Department of Clinical Epidemiology

Aarhus University and Aarhus University Hospital
Olof Palmes Allé 43-45
8200 Aarhus N
Denmark
Tel.: +45 87167212
www.kea.au.dk

HbA1c-defined prediabetes in Denmark - Occurrence, trajectories, progression, and treatment eligibility

AARHUS UNIVERSITY

HbA1c-defined prediabetes in Denmark

Occurrence, trajectories, progression, and treatment eligibility

PhD dissertation

Sia Kromann Nicolaisen

Report no. 167

AARHUS UNIVERSITY
DEPARTMENT OF CLINICAL EPIDEMIOLOGY

midt
Central Denmark Region
Aarhus University Hospital

167

DEPARTMENT OF CLINICAL EPIDEMIOLOGY

HbA1c-defined prediabetes in Denmark

Occurrence, trajectories, progression, and treatment eligibility

PhD dissertation

Sia Kromann Nicolaisen

Health
Aarhus University
Department of Clinical Medicine

&

Aarhus University and Aarhus University Hospital
Department of Clinical Epidemiology

2023

Author

Sia Kromann Nicolaisen, MSc
Department of Clinical Epidemiology
Aarhus University Hospital
Olof Palmes Allé 43–45
8200 Aarhus N
Denmark
E-mail: sknc@clin.au.dk

Supervisors

Lars Pedersen, MSc, PhD (main supervisor)
Professor, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

Reimar W. Thomsen, MD, PhD
Professor, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

Henrik Toft Sørensen, MD, PhD, DMSc, DSc
Clinical Professor and Chair, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

Collaborators

Saskia le Cessie, MSc, PhD
Professor, Department of Clinical Epidemiology & Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

Olaf M. Dekkers, MD, PhD
Professor, Department of Clinical Epidemiology & Department of Endocrinology and Metabolism, Leiden University Medical Center, Leiden, The Netherlands

Irene Petersen, MSc, PhD
Professor of Epidemiology and Health Informatics, Department of Primary Care and Population Health, University College London, London, United Kingdom
Professor, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

Daniel R. Witte, MSc, PhD
Professor of Diabetes Epidemiology, Steno Diabetes Center Aarhus & Department of Public Health, Aarhus University, Aarhus, Denmark

The assessment committee

Erik Thorlund Parner, MSc, PhD (chairman and moderator of the defence)
Professor, Department of Public Health, Aarhus University, Aarhus, Denmark

Elizabeth Williamson, BA, MSc, PhD
Professor of Biostatistics and Health Data Science, London School of Hygiene & Tropical Medicine, London, United Kingdom

Frederik Persson, MD, DMSc
Senior Researcher, Steno Diabetes Center Copenhagen, Herlev, Denmark

Financial support

This PhD project was made possible with financial support from:

- Aarhus University
- Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University
- The Danish Diabetes Association
- ICPE 2022 Scholarship

Acknowledgement

You are now holding my PhD dissertation — the condensed result of three years as a PhD student at the Department of Clinical Epidemiology (DCE). The dissertation includes a lot of pages, but you do not have to read it all. I am an advocate of including all the necessary details such that you can always find the details you need, no matter whether you are an epidemiologist or a statistician, a clinician or a theorist, a professional or an amateur, a doctor or a patient.

I want to express my sincere gratitude to my main supervisor, *Lars Pedersen*. You have taught me so much during the past seven years and I am grateful for all our theoretical discussions. Your door is always open and I am impressed with your patience and your ability to always reduce my many questions, intangible thoughts, and crazy ideas to something specific. Research is about asking the right questions and you have certainly helped me do that.

I also want to thank my co-supervisors, *Reimar W. Thomsen* and *Henrik Toft Sørensen*. Reimar, thank you for always reminding me that there is a real world out there, with real patients and real doctors and that they should at some point also benefit from our work. Your knowledge about diabetes is immense and I am grateful for being part of your research group. Henrik, thank you for always keeping the project on track and ensuring each study has the right focus. Thank you for asking the simple, yet fundamental questions.

Irene Petersen, I want to thank you and your colleagues in London for welcoming me in your department and sharing your knowledge. Same goes for *Daniel R. Witte*, thank your for sharing your knowledge and pinpointing the gaps in the literature.

A "dankjewel" goes to everyone in the Department of Clinical Epidemiology at Leiden University Medical Center (LUMC). You made me feel welcome from my very first day. *Olaf M. Dekkers*, thank you for your valuable input about diabetes. *Saskia le Cessie*, I am grateful for all your help and all our discussions. Thanks you for keeping an eye on me and supporting me and thank you for agreeing with me that statistics and formulas are not just a necessary evil. I hope to be able to continue my work with you and I will be pleased to come back to the LUMC.

A special thanks to all my colleagues at the DCE for great company and support, especially in the PhD group, the D&M group, and among the statisticians. *Tina Christensen*, thank you for valuable feedback and editorial assistance with this dissertation.

Last but not least, I want to thank my friends and my family for always believing in me.

Thank you.
Sia Kromann Nicolaisen
Aarhus, January 2023

List of papers

This PhD dissertation is based on the following four studies:

Study I: Descriptive study

HbA1c-defined prediabetes and progression to type 2 diabetes in Denmark during 2012-2018: a study based on laboratory data from routine care

Nicolaisen SK, Pedersen L, Witte DR, Sørensen HT, Thomsen RW.
Submitted.

Study II: Prediction study

Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark.

Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L.
BMJ Open Diabetes Research & Care 2022;10(5):e002946.

Study III: Trajectory study

Longitudinal HbA1c patterns before first treatment of diabetes in everyday clinical practice: A latent class trajectory analysis.

Nicolaisen SK, le Cessie S, Thomsen RW, Witte DR, Dekkers OM, Sørensen HT, Pedersen L.
In draft.

Study IV: Regression discontinuity design study

Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study.

Petersen I, Nicolaisen SK, Ricciardi F, Sharma M, Thomsen RW, Baio G, Pedersen L.
Clinical Epidemiology 2020;12:569-577.

Abbreviations

2hPG	2-hour plasma glucose
ACE	Average causal effect
ADA	American Diabetes Association
AIC	Akaike information criterion
ATC	Anatomical Therapeutic Chemical
AUCt	Time-dependent area under the curve
AUPD	Aarhus University Prescription Database
BIC	Bayesian information criterion
BMI	Body mass index
CCI	Charlson Comorbidity Index
CI	Confidence interval
CVD	Cardiovascular disease
DCCT	Diabetes Control and Complications Trial
DCRS	Danish Civil Registration System
DNHS	Danish National Health Survey
DNPR	Danish National Patient Registry
DPP	Diabetes Prevention Program
DPPOS	Diabetes Prevention Program Outcomes Study
DPS	Finnish Diabetes Prevention Study
DRMPS	Danish Register of Medicinal Product Statistics
EDIC	Epidemiology of Diabetes Interventions and Complications
FPG	Fasting plasma glucose
HbA1c	Glycated haemoglobin
HR	Hazard ratio
ICD-10	International Statistical Classification of Diseases and Related Health Problems, Tenth Revision
IDF	International Diabetes Federation
IEC	International Expert Committee
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IFG	Impaired fasting glucose
IGT	Impaired glucose tolerance
IPA	Index of prediction accuracy
IQR	Interquartile range
ITT	Intention to treat
IV	Instrumental variable

LABKA	Clinical laboratory information system
LADA	Latent autoimmune diabetes of adulthood
LASSO	Least absolute shrinkage and selection operator
MODY	Maturity onset diabetes of the young
NGSP	National Glycohemoglobin Standardization Program
NPU	Nomenclature for Properties and Units
OGTT	Oral glucose tolerance test
RDD	Regression discontinuity design
RLRR	Register of Laboratory Results for Research
ROCt	Time-dependent receiver operating characteristic
THIN	The Health Improvement Network
UKPDS	United Kingdom Prospective Diabetes Study
WHO	World Health Organization

Contents

1 Introduction

In 2011, the World Health Organization (WHO) introduced glycated haemoglobin (HbA1c) as a diagnostic criterion for diabetes[1] based on recommendations from the International Expert Committee (IEC)[2]. Previously, HbA1c was mainly used to guide management and adjust treatment of diabetes[2]. However, as HbA1c met most of the criteria for an ideally perfect test (being accurate, specific, standardized/standardizable, handy, and inexpensive[3]), clinicians had wondered whether this test could also be used to diagnose diabetes as it had various advantages over the then used measures; fasting plasma glucose (FPG) measured using a fasting blood test and 2-hour plasma glucose (2hPG) measured using an oral glucose tolerance test (OGTT). The Danish Health Authority introduced HbA1c in the Danish recommendations in 2012[4] and HbA1c testing has increased substantially in the Danish population since then. Denmark has a tax-supported health care system ensuring unfettered access to medical care for all residents[5] and HbA1c measurements have been recorded in routine care laboratory databases for more than 20 years. As all Danes (approximately 5.9 million individuals in 2022[6]) are assigned a unique personal identification number, individual linkage between registries is possible[7]. This offers a unique possibility to conduct epidemiologic research[8] as all Danes can be followed from cradle to grave in the registries[9].

This PhD project uses data from the Danish medical registries and aims to contribute new knowledge about HbA1c as a prognostic marker for diabetes development, the individual HbA1c values prior to initiating any glucose-lowering treatment, as well as the clinical outcomes related to treatment eligibility. The main focus is on the statistical analyses and the methodological challenges within the field of prediabetes and diabetes.

This PhD dissertation includes a description of prediabetes and diabetes in Denmark and introduces the concepts of the glycemic burden and HbA1c trajectories. This is followed by a specification of the aims for the studies, along with the methods used in each study. The methods section is supported by methodological appendices with further details about the methods, including relevant SAS or R codes. The results from each of the studies are presented and discussed, followed by a conclusion and a section about future perspectives. Each of the studies are enclosed separately in an appendix.

2 | Background

It has been known since the 1970s that values of HbA1c are higher in individuals with diabetes compared to those without diabetes[10]. HbA1c depicts the average glycemic level in the blood over the past 8–12 weeks[11] and is therefore a measure of the long-term glycemic level rather than the level at a single point in time. When HbA1c was introduced in the diagnostic definition of diabetes, it was emphasized that the current methods/measures were still applicable[4]. However, measuring the HbA1c quickly turned into the most frequently used method to diagnose diabetes[12] as it is a much more convenient alternative to the former methods; it is more stable than FPG and 2hPG, it does not require the individual to be fasting, and the test can be conducted at any time during the day[13]. Measuring the HbA1c is recommended as the primary diagnostic method in Denmark[14] and FPG and 2hPG should primarily be used when values of HbA1c are not reliable (e.g., in cases with sickle cell disease, pregnancy, glucose-6-phosphate dehydrogenase deficiency, HIV, hemodialysis, recent blood loss or transfusion, or erythropoietin therapy[15])[14]. Values of HbA1c are reported in the SI unit of the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC), mmol/mol without decimals, along with the derived National Glycohemoglobin Standardization Program (NGSP)/Diabetes Control and Complications Trial (DCCT) unit, % reported with one decimal. The units are converted using the rounded equations[13,16] (Figure 2.1)

$$\begin{aligned} \mathrm{HbA1c}_{\mathrm{mmol/mol}} &= 10.93 \cdot \mathrm{HbA1c}_{\%} - 23.5 \\ \mathrm{HbA1c}_{\%} &= 0.0915 \cdot \mathrm{HbA1c}_{\mathrm{mmol/mol}} + 2.15. \end{aligned}$$

In the Danish registries, the DCCT unit, %, was mainly used before 2014 and the IFCC unit, mmol/mol, has been increasingly used since 2008. HbA1c is currently one of the most ordered blood tests in routine clinical care in Denmark[17] and after 2012, 95% of those who either had hospital-diagnosed diabetes or initiated glucose-lowering treatment had HbA1c measured at least once in the year prior to their diabetes diagnosis[12]. The proportion of individuals with a plasma glucose measurement or an OGTT in the year prior to diagnosis quickly decreased after 2012, whereas the proportion with an HbA1c measurement increased[12,18].

2.1 Prediabetes

Prediabetes is a stage where glucose levels in the blood are too high to be considered normal, but too low to have crossed the threshold for diabetes[1,2,20,21].

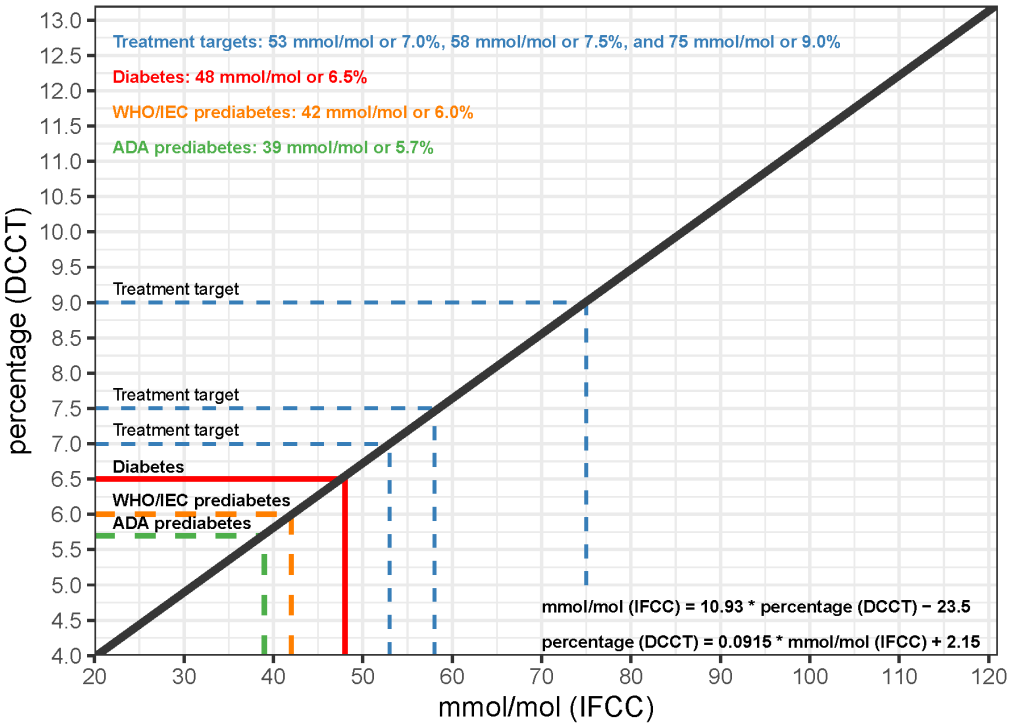


Figure 2.1: Conversion of units for HbA1c. Prediabetes is either defined as HbA1c 39–47 mmol/mol or 42–47 mmol/mol. Diabetes is defined as HbA1c ≥ 48 mmol/mol. Diabetes treatment should be initialized when HbA1c ≥ 48 mmol/mol[14,19]. HbA1c should preferably be below 48 mmol/mol at least in the first years of the disease[14,19]. In the long run, it is recommended that HbA1c does not exceed 53 mmol/mol[14,19]. For some individuals, this limit is set to 58 mmol/mol or even 75 mmol/mol if the only treatment purpose is to reduce symptoms[14,19].

As prediabetes is defined based on values below the threshold for diabetes, changing the definition of diabetes inevitably also changes the definition of prediabetes. The definition of prediabetes is rather unclear[21,22] (Table 2.1), but the general consensus is that individuals with prediabetes are at increased risk of developing diabetes and diabetes-related complications[15,21]. The WHO refrains from using the term "prediabetes" and argues it might give the impression that no intervention is necessary because diabetes is inevitably the next stage[20]. Instead, they use the term "intermediate hyperglycemia" to indicate that not all individuals with prediabetes progress to diabetes[20] and that it is an intermediate condition in the progression from a state with normal plasma glucose levels to diabetes. The WHO defines intermediate hyperglycemia using either impaired glucose tolerance (IGT) defined as 2hPG 7.8–11.0 mmol/L (without diabetes) or impaired fasting glucose (IFG) defined as FPG 6.1–6.9 mmol/L and (if it is measured) 2hPG in the normal range (i.e., < 7.8 mmol/L)[23,24] (Table 2.1). Values ≥ 7.0 mmol/L for FPG or > 11.0 mmol/L for 2hPG define diabetes[23]. Since 2011, HbA1c has also been part of the definition of diabetes, with HbA1c ≥ 48 mmol/mol as the threshold[1]. The WHO has not profoundly accepted HbA1c as part of their definition of prediabetes[1,2], but acknowledges the threshold proposed by the IEC who defined a "highest-risk group for diabetes", i.e., HbA1c 42–47 mmol/mol[2]. The risk of

diabetes is a continuum and there is a gradual increase in risk with increasing glucose levels. Therefore, there is no clear definition of "normoglycemia" (i.e., values below the threshold for prediabetes) and consequently no clear threshold to define prediabetes[2]. The WHO and the American Diabetes Association (ADA) agree on the threshold for diabetes, and partly also on the definition of IGT, but the ADA uses a slightly lower threshold for IFG[15]. The ADA defines IFG using FPG 5.6–6.9 mmol/L. In addition, the ADA has included HbA1c 39–47 mmol/mol in the definition of prediabetes along with IGT and IFG (Table 2.1). Unlike the WHO, the ADA uses the term "prediabetes"[15].

Definition	American Diabetes Association (ADA) Prediabetes	World Health Organization (WHO) Intermediate hyperglycemia	International Expert Committee (IEC) Highest-risk group for diabetes
Impaired glucose tolerance (IGT): 2-hour plasma glucose (2hPG) after a 75 g oral glucose tolerance test (OGTT)	7.8–11.0 mmol/L (140–200 mg/dL)	7.8–11.0 mmol/L (140–200 mg/dL)	—
Impaired fasting glucose (IFG): fasting plasma glucose (FPG) concentration	5.6–6.9 mmol/L (100–125 mg/dL)	6.1–6.9 mmol/L (110–125 mg/dL)	—
Subdiabetic glycated haemoglobin (HbA1c): HbA1c	39–47 mmol/mol (5.7–6.4 %)	—	42–47 mmol/mol (6.0–6.4 %)

Table 2.1: Values of glucose concentrations (using the units mmol/L and mg/dL) and HbA1c (using the units mmol/mol and %) used to define prediabetes[2,15,24]. The ADA defines IFG, IGT, and subdiabetic HbA1c as three separate groups where the presence of at least one of them defines prediabetes as long as diabetes is not present. The WHO defines IFG and IGT as interdependent groups, where the IGT can only be present as long as FPG is below the threshold for diabetes (i.e., FPG < 7.0 mmol/L) and IFG is present if the values of 2hPG are either not measured or below the threshold for IGT.

The various definitions of prediabetes (Table 2.1) identify different populations and even though they are overlapping, they have distinct characteristics[25]. Studies based on the Whitehall II cohort of more than 5,000 British civil servants aged 50–80 years without diabetes at the time of enrolment compared the overlap of definitions using FPG and HbA1c measured simultaneously at a clinical examination[25]. A total of 402 individuals had prediabetes based on FPG in the interval 6.1–6.9 mmol/L and 288 had prediabetes defined as HbA1c

42–47 mmol/mol. Only 10% (62 individuals) had prediabetes according to both criteria[25]. Compared to those with prediabetes defined by the FPG criteria, those with prediabetes defined by HbA1c were more likely to be women (34.1% vs. 19.9%), were older (mean 65.2 years vs. 62.6 years), and had a higher rate of previous cardiovascular disease (CVD) (18.4% vs. 11.9%). An important feature is also the fact that they had higher levels of HbA1c (43.2 mmol/mol vs. 38.1 mmol/mol), higher values of 2hPG (7.9 mmol/L vs. 7.1 mmol/L), and lower values of FPG (5.7 mmol/L vs. 6.3 mmol/L). This illustrates the inability of the definitions to identify the same individuals. In addition, the different definitions also show differences in the risks of progression to diabetes[26–28], reversion to normal glucose values[29], and risks of diabetes complications and all-cause mortality[25,30,31]. Most of the current evidence regarding interventions to diabetes is based on trials and studies, where prediabetes is defined by IFG or IGT[15,28,32], and there is a general lack of evidence regarding prediabetes defined using HbA1c[15,28].

2.1.1 Prediabetes in Denmark

It has been estimated that around 310,000–320,000 adult Danes (in 2017) had prediabetes defined using HbA1c 42–47 mmol/mol and this corresponds to around 7% of the entire Danish adult population (age 20–100 years)[33]. A slightly higher estimate of around 360,000–375,000 Danes with prediabetes (in 2017)[34] relies on the fact that the number of individuals with prediabtes has been estimated to correspond to one and a half times the number of individuals with type 2 diabetes. It has been predicted that 467,000 Danes will have type 2 diabetes in 2030[35]. This yields 700,000 Danes with prediabetes in 2030. The incidence of prediabetes has been studied in different selected cohorts[36–41], but rarely based on HbA1c[42–44] and virtually never exclusively based on the WHO/IEC definition used in Denmark (42–47 mmol/mol). Based on a survey, the Danish Inter99 Study[45] included 3,187 individuals (52% women, median age 46 years, and mean body mass index (BMI) 26 kg/m^2) who were initially free of both diabetes and prediabetes. They followed them for five years and found that 303 individuals had developed prediabetes (defined based on IFG and IGT) during 16,328 years of follow-up, i.e., a prediabetes incidence of 18.6 per 1,000 person-years. The true incidence and prevalence of prediabetes in the general Danish population is unknown as there has been no systematic population-wide screening for prediabetes or type 2 diabetes in Denmark. Nevertheless, population-based laboratory databases could now potentially be a tool to study prediabetes epidemiology due to the substantially increased use of HbA1c testing after 2012[46].

2.1.2 Diabetes in Denmark

In 2017, almost 5% of the entire Danish population, corresponding to 280,000 individuals, had diabetes[47]. Of these, approximately 10% were classified as having type 1 diabetes[47], leaving around 250,000 individuals with type 2 diabetes[33,47]. The type 2 diabetes prevalence depends on age and sex, with slightly

lower prevalence among women than men and an increasing prevalence as age progresses[47]. Even though around 90% of those with diabetes are classified as having type 2 diabetes, only around 80–85% truly have type 2 diabetes[14,34]. A group of around 5–10% have so-called latent autoimmune diabetes of adulthood (LADA), a type usually classified as type 2 diabetes, yet with a phenotype more similar to type 1 diabetes[34]. An even smaller group of 1–2% have maturity onset diabetes of the young (MODY), a genetically inherited type of diabetes[34]. Other types of diabetes include, e.g., gestational diabetes, medically induced diabetes, neonatal diabetes, and secondary diabetes (often caused by pancreatic diseases). These types are rare and together with LADA and MODY, they account for the approximately 10% of diabetes cases that are neither type 1 nor type 2 diabetes[48].

After the introduction of HbA1c as a diagnostic criterion, the epidemiology of type 2 diabetes changed and an immediate decrease in the incidence of type 2 diabetes was observed due to non-overlapping definitions[12,47]. A Danish biobank study evaluated the change in the number of individuals with diabetes identified by HbA1c compared to FPG[13]. The study included 4,239 individuals (44% women, aged 25–75 years) from the background population of the former County of Vejle, selected to match (by gender and age in 10-year age groups) the diabetes population in Vejle. FPG and HbA1c were measured at study inclusion and whereas 3.6% of the individuals from the background population had diabetes based on FPG $\geq$ 7.0 mmol/L, only 1.9% had diabetes based on HbA1c $\geq$ 48 mmol/mol[13]. The prevalence of diabetes was therefore reduced by 46% when changing the diagnostic method from one FPG measurement to one HbA1c measurement[13]. The majority of the individuals who are now diagnosed based on HbA1c $\geq$ 48 mmol/mol will, if measured, most likely also have FPG or 2hPG values indicating diabetes[14]. On the other hand, individuals who would previously have been diagnosed by means of FPG or 2hPG, will now mainly have measurements of HbA1c. For a large group, the HbA1c value will not be above 48 mmol/mol and they will therefore remain undiagnosed and untreated[13,14]. According to the guidelines, the diagnosis of diabetes should be confirmed by an additional measurement of the same type (HbA1c, FPG, or 2hPG), unless there are obvious signs of diabetes[14]. Whereas only 66% of the individuals from the Danish biobank study[13] who had FPG-defined diabetes also had FPG $\geq$ 7.0 mmol/L when their measurement was repeated within one year, 91% of the individuals diagnosed using an HbA1c measurement also had HbA1c $\geq$ 48 mmol/mol the second time[13].

After 2014, the incidence of type 2 diabetes increased again[12,47] and in recent years, the incidence rate has been around 4 per 1,000 person-years[12,47]. The life-time risk of type 2 diabetes is 25% (26% for men and 21% for women)[49] and despite the improvement in patient care during the last decades, hence a better individual future prospect, the accumulated future burden on the population is increasing[49]. The increasing prevalence of type 2 diabetes is mainly driven by an increase in incidence[35] and preventing new cases of diabetes is therefore crucial as the number of individuals in need of diabetes care will only increase in the future[35].

2.2 Progression from prediabetes to diabetes

Prediabetes is defined as "increased risk of type 2 diabetes" and HbA1c levels above the threshold for prediabetes increase the risk of future type 2 diabetes compared to normal levels of HbA1c[1,2,21,27,50–53]. However, as is also evident from the discussion about the term "prediabetes", not everyone with prediabetes will progress to overt diabetes. Reported 5-year risk of progression from prediabetes to type 2 diabetes vary based on the definition of prediabetes — from 18% for IFG based on the ADA definition to 50% when requiring both IFG and IGT[28]. For prediabetes defined using HbA1c 42–47 mmol/mol, the 5-year risk has been estimated at 38% in a 2018 Cochrane Review[28], but the evidence is sparse when prediabetes is defined using only HbA1c[15,26,28]. A meta-analysis showed a progression rate from prediabetes defined using HbA1c 42–47 mmol/mol to diabetes defined using either FPG or HbA1c of 36 per 1,000 person-years. When diabetes was defined using only HbA1c, the progression rate was only 26 per 1,000 person-years. In a British study using data from the The Health Improvement Network (THIN) database[54], almost 400,000 individuals with prediabetes were identified using Read codes, FPG, or HbA1c to define prediabetes. Almost 19% progressed to diabetes (defined using Read codes) during a median 2.7 years, yielding a progression rate of 53.5 per 1,000 person-years. As is evident, the estimates for progression vary greatly across definitions of prediabetes/diabetes, but it also depends on the characteristics of the individuals[20,24]. Some important risk factors for developing type 2 diabetes include age, sex, BMI, pre-existing comorbidities, ethnicity, smoking, unhealthy diet and lifestyle, family history of diabetes, and gestational diabetes[20]. As type 2 diabetes can potentially be prevented (or delayed), understanding the progression from prediabetes to type 2 diabetes is crucial[24].

2.3 Treatment guidelines

The Danish Ministry of Health has published a health initiative for diabetes (Den Nationale Diabeteshandlingsplan, 2017–2020)[55], in which 12 initiatives are listed. Some of the main points concern prevention of diabetes and actions for earlier detection[55]. The general practitioners are encouraged to pay more attention to prediabetes to potentially prevent future type 2 diabetes[55], but there are no official guidelines for screening for prediabetes or diabetes in Denmark. According to the ADA, screening for prediabetes and type 2 diabetes is recommended in order to guide on whether to do any diagnostic testing (e.g., testing for elevated HbA1c) and includes an assessment of risk factors such as high BMI, old age, or physical inactivity[15]. In general, this should be repeated at a minimum of 3-year intervals, but at least once per year for individuals with prediabetes[15]. There are currently no Danish guidelines for pharmacological treatment of prediabetes. However, as the individuals with prediabetes are at increased risk of diabetes, they are encouraged to, e.g., loose weight, exercise, have a healthy diet, and quit smoking[56]. When type 2 diabetes is diagnosed, a treatment plan assessing resources, lifestyle, complications, and risk factors should be developed[14]. It is emphasized that once

type 2 diabetes is diagnosed (i.e., when HbA1c $\geq$ 48 mmol/mol) it will always be present and individuals diagnosed with type 2 diabetes should always be treated according to the guidelines — even if HbA1c has been brought down to a level below 48 mmol/mol[14]. The purpose of the treatment is to increase quality of life, remove or reduce symptoms, prevent or postpone complications, and lower the mortality[14]. This is done through lifestyle interventions and treatment of cardiovascular risk factors (e.g., dyslipidemia, microalbinuria, and hypertension)[18]. It can be difficult to achieve good glycemic control, and thus reduce the risk of microvascular and macrovascular complications to a low level, without glucose-lowering treatment[14,19]. It is recommended that the treatment is initiated (in combination with lifestyle interventions) early after diagnosing type 2 diabetes and even before having seen the effect of the lifestyle interventions[14]. As a general rule, treatment should be initiated within 3–6 months after the diagnosis of type 2 diabetes and should not depend on the HbA1c level. The purpose of the pharmacologic treatment is to lower the glucose level as damage caused by dysregulated glycemic levels is irreversible[14]. The preferred first-line treatment option is currently metformin, but some individuals will benefit from a combination of drugs[14,19].

2.3.1 Treatment targets for HbA1c

The treatment target for levels of HbA1c (Figure 2.1) depends on the individual, although it is generally recommended that HbA1c is lowered as much as possible and preferably to a level below 48 mmol/mol[14,19]. This is the aim in the first years after the diagnosis for individuals diagnosed with relatively low values of HbA1c (i.e., HbA1c < 70 mmol/mol). For the majority of the individuals, it is not possible to maintain HbA1c < 48 mmol/mol over time and it is then recommended that HbA1c at least stays below 53 mmol/mol. For individuals with unstable levels of HbA1c or initial values above 70 mmol/mol, levels below 58 mmol/mol rather than 53 mmol/mol can be preferred. This is also the case for fragile individuals (e.g., individuals with old age, long duration of diabetes, or a large burden of comorbidites) for whom strict glycemic control may increase the risk of hypoglycemia and death[14]. If the primary goal is to reduce symptoms, a level between 58 mmol/mol and 75 mmol/mol can be accepted[14,19]. It is emphasized that the treatment plan should be developed together with the individual and that all aspects should be taken into account. Still, the overall aim is to lower the glycemic level and it is generally better to be proactive and initiate and intensify treatment sooner rather than later instead of withholding the treatment and then later try to revert a case of dysregulated diabetes[14].

2.4 The glycemic burden and HbA1c trajectories

Intervention (be it pharmacological treatment or lifestyle changes) should be initiated soon after type 2 diabetes is diagnosed and the overall purpose of the intervention is to lower the glucose levels and maintain glycemic control[14,19]. Ideally, intervention should be initiated immediately after the level of HbA1c

has crossed the threshold for diabetes. As the HbA1c value is often much higher than 48 mmol/mol when treatment is initiated this indicates therapeutic inertia. Around 50% of those who initiated treatment in Denmark in the period 2012–2017 had HbA1c values above 53 mmol/mol and almost 20% had HbA1c above 75 mmol/mol[46]. Evidence from the United Kingdom Prospective Diabetes Study (UKPDS) shows that even small increases in HbA1c in the years after the diagnosis of type 2 diabetes increase the risk of death or myocardial infarction[57]. This highlights the need for early implementation of glycemic control among individuals with newly diagnosed type 2 diabetes[57,58]. The UKPDS and the DCCT/Epidemiology of Diabetes Interventions and Complications (EDIC) trials have shown beneficial effects of periods with glycemic control and they both indicate that there might be a beneficial effect from intensive glucose control compared to standard treatment — even after glucose levels have reverted to the same levels[57–60] (Figure 2.2). This has been denoted "metabolic memory" (type 1 diabetes) and "legacy effect" (type 2 diabetes) and indicates that not only the current glucose levels but also the prior glucose levels have an effect on the long-term outcomes related to diabetes[57–61]. The focus should not only be on HbA1c at a single point in time, but rather on the development of the HbA1c levels over time, i.e., the HbA1c trajectories, to also include information about the longitudinal and historical levels. Evidence from three small studies indicates that the progression from normal glucose, via prediabetes, to diabetes might be characterized by slowly increasing glucose levels followed by a rapid increase before the onset of diabetes[62–64] (Figure 2.3). A considerable glycemic burden may be present even before diabetes has been diagnosed. The studies are, however, limited by small sample sizes (2hPG trajectories from 55 pima indians without diabetes[63], 2hPG and FPG trajectories from 505 individuals without diabetes from the Whitehall II cohort[62], and FPG and HbA1c trajectories from 193 Japanese individuals without diabetes[64]) and evidence from population-based studies is scarce.

As most of the individuals with prediabetes (and type 2 diabetes) are currently identified based on HbA1c and not FPG or OGTT, updated research is essential to fully understand the prediabetes epidemiology and the progression from prediabetes to diabetes. In order to prevent of new cases of type 2 diabetes, interventions should be initiated before hyperglycemia causes any potentially irreversible damage.

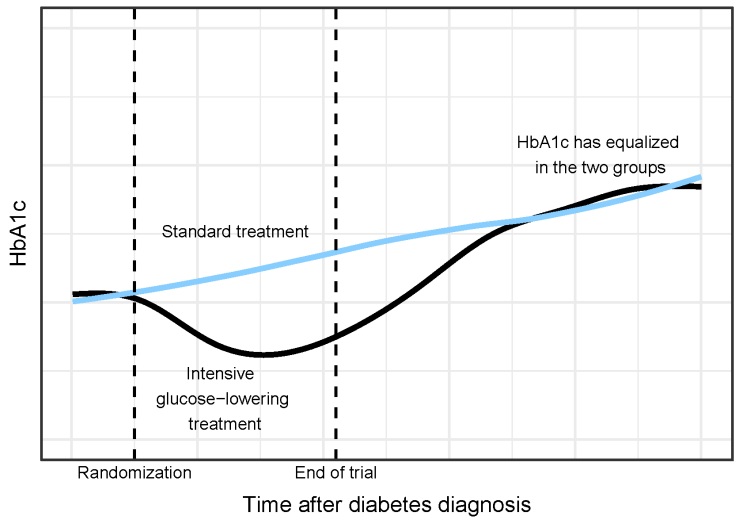


Figure 2.2: In trials, individuals with diabetes have been randomized to either standard treatment or intense glucose-lowering treatment[57–60]. During the randomization period, where one group was treated with standard treatment and one group with intense glucose-lowering treatment, the HbA1c for the two groups differed. After the trials, where all patients went back to receiving the standard treatment, HbA1c from the two groups equalized. The concepts of "metabolic memory" and "legacy effects" were originally described in regards to the DCCT/EDIC (type 1 diabetes) and the UKPDS (type 2 diabetes) trials and indicate that there is a beneficial effect of intensive treatment, where glucose levels are lowered, even after glucose levels have reverted back to the same level as for those without the intensive treatment[57–60].

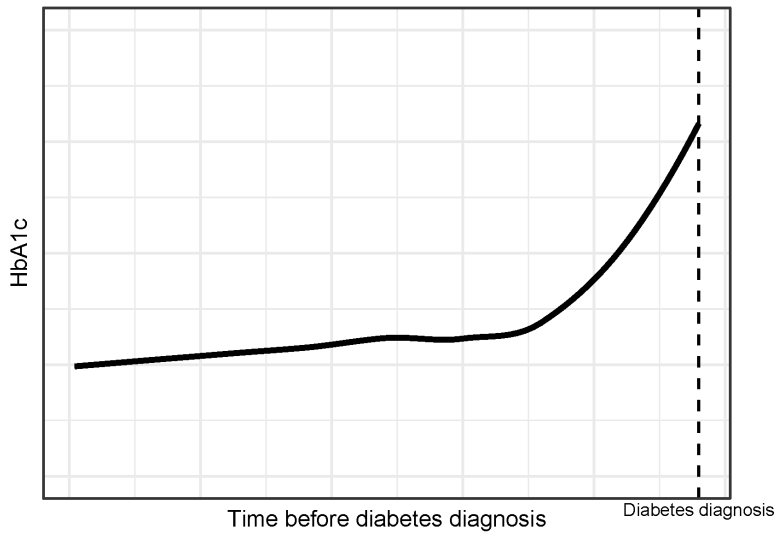


Figure 2.3: A hypothetical representation of an HbA1c trajectory prior to the diagnosis of diabetes. The progression from normal glucose levels to diabetes is believed to be characterized by slowly increasing glucose levels followed by a rapid increase before the onset of diabetes[62–64]. The increased glucose levels prior to the diabetes diagnosis may contribute to the glycemic burden even before diabetes has been diagnosed.

3 | Aims

The overall aim of this PhD project is to use data from the Danish medical registries and model the clinical course of prediabetes to understand the progression towards type 2 diabetes. We will explore HbA1c trajectories and study the association between treatment initiation and long-term clinical outcomes. We will assess whether patients with prediabetes can be classified by their HbA1c trajectories as this may contribute importantly to the debate on when to initiate glucose-lowering treatment. This project includes four studies with the following specific aims:

Study I: Descriptive study

To examine the use of HbA1c measurements in Denmark after 2012 and exploit the Danish laboratory registry data to identify individuals with incident HbA1c-defined prediabetes and examine their prevalence, incidence rate, characteristics, mortality, and risk of progression to diabetes.

Study II: Prediction study

To develop and validate prediction models for the 5-year risk of progressing to type 2 diabetes among individuals with incident HbA1c-defined prediabetes. The study will be conducted in a competing risk setting and the validation measures will account for the competing risk of death.

Study III: Trajectory study

To use the longitudinal HbA1c data available in the Danish laboratory databases using a large cohort of real-world individuals who initiated glucose-lowering treatment. We will examine whether there are different HbA1c trajectories and thereby whether there is heterogeneity in levels of HbA1c. We will assess the therapeutic inertia in the period prior to first-ever glucose-lowering treatment initiation and we will use latent class trajectory analyses to group individuals based on their HbA1c trajectories.

Study IV: Regression discontinuity design study

To examine the effect of being eligible for type 2 diabetes treatment on mortality and cardiovascular events. We will use the regression discontinuity design (RDD) and therefore potentially eliminate confounding by design as we will compare individuals with levels of HbA1c just below and individuals with levels just above the threshold of HbA1c 48 mmol/mol.

4 | Methods

Denmark has a long history of registering and the Danish National Archives holds information about Denmark and the Danes dated back to the 1100s[65]. In 1968, the CPR number was introduced as a personal identifier for all Danes. Since then, almost all personal records have been registered based on the CPR number[7] (Figure 4.1). This offers almost unique possibilities within the field of epidemiology, because of the availability of the nationwide health care registries with routinely collected administrative, health, and clinical quality data[5], where the CPR number makes individual linkage between the registries possible. The Danish health care system is universally-covering and tax-supported and offers access to general practitioners and hospitals along with partial reimbursement for prescribed drugs. Since 2007, the health care system has mainly been administered by the five Danish administrative regions and they have the primary responsibility for the somatic and psychiatric hospitals, the general practitioners, and the specialists in private practice[5] (Figure 4.2).

Figure 4.1: Almost all individual contacts with the Danish society are based on the CPR number. As an example, the health insurance card, "the yellow card", which ensures the right to health care in Denmark, includes information about the CPR number (280790-XXXX), name (Sia Kromann Nicolaisen), address (*Adresse*), municipality (Aarhus Kommune), and region (Region Midtjylland) of a Danish resident, together with the name, address, and phone number of the general practitioner.

Individual-level information from the following data sources were used for the studies in this PhD project.

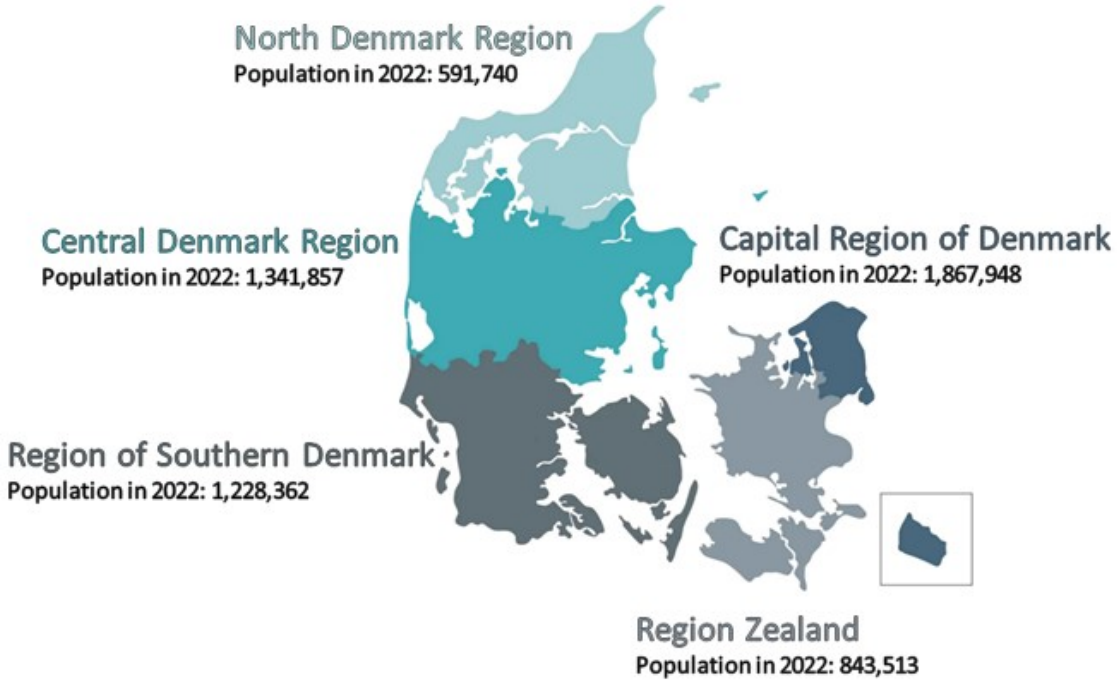


Figure 4.2: The five Danish administrative regions; North Denmark Region, Central Denmark Region, Region of Southern Denmark, Region Zealand, and Capital Region of Denmark[6,66]. Adapted from Nicolaisen et al.[67] and Region Nordjylland[66].

4.1 Administrative and health care databases

The Danish Civil Registration System (DCRS)[7] was established in 1968 and apart from being the original setting of the CPR number, the registry includes vital status and date of death for the entire Danish population. The registry is updated on a daily basis and includes dates of birth and death, immigration/emigration, civil status and place of residence.

The Danish National Patient Registry (DNPR)[68] contains all inpatient discharge diagnoses from all hospitals since 1977 and from emergency room and outpatient specialist clinic contacts since 1995. The registry provides information about primary and secondary diagnoses for both inpatient admissions and outpatient visits and includes both admission and discharge dates. Since 1993, diagnoses have been recorded using the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) codes.

The Danish Register of Medicinal Product Statistics (DRMPS)[69–71] includes complete prescription information from all pharmacies outside hospitals since 1994. Prescriptions are recorded using the Anatomical Therapeutic Chemical (ATC) classification system. In the RDD study, Study IV, the prescriptions were from the Aarhus University Prescription Database (AUPD)[72], which includes data only on reimbursed medications from North Denmark Region and Central Denmark Region.

In addition to these data sources, the studies also use data from socioeconomic registries maintained by Statistics Denmark. These registries include data

on family and household socioeconomics, ethnic origin, educational level, employment status, and income[73–75].

4.1.1 Danish National Health Survey

In the prediction study, Study II, data from the Danish National Health Survey (DNHS)[76] are also included. The DNHS survey includes self-reported information from approximately 300,000 representatively sampled Danes in each of the years 2010, 2013, and 2017. The surveys were implemented as part of a public-health surveillance program in order to monitor the health status over time[76]. Data from the DNHS complement the administrative and health care databases as the surveys facilitate the collection of individual-level self-reported information on variables not covered by the databases. The DNHS includes information on BMI, alcohol consumption, smoking status, and dietary habits, as well as self-rated health, lifestyle, and quality of life.

4.2 Laboratory data

HbA1c measurements are identified from two routine care laboratory databases: the nationwide Register of Laboratory Results for Research (RLRR)[17] and the regional clinical laboratory information system (LABKA) research database at Aarhus University[17,77] including North Denmark Region and Central Denmark Region. When combined, the registries contain virtually all laboratory measurements from both hospitals and general practitioners for the entire Danish population as biological samples are usually analysed in public hospital laboratories and measurement results are therefore recorded in the registries[17]. HbA1c values are identified using the Nomenclature for Properties and Units (NPU) coding system[78], requiring information about the part of the human body under examination, the measured component, relevant kind-of-property, and measurement unit[78,79]. To identify the HbA1c measurements in the registries, the NPU codes NPU03835 and NPU27300 are used, where the first was mainly used before 2014 using the DCCT unit % and the latter is mainly used after 2008 with the IFCC unit mmol/mol. In the current studies, all measurements are converted into mmol/mol and rounded to nearest integer. In the transition period, measurements were recorded using both units and in order to avoid including the same measurement twice, only one measurement per day is included in the studies. If multiple measurements are registered on the same day, the mean of these is included. An individual's region of residence at the time of the measurement is considered a proxy for the region in which the measurement was conducted. Codes for point-of-care measurements are not included, as these cannot be used to diagnose diabetes.

In 3 out of 5 Danish administrative regions (North Denmark Region, Central Denmark Region, and Capital Region of Denmark), laboratory measurements have been available since the 1990s or the mid 2000s. In Region Zealand, measurements became available in 2014 and in Region of Southern Denmark, measurements became available during 2015[17,77]. In the current years, the combined laboratory

databases have full nationwide coverage. The RDD study, Study IV, is restricted to Central Denmark Region and includes only data from LABKA. Additional NPU codes and/or analysis numbers are used in this study, but they represent less than 5% of data.

4.3 Study populations

The studies I, II, and III, are conducted based on the same data extraction hosted at Statistics Denmark. Data include full information (until end of 2018) on all individuals born before 2005 residing in Denmark at some point hereafter. When restricted to residence during the study period 2012–2018, this includes 5,483,467 individuals.

In the descriptive study, Study I, we outlined the use of HbA1c measurements in Denmark after 2012 to justify the use of the registry data. All available measurements from the adult Danish population (ages 20–100 years) in the period 2012–2018 were included for this analysis (Table 4.1).

The prevalence of HbA1c-defined prediabetes was calculated among individuals who were alive and living in Denmark per 31 December 2018 and had at least five years of residence in a Danish region with available laboratory data.

The incidence of HbA1c-defined prediabetes was estimated among individuals with at least five years of residence in a Danish region with available laboratory data during 2012–2018. During the 5-year period, we required no indications of pre-existing prediabetes (HbA1c 42–47 mmol/mol) or diabetes (HbA1c $\geq$ 48 mmol/mol, hospital contact with a diagnosis of diabetes, or redemption of a prescription for glucose-lowering medication).

The 5-year cumulative incidence of progression from incident HbA1c-defined prediabetes to diabetes was estimated among individuals with at least five years of residence in Denmark and at least one year of residence in a Danish region with available laboratory data during 2012–2018. During the 1-year period, we required no indications of pre-existing prediabetes (HbA1c 42–47 mmol/mol) or HbA1c-defined diabetes (HbA1c $\geq$ 48 mmol/mol) and during the 5-year period, we required no hospital contact with a diagnosis of diabetes and no redemption of a prescription for glucose-lowering medication.

In the prediction study[67], Study II, the individuals with incident prediabetes (using a 1-year of look-back for laboratory data) were further restricted to those who had participated in the DNHS survey in at least one of the rounds. In this study, only those with prediabetes after the age of 30 years were included. Individuals were then randomly split into a development sample (80% of the individuals) used for model development and a validation sample (20% of the individuals) used to estimate external model performance.

In the trajectory study, Study III, data were geographically restricted to Central Denmark Region, North Denmark Region, and Capital Region of Denmark. We included individuals with at least five years of residence in these three regions who initiated any glucose-lowering treatment in the years 2017–2018 and had at least one HbA1c measurement in the previous five years.

In the RDD study[80], Study IV, the study period was 2006–2014 and data were geographically restricted to include only Central Denmark Region. The study population comprised those with a first-ever HbA1c measurement in the period 2006–2014, who were 40–80 years of age at the time of the measurement, and had been living in the region for at least one year. The date of the measurement was set as the index date. Individuals who were lost to follow-up during the study period were excluded, as were those with a history of diabetes (hospital contact with a diagnosis of diabetes) and those with any glucose-lowering treatment prior to the index date. As the primary outcome in this study was a composite of death and CVD (including myocardial infarction and stroke), those with a CVD prior to the index date were excluded (Table 4.1).

	Study I				Study II	Study III	Study IV
Study population	HbA1c measurements	Prevalent prediabetes	Incident prediabetes	Incident prediabetes	Incident prediabetes	Treatment initiators	First-ever HbA1c measurement (42–53 mmol/mol)
Study estimate	Description	Prevalence	Incidence	Cumulative incidence	Prediction	Trajectory	Treatment eligibility
Year (index)	2012–2018	2012–2018	2012–2018	2012–2018	2012–2018	2017–2018	2006–2014
Geographical restriction	Denmark	Denmark	Denmark	Denmark	Denmark	Central Denmark Region, North Denmark Region, and Capital Region of Denmark	Central Denmark Region
Residence in Denmark	—	5 years	5 years	5 years	5 years	5 years	1 year
Residence in region with laboratory data	—	5 years	5 years	1 year	1 year	5 years	1 year
Age	20–100 years	20–100 years	20–100 years	20–100 years	≥30 years	—	40–80 years
Number of individuals in the final study population	$N = 4,979,590$ (12,762,320 HbA1c measurements)	$N = 234,056$	$N = 180,923$	$N = 366,752$	$N = 26,007$ ($N = 20,806$)	$N = 20,733$	$N = 43,070$

Table 4.1: Temporal and geographical restrictions used in the study populations. Residence in Denmark ensures availability of look-back data on hospitalizations and prescriptions. The geographical differences in the availability of the laboratory data (Section 4.2) requires additional restrictions to ensure look-back data regarding HbA1c measuresments. Study I includes two different populations with incident prediabetes; one requiring five years of residence in a region with laboratory data and one requiring only one year. To estimate the incidence rate of prediabetes, including as much look-back time as possible was preferred in order to ensure the prediabetes was truly incident. In order to have adequate data during follow-up, only one year of look-back was required in data used to estimate the cumulative incidence. The study population in Study II also included incident prediabetes with one year of look-back. In this study, the population was restricted to those with Danish National Health Survey (DNHS) data. The development sample ($N = 20,806$) constituted 80% of the full population ($N = 26,007$). Study III included everyone who initiated glucose-lowering treatment. Study IV included everyone with a first-ever HbA1c in the interval 42–53 mmol/mol.

4.4 Statistical analyses

Further details about each of the four studies are reported in the full versions of the research papers provided in Appendices I, II, III, and IV. Details about the statistical methods used in each study are outlined in Appendices A, B, C, and D.

4.4.1 Study I: Descriptive study

The descriptive study was conducted to study the occurrence of HbA1c-defined prediabetes in Denmark after the introduction of HbA1c as a diagnostic criterion for diabetes (Appendix I).

Prevalent HbA1c-defined prediabetes per 31 December 2018 was defined as HbA1c in the interval 42–47 mmol/mol during 2012–2018 and no indications of diabetes (HbA1c ≥ 48 mmol/mol, hospital contact with a diagnosis of diabetes, or redemption of a prescription for glucose-lowering medication) during 2012–2018. The prevalence was calculated as the number of individuals with prevalent HbA1c-defined prediabetes per 31 December 2018 divided by the total eligible population per 31 December 2018.

Incident HbA1c-defined prediabetes was defined as the first HbA1c measurement in the interval 42–47 mmol/mol. To estimate the incidence rate of prediabetes, all individuals were followed from the start of follow-up (1 January 2012 or the first date when they fulfilled the inclusion criteria, whichever came last) to the date of incident HbA1c-defined prediabetes, emigration, study end (31 December 2018) or death, whichever came first. The incidence rate was calculated as the number of individuals with incident HbA1c-defined prediabetes per 1,000 person-years at risk.

Both the prevalence and the incidence rates were sex- and age-standardized based on the WHO World 2000–2025 Standard Population Distribution[81,82].

To estimate the 5-year cumulative incidence of progression from incident HbA1c-defined prediabetes to diabetes or death, individuals with incident HbA1c-defined prediabetes were followed from the date they had their first HbA1c measurement in the interval 42–47 mmol/mol to the time of diabetes, emigration, study end (31 December 2018), end of follow-up (maximum five years after incident HbA1c-defined prediabetes), or death, whichever came first. Diabetes during follow-up was defined as either HbA1c-defined diabetes (HbA1c ≥ 48 mmol/mol), hospital contact with a diabetes diagnosis, or redemption of a prescription for glucose-lowering medication.

The 5-year cumulative incidence of progression to diabetes was estimated using the non-parametric estimate of the cause-specific cumulative incidence function with death as a competing event[83] (Appendix A). The 5-year cumulative incidence of death was estimated based on the Kaplan-Meier estimate.

4.4.2 Study II: Prediction study

The prediction model developed and validated in the prediction study[67] is a prognostic model developed to inform about the individual instantaneous risk of

progressing from HbA1c-defined prediabetes to HbA1c-defined diabetes (HbA1c ≥ 48 mmol/mol) within five years[67] (Appendix II). The risk of death within five years was included as a competing risk in the analyses and thus taken into account both in model development and model validation (Appendix B).

The main prognostic prediction model was developed in two steps. First, a Fine-Gray survival model with the least absolute shrinkage and selection operator (LASSO) was fitted in the development sample to perform variable selection among all potential predictors. The tuning parameters were selected using 1,000 iterations and the Bayesian information criterion (BIC)[84]. Second, a Fine-Gray survival model was refitted in the development sample using the selected variables. For comparison purposes, a Fine-Gray survival model including only age and sex (minimum model) was fitted with no variable selection.

Both models (main and minimum) were then applied to the validation sample and 5-year risks were estimated for each individual. The discrimination of the models was assessed using time-dependent receiver operating characteristic (ROCt) curves and the time-dependent area under the curve (AUCt)[85,86]. Both were estimated after five years. The AUCt was estimated using inverse probability of censoring weighting with Kaplan-Meier estimated weights. The calibration of the models was visually assessed using the calibration curves. In addition, the calibration was assessed using the Brier score and a scaled version of the Brier score, the index of prediction accuracy (IPA), was used to consider calibration and discrimination simultaneously[87] (Appendix B).

4.4.3 Study III: Trajectory study

The analyses in the trajectory study combined elements from mixed-effects modelling, spline regression, latent class analyses, and stochastic smoothing (Appendix C, Appendix III). We used the latent class trajectory analyses to identify distinct trajectories of HbA1c in the five years prior to treatment initiation[88] (Appendix C).

The trajectory models were fitted using a three-step approach. In the first step, standard linear mixed-effects models were fitted to find the best model parametrization. In the second step, the best performing model parametrization was then used to fit latent class models with 2, 3, 4, and 5 latent classes. In the third step, the trajectories were smoothed in each of the latent classes defined by the best performing model. HbA1c was modelled as a function of days before treatment initiation and time 0 was the day of treatment initiation.

Standard linear mixed-effects models (i.e., 1-class models without latent classes) including a random intercept and a random slope (time) were fitted with an unspecified variance-covariance structure to account for repeated measurements from the same individual and to allow measurements to be correlated as a function of time (Appendix C). One model was fitted with a fixed linear, squared, and cubic time term. Additional models were specified using restricted cubic splines[89] (Appendix C). The knots were either determined based on quantiles of data or were determined to split time in equally-spaced intervals. Both options were fitted with 3, 4, and 5 knots. To identify the best performing model parametrization, the seven standard linear mixed-effects models were then

compared using a visual comparison of the mean curves, the maximum value of the log-likelihood function, the BIC, and the Akaike information criterion (AIC).

In the second step, the best performing model parametrization was used to fit latent class models with 2, 3, 4, and 5 latent classes. To ease the computation, these models were fitted based on a random third of the individuals. Class-membership probabilities were assumed not to depend on any additional covariates and we allowed covariates to vary across latent classes. To find the optimal latent class model, the models were compared using the BIC and the AIC. The size of the classes, the number of parameters in the model, the clinical relevance of the classes, and the individual posterior class-membership probabilities were also taken into account. We did not further increase the number of classes if at least one class included less than 1% of the individuals.

Based on the optimal latent class model, the posterior class-membership probabilities were calculated for all individuals and individuals were then assigned to the latent class with the highest posterior class-membership probability[88] (Appendix C). The mean trajectories were refitted using non-parametric stochastic smoothing in each of the classes defined by the best performing model. The trajectories were smoothed using a random walk of order 2 with rounded month as time unit and individual random intercepts[90,91].

4.4.4 Study IV: Regression discontinuity design study

In the RDD study[80], we used the regression discontinuity design to estimate the effect of being *eligible* for type 2 diabetes treatment on CVD and death[80] (Appendix D, Appendix IV).

According to the treatment guidelines (Section 2.3), glucose-lowering treatment should be initiated if an individual has HbA1c $\geq$ 48 mmol/mol. If the HbA1c is 47 mmol/mol or below, the individual does not have diabetes and should not initiate glucose-lowering treatment. As with every other measurement, the HbA1c measurements will inevitably be affected by measurement errors and other factors that can potentially result in small changes from the actual to the measured value. These small changes can also occur close to the threshold, HbA1c 48 mmol/mol, and can result in individuals being "randomly" assigned to have values either above or below the threshold. This is exactly what the RDD uses as the pseudo-random treatment allocation where the threshold acts as a randomizing device for individuals close to the threshold. Consequently, individuals with values just below and individuals with values just above the threshold are expected to be similar. The estimate from the RDD is the *local* effect of being eligible for treatment.

Some assumptions must be checked for the estimates to be reliable. First, to ensure that neither patients nor doctors deliberately changed the value of the HbA1c measurement (and hereby changed treatment allocation group), histograms of the HbA1c values from study population were checked to warrant that there was no discontinuity around the threshold. Second, to check that individuals with values just below and individuals with values just above the threshold were in fact similar, the distribution of covariates were compared. In

our main analysis, we included individuals with HbA1c 42–53 mmol/mol, i.e., we compared the groups 42–47 mmol/mol (prediabetes, not eligible for treatment) and 48–53 mmol/mol (type 2 diabetes, eligible for treatment).

The effect of treatment eligibility on CVD and death was estimated based on a Cox regression model with the following parametrization (Appendix D):

$$log(h(y|X)) = \beta_0 + \beta_1(X - x_0) + \beta_2 \mathbb{1}_{(X<x_0)} + \beta_3 \mathbb{1}_{(X>x_0)}(X - x_0) \tag{4.1}$$

where X denotes the vector of HbA1c values, Y the vector of outcomes, x_0 denotes the threshold (48 mmol/mol), β_1 is the slope of the line below the threshold, $\beta_1 + \beta_3$ is the slope of the line above the threshold, and β_2 is the difference in intercepts (discontinuity) in the threshold, i.e., the desired estimator of the intention to treat (ITT) effect of treatment eligibility on CVD or death (Appendix D).

4.5 Additional information

The Danish registry data used in the four studies are available to researchers from research environments pre-approved by Statistics Denmark upon project approval by the Danish Data Protection Agency and Statistics Denmark. Researchers can apply for access to data after their request is approved by the Danish Data Protection Agency (https://www.datatilsynet.dk). The participants in the DNHS gave informed consent for their survey responses to be used in research. No additional ethical approval was needed for these studies. All data originated from registries and/or health surveys and none were specifically collected for these studies.

All statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc, Cary, North Carolina) and R version 4.0.2 (R Core Team, 2020).

5 | Results

The main findings from the four studies are presented in this chapter. More details are presented in the full versions of the research papers provided in Appendices I, II, III, and IV.

5.1 Study I: Descriptive study

A total of 4,979,590 adult Danes (ages 20–100 years) were living in Denmark in a region with available laboratory data during 2012–2018 (Table 4.1, Appendix I). For these adult Danes, a total of 12,762,320 HbA1c measurements (Figure 5.1) from 3,036,866 (61.0%) distinct individuals were available in the laboratory database in the years 2012–2018. We estimated that 70.8% (95% confidence interval (CI) 70.8–70.9) of the entire adult Danish population had been measured at least once during the study period, 2012–2018. For people above 70 years of age, more than two-thirds had at least one HbA1c measurement in 2018 (Appendix I).

Among the 3,302,759 adult Danes (ages 20–100 years) alive and living in Denmark per 31 December 2018, a total of 234,056 individuals were identified with prediabetes during 2012–2018 without any indication of diabetes (Table 4.1). This yields an overall prevalence of 7.1% (95% CI 7.1–7.1) among adult Danes, whereas the sex- and age-standardized prevalence was 4.4% (95% CI 4.3–4.4).

During 2012–2018, 3,226,748 individuals were at risk of incident HbA1c-defined prediabetes and contributed more than 12 million person-years at risk (Appendix I). A total of 180,923 (5.6%) individuals were identified with incident HbA1c-defined prediabetes during the time at risk. The overall incidence rate of HbA1c-defined prediabetes in Denmark in the period 2012-2018 was 14.2 (95% CI 14.1–14.3) per 1,000 person-years. When stratified by age (Figure 5.2), the incidence for women were generally lower than for men. The overall sex- and age-standardized incidence rate was 10.4 (95% CI 10.4–10.5) per 1,000 person-years (Appendix I).

Among the 180,923 individuals with incident HbA1c-defined prediabetes in the years 2012–2018 (Table 4.1), 93,790 (51.8%) were women, the median age was 66.9 years (interquartile range (IQR) 56.7–75.7), and, based on the Charlson Comorbidity Index (CCI) score, 52,629 (29.1%) already had comorbidities at the time of the prediabetes diagnosis (Table 5.1, Appendix I). The median HbA1c was 43 mmol/mol (IQR 42–44) and 12,875 (7.1%) of the individuals were measured during an inpatient hospitalization (or in the emergency department), 46,270 (25,6%) were followed in the outpatient hospital specialist clinics at the time of

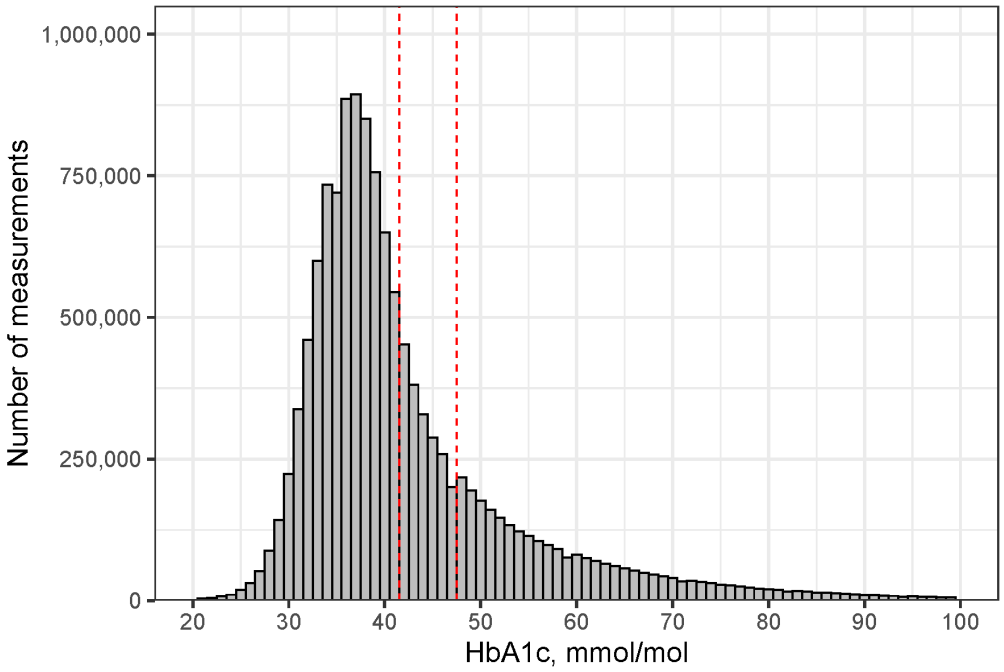


Figure 5.1: Histogram showing the distribution of the $N = 12,762,320$ HbA1c measurements available in the laboratory database in the years 2012–2018. The red dashed lines indicate the interval for prediabetes, 42–47 mmol/mol, (Appendix I).

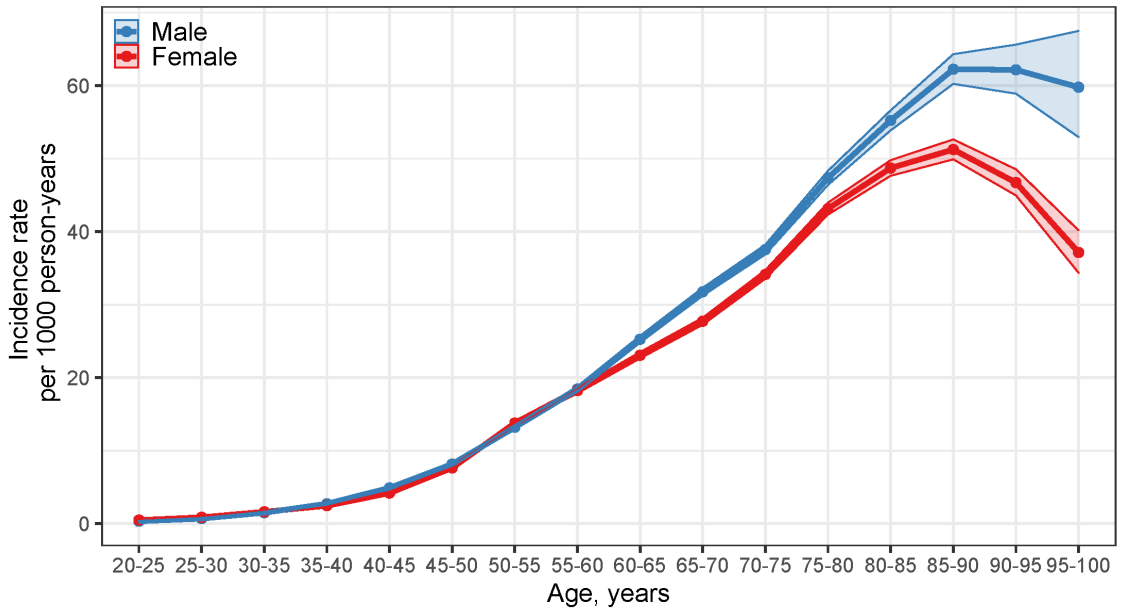


Figure 5.2: Incidence rates of HbA1c-defined prediabetes in the adult Danish population 2012–2018 per 1,000 person-years by age and sex (Appendix I).

the prediabetes diagnosis, and 121,778 (67.3%) were most likely measured at the general practitioner (Appendix I). At the time of the prediabetes diagnosis, 64,015 (35.4%) were living alone and 66,477 (36.7%) had no education or had basic education or primary school as highest achieved education. A total of 118,907 (65.7%) were unemployed or not part of the workforce (also including retirement).

Despite the slight differences in definition of incident prediabetes (Table 4.1), the baseline characteristics were generally similar (Table 5.1).

During a median of 2.7 years (IQR 1.6–4.5) of follow-up among the 366,752 individuals with incident prediabetes based on one year of look-back (Table 4.1), 41,350 (11.3%) died and it was observed that 49,855 (13.6%) developed diabetes. Among the 75,907 (20.1%) individuals who were followed for all 5 years, 18,359 (24.2%) developed diabetes during their follow-up. The estimated overall 5-year cumulative incidence for death was 17.5% (95% CI 17.3–17.7) and the estimated overall 5-year cumulative incidence for diabetes was 21.3% (95% CI 21.1–21.5) (Figure 5.3, Appendix A).

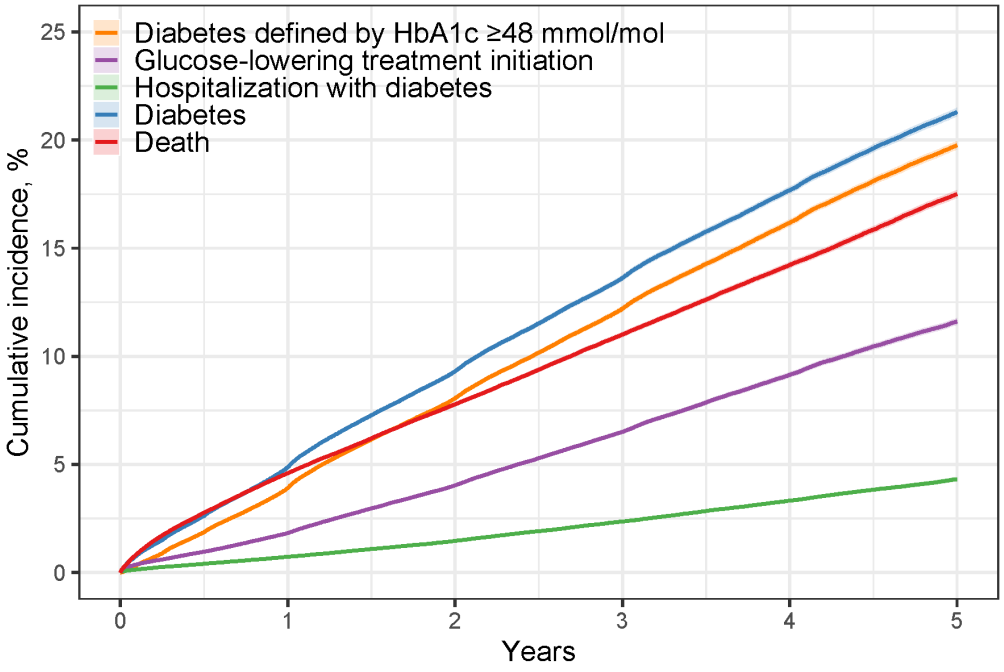


Figure 5.3: Cumulative incidences for progression from HbA1c-defined prediabetes to diabetes or death. Diabetes is defined as a composite outcome defined using either HbA1c ≥ 48 mmol/mol, hospital contact with a diagnosis of diabetes, or redemption of a prescription for glucose-lowering medication. The analyses are based on $N = 366,757$ individuals with incident HbA1c-defined prediabetes using the less restrictive definition with one rather than five years of look-back for laboratory data (Table 4.1). Adapted from Appendix I.

Variable, N (%) or Median (IQR)	Study I (5 years of look-back)	Study I (1 year of look-back)	Study II (1 year of look-back, DNHS data)
Total	180,923 (100.0%)	366,752 (100.0%)	20,806 (100.0%)
Sex			
Female	93,790 (51.8%)	191,877 (52.3%)	10,792 (51.9%)
Male	87,133 (48.2%)	174,875 (47.7%)	10,014 (48.1%)
Age at index	66.9 (56.7-75.7)	67.8 (57.9-76.3)	69.6 (61.0-77.1)
Body mass index (kg/m^2)	—	—	26.7 (24.1-29.8)
Prediabetes-defining HbA1c	43.0 (42.0-44.0)	43.0 (42.0-44.0)	43.0 (42.0-44.0)
Prescription drug use			
Any hypolipidemic treatment	58,415 (32.3%)	127,206 (34.7%)	7,808 (37.5%)
Loop-diuretics	18,540 (10.2%)	39,844 (10.9%)	2,229 (10.7%)
Any potential antihypertensive drug	97,977 (54.2%)	210,522 (57.4%)	12,591 (60.5%)
Beta-blockers	40,764 (22.5%)	86,687 (23.6%)	5,175 (24.9%)
ACE inhibitors or ATII antagonists	64,111 (35.4%)	138,194 (37.7%)	8,371 (40.2%)
Oral steroids	12,703 (7.0%)	24,394 (6.7%)	1,505 (7.2%)
Antibiotics	46,599 (25.8%)	94,979 (25.9%)	5,270 (25.3%)
Antidepressants	25,930 (14.3%)	52,776 (14.4%)	2,569 (12.3%)
Comorbidities			
Charlson Comorbidity Index score			
0	128,294 (70.9%)	255,318 (69.6%)	14,437 (69.4%)
1–2	42,538 (23.5%)	90,072 (24.6%)	5,107 (24.5%)
≥3	10,091 (5.6%)	21,362 (5.8%)	1,262 (6.1%)
Cardiovascular disease	36,877 (20.4%)	78,091 (21.3%)	4,530 (21.8%)
Hypertension	27,513 (15.2%)	62,974 (17.2%)	3,561 (17.1%)
Cancer	12,530 (6.9%)	25,644 (7.0%)	1,692 (8.1%)
Kidney disease	2,771 (1.5%)	6,206 (1.7%)	373 (1.8%)
Liver disease	1,153 (0.6%)	2,516 (0.7%)	93 (0.4%)
Dementia	2,703 (1.5%)	5,839 (1.6%)	212 (1.0%)
Pancreatic disease	585 (0.3%)	1,271 (0.3%)	67 (0.3%)
Admission to an intensive care unit	6,401 (3.5%)	13,266 (3.6%)	719 (3.5%)
Socioeconomic variables			
Highest education achieved			
None, basic education, or primary school	66,477 (36.7%)	137,452 (37.5%)	7,209 (34.6%)
Youth education, high school or similar	72,835 (40.3%)	147,890 (40.3%)	8,687 (41.8%)
Higher education	36,833 (20.4%)	71,186 (19.4%)	4,485 (21.6%)
Employment status			
Employed	62,016 (34.3%)	116,466 (31.8%)	6,398 (30.8%)
Unemployed or not part of the workforce	118,907 (65.7%)	250,286 (68.2%)	14,408 (69.2%)
Type of household			
Living alone	64,015 (35.4%)	131,263 (35.8%)	6,327 (30.4%)
Not living alone	116,889 (64.6%)	235,455 (64.2%)	14,479 (69.6%)

Table 5.1: Baseline characteristics of all the individuals in the three populations with incident HbA1c-defined prediabetes. The two populations from Study I are defined based on five years of look-back ($N = 180,923$) in the laboratory data for the incidence analysis and based on one year of look-back ($N = 366,752$) for the progression analysis. The development sample from Study II ($N = 20,806$) is also defined using one year of look-back and is restricted to individuals with Danish National Health Survey (DNHS) data. Prescription drug use is defined as prescription redemption during the prior 180 days. Comorbidities are based on hospital contacts the prior five years. Adapted from Appendix I, Appendix II, and Nicolaisen et al.[67].

5.2 Study II: Prediction study

The prediction study, Study II, was based on the 486,495 individuals from the DNHS survey[76] among whom 26,007 (5.3%) were identified as having incident HbA1c-defined prediabetes[67] (Appendix II). The 26,007 individuals were randomly divided into a development sample ($N = 20,806$) and a validation sample ($N = 5,201$).

The prediction model was developed using the development sample. Components from 11 of the potential predictors were selected based on the LASSO regression; age (years), sex (female vs. male), prediabetes-defining HbA1c measurement (mmol/mol), BMI (kg/m^2), any antihypertensive drug (Yes vs. No), pancreatic disease (Yes vs. No), cancer (Yes vs. No), unhealthy diet (Overall diet: Unhealthy vs. Average or Healthy), doctor's advice to lose weight or change dietary habits (Advised by general practitioner to lose weight or change dietary habits during the past 12 months: Yes vs. No), not having anyone to talk to when in need of support (Availability of someone to talk to if problems occur or support is needed: No, never or almost never vs. Yes, often, mostly, or sometimes), and good self-rated health (Overall self-rated health: Good vs. Fair/poor or Excellent/very good)[67].

The prediction model was applied in the validation sample along with the minimum model including only sex and age. Comparing the estimated probabilities from the two models showed that the main model assigned higher probabilities to a large subgroup of the individuals who progressed to HbA1c-defined diabetes without overestimating the probabilities for those without the outcome (Figure 5.4). The main model had a better discriminative ability than the minimum model indicated by a higher AUCt (72.7 (95% CI 71.2–74.3) in the main model vs. 68.2 (95% CI 66.7–69.7) in the minimum model) (Figure 5.4). The calibration curves generally showed good calibration for both models (Figure 5.4). Compared to the minimum model, the main model had a lower Brier score (10.7 (95% CI 8.8–12.6) for the main vs. 12.8 (95% CI 10.7–14.8) for the minimum model) and a higher IPA (18.2 for the main model vs. 2.8 for the minimum model) indicating better overall performance when calibration was taken into consideration[67] (Appendix II).

5.3 Study III: Trajectory study

In 2017–2018, 2,964,032 individuals had lived in Central Denmark Region, North Denmark Region, and/or Capital Region of Denmark for at least five years. A total of 20,733 individuals initiated their first-ever glucose-lowering treatment while having at least one HbA1c measurement in the prior five years (Appendix III). The individuals had a median of 4 measurements (IQR 2–7) in the five years prior to treatment initiation and contributed a total of 105,211 measurements (Figure 5.5).

The trajectory models were fitted in three steps. In the first step, all model types were fitted as standard linear mixed-effects models (i.e., 1-class models). Based on the BIC, the AIC, and the maximum value of the log-likelihood function, the best model parametrization was the restricted cubic splines model with five

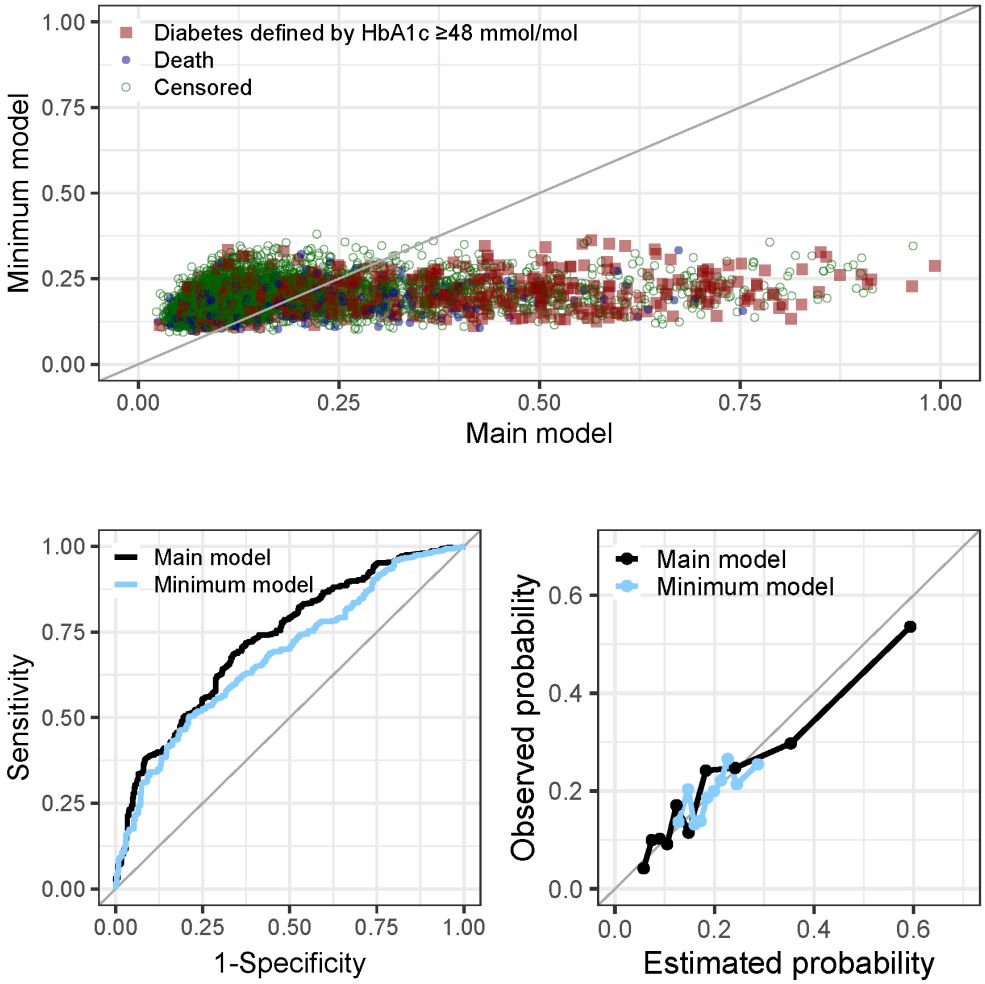


Figure 5.4: The upper plot is a comparison of the estimated probabilities of HbA1c-defined diabetes from the two prediction models. The graph is colored by observed outcome: HbA1c-defined diabetes (outcome), death, or censored (i.e., emigration, study end (31 December 2018), or end of follow-up (five years after index date)). The lower plots show the ROCt curves comparing the discriminative ability of the main model and the minimum model and the calibration curves comparing the estimated and observed probabilities for the two models. The estimates for the observed probabilities were defined based on quantiles of the estimated probabilities. The main model included the following predictors: age, sex, HbA1c, BMI, any antihypertensive drug, pancreatic disease, cancer, unhealthy diet, doctor's advice to lose weight or change dietary habits, not having anyone to talk to when in need of support, and good self-rated health. The minimum model included only age and sex. Adapted from Nicolaisen et al.[67] (Appendix II).

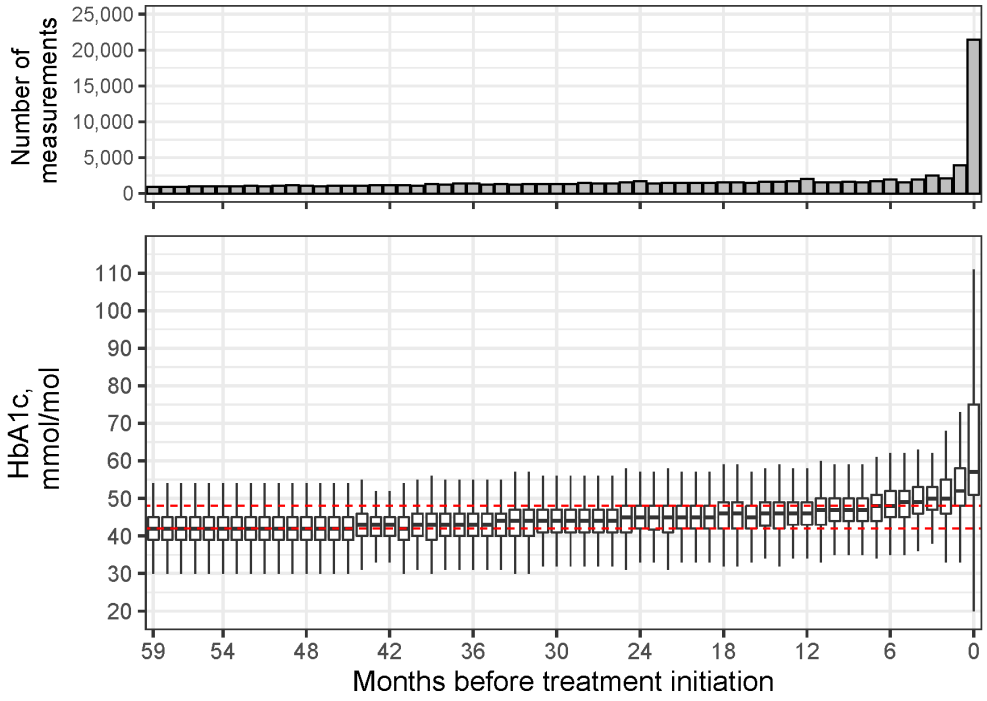


Figure 5.5: The upper plot is a histogram of all $N = 105,211$ measurements. The majority of the measurements are conducted in the month prior to treatment initiation. The lower plot shows boxplots for all measurements in each month. The red dashed lines indicate the interval for prediabetes, 42–47 mmol/mol. Adapted from Appendix III.

knots placed based on quantiles of data. However, as the mean curves showed a similar fit to data, the restricted cubic splines with only three knots placed based on quantiles of data was preferred in order to keep the number of parameters low (Appendix III).

In the second step, the 3-knot model was fitted with 2, 3, 4, and 5 latent classes. Based on the BIC and the AIC, the 5-class model showed the best fit. As one of the classes in the 5-class model included only 0.7% of the individuals the 4-class model was preferred (Appendix III). Individuals were assigned to the class with the largest posterior probability. The four classes were; Class 1, 15,283 (73.7%) individuals with 82.4% of the measurements; Class 2, 2,862 (13.8%) individuals with 11.1% of the measurements; Class 3, 2,164 (10.4%) individuals with 5.6% of the measurements; and Class 4, 424 (2.1%) individuals with 1.0% of the measurements (Table 5.2, Appendix III).

In the third step, all measurements in each of the latent classes were used to separately smooth the trajectories for the latent classes (Figure 5.6). The smoothed trajectory for the largest class (Class 1) had slowly increasing levels of HbA1c during the entire 5-year period, yet with slightly faster increasing levels in the last year up to first-ever glucose-lowering treatment initiation. Based on the parametrization, the mean trajectory for Class 1 exceeded the threshold for diabetes (48 mmol/mol) 8.8 months before treatment was initiated and the mean HbA1c was 51.9 mmol/mol (95% CI 51.6–52.2) at the time of treatment initiation (Appendix III). The smoothed trajectories for Class 2 and Class 3 had

slowly increasing levels of HbA1c until 1 to 1.5 years prior to treatment initiation, where the HbA1c levels started to increase much more rapidly than for Class 1 (Figure 5.6). The mean trajectory for Class 2 exceeded 48 mmol/mol 16.7 months prior to treatment initiation and reached 78.8 mmol/mol (95% CI 77.7–79.8) at the time of treatment initiation (Appendix III). The mean trajectory for Class 3 exceeded 48 mmol/mol 16.2 months prior to treatment initiation and reached 104.7 mmol/mol (95% CI 103.5–105.9) at the time of treatment initiation (Appendix III). For Class 1 and Class 2, HbA1c levels in the smoothed trajectories were above 42 mmol/mol (i.e., the threshold for prediabetes) for most of the five years prior to treatment initiation (Figure 5.6). The levels were slightly lower for Class 3 (Figure 5.6). Class 4 included only 2% of the individuals. The smoothed trajectory had more unstable levels of HbA1c, tended to be lower than in the other classes, and had a very steep increase prior to treatment initiation. The mean HbA1c level at the time of treatment initiation was 137.0 mmol/mol (95% CI 134.7–139.2) and the levels exceeded 48 mmol/mol 15.1 months prior to treatment initiation (Figure 5.6, Appendix III).

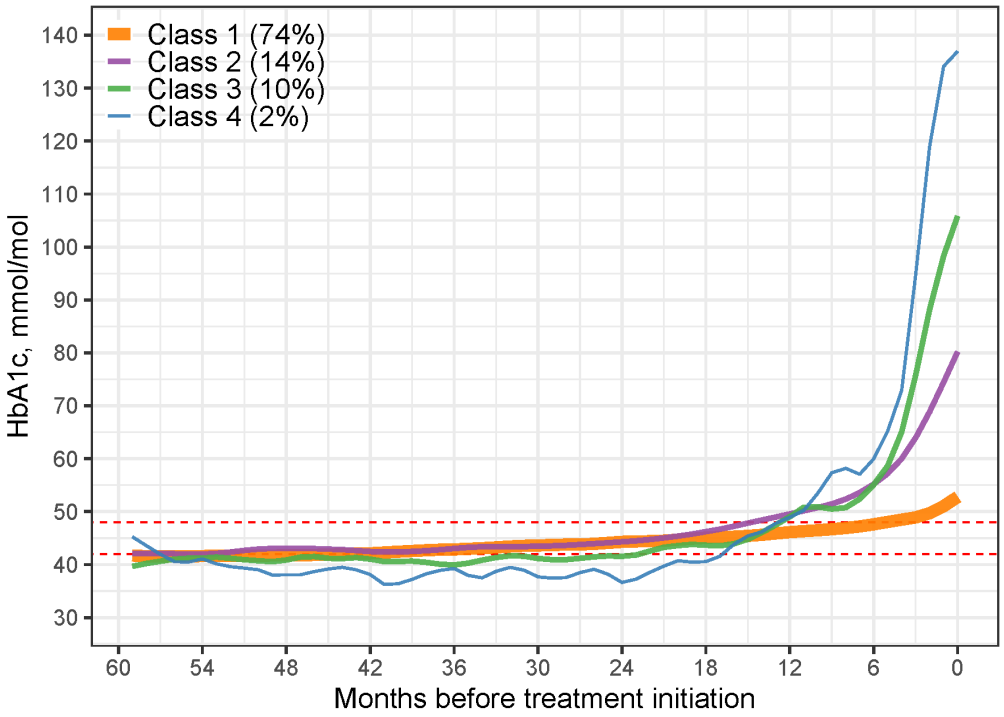


Figure 5.6: Smoothed trajectories in each of the latent classes. The thickness of the trajectories represents the number of individuals assigned to the latent class, i.e., as 74% of the individuals are assigned to Class 1, the line representing the smoothed trajectory for Class 1 (orange) is thicker than the lines for Class 2 (purple) and Class 3 (green) with 14% and 10% of the individuals. Class 4 (blue) only includes 2% of the individuals. The red dashed lines indicate the threshold for prediabetes, i.e., HbA1c ≥ 42 mmol/mol, and the threshold for diabetes, i.e., HbA1c ≥ 48 mmol/mol. Adapted from Appendix III.

5.3.1 Baseline characteristics

The individuals in Class 1 had HbA1c measured more often than those in the other three classes and had a median of 5 measurements (IQR 3–8) during the five years before treatment (Table 5.2, Appendix III). In comparison, Class 2 had median 3 measurements (IQR 2–5). Class 3 and Class 4 both had only median 2 measurements (IQR 1–3) (Table 5.2) and 34.3% in Class 3 and 38.9% in Class 4 had only 1 measurement during the five years. In all classes, the most recent HbA1c measurement was conducted close to the time of treatment initiation (between median 4 and 11 days before, (Appendix III)). Class 1 included more women than the other classes and the median age decreased from Class 1 to Class 4 (61.6 years (IQR 50.8–71.4) in Class 1, 57.3 years (IQR 47.8–67.6) in Class 2, 54.5 years (IQR 44.8–63.7) in Class 3, and 52.6 years (IQR 34.0–63.2) in Class 4). Of note, Class 4 included the largest proportion of individuals below 30 years of age (5.6% in Class 1, 4.0% in Class 2, 8.6% in Class 3, and 20.5% in Class 4). Individuals in Class 4 also had higher prevalence of low education, low income, and the highest prevalence of being unmarried. Individuals from Class 3 and Class 4 had much more often received a hospital diagnosis of diabetes (31.3% and 62.0% in Class 3 and 4, versus only 8.1% in Class 1 and 18.4% in Class 2). In particular Class 4 stood out with a high use of insulin monotherapy (40.6%) or any polytherapy initiation (24.8%) whereas almost all (94.3%) in the large Class 1 started non-insulin monotherapy (typically with metformin). Pre-existing cardiovascular disease or cardiovascular drug use was much more common in Class 1 than in the smaller classes, whereas Class 4 had more alcoholism-related diagnoses.

Variable, N (%) or Median (IQR)	All individuals	Class 1	Class 2	Class 3	Class 4
Total	20,733 (100.0%)	15,283 (100.0%)	2,862 (100.0%)	2,164 (100.0%)	424 (100.0%)
Sex					
Female	9,330 (45.0%)	7,561 (49.5%)	940 (32.8%)	685 (31.7%)	144 (34.0%)
Male	11,403 (55.0%)	7,722 (50.5%)	1,922 (67.2%)	1,479 (68.3%)	280 (66.0%)
Age at index	60.0 (49.6-70.3)	61.6 (50.8-71.4)	57.3 (47.8-67.6)	54.5 (44.8-63.7)	52.6 (34.0-63.2)
Days since last HbA1c measurement	8 (4-22)	11 (6-32)	6 (2-11)	4 (1-9)	7 (2-13)
Number of HbA1c measurements	4 (2-7)	5 (3-8)	3 (2-5)	2 (1-3)	2 (1-3)
Socioeconomic variables					
Marital status					
Married	10,519 (50.7%)	7,977 (52.2%)	1,406 (49.1%)	989 (45.7%)	147 (34.7%)
Divorced	3,400 (16.4%)	2,517 (16.5%)	499 (17.4%)	315 (14.6%)	69 (16.3%)
Widow/widower	1,972 (9.5%)	1,601 (10.5%)	219 (7.7%)	136 (6.3%)	16 (3.8%)
Unmarried	4,842 (23.4%)	3,188 (20.9%)	738 (25.8%)	724 (33.5%)	192 (45.3%)
Highest education achieved					
None, basic education, or primary school	6,976 (33.6%)	5,167 (33.8%)	932 (32.6%)	711 (32.9%)	166 (39.2%)
Youth education, high school, or similar	8,893 (42.9%)	6,508 (42.6%)	1,256 (43.9%)	953 (44.0%)	176 (41.5%)
Higher education	4,253 (20.5%)	3,145 (20.6%)	586 (20.5%)	454 (21.0%)	68 (16.0%)
Income					
Lowest income group	3,258 (15.7%)	2,443 (16.0%)	398 (13.9%)	320 (14.8%)	97 (22.9%)
Low-to-medium income	7,427 (35.8%)	5,757 (37.7%)	934 (32.6%)	616 (28.5%)	120 (28.3%)
Medium-to-high income	5,473 (26.4%)	4,017 (26.3%)	744 (26.0%)	605 (28.0%)	107 (25.2%)
Highest income group	4,404 (21.2%)	2,991 (19.6%)	752 (26.3%)	581 (26.8%)	80 (18.9%)
Comorbidities and prescription drug use					
Statins	6,396 (30.8%)	5,500 (36.0%)	575 (20.1%)	272 (12.6%)	49 (11.6%)
Any potential antihypertensive drug	10,527 (50.8%)	8,559 (56.0%)	1,209 (42.2%)	637 (29.4%)	122 (28.8%)
Diabetes	2,709 (13.1%)	1,243 (8.1%)	526 (18.4%)	677 (31.3%)	263 (62.0%)
Cardiovascular disease	3,730 (18.0%)	2,968 (19.4%)	479 (16.7%)	241 (11.1%)	42 (9.9%)
Alcoholism-related diagnosis or medication	499 (2.4%)	330 (2.2%)	91 (3.2%)	48 (2.2%)	30 (7.1%)
Admission to an intensive care unit	849 (4.1%)	611 (4.0%)	132 (4.6%)	68 (3.1%)	38 (9.0%)
First-ever glucose-lowering treatment					
Monotherapy	19,810 (95.5%)	15,058 (98.5%)	2,641 (92.3%)	1,792 (82.8%)	319 (75.2%)
Non-insulin based monotherapy	18,400 (88.7%)	14,407 (94.3%)	2,407 (84.1%)	1,439 (66.5%)	147 (34.7%)
Insulin based monotherapy	1,410 (6.8%)	651 (4.3%)	234 (8.2%)	353 (16.3%)	172 (40.6%)
Polytherapy	923 (4.5%)	225 (1.5%)	221 (7.7%)	372 (17.2%)	105 (24.8%)

Table 5.2: Baseline characteristics of the individuals in the four classes. Comorbidities are based on hospital contacts the prior five years. Prescription drug use is defined as prescription redemption during the prior 180 days. The population is defined based on first-ever glucose-lowering treatment and the treatment includes the prescriptions at index date and up to 14 days after the index date to ensure the full baseline glucose-lowering treatment regimen is captured. Adapted from Appendix III.

5.4 Study IV: Regression discontinuity design study

Among the 1,641,615 individuals living in Central Denmark Region during 2005–2014, 654,208 had their first-ever HbA1c measurement during 2006–2014 and 290,333 complied with the inclusion criteria, i.e., they were 40–80 years old, had no prior CVD, no prior glucose-lowering treatment, and no prior diabetes hospitalizations[80] (Appendix IV). A total of 43,070 had their first-ever HbA1c in the interval 42 mmol/mol to 53 mmol/mol; 36,360 (84.4%) in the interval 42–47 mmol/mol, i.e., below the threshold for diabetes (prediabetes) and therefore not eligible for treatment, and 6,710 (15.6%) in the interval 48–53 mmol/mol, i.e., above the threshold for diabetes and therefore eligible for treatment. Despite some differences in the use of medications (e.g., 24.3% receiving diuretics among those below the threshold vs. 32.0% among those above the threshold and 46.7% with antihypertensive drugs among those below and 53.8% among those above the threshold), the two groups were in general comparable and there were no signs of discontinuity in the HbA1c values around the threshold[80] (Appendix IV).

The overall median follow-up time was 4.4 years (IQR 2.3–6.9); 4.2 years (IQR 2.2–6.8) for those below the threshold and 5.4 years (IQR 2.8–7.4) for those above the threshold. CVD and death were in general more frequent among those above the threshold with 27.5 (95% CI 26.7–28.3) outcomes per 1,000 person-years in those with HbA1c 42–47 mmol/mol and 37.2 (95% CI 35.1–39.3) outcomes per 1,000 person-years for those with HbA1c 48–53 mmol/mol. However, when the hazard ratio (HR) was estimated using the parametrization in Equation (4.1) (Appendix D), and thus estimating the discontinuity in the threshold, individuals with a first HbA1c just above the threshold had on average 21% lower rate of death or CVD during follow-up compared to those with a first HbA1c just below the threshold (HR 0.79 (95% CI 0.69–0.90)) (Figure 5.7).

Among the 43,070 individuals, 13,406 (31.1%) were included during the years 2012–2014, whereof 12,056 (89.9%) had values below the threshold and 1,350 (10.1%) had values above the threshold. Restricting the analyses to the individuals from 2012–2014 showed a similar result (HR 0.78 (95% CI 0.54–1.12)), but the analysis was limited by only 512 outcomes (Appendix IV).

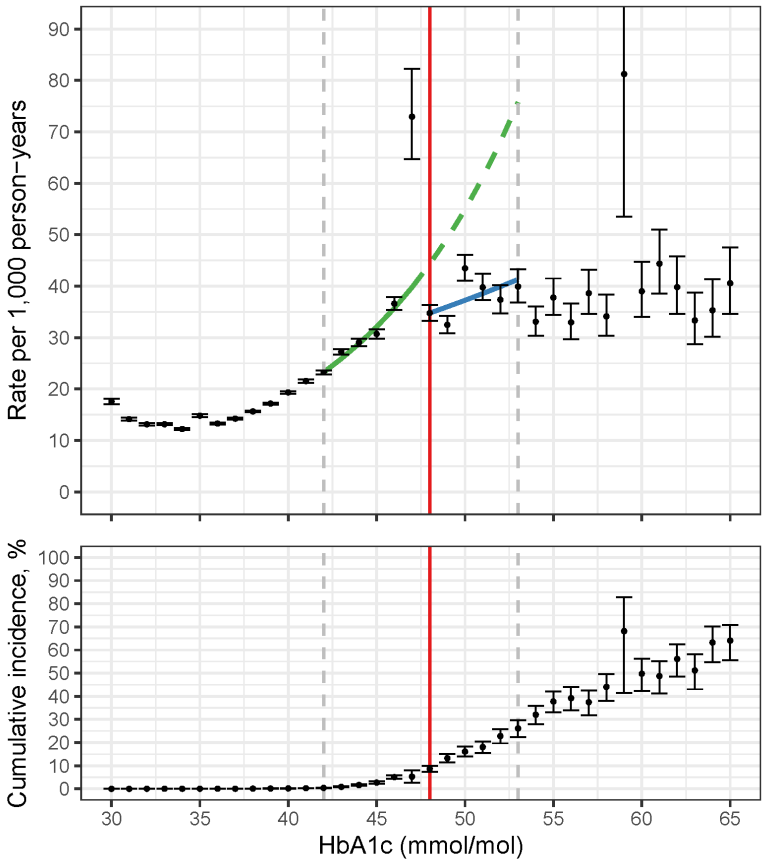


Figure 5.7: Regression discontinuity for cardiovascular disease (CVD) or death according to first-ever HbA1c measurement. In the upper plot, the dots show the incidence rate per 1,000 person-years of CVD or death for each value of first-ever HbA1c. The red line indicates the HbA1c threshold of 48 mmol/mol whereas the range used for the RDD analysis is indicated by the grey dashed lines, 42–53 mmol/mol. The green line shows the estimated RDD regression line (hazard rate) below the threshold and the blue line shows the estimated RDD regression line above the threshold. The difference between the green and blue line in the threshold is the effect of threshold. The green dashed line is a hypothetical expansion of the regression below the threshold and is a representation of what would have happened had the threshold not existed (Appendix D). The baseline hazard is unknown but the lines are plotted such that the regression fits with the incidence rate in the threshold, 48 mmol/mol. The lower plots shows the 3-month cumulative incidence of initiating any glucose-lowering treatment. Adapted from Petersen et al.[80] (Appendix IV).

6 | Discussion

HbA1c testing in the general Danish population increased after the introduction of HbA1c as a tool for diagnosing diabetes[12]. We found that more than 70% of the entire Danish adult population had at least one measurement during the years 2012–2018. For people above 70 years of age, more than two-thirds had at least one measurement of HbA1c in 2018. Laboratory databases with complete population coverage therefore offer a valuable tool to study the epidemiology of prediabetes. More than 1.4 out of 100 adult Danes were identified with incident prediabetes each year and they had values of HbA1c in the lower end of the range of prediabetes. Within five years after the diagnosis of prediabetes, more than 1 in 5 progressed to diabetes and almost 1 in 6 died. In addition to well-known predictors (e.g., age, sex, and BMI), self-rated lifestyle, health, and social support were important and modifiable predictors for type 2 diabetes. Although our model was well calibrated and identified some individuals with prediabetes who were at high risk of progression, the discriminative ability was modest as AUCt was only 73 for HbA1c-defined type 2 diabetes. Prior to first-ever glucose-lowering treatment initiation, levels of HbA1c were slowly increasing for a long time before more rapidly increasing levels prompted the initiation of treatment. We identified four distinct trajectories and they had mean HbA1c levels above 48 mmol/mol around 9 to 17 months before treatment was initiated. Finally, we found that individuals with HbA1c just above the diabetes diagnosis threshold of 48 mmol/mol, and thus eligible for glucose-lowering treatment, had a lower rate of CVD and death compared to those whose HbA1c was just below the treatment threshold.

6.1 Prediabetes epidemiology (Study I)

Our studies are among the first to examine the occurrence and prognosis of prediabetes based on HbA1c measurements from laboratory databases with complete population coverage. Direct comparison of our findings with previous studies is therefore challenging, given the use of varying thresholds and definitions of prediabetes (Table 2.1).

6.1.1 Prevalence of prediabetes

Using HbA1c 42–47 mmol/mol as the definition of prediabetes, we estimated the overall prevalence in Denmark to be 7%. This is in line with the prevalence

estimated in a previous Danish study, which also reported that 7% of the entire adult Danish population (aged 20–100 years) had prediabetes (42–47 mmol/mol)[33]. In December 2022, 4,637,430 individuals aged 20-100 years were living in Denmark[6] and a prevalence of 7% indicates that almost 330,000 adult Danes currently have HbA1c-defined prediabetes. Our sex- and age-standardized prevalence of 4% was lower than the worldwide prevalences (10.6% for IGT and 6.2% for IFG) reported in the International Diabetes Federation (IDF) Diabetes Atlas[21,81] but reasonably close to the Europe-only estimates (7.1% for IGT and 3.3% for IFG). HbA1c was not included in the IDF definition of prediabetes.

6.1.2 Incidence of prediabetes

We found an overall incidence rate of HbA1c-defined prediabetes of 14.2 per 1,000 person-years. The sex-and age-standardized incidence rate was 10.4 per 1,000 person-years. Two other population-based studies calculated incidence rates based on registry data: a Canadian study of immigrants and long-term Canadian residents including more than 300,000 individuals with prediabetes (IFG, IGT, or HbA1c 42–47 mmol/mol)[39] and a British study based on electronic general practice health records[44], including more than 400,000 individuals with incident prediabetes (HbA1c 42–47 mmol/mol or Read codes for IFG, IGT, or prediabetes). The Canadian study found an incidence rate of 27.9 per 1,000 person-years among Western European immigrants[39], which is higher than our estimate. The British study found much lower incidence rates; 4.5 per 1,000 person-years for men and 4.7 per 1,000 person-years for women, compared to 13.9 per 1,000 person-years for men and 14.4 per 1,000 person-years for women in our study population and 10.8 per 1,000 person-years for men and 10.0 per 1,000 person-years for women after standardization. The low incidence rates in the British study were probably due to prediabetes being incompletely registered in the UK primary care records[44]. However, in line with our results, they found that women had an overall higher incidence rate than men, but that men had higher prediabetes incidence rates than women across all age groups above age 55. After standardization, our overall results showed higher incidences among men than women, which highlights the need for standardized results for comparison purposes as the incidence rate is highly dependent on sex and specifically on age.

Our overall findings are more consistent with those from the Danish Inter99 study[45] where the prediabetes (IFG and IGT) incidence was 18.6 per 1,000 person-years. However, HbA1c was not included in the definition of prediabetes in this study. In two small studies, incidence rates were estimated based on definitions solely using HbA1c; an incidence of 50.0 per 1,000 person-years in a cohort study based on general healthcare records of "high-need" patients in New Zealand with prediabetes defined as HbA1c 41–49 mmol/mol[43], and an incidence of 36.9 per 1,000 person-years based on results from the Singapore Malay Eye Study with prediabetes defined as HbA1c 39–47 mmol/mol[42]. Both had substantially higher incidence rates compared to our results, but they used wider intervals of HbA1c and did therefore include individuals with incident prediabetes not classified as prediabetes in our study.

6.2 Defining incident prediabetes

In our studies, we argue that the overall quality of the laboratory data, the widespread use of HbA1c measurements in the general population, and the Danish registry data made it feasible to follow the adult Danish population in the registries and calculate the incidence rate of prediabetes based on laboratory data. The Danish population has not been screened for prediabetes or diabetes and we therefore had to rely on the massive testing in Denmark. We have definitely not identified *all* Danes with prediabetes and have *under*estimated the incidence rate as we have assumed that those without an HbA1c measurement did not have prediabetes. HbA1c measured as part of diabetes management would often yield values in the higher range (i.e., ≥ 48 mmol/mol) or close to the threshold for diabetes. However, as is evident from the histogram of all HbA1c measurements in the laboratory data (Figure 5.1), the majority of the measurements were within the normal area. This indicates that HbA1c is not only measured among individuals with (suspected) diabetes. In addition, the median value of all the included measurements decreased over time, indicating a shift from predominantly using HbA1c as part of diabetes monitoring before 2012 to a more widespread use (Appendix I). The exact indication or reason for ordering a blood test is unknown in our data, but testing relies on clinical decisions and requires a contact with a general practitioner or a hospital visit. As diabetes is a lifestyle-related disease, and also closely related to other metabolic diseases, the individuals with prediabetes or diabetes would probably have at least some contact with the healthcare system and would most likely also have at least one HbA1c measurement.

The individuals we identified with prediabetes included individuals with incident prediabetes and individuals with prevalent prediabetes. In order to ensure the individuals truly had *incident* prediabetes, we included a 5-year look-back period in the laboratory data to exclude pre-existent prediabetes (HbA1c ≥ 42 mmol/mol) and diabetes (HbA1c ≥ 48 mmol/mol). As HbA1c testing was less frequent before 2012, some individuals were included with prevalent prediabetes or diabetes in the early study years, simply because they did not have an HbA1c measurement in the 5-year look-back period. Others might have had prediabetes or diabetes based on IGT or IFG, but neither OGTT nor FPG measurements were available in our data. The incidence rate was relatively high in 2012 (14.3 per 1,000 person-years in 2012 vs. 6.4 per 1,000 person-years in 2018 (Appendix I)), which is a consequence of including prevalent prediabetes and because of the changes in the diagnostic criteria. Although the use of FPG and OGTT declined rapidly after 2012[12], some individuals might still have been diagnosed with prediabetes or diabetes based on these measures after 2012. However, our main focus was to use HbA1c to mimic what happens in everyday clinical practice.

6.3 Progression to diabetes (Study I and Study II)

The WHO refrains from using the term "prediabetes" to prevent people from thinking that everyone with prediabetes will progress to diabetes[20]. Despite the controversy in the use of the terms "prediabetes" vs. "intermediate hyperglycemia"

(Section 2.1), the general consensus is that individuals with prediabetes are at increased risk of developing diabetes compared to those without prediabetes. In Study I and Study II, we estimated 5-year cumulative incidences for progression from incident HbA1c-defined prediabetes to diabetes or death (Figure 5.3). We found that 21% developed diabetes within five years after the prediabetes diagnosis and 18% died.

One of the major prediabetes trials, the Diabetes Prevention Program (DPP)[61], included 3,234 individuals with prediabetes during 1996–1999 and studied whether lifestyle interventions or treatment with metformin would prevent or delay the progression to diabetes compared to placebo (standard lifestyle recommendations plus placebo twice daily). Inclusion was based on age (≥ 25 years) and high BMI (BMI ≥ 24 kg/m^2) and prediabetes was defined using FPG 5.3–6.9 mmol/L and 2hPG 7.8–11.0 mmol/L. Among the 1,082 individuals assigned to placebo, 69% were women, the mean age was 50 years, and the mean BMI was 34 kg/m^2. In the subsequent Diabetes Prevention Program Outcomes Study (DPPOS), 2,776 of the original study participants were followed for a total of 15 years[92,93] and they reported a 5-year cumulative incidence of diabetes of 35% among those originally assigned to placebo. Diabetes was defined using FPG ≥ 7.0 mmol/L or 2hPG > 11.0 mmol/L and even though HbA1c was measured at study inclusion, it was neither used as a criterion for study inclusion nor as an outcome[51]. At study inclusion, the mean HbA1c was 41 mmol/mol[92,93], i.e., below the threshold for prediabetes and only 26% of the DPPOS participants with diabetes had HbA1c levels ≥ 48 mmol/mol at the time of the diabetes diagnosis[51]. In addition to the different baseline characteristics, this may explain why their 5-year cumulative incidence rate was different from ours.

Similar inclusion criteria were used in the Finnish Diabetes Prevention Study (DPS)[94,95] (BMI ≥ 25 kg/m^2, 40–65 years of age, and prediabetes defined as 2hPG 7.8–11.0 mmol/L among those with FPG < 7.8 mmol/L), where 522 individuals were randomised during 1993–1998 to intensive lifestyle interventions or to a control group receiving general lifestyle information. The 5-year cumulative incidence of diabetes was 34% among the 257 individuals who had been assigned to placebo treatment. However, diabetes was defined as FPG ≥ 7.8 mmol/L or 2hPG ≥ 11.0 mmol/L and HbA1c was not included in the study[94,95].

The progression from HbA1c-defined prediabetes (HbA1c 42–47 mmol/mol) to diabetes defined as FPG ≥ 7.0 mmol/L, HbA1c ≥ 48 mmol/mol, self-reported diabetes, or glucose-lowering treatment initiation was reported in a 2018 Cochrane Review based on prospective cohort studies (excluding intervention trials and studies on cohorts with additional comorbidities)[28]. The estimated 5-year risk of progression was 38%, but the review included only three small studies with few outcomes (1,103 individuals with 322 outcomes[96], 203 individuals with 100 outcomes[97], and 156 individuals with 58 outcomes[98]) and all studies were based on selected populations of individuals seen at repeat health examinations. In general, only seven small studies including a total of 2,529 individuals with HbA1c-defined prediabetes were included in the analysis of the cumulative incidences (3-year cumulative incidence of 7% based on one study with 370 individuals, 4-year cumulative incidence of 44% based on two studies with 627 participants, 5-year cumulative incidence of 38% based on three studies with

1,462 individuals, and 15-year cumulative incidence of 29% based on one study with 70 individuals). In contrast, our estimates are based on population-based registry data for a more than 300,000 individuals with incident HbA1c-defined prediabetes.

6.3.1 Progression estimate

We defined prediabetes based on HbA1c only and included HbA1c ≥ 48 mmol/mol, hospital contact with a diagnosis of diabetes, and redemption of a prescription for glucose-lowering medication in the definition of diabetes. In our progression analyses, prediabetes was defined using one rather than five years of look-back to detect prior diabetes or prediabetes in the laboratory data (we still used five years to exclude any prior treatment or hospitalization). This may have prompted inclusion of individuals with prevalent rather than incident prediabetes, but the baseline characteristics were fairly similar (Table 5.1). Those with prevalent prediabetes might either be individuals already on a path towards diabetes (hence causing an *over*estimation) or be individuals who have already managed to control their risk factors and therefore will either never progress to diabetes or will eventually revert back to normal glucose levels (hence causing an *under*estimation). As the median HbA1c was 42 mmol/mol (IQR 42–44) at study inclusion and the vast majority (86%) had values in the lower half of the prediabetes interval (42–44 mmol/mol), we believe the majority were identified with early prediabetes.

All elements of our definitions are based on real-world data and rely on individual contacts with a general practitioner or a hospital. According to the Danish Diabetes Association, a quarter of the total number of diabetes cases are not aware they have diabetes[34]. These individuals are not included in our cumulative incidence and cause an *under*estimation as unknown diabetes will of course not be recorded in the registries. As our population includes those with *known* prediabetes, the individuals should, at least to some degree, be followed by the general practitioner and have their HbA1c measured. In addition, we showed in Study I that the majority of all adult Danes (71%) had at least one HbA1c measurement during 2012–2018, and for older people, a large proportion had at least one HbA1c measurement in any given year. This would only increase the chance to detect diabetes.

6.4 Personalized probabilities (Study II)

We have estimated that 1 in 5 will progress from prediabetes to diabetes within five years. However, some individuals have a risk lower than 20% and some individuals have a risk higher than 20%. In Study II, we developed and validated a prediction model in order to use individual characteristics to estimate individual probabilities and identify important predictors of diabetes[67] (Appendix II). The predictors in our model were selected based on variable selection. We initially included multiple potential predictors in our model selection: demographic variables (age, sex, and ethnic origin), HbA1c measures (e.g., the value of the prediabetes-defining HbA1c measurement), current prescription drug use (e.g.,

statins, opioids, or antihypertensives), pre-existing comorbidities (e.g., CVD, cancer, or pancreatic disease), and socioeconomic variables (e.g., income and education). We also included variables from the DNHS questionnaire about self-reported lifestyle and health indicators[67,76]. After variable selection, our model included some well-known diabetes predictors such as age, sex, BMI, and pre-existing comorbidities[20,67]. In addition, our set-up also showed that self-rated health, self-reported doctor's advice regarding lifestyle problems, and measures of lack of a strong social network were important predictors for diabetes[67]. We did not have information about, e.g., blood pressure, waist circumference, or family history of diabetes[67]. Our prediction model could potentially be improved by including these rather easily available (in clinical practice) covariates as potential predictors. Unfortunately, HbA1c was the only biomarker available in our data. We could therefore not include, e.g., lipids, glucose levels, or estimates of insulin resistance and/or beta-cell function[67]. These variables could further improve the model.

Based on details about the individual characteristics, the model estimates individual probabilities and identifies individuals at high risk of progressing from prediabetes to diabetes. The model can essentially be used to target preventive intervention. It provides a snapshot of the current risk of progression[67], but to also inform about the risk of progression under certain preventive interventions or treatment strategies (e.g., to guide on choice of treatment), these interventions should be included in the model and be part of the risk assessment[67].

6.4.1 Ultimate prediction

We estimated that 1 in 5 will progress from prediabetes to diabetes within five years, meaning that if we have five random Danish adults with prediabetes, one will get diabetes. Am I this one individual, or am I among the four individuals who will not get diabetes? With the prediction model, we can now for each specific individual calculate a personalized risk rather than a population average. The problem is though that the risk is still a probability, meaning that if the individual risk is say 5%, 1 out of 20 "identical copies" of the individual will progress to diabetes — but the model cannot tell the individual which one of the copies he or she actually is[99]. An ultimate prediction model would be able to predict with a probability of 1, whether the individual progresses to diabetes or not, but this is not feasible in practice[99]. Also, the underlying purpose of medicine is to intervene and change the course of a disease, not just predict future disease states[99].

We validated our model using split-sample validation (our original sample was randomly split into a development sample consisting of 80% of the individuals and a validation sample consisting of 20% of the individuals), which causes an *over*estimation of the validity[67]. In spite of this, the discriminative ability, reflected by the AUCt, was modest at only 73 for HbA1c-defined type 2 diabetes[67]. It is better than chance, yet far from perfect. An American study[52] assessed the performance of HbA1c in predicting long-term diabetes (glucose-lowering treatment, FPG $\geq$ 7.0 mmol/L, HbA1c $\geq$ 48 mmol/mol, or self-reported diabetes) using prediction models with and without HbA1c for individuals without diabetes. They reported an area under the curve of 66 (95% CI 63–68) for a model including

only HbA1c, age, and sex, and 86 (95% CI 84–89) for a model in which fasting laboratory tests and clinical visits were added[52,67]. These estimates are similar to ours, but our main model containing input from multiple predictors had only slightly better discrimination than our model including just age and sex[52,67].

Our models are prognostic rather than diagnostic models and we predicted the 5-year risk of diabetes rather than the risk of current diabetes[100,101]. No one knows what will happen in the future — not even the ultimate prediction model[102]. "All models are wrong, but some are useful" is a phrase often heard in statistics[103]. Our model is also wrong simply because it is a model. It reflects our attempt to approximate the probability of subsequent progression to diabetes based on the data sources we had available. Adding more variables (e.g., biomarkers or genetic risk scores), choosing different variable selection strategies, or stratifying within prediabetes and diabetes phenotypes may potentially improve model validity[67] but the underlying problems would still be the same: the model is not perfect, we do not know what happens in the future, and we cannot make a perfect individual prediction (probabilities equal to 1). One of the strengths of our study is that we have included all relevant information in the publication[67,104]. This makes it possible for others to validate and not least calibrate our models to fit any specific setting where a customised model might be needed.

6.5 HbA1c Trajectories (Study III)

Our prediction models in Study II confirmed that higher vs. lower values of HbA1c at the time of prediabetes diagnosis were associated with an increased risk of future diabetes and treatment initiation. In Study III, we looked at what happened prior to treatment initiation, modelled the HbA1c trajectories, and identified distinct longitudinal patterns of HbA1c prior to glucose-lowering treatment initiation. In line with current knowledge[62–64] (Figure 2.3), we found slowly increasing HbA1c levels in the early stage of disease development, followed by rapidly increasing levels immediately prior to the diagnosis of diabetes (and diabetes treatment initiation). Our analysis extends the previous knowledge by providing evidence that the increase in the last phase of disease progression differed considerably between individuals and we identified four distinct latent classes. One of the major difficulties related to using the latent class trajectory analysis is to determine the number of classes. Other models and number of classes might have been possible, but we used both polynomial regressions and spline models in an attempt to capture as much of the variability as possible. Furthermore, we fitted the models for an increasing number of classes in order to find the optimal number of classes. As the trajectories mainly differ immediately before treatment initiation, a simple cluster analysis based on only one measurement per individual (e.g., k-means clustering using the last measurement before treatment initiation) could potentially have yielded the same classes. We put no restrictions on the timing of the HbA1c measurements (hence no selection) in our model and wanted to use all possible measurements to include as much information as possible. We estimated *both* the classes *and* the trajectories within one model (Appendix C).

Individuals in the largest of the four classes (Class 1, including 74% of the individuals) were most probably treated according to the guidelines (e.g., waiting for a confirmatory elevated HbA1c measurement or potentially initiating non-medical interventions (Section 2.3)). Still, based on the parametrization, we estimated that they had HbA1c $\geq$ 48 mmol/mol for almost nine months before treatment was initiated. Individuals in the remaining three classes (Class 2, Class 3, and Class 4), including one quarter of the individuals, had very high HbA1c levels at treatment initiation and had increased HbA1c levels above the threshold for diabetes ($\geq$ 48 mmol/mol) for an even longer period prior to treatment initiation. For all four classes, treatment could potentially have been initiated earlier. If treatment had been initiated before the rapidly increasing HbA1c levels, the onset of diabetes could potentially have been delayed or even prevented[63,95].

Before studying the effect of treatment and treatment regimens, it is important to recognize the bias introduced by diabetes being defined using a threshold value (threshold bias[63]). Measurements conducted close to the date when treatment was initiated will inevitably be higher than measurements conducted earlier[63] (if measurements were higher earlier, treatment would most likely also have been initiated earlier) and measurement error can potentially "cause" the disease. Having a random high measurement can prompt treatment initiation and simply because of regression to the mean, the HbA1c will most likely be lower after treatment initiation. Of note, as HbA1c depicts the average glycemic level during the past 8–12 weeks[11] (Chapter 2), it is not as a affected by measurement error as the measures of glucose (FPG and 2hPG)[13]. The increasing levels of HbA1c prior to treatment initiation is a result of a combination of naturally increasing values and the methodology as it depends on the threshold[63].

6.6 Treatment initiation (Study IV)

Whereas the measurement error and threshold value potentially caused problems in Study III, this is exactly what was utilized as an advantage in Study IV. Comparing individuals with HbA1c values close to the threshold showed a beneficial effect of the treatment threshold. The effect is not entirely described by treatment initiation, as almost none of those not eligible for treatment and only around 20% of those eligible for treatment actually initiated treatment (within three months). Lifestyle interventions, including physical activity, healthy diet, and weight loss, as well as intense cardiovascular risk-factor management also play a role in achieving the beneficial effect[25]. Because we used the RDD, individuals were "randomized" and therefore similar (in expectation) in each group. Confounding, be it known or unknown, measured or unmeasured, was thus not expected to be an issue in this study.

Our results have subsequently been confirmed by another Danish study[105] showing an increasing risk of CVD with increasing values of HbA1c, but with a drop in risk for individuals with HbA1c just above the threshold (48–49 mmol/mol). Like in our study, it is suggested that treatment for CVD risk factors should be initiated earlier and that individuals with prediabetes should undergo more intense monitoring[105]. The prediabetes trials (DPP, DPPOS,

and DPS) all showed a beneficial effect of intensive treatment in individuals with prediabetes[61,92–95] and a large group of Danes (around 330,000 adult Danes had prevalent prediabetes in December 2022) could potentially benefit from intense clinical attention and earlier treatment initiation.

6.6.1 Treatment threshold

The threshold for diabetes was originally determined to identify the individuals with increased prevalence of retinopathy[2] and the HbA1c value 6.5 % was set in order to mimic this threshold. HbA1c has previously been recorded using the DCCT unit, %, whereas it is currently mainly recorded using the IFCC unit, mmol/mol. Values are converted from one unit to the other and the threshold is converted from 6.5 % to 48 mmol/mol (Figure 2.1). However, the conversion formula is not perfect and the HbA1c value 47 mmol/mol is one of the problematic values. The formulas convert 6.4 % to 46 mmol/mol and 6.5 % to 48 mmol/mol. Thus, no value is converted to 47 mmol/mol. This is also depicted by fewer measurements with this value (Figure 5.1) and markedly less precise point estimates in the RDD study (Figure 5.7). When converting the other way, 47 mmol/mol, which is below the threshold for diabetes, is converted into 6.5 %, which is above the threshold for diabetes. The value 48 mmol/mol is correctly converted into 6.5 %. In the RDD study, we ensured to include individuals below or above the threshold according to the unit in which they were originally recorded. We also conducted sensitivity analyses excluding individuals whose HbA1c value was 47 mmol/mol, which did not change the results (HR 0.79 (95% CI 0.70–0.91)). It is generally problematic to have a conversion formula where individuals without diabetes will all of a sudden have diabetes if the value is converted to another unit. This only adds to the already opaque debate about prediabetes.

6.7 Clinical implications

In Denmark, HbA1c was introduced as a diagnostic criterion for prediabetes and diabetes in 2012. In the guidelines, the definitions were expanded as the HbA1c criterion was added to definitions. In practice, HbA1c substituted the old measurement methods and the definitions were changed. This change in the (practical) definitions of prediabetes and diabetes has also changed the epidemiology as it is no longer the same individuals that are identified and the severity of the disease at the time of diagnosis has changed. We used data from the Danish medical registries and modelled the clinical course of prediabetes as it is reflected in everyday clinical practice. We developed a prediction model in order to estimate personalized risks and effectively target preventive interventions. Furthermore, we modelled distinct HbA1c trajectories to address the heterogeneity in levels of HbA1c. We assessed the treatment inertia and showed that being eligible for treatment lowered the rate of future CVD and death, as we compared individuals with HbA1c just above the treatment threshold and individuals with HbA1c just below the treatment threshold.

Our findings suggest that suggest that prediabetes is frequent in Denmark and individuals with prediabetes might have unmet needs for preventive interventions. Treatment eligibility per se was associated with a reduced rate of CVD and death and acknowledging the heterogeneity in the way prediabetes progresses to diabetes could possibly lead to more individualized prevention strategies and therapy in the future.

7 | Conclusion

We used data from the Danish medical registries and showed that HbA1c-defined prediabetes can indeed be studied based on real-world data from routine clinical care. We identified individuals with HbA1c-defined prediabetes soon after their HbA1c levels progressed from the normal range into the range of prediabetes, allowing the estimation of updated prevalence and incidence rates for HbA1c-defined prediabetes in Denmark. We showed that 1 in 5 individuals with prediabetes progresses to diabetes within five years, but knowing individual-level risks for progression from prediabetes to type 2 diabetes is essential to effectively target preventive interventions. Metabolic factors and pre-existing comorbidities were confirmed as predictors of progression to type 2 diabetes, yet even after establishing self-rated health, lifestyle, and existence of a social network as important predictors, it is still difficult to identify the exact individuals who will progress to diabetes. By acknowledging the heterogeneity, not only in diabetes but also in the way prediabetes progresses to diabetes, treatment and treatment guidelines could possibly be improved to ensure all individuals receive the best suitable treatment. We showed a beneficial effect of being eligible for type 2 diabetes treatment and some individuals might therefore benefit from earlier monitoring and intervention to avoid having extremely high levels of HbA1c before treatment is initiated.

7.1 Future perspectives

Research is about asking the right questions. As prediabetes is vaguely defined, the terminology is inconsistent, and the conversion from one unit to another can give rise to problems, the research field of prediabetes leaves much to be desired. If diabetes is determined based on an increased prevalence of retinopathy[2] and prediabetes is defined as an increased risk of diabetes, prediabetes is actually "increased risk of increased prevalence of retinopathy". In addition, a transatlantic trip may cure or cause prediabetes simply because of differences in the diagnostic criteria[106]. Our studies were based on Danish real-world data and we have tried to question the current definitions and treatment guidelines, focusing on statistical analyses and methodological challenges within the field of prediabetes and type 2 diabetes. We showed that treatment eligibility had a beneficial effect, suggesting that the treatment threshold (48 mmol/mol) could potentially be lowered. Extrapolating the results from Study IV suggests an optimal value of 44.6 mmol/mol (Appendix D). Based on the trajectories from Study III, this value

is reached around 1.5 years before treatment is currently initiated. Identifying the individuals for whom intense monitoring and early treatment initiation could be an effective treatment regimen requires a satisfactory prediction model (i.e., better than the model from Study II) with accurate personalized probabilities of progression. We showed in Study I that one out of five adults with prediabetes will progress to diabetes within five years. Our results add to the debate about prediabetes, glucose-lowering treatment initiation, and HbA1c trajectories, yet more research is needed.

8 | English summary

Since 2012, where HbA1c was introduced as a diagnostic criterion for diabetes in Denmark, HbA1c testing has increased substantially in the Danish population. HbA1c testing is currently the most frequently used method to diagnose both diabetes and prediabetes. This offers unique possibilities to exploit Danish population-based laboratory data to better understand the epidemiology of prediabetes. The aim of this PhD project was to use data from the Danish medical registries to study prediabetes occurrence, to model the course of prediabetes and understand the progression towards type 2 diabetes, and, finally, to study the effect of being eligible for diabetes treatment initiation on clinical outcomes.

In a descriptive study (Study I), we included 12,762,32 HbA1c measurements available from the Danish laboratory databases in the years 2012—2018. We estimated that 71% of the entire Danish adult population had at least one HbA1c measurement in this period. Among adults, we estimated an incidence rate of prediabetes of 14 per 1,000 person-years and a prevalence of 7% in 2018. Five years after the diagnosis of prediabetes, more than 1 in 5 had progressed to diabetes.

In a prediction study (Study II), we developed and validated a prediction model for the progression towards type 2 diabetes based on 26,007 individuals with incident HbA1c-defined prediabetes who had participated in a detailed health survey. In addition to age, sex, metabolic factors, and comorbidities, we found that self-rated health, lifestyle, and existence of a social network were important predictors of the progression to type 2 diabetes. The discriminative ability of our model was modest and the model should therefore be improved prior to being used in clinical practice.

In a trajectory study (Study III), we included 20,733 individuals who initiated their first-ever glucose-lowering treatment in 2017-–2018. Based on a total of 105,211 HbA1c measurements from the five years prior to treatment initiation, we identified four distinct trajectories. In all four trajectories, levels of HbA1c were slowly increasing for a long time before more rapidly increasing levels prompted the initiation of treatment. Despite differences in the steepness of the increase, the four trajectories showed mean HbA1c levels above the diagnostic threshold for diabetes for 9 to 17 months before treatment was finally initiated.

Using the regression discontinuity design (Study IV), we studied the effect of treatment eligibility among 43,070 individuals who had HbA1c values close to the diabetes treatment threshold (48 mmol/mol). We found that those who had an HbA1c value just above the diabetes threshold had a 21% lower rate of cardiovascular disease and death, compared to those whose HbA1c value was just

below the threshold.

The findings in this PhD project suggest that prediabetes is frequent in Denmark and that individuals with prediabetes might have unmet needs for preventive interventions. We showed that diabetes treatment eligibility per se was associated with a reduced rate of cardiovascular disease and death. Acknowledging the heterogeneity in the way prediabetes progresses to diabetes could possibly lead to more individualized prevention strategies and therapy in the future.

9 | Dansk resumé (Danish summary)

Siden 2012, hvor HbA1c blev introduceret som diagnostisk kriterium for diabetes i Danmark, er antallet af HbA1c-målinger i befolkningen steget væsentligt. I dag er måling af HbA1c den mest anvendte metode til at diagnosticere både diabetes og prædiabetes. Dette giver unikke muligheder for at udnytte data fra de danske laboratoriedatabaser til at undersøge prædiabetes epidemiologi. Formålet med dette ph.d.-projekt var at bruge danske registerdata til at undersøge forekomsten af prædiabetes, til at modellere forløbet og forstå progressionen fra prædiabetes til type 2 diabetes, samt at undersøge hvilken effekt det at opfylde det diagnostiske kriterium for diabetes, og dermed opfylde indikationen for at komme i diabetesbehandling, har på fremtidige kliniske udfald.

I et deskriptivt studie (Studie I) inkluderede vi 12.762.32 HbA1c målinger fra de danske laboratoriedatabaser i årene 2012—2018. Vi estimerede at 71% af den voksne danske befolkning havde mindst én HbA1c måling i denne periode. Blandt alle voksne danskere estimerede vi en prædiabetes incidensrate på 14 per 1.000 personår samt en prævalens af prædiabetes på 7% i 2018. I løbet af de første 5 år efter prædiabetes diagnosticeres, vil mere end 1 ud af 5 have udviklet diabetes.

I et prædiktionsstudie (Studie II) udviklede og validerede vi en prædiktionsmodel for at estimere risikoen for at udvikle type 2 diabetes. Vi inkluderede 26.007 individer med incident HbA1c-defineret prædiabetes, som havde deltaget i et spørgeskema om helbredsforhold. Udover alder, køn, metaboliske faktorer og komorbiditet, så er selvrapporteret helbred, livsstil og det at have et socialt netværk vigtige prædiktorer for udviklingen af type 2 diabetes. Modellens diskriminative evne var moderat, og modellen skal derfor forbedres før den kan implementeres i klinisk praksis.

I et ”trajectory” studie (Studie III), hvor vi undersøgte udviklingen i HbA1c-værdier i en periode før diabetesbehandling blev igangsat, inkluderede vi 20.733 individer som havde påbegyndt deres første diabetesbehandling i årene 2017—2018. Baseret på 105.211 HbA1c-målinger fra de 5 år før diabetesbehandlingen blev igangsat, fandt vi i alt fire forskellige trajectories (udviklingsforløb). Alle trajectories viste langsomt stigende HbA1c-niveauer i en længere periode, før en markant stigning gjorde diabetesbehandling nødvendig. På trods af forskelle i hvor stejlt HbA1c steg, så havde de fire trajectories HbA1c-niveauer over den diagnostiske grænse for diabetes i 9 til 17 måneder før diabetesbehandlingen blev igangsat.

Ved at bruge et "regression discontinuity design" som studiedesign (Studie IV) undersøgte vi effekten af at opfylde indikationen for at komme i diabetesbehandling. Vi inkluderede 43.070 individer der havde HbA1c-værdier

tæt på grænseværdien for hvornår diabetesbehandling ifølge guidelines skal igangsættes (48 mmol/mol). Vi fandt at de, der havde HbA1c-værdier lige over den diagnostiske grænse, havde en 21% lavere rate af hjertekarsygdom og død, sammenlignet med individer hvis HbA1c-værdi lå lige under grænsen.

Resultaterne i dette ph.d.-projekt indikerer, at prædiabetes er hyppigt forekommende i Danmark, og der kan være et potentiale i forebyggende interventioner blandt individer med prædiabetes. Vi viste, at det at opfylde indikationen for at komme i diabetesbehandling i sig selv associerer med en reduceret rate af hjertekarsygdom og død. Bedre forståelse af den heterogenitet hvormed prædiabetes udvikler sig til diabetes over tid, kan fremover muliggøre mere individualiserede forebyggelses- og behandlingsstrategier.

References

1. World Health Organization: WHO Guidelines Approved by the Guidelines Review Committee. *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation*. WHO Guidelines Approved by the Guidelines Review Committee. Geneva: World Health Organization, 2011.
2. The International Expert Committee: International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. Diabetes Care 2009;**32**:1327–34.
3. Fonseca V, Inzucchi SE, and Ferrannini E: Redefining the diagnosis of diabetes using glycated hemoglobin. Diabetes Care 2009;**32**:1344–5.
4. Sundhedsstyrelsen: Notat vedrørende diagnostik af diabetes mellitus med hba1c. [Accessed 29 September 2022]. 2012. URL: https://www.sst.dk/da/udgivelser/2012/notat-vedroerende-diagnostik-af-diabetes-mellitus-med-hba1c.
5. Schmidt M, Schmidt SAJ, Adelborg K, Sundbøll J, Laugesen K, Ehrenstein V, and Sørensen HT: The Danish health care system and epidemiological research: from health care contacts to database records. Clin Epidemiol 2019;**11**:563–91.
6. Danmarks Statistik: Statistikbanken, Folketal. [Accessed 15 January 2023]. 2021. URL: https://%5C%5C%20statistikbanken.dk/statbank5a/default.asp?w=1856.
7. Schmidt M, Pedersen L, and Sørensen HT: The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 2014;**29**:541–9.
8. Rothman K: Epidemiology: An Introduction. Oxford University Press, 2002.
9. Frank L: Epidemiology. When an entire country is a cohort. Science 2000;**287**:2398–9.
10. Penttilä I, Penttilä K, Holm P, Laitinen H, Ranta P, Törrönen J, and Rauramaa R: Methods, units and quality requirements for the analysis of haemoglobin A1c in diabetes mellitus. World J Methodol 2016;**6**:133–42.
11. Nathan DM, Turgeon H, and Regan S: Relationship between glycated haemoglobin levels and mean glucose levels over time. Diabetologia 2007;**50**:2239–44.

12. Knudsen JS, Knudsen SS, Hulman A, Witte DR, Gregg EW, Lauritzen T, Pedersen L, Sørensen HT, and Thomsen RW: Changes in type 2 diabetes incidence and mortality associated with introduction of HbA1c as diagnostic option: A Danish 24-year population-based study. Lancet Reg Health Eur 2022;**14**:100291.

13. Nielsen AA, Petersen PH, Green A, Christensen C, Christensen H, and Brandslund I: Changing from glucose to HbA1c for diabetes diagnosis: predictive values of one test and importance of analytical bias and imprecision. Clin Chem Lab Med 2014;**52**:1069–77.

14. Dansk Selskab for Almen Medicin: Type 2-diabetes – opfølgning og behandling. Report. 2019. URL: https://vejledninger.dsam.dk/type2/.

15. American Diabetes Association Professional Practice Committee: 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2022. Diabetes Care 2022;**45**:S17–S38.

16. Lægehåndbogen: Hæmoglobin A1c (HbA1c). [Accessed 06 February 2021]. 2020. URL: https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/undersoegelser-og-proever/klinisk-biokemi/blodproever/haemoglobin-a1c-hba1c/.

17. Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, and Adelborg K: Existing Data Sources in Clinical Epidemiology: Laboratory Information System Databases in Denmark. Clin Epidemiol 2020;**12**:469–75.

18. Gyldenkerne C, Kahlert J, Olesen KK, Thrane PG, Sørensen HT, Thomsen RW, and Maeng M: Twenty-Year Temporal Trends in Risk of Ischemic Stroke in Incident Type 2 Diabetes: A Danish Population-Based Cohort Study. Diabetes Care 2022;**45**:2144–51.

19. Hansen KB, Kristensen JK, Balasubramaniam K, et al.: Farmakologisk behandling af type 2-diabetes - Fælles retningslinje fra DES og DSAM. Report. Dansk Endokrinologisk Selskab (DES), Dansk Selskab for Almen Medicin (DSAM), 2022. URL: https://vejledninger.dsam.dk/fbv-t2dm/.

20. Tabák AG, Herder C, Rathmann W, Brunner EJ, and Kivimäki M: Prediabetes: a high-risk state for diabetes development. Lancet 2012;**379**:2279–90.

21. Hostalek U: Global epidemiology of prediabetes - present and future perspectives. Clin Diabetes Endocrinol 2019;**5**:5.

22. Makaroff LE: The need for international consensus on prediabetes. Lancet Diabetes Endocrinol 2017;**5**:5–7.

23. World Health Organization and International Diabetes Federation: Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia : report of a WHO/IDF consultation. 2006.

24. World Health Organization: Global report on diabetes. World Health Organization, 2016.

25. Vistisen D, Witte DR, Brunner EJ, Kivimäki M, Tabák A, Jørgensen ME, and Færch K: Risk of Cardiovascular Disease and Death in Individuals With Prediabetes Defined by Different Criteria: The Whitehall II Study. Diabetes Care 2018;**41**:899–906.

26. Morris DH, Khunti K, Achana F, Srinivasan B, Gray LJ, Davies MJ, and Webb D: Progression rates from HbA1c 6.0-6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. Diabetologia 2013;**56**:1489–93.

27. Lee CMY, Colagiuri S, Woodward M, et al.: Comparing different definitions of prediabetes with subsequent risk of diabetes: an individual participant data meta-analysis involving 76 513 individuals and 8208 cases of incident diabetes. BMJ Open Diabetes Res Care 2019;**7**:e000794.

28. Richter B, Hemmingsen B, Metzendorf MI, and Takwoingi Y: Development of type 2 diabetes mellitus in people with intermediate hyperglycaemia. Cochrane Database Syst Rev 2018;**10**:CD012661.

29. Vistisen D, Kivimäki M, Perreault L, Hulman A, Witte DR, Brunner EJ, Tabák A, Jørgensen ME, and Færch K: Reversion from prediabetes to normoglycaemia and risk of cardiovascular disease and mortality: the Whitehall II cohort study. Diabetologia 2019;**62**:1385–90.

30. Schlesinger S, Neuenschwander M, Barbaresko J, Lang A, Maalmi H bibinitperiod, Rathmann W, Roden M, and Herder C: Prediabetes and risk of mortality, diabetes-related complications and comorbidities: umbrella review of meta-analyses of prospective studies. Diabetologia 2022;**65**:275–85.

31. Cai X, Zhang Y, Li M, Wu JH, Mai L, Li J, Yang Y, Hu Y, and Huang Y: Association between prediabetes and risk of all cause mortality and cardiovascular disease: updated meta-analysis. BMJ 2020;**370**:m2297.

32. Jonas DE, Crotty K, Yun JD, et al.: U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. *Screening for Prediabetes and Type 2 Diabetes Mellitus: An Evidence Review for the U.S. Preventive Services Task Force*. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US), 2021.

33. Jørgensen ME, Ellervik C, Ekholm O, Johansen NB, and Carstensen B: Estimates of prediabetes and undiagnosed type 2 diabetes in Denmark: The end of an epidemic or a diagnostic artefact? Scand J Public Health 2020;**48**:106–12.

34. Diabetesforeningen: Diabetes i Danmark. [Accessed 8 October 2022]. URL: https://diabetes.dk/forskning/viden-om-diabetes/diabetes-i-danmark.

35. Carstensen B, Rønn PF, and Jørgensen ME: Components of diabetes prevalence in Denmark 1996-2016 and future trends until 2030. BMJ Open Diabetes Res Care 2020;**8**.

36. Gujral UP, Narayan KMV, Kandula NR, Liu K, and Kanaya AM: Incidence of diabetes and prediabetes and predictors of glycemic change among South Asians in the USA: the MASALA study. BMJ Open Diabetes Res Care 2020;**8**.

37. Sadeghi M, Talaei M, Parvaresh Rizi E, Dianatkhah M, Oveisgharan S, and Sarrafzadegan N: Determinants of incident prediabetes and type 2 diabetes in a 7-year cohort in a developing country: The Isfahan Cohort Study. J Diabetes 2015;**7**:633–41.

38. Latifi SM, Karandish M, Shahbazian H, and Hardani Pasand L: Incidence of Prediabetes and Type 2 Diabetes among People Aged over 20 Years in Ahvaz: A 5-Year Perspective Study (2009-2014). J Diabetes Res 2016;**2016**:4908647.

39. Fazli GS, Moineddin R, Bierman AS, and Booth GL: Ethnic differences in prediabetes incidence among immigrants to Canada: a population-based cohort study. BMC Med 2019;**17**:100.

40. Vijayakumar G, Manghat S, Vijayakumar R, Simon L, Scaria LM bibinitperiod, Vijayakumar A, Sreehari GK, Kutty VR, Rachana A, and Jaleel A: Incidence of type 2 diabetes mellitus and prediabetes in Kerala, India: results from a 10-year prospective cohort. BMC Public Health 2019;**19**:140.

41. Hadaegh F, Derakhshan A, Zafari N, Khalili D, Mirbolouk M, Saadat N, and Azizi F: Pre-diabetes tsunami: incidence rates and risk factors of pre-diabetes and its different phenotypes over 9 years of follow-up. Diabet Med 2017;**34**:69–78.

42. Man REK, Charumathi S, Gan ATL, Fenwick EK, Tey CS, Chua J, Wong TY, Cheng CY, and Lamoureux EL: Cumulative incidence and risk factors of prediabetes and type 2 diabetes in a Singaporean Malay cohort. Diabetes Res Clin Pract 2017;**127**:163–71.

43. Gu Y, Warren J, Kennelly J, Walker N, and Harwood M: Incidence Rate of Prediabetes: An Analysis of New Zealand Primary Care Data. Stud Health Technol Inform 2015;**214**:81–6.

44. Pal K, Horsfall L, Sharma M, Nazareth I, and Petersen I: Time trends in the incidence of clinically diagnosed type 2 diabetes and pre-diabetes in the UK 2009-2018: a retrospective cohort study. BMJ Open Diabetes Res Care 2021;**9**.

45. Engberg S, Vistisen D, Lau C, Glümer C, Jørgensen T, Pedersen O, and Borch-Johnsen K: Progression to impaired glucose regulation and diabetes in the population-based Inter99 study. Diabetes Care 2009;**32**:606–11.

46. Knudsen JS, Hulman A, Rønn PF, Lauritzen T, Sørensen HT, Witte DR, and Thomsen RW: Trends in HbA(1c) and LDL Cholesterol in Patients With Type 2 Diabetes Receiving First-Time Treatment in Northern Denmark, 2000-2017: Population-Based Sequential Cross-Sectional Analysis. Diabetes Care 2020;**43**:e17–e19.

47. Carstensen B, Rønn PF, and Jørgensen ME: Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996-2016. BMJ Open Diabetes Res Care 2020;**8**.

48. Diabetesforeningen: Andre typer diabetes. [Accessed 8 November 2022]. URL: https://diabetes.dk/andre-diabetestyper.

49. Carstensen B, Rønn PF, and Jørgensen ME: Lifetime risk and years lost to type 1 and type 2 diabetes in Denmark, 1996-2016. BMJ Open Diabetes Res Care 2021;**9**.

50. DeJesus RS, Breitkopf CR, Rutten LJ, Jacobson DJ, Wilson PM, and Sauver JS: Incidence Rate of Prediabetes Progression to Diabetes: Modeling an Optimum Target Group for Intervention.
Popul Health Manag 2017;**20**:216–23.

51. Diabetes Prevention Program Research Group: HbA1c as a Predictor of Diabetes and as an Outcome in the Diabetes Prevention Program:
A Randomized Clinical Trial. Diabetes Care 2015;**38**:51–8.

52. Leong A, Daya N, Porneala B, Devlin JJ, Shiffman D, McPhaul MJ, Selvin E, and Meigs JB: Prediction of Type 2 Diabetes by Hemoglobin A(1c) in Two Community-Based Cohorts. Diabetes Care 2018;**41**:60–8.

53. Warren B, Pankow JS, Matsushita K, Punjabi NM, Daya NR, Grams M, Woodward M, and Selvin E: Comparative prognostic performance of definitions of prediabetes: a prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) study. Lancet Diabetes Endocrinol 2017;**5**:34–42.

54. Gardner MP, Wang J, Hazlehurst JM, Sainsbury C, Blissett J, Ñirantharakumar K, Thomas N, and Bellary S: Risk of progression from pre-diabetes to type 2 diabetes in a large UK adult cohort. Diabet Med 2022:e14996.

55. Sundheds- og Ældreministeriet: Den nationale diabetes-handlingsplan. 2017. URL: https://sum.dk/publikationer/2017/november/den-nationale-diabetes-handlingsplan.

56. Videncenter for Diabetes: Prædiabetes. [Accessed 10 October 2022]. 2022. URL: https://%5C%5C%20videncenterfordiabetes.dk/viden-om-diabetes/type-2-diabetes/hvad-er-type-2-diabetes/praediabetes?utm_campaign=googleads&utm_medium=ad&utm_source=google.

57. Lind M, Imberg H, Coleman RL, Nerman O, and Holman RR: Historical HbA(1c) Values May Explain the Type 2 Diabetes Legacy Effect: UKPDS 88. Diabetes Care 2021;**44**:2231–7.

58. Riddle MC, Gerstein HC, and Home PD: Lingering Effects of Hyperglycemia in Recently Diagnosed Diabetes During Long-term Follow-up of the DCCT/EDIC and UKPDS Cohorts: More Evidence That Early Control Matters. Diabetes Care 2021;**44**:2212–5.

59. Lachin JM and Nathan DM: Understanding Metabolic Memory: The Prolonged Influence of Glycemia During the Diabetes Control and Complications Trial (DCCT) on Future Risks of Complications During the Study of the Epidemiology of Diabetes Interventions and Complications (EDIC). Diabetes Care 2021;**44**:2216–24.

60. Lachin JM, Bebu I, and Nathan DM: The Beneficial Effects of Earlier Versus Later Implementation of Intensive Therapy in Type 1 Diabetes. Diabetes Care 2021;**44**:2225–30.

61. Diabetes Prevention Program Research Group: Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 2002;**346**:393–403.

62. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, and Witte DR: Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. Lancet 2009;**373**:2215–21.

63. Mason CC, Hanson RL, and Knowler WC: Progression to type 2 diabetes characterized by moderate then rapid glucose increases.
Diabetes 2007;**56**:2054–61.

64. Heianza Y, Arase Y, Fujihara K, et al.: Longitudinal Trajectories of HbA1c and Fasting Plasma Glucose Levels During the Development of Type 2 Diabetes: The Toranomon Hospital Health Management Center Study 7 (TOPICS 7). Diabetes Care 2012;**35**:1050–2.

65. Rigsarkivet: Om arkivalierne. [Accessed 17 November 2022]. 2022. URL: https://www.sa.dk/daisy/om_daisy.

66. Region Nordjylland: Regioner i Danmark. [Accessed 16 June 2021]. Modified. 2021. URL: https://rn.dk/genveje/fakta-om-nordjylland/regioner-i-danmark.

67. Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, and Pedersen L:
Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. BMJ Open Diabetes Res Care 2022;**10**.

68. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, and Sørensen HT: The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015;**7**:449–90.

69. Kildemoes HW, Sørensen HT, and Hallas J: The Danish National Prescription Registry. Scand J Public Health 2011;**39**:38–41.

70. Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, and Sørensen HT: Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. Clin Epidemiol 2012;**4**:303–13.

71. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, and Schmidt M: Data Resource Profile: The Danish National Prescription Registry. Int J Epidemiol 2017;**46**:798–798f.

72. Ehrenstein V, Antonsen S, and Pedersen L: Existing data sources for clinical epidemiology: Aarhus University Prescription Database. Clin Epidemiol 2010;**2**:273–9.

73. Thygesen LC, Daasnes C, Thaulow I, and Brønnum-Hansen H: Introduction to Danish (nationwide) registers on health and social issues: Structure, access, legislation, and archiving. Scand J Public Health 2011;**39**:12–6.

74. Baadsgaard M and Quitzau J: Danish registers on personal income and transfer payments. Scand J Public Health 2011;**39**:103–5.

75. Jensen VM and Rasmussen AW: Danish Education Registers. Scand J Public Health 2011;**39**:91–4.

76. Christensen AI, Lau CJ, Kristensen PL, Johnsen SB, Wingstrand A, Friis K, Davidsen M, and Andreasen AH: The Danish National Health Survey: Study design, response rate and respondent characteristics in 2010, 2013 and 2017. Scand J Public Health 2022;**50**:180–8.

77. Grann AF, Erichsen R, Nielsen AG, Froslev T, and Thomsen RW: Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. Clin Epidemiol 2011;**3**:133–8.

78. Pontet F, Petersen UM, Fuentes-Arderiu X, et al.: Clinical laboratory sciences data transmission: the NPU coding system. Stud Health Technol Inform 2009;**150**:265–9.

79. Sundhedsdatastyrelsen: NPU terminologi for laboratorier, LabTerm. [Accessed 19 January 2023]. 2021. URL: https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-terminologi/npu.

80. Petersen I, Nicolaisen SK, Ricciardi F, Sharma M, Thomsen RW, Baio G, and Pedersen L: Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study. Clin Epidemiol 2020;**12**:569–77.

81. International Diabetes Federation: IDF Diabetes Atlas. *IDF Diabetes Atlas*. International Diabetes Federation, 2021.

82. Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, and Inoue M: Age standardization of rates: A new WHO standard.

83. SAS Institute Inc.: SAS/STAT 14.1 User's Guide The LIFETEST Procedure. SAS institute Incorporated, 2015.

84. Fu Z, Parikh CR, and Zhou B: Penalized variable selection in competing risks regression. Lifetime Data Anal 2017;**23**:353–76.

85. Blanche P, Kattan MW, and Gerds TA: The c-index is not proper for the evaluation of t-year predicted risks. Biostatistics 2019;**20**:347–57.

86. Blanche P, Dartigues JF, and Jacqmin-Gadda H: Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 2013;**32**:5381–97.

87. Kattan MW and Gerds TA: The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res 2018;**2**:7.

88. Proust-Lima C, Philipps V, and Liquet B: Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. J Stat Softw 2017;**78**:1–56.

89. Harrell F. E. J: Regression Modeling Strategies. 2nd ed. Vol. 2. Springer Series in Statistics. Springer, 2015.

90. Gómez-Rubio V: Bayesian inference with INLA. Chapman & Hall/CRC Press, 2020 (online version 2021-08-29).

91. Rue H: R-INLA Project. [Accessed 7 June 2022]. 2022. URL: https://www.r-inla.org.

92. Diabetes Prevention Program Research Group: Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. Lancet Diabetes Endocrinol 2015;**3**:866–75.

93. Diabetes Prevention Program Research Group: 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet 2009;**374**:1677–86.

94. Tuomilehto J, Lindström J, Eriksson JG, et al.: Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 2001;**344**:1343–50.

95. Lindström J, Peltonen M, Eriksson JG, Ilanne-Parikka P, Aunola S bibinitperiod, Keinänen-Kiukaanniemi S, Uusitupa M, and Tuomilehto J: Improved lifestyle and decreased diabetes risk over 13 years: long-term follow-up of the randomised Finnish Diabetes Prevention Study (DPS). Diabetologia 2013;**56**:284–93.

96. Kim CH, Kim HK, Kim EH, Bae SJ, Choe J, and Park JY: Risk of progression to diabetes from prediabetes defined by HbA1c or fasting plasma glucose criteria in Koreans. Diabetes Res Clin Pract 2016;**118**:105–11.

97. Heianza Y, Arase Y, Fujihara K, et al.: Screening for pre-diabetes to predict future diabetes using various cut-off points for HbA(1c) and impaired fasting glucose: the Toranomon Hospital Health Management Center Study 4 (TOPICS 4). Diabet Med 2012;**29**:e279–85.

98. Nakagami T, Tanaka Y, Oya J, Kurita M, Isago C, Hasegawa Y, Ito A, Hirota N, Tsuzura R, and Uchigata Y: Associations of HbA1c and fasting plasma glucose with incident diabetes: Implications for pre-diabetes thresholds in a Japanese population. Prim Care Diabetes 2016;**10**:407–14.

99. Dekkers OM and Mulder JM: When will individuals meet their personalized probabilities? A philosophical note on risk prediction. Eur J Epidemiol 2020;**35**:1115–21.

100. Gray LJ, Taub NA, Khunti K, Gardiner E, Hiles S, Webb DR, Srinivasan BT, and Davies MJ: The Leicester Risk Assessment score for detecting undiagnosed Type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. Diabet Med 2010;**27**:887–95.

101. Diabetesforeningen: Test din risiko. [Accessed 19 december 2022]. URL: https://diabetes.dk/diabetes-2/fakta-om-diabetes-2/test-din-risiko.aspx.

102. Vandenbroucke JP: Cerebral sinus thrombosis and oral contraceptives. There are limits to predictability. BMJ 1998;**317**:483–4.

103. Wikipedia: All models are wrong. [Accessed 16 December 2022]. 2022. URL: https://en.wikipedia.org/wiki/All_models_are_wrong.

104. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, and Collins GS: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration.
Annals of Internal Medicine 2015;**162**:W1–W73.

105. Yahyavi SK, Snorgaard O, Knop FK, Schou M, Lee C, Selmer C, Gislason G, Torp-Pedersen C, Jensen MB, and Bonde AN: Prediabetes Defined by First Measured HbA(1c) Predicts Higher Cardiovascular Risk Compared With HbA(1c) in the Diabetes Range: A Cohort Study of Nationwide Registries. Diabetes Care 2021;**44**:2767–74.

106. Borch-Johnsen K and Colagiuri S: Diagnosing diabetes–time for a change? Diabetologia 2009;**52**:2247–50.

Appendices

Methodological appendices

Appendix A: Cumulative incidence

Appendix B: Prediction and competing risk

Appendix C: Latent class trajectory analyses

Appendix D: Regression discontinuity design

Appendix A | Cumulative incidence

The cumulative incidence is the probability of an outcome, Y, in a specified time period[1,2]. In our studies, we were interested in the progression from HbA1c-defined prediabetes to diabetes during a 5-year period, i.e., for $t = 5$ years, $CIF(t) = P(T \leq t)$, with T being the time to diabetes[2]. In the ideal situation, all individuals would be followed for five years and we would therefore know whether or not the individuals progress to diabetes. The estimated probability of diabetes would simply be the number of individuals with diabetes after five years, D, divided by the total number of individuals with prediabetes, N. Even so, this is almost never the case as individuals might die during the 5-year period, they might emigrate and be lost to follow-up in our registry data, or we might simply not have enough follow-up time because of the data structure or because our study period has ended. This needs to be taken into account. Individuals for whom we do not have data available after five years are *censored* whenever we can no longer follow them. This means they are still at risk of the outcome (with the same probability as if we had actually had their data), but we simply cannot tell. Individuals who die during follow-up will no longer be at risk of diabetes and death is therefore a *competing event*.

In a situation where there are no competing events, for example when death is the outcome of interest, the probability of the outcome (i.e., dying) is the same as 1 minus the probability of not having the outcome (i.e., survining). This means that $CIF(t) = 1 - S(t)$, where $S(t)$ denotes the survival function, i.e., the probability of surviving time t. The survival function can be estimated using the Kaplan-Meier estimator[2,3]. Let $t_1 < t_2 < \cdots < t_M$ be the distinct event times. For $i = 1, \ldots, M$ let n_i be the number of surviving individuals just prior to t_i (i.e., individuals at risk) and d_i the number of individuals with the event at t_i. Then the Kaplan-Meier estimate is[3]

$$\hat{S}_{KM}(t_i) = \prod_{j=1}^{i} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j=1}^{i} \left(1 - \frac{d_j}{n_j} \right).$$

The cumulative incidence function describes the probability of dying until time t, the survival function the probability of not dying before t, and the hazard function, $h(t)$ expresses the probability of dying at time t, i.e., the instantaneous rate (per time unit) at time t given survival until t:[4,5]

$$h(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t). \tag{A.1}$$

The relation is described as $S(t) = \exp(-H(t))$ or similarly,
$CIF(t) = 1 - \exp(-H(t))$, where $H(t)$ is the cumulative hazard, which can be

estimated using the Nelson-Aalen estimator[2]

$$\hat{H}_{NA}(t_i) = \sum_{j=1}^{i} \left(\frac{d_j}{n_j} \right).$$

However, our outcome of interest is diabetes and not death. Death should therefore be considered a competing event. The problem with competing events is that $CIF(t) = 1 - S(t)$ is no longer true and the cumulative incidence cannot be estimated directly based on $S(t)$ or $H(t)$. Estimating the cumulative incidence based on the Kaplan-Meier estimator corresponds to treating death as independent censoring. Instead, versions of the Kaplan-Meier and Nelson-Aalen can be combined into the Aalen-Johansen estimator[3]

$$\widehat{CIF}_k(t_i) = \sum_{j=1}^{i} \frac{d_{kj}}{n_j} \hat{S}^*_{KM}(t_{j-1}),$$

where k denotes the cause (e.g., either diabetes or death), n_j the number of individuals at risk at time j, d_{kj} the number of failures at at time j due to cause k, and $\hat{S}^*_{KM}(t_i)$ is the Kaplan-Meier estimate where all causes are combined.

In Study I, we identified 366,752 individuals with incident prediabetes, whereof 41,350 (11.3%) died during follow-up (Section 5.1). Using the Kaplan-Meier estimator, the estimated overall 5-year cumulative incidence for death was 17.5% (95% CI 17.3–17.7). We observed that 49,855 (13.6%) developed diabetes during follow-up. We noted that among those who were followed for all five years, 24.2% developed diabetes during follow-up. If death was ignored as a competing event (hence treated as a censoring event), the 5-year cumulative incidence of diabetes would be 23.3% (95% CI 23.1–23.5). When death was considered a competing event, the estimated overall 5-year cumulative incidence for diabetes was 21.3% (95% CI 21.1–21.5) (Figure A.1). The SAS code to estimate the cumulative incidence is provided in SAS code A.1.

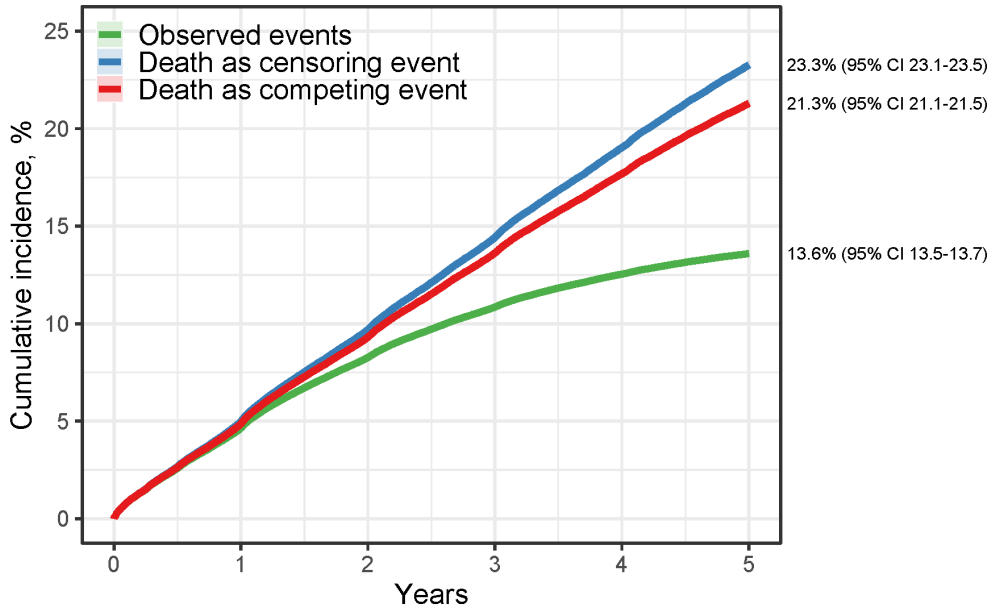


Figure A.1: Three estimates of the cumulative incidence curve. As in Study I, diabetes during follow-up was defined as either HbA1c-defined diabetes (HbA1c ≥ 48 mmol/mol), hospital contact with a diabetes diagnosis, or redemption of a prescription for glucose-lowering medication. The lowest estimate of the 5-year cumulative incidence, 13.6%, is estimated as $\sum_{j=1}^{i} \frac{d_i}{N}$ after five years. This estimate illustrates the number of individuals where we have *observed* the outcome. It does not take time into account and all individuals who are either censored or die during follow-up are considered free of diabetes after five years. In the Kaplan-Meier estimate of the 5-year cumulative incidence, 23.3%, censoring is taken into account. The cumulative incidence is overestimated because individuals with the competing event, death, are treated as if they were censored. Here, the most accurate estimate is the Aalen-Johansen estimate, 21.3%, where censoring is taken into account and death is considered a competing event. Adapted from Appendix I.

```
/* Censoring (Kaplan-Meier) */
/* cl using error = Aalen method and conftype = loglog */
proc lifetest data=data outcif=cif_diab plots=cif(test cl) atrisk;
        /*0 censoring, 1 outcome, 2 competing event */
        time CensYears*Outcome(0 2)/eventcode=1;
run;

/* Competing event (Aalen-Johansen) */
proc lifetest data=data outcif=cif_diab plots=cif(test cl) atrisk;
/*0 censoring, 1 outcome, 2 competing event */
/* (2 is not specified in the code, i.e., included as competing event) */
        time CensYears*Outcome(0)/eventcode=1;
run;
```

SAS code A.1: SAS code to estimate cumulative incidences. The variable `CensYears` includes time to outcome in years and `Outcome` is the outcome variable with `0` denoting censoring, `1` denoting the outcome of interest, and `2` denoting the competing event.

References

1. Wolbers M, Koller MT, Witteman JC, and Steyerberg EW: Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009;**20**:555–61.
2. Andersen PK, Geskus RB, Witte T de, and Putter H: Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol 2012;**41**:861–70.
3. SAS Institute Inc.: SAS/STAT 14.1 User's Guide The LIFETEST Procedure. SAS institute Incorporated, 2015.
4. Sutradhar R and Austin PC: Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. Ann Epidemiol 2018;**28**:54–7.
5. Austin PC and Fine JP: Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med 2017;**36**:4391–400.

Appendix B | Prediction and competing risk

The cumulative incidence estimated in Study I and Appendix A is a non-parametric estimate. In the prediction study[1], Study II, we aimed at estimating individualized probabilities by including how the cumulative incidence is associated with a set of explanatory variables, **X**. The cumulative incidence, $CIF(t) = 1 - \exp(-H(t))$, can be estimated using the Cox proportional hazards estimate for the hazard function, $h(t) = \lambda_0(t)\exp(\beta X)$, but this does not allow for death as a competing event. Instead, by extending the cause-specific hazard function in Equation (A.1), the Fine-Gray regression model can be used to link the cumulative incidence to the explanatory variables via the subdistribution hazard[2,3]

$$h_k^{sd}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t, K = k | T \geq t, \cup (T < t \cap K \neq k)) \tag{B.1}$$

and it is suitable for predictive modelling in a competing risk setting[4]. The subdistribution hazard function is the instantaneous rate of a given outcome among those who have not yet experienced an outcome of that specific type. The risk set includes the individuals who are still at risk of the event of interest and those who have already failed from the competing event prior to t[5]. The one-to-one relation to the cumulative incidence function, $CIF(t)$, has been re-established in the competing risk setting[3] and the cumulative incidence follows directly from the subdistribution hazard function.

Variable selection

We included a wide range of potential predictors (i.e., X) in our prediction study. We then used the least absolute shrinkage and selection operator (LASSO) to select a set a predictors[5] to include in the model. Similar to the Cox proportional hazards model, the subdistribution hazard function can be estimated using

$$\widehat{h_1^{sd}}(t|X) = \lambda_{01}(t)\exp(X\beta),$$

where $k = 1$ denotes the event of interest (diabetes), $\lambda_{01}(t)$ the baseline hazard for $k = 1$, β is the $d \times 1$ vector of regression coefficients and **X** the $N \times d$ matrix including the individual explanatory variables for each of the N individuals. The estimates for β can be found via maximum likelihood estimation, $\hat{\beta} = \arg\max l(\beta)$, where l denotes the log likelihood function[5]. As we wanted to select only a subset

of the variables, we used the penalization function

$$Q(\beta) = l(\beta) + N \sum_{j=1}^{d} \lambda |\beta_j|$$

where λ is the tuning parameter (selected using the BIC and 1,000 iteration). The estimate is obtained using $\tilde{\beta} = \arg\max Q(\beta)$. We then refitted the subdistribution hazards function estimating using only the predictors for which $\tilde{\beta} > 0$. The variable selection was performed using the R package `crrp`[5]. A simplified version of the code is provided in R code B.1.

```
######################
# Variable selection #
######################
LASSO_48 <- crrp(
  time = LASSO_data$fu_years_48,
  fstatus = LASSO_data$fu_status_48,
  failcode = 1,
  cencode = 2,
  X = LASSO_data,
  penalty = "LASSO"
)

# Find the right coefficients - use BIC to select tuning parameters
which.min(LASSO_48$BIC)
beta_48 <- LASSO_48$beta[, which.min(LASSO_48$BIC)]
data.frame(beta_48[!data.frame(beta_48) == rep(0, length(beta_48))])
# save only those not equal to 0
design_mat_48_pen <- names(which(!beta_48 == rep(0, length(beta_48))))


# Use the variables to fit a new model (from which prediction is possible).
# Input in FGR is a formula
# - create formula based on the variables in the selection
LASSO_formula_48 <- as.formula(
  paste("Hist(fu_years_48,␣fu_status_48,␣cens.code='2')",
        paste(sort(design_mat_48_pen)[!sort(design_mat_48_pen)
        %in%
        tolower(c("cpr", "fu_status_48", "fu_years_48", "fu_status_a10",
                "fu_years_a10"))], collapse = "+"), sep = "␣~␣")
)

# Fit FG main model
valid_FG_main_48 <- FGR(
  formula = LASSO_formula_48,
  data = LASSO_data,
  cause = 1
)
```

R code B.1: R code to perform variable selection using the `crrp` package.

Using the estimated set of β-values from the refitted subdistribution hazards function along with the Breslow-type estimate of the underlying subdistribution hazard ($\beta_0 = \lambda_{01}(t)$) gives the prediction model[1]

$$P(T \leq t) = 1 - \exp\left(-\hat{\beta}_0 \exp\left(\hat{\beta}_1 X_1 + \cdots + \hat{\beta}_b X_b\right)\right)$$

including b predictors for the $t = 5$ year risk of developing diabetes, when death is considered a competing event.

Validation

As with any other prediction model, the prediction model developed in Study II should be validated before it can potentially be used in clinical practice. The competing risk of death must also be taken into account in the validation[6]. We used the time-dependent receiver operating characteristic (ROCt) curves and the time-dependent area under the curve (AUCt) to assess the discrimination[7,8]. In addition to a visual assessment of the calibration (using a calibration curve), we used the Brier score and the index of prediction accuracy (IPA) to consider the calibration and discrimination at the same time[8,9]. All measures were evaluated at the fixed time, $t = 5$.

The AUCt compares the probabilities for diabetes, p_i and p_j, for two individuals, i and j, where the individual i (case) had the outcome of interest (diabetes) and individual j (control) did not. For individual j, we must take into account whether it was potentially because of censoring (after which we were not able to observe a potential case of diabetes) or potentially because of the competing event (death) during follow-up. As in Equation (B.1), the risk set must consist of those who are either still at risk of diabetes or who has already experienced the competing event. Using this definition, AUCt can be defined as[8]

$$AUCt(t) = P(p_i > p_j | T_i \leq t, k = 1, \{T_j > t\} \cup \{T_j \leq t, k \neq 1\}),$$

and individuals with the competing event are thus considered part of the controls in AUCt. As the event time, T_j, is unknown for individuals who are censored, the AUCt is estimated using inverse probability of censoring weighting with Kaplan-Meier estimated weights[6,8].

The Brier score is a measure of accuracy and it reflects both calibration and discrimination[9]. It is the expected quadratic distance between then event indicator and the estimated risk, $\mathbb{E}\left[\mathbb{1}_{(T \leq t, k=1)} - p\right]^2$, estimated using the inverse probability of censoring weights[6]. The Brier score can be rescaled into the index of prediction accuracy (IPA)[9], a measure of average performance, where the Brier score for the prediction model is related to the model without any covariates[6]. For further details, a comprehensive overview of performance measures for prediction models in the competing event setting