Lanka-gruppen, havde indvandrergupperne tilsvarende eller højere risiko end indfødte danskere for ikke at opfylde behandlingsanbefalinger på tværs af alle indikatorer for rutinemæssig kontrolundersøgelser (rå RR fra 0,64 [0,51-0,80] (hæmoglobin-A1c (HbA1c-undersøgelse, Sri Lanka-gruppen) til 1,62 [1,36-1,95] (HbA1c-undersøgelse, Somalia-gruppen)), blodsukkerniveau (rå RR fra 1,27 [1,24-1,30] (Mellemøsten-gruppen) til 1,46 [1,41-1,52] (Pakistan-gruppen)) og kolesterolniveau (rå RR fra 1,08 [1,03-1,14] (Eksjugoslavien-gruppen) til 1,78 [1,67-1,90] (Somalia-gruppen)), mens der ikke sås nogen gennemgående forskelle for tilstedeværelse af medicinsk behandling (rå RR fra 0,61 [0,46-0,80] (GLD, Sri Lanka-gruppen) til 1,67 [1,34-2,09] (GLD, Somalia-gruppen)). I studie III var indvandrere mindre tilbøjelige til at få GLD kombinationsbehandling end indfødte danskere (fuldt justeret RR fra 0,77 [0,71-0,85] (Somalia-gruppen) til 1,00 [0,97-1,04] (Eksjugoslavien-gruppen)). Indvandrere var mere tilbøjelige til at benytte orale GLD typer (fuldt justeret RR (brug af ethvert oralt GLD) fra 0,99 [0,97-1,01] (Europa-gruppen) til 1,09 [1,06-1,11] (Sri Lanka-gruppen)), men mindre tilbøjelige til at benytte injektionsbaserede GLD typer så som insulin (fuldt justeret RR fra 0,66 [0,62-0,71] (Sri Lanka-gruppen) til 0,94 [0,89-0,99] (Europa-gruppen)) og især glucagon-like peptide-1 receptor agonister (GLP1RA) (fuldt justeret RR fra 0,29 [0,22-0,39] (Somalia-gruppen) til 0,95 [0,89-1,01] (Europa-gruppen)).

**Konklusion og perspektiver:** Denne ph.d.-afhandling kunne præcist identificere befolkninger med T1D og T2D i den brede danske befolkning ud fra registerdata. Diabetesforskere kan anvende OSDC og RUKS til at danne gyldige studiepopulationer ud fra de danske registre, men bør være opmærksomme på deres begrænsninger, herunder blandt patienter med atypisk alder for sygdomsdebut samt potentielle udfordringer som følge af udviklingen i brugen af GLD i fremtiden. Afhandlingen fandt plads til forbedringer i behandlingen af T2D i Danmark, og viste flere uligheder mellem indvandrere og indfødte danskere i behandlingen af T2D, som kan medføre øget risiko for komplikationer og død blandt indvandrere - særligt blandt indvandrere fra Somalia, som havde den laveste behandlingskvalitet af alle grupper. Den store forskel i brugen af GLP1RA mellem indvandrere og indfødte danskere kan muligvis være et mål for fremtidige interventioner til at øge behandlingskvaliteten og mindske komplikationsrisikoen blandt indvandrere med T2D.



## **A. Supplementary validation analyses**

### **A.0.1. Validation of unspecified diabetes in the Health In Central Denmark survey and in the National Health Survey.**

Validation analyses performed in the HICD study population from study I:

Validation analyses performed among respondents in the National Health Survey 2017 (individuals reporting diabetes as a current disease were considered diabetes cases).

Table A.1.: **Supplementary: Validation of unspecified diabetes in the HICD study population**

(a) The Open-Source Diabetes Classifier (OSDC)

OSDC	Survey: +DM	Survey: -DM	Total N
OSDC: +DM	2,512	221	2,733
OSDC: -DM	121	26,537	26,658
Total N	2,633	26,758	29,391
Sensitivity	0.954 (0.945, 0.962)		
Specificity	0.992 (0.991, 0.993)		
PPV	0.919 (0.908, 0.929)		
NPV	0.995 (0.995, 0.996)		

(b) The Register for Selected Chronic Diseases (RSCD)

RSCD	Survey: +DM	Survey: -DM	Total N
RSCD: +DM	2,422	122	2,544
RSCD: -DM	211	26,636	26,847
Total N	2,633	26,758	29,391
Sensitivity	0.920 (0.909, 0.930)		
Specificity	0.995 (0.995, 0.996)		
PPV	0.952 (0.943, 0.960)		
NPV	0.992 (0.991, 0.993)		

Table A.2.: **Supplementary: Validation of unspecified diabetes in the National Health Survey**

(a) The Open-Source Diabetes Classifier (OSDC)

OSDC	Survey: +DM	Survey: -DM	Total N
<b>OSDC: +DM</b>	1,630	258	1,888
<b>OSDC: -DM</b>	87	26,460	26,547
<b>Total N</b>	1,717	26,718	28,435
<b>Sensitivity</b>	0.95 (0.94, 0.96)		
<b>Specificity</b>	0.99 (0.99, 0.99)		
<b>PPV</b>	0.86 (0.85, 0.88)		
<b>NPV</b>	1.00 (1.00, 1.00)		

(b) The Register for Selected Chronic Diseases (RSCD)

RSCD	Survey: +DM	Survey: -DM	Total N
<b>RSCD: +DM</b>	1,546	184	1,730
<b>RSCD: -DM</b>	171	26,534	26,705
<b>Total N</b>	1,717	26,718	28,435
<b>Sensitivity</b>	0.90 (0.89, 0.91)		
<b>Specificity</b>	0.99 (0.99, 0.99)		
<b>PPV</b>	0.89 (0.88, 0.91)		
<b>NPV</b>	0.99 (0.99, 0.99)		

## **B. Full papers**

### **B.1. Paper I**

## B.2. Paper II

### B.3. Paper III



## C. Addenda and errata

This chapter contains an overview of:

- Addenda: contents added since the submitted version of the dissertation (commit db88939).
- Errata: corrections to errors discovered since submission of the dissertation.

### C.1. Addenda

- 5 May 2023:
  - Study I paper published. List of studies updated with citation and link to doi.
- 8 August 2023:
  - Study III paper published. List of studies updated with citation and link to doi.

### C.2. Errata

- 2 May 2023:
  - On lines 5-10, page 81, section 5.1.5., the following excerpt on the interpretation of Figure 5.4 states:
  - “Before 2016, when HbA1c data was not available from all regions of Denmark, a spike of inclusions is seen for women with T2D included at age 40 years. This spike is absent in the subsequent years, but the distribution is slightly skewed with a plateau in the age range 40-45 years. This is likely a residual effect of women in this age range having been prematurely included in the preceding years.”
    - \* This statement misinterprets some elements of the figure, which are addressed below:

- \* The density plots are smoothed, and spikes in single years are less visible compared to a histogram of observed counts (due to statistical disclosure requirements, this data could not be visualised using histograms). While the spike in included OSDC T2D-cases among women age 40 is indeed much lower in the years 2016-2018 than in the previous years, it is still present.
- \* Most likely, the spike is caused by women with onset of T2D before the age of 40 years during a period of time without HbA1c data. These women may have initiated metformin treatment before HbA1c data became available, and if they achieved stable euglycaemia using only metformin, they would not be captured by the HbA1c inclusion criterion after the HbA1c data became available. Thus, their inclusion into the cohort would be delayed until the censoring of metformin purchases ceases at age 40 years, causing a spike at this age of inclusion. In the post-2016 data, women living in the Southern Denmark Region (which did not supply HbA1c data prior to 2015, compared to the other regions, which supplied data from 2011 onwards) are likely the major source of this spike.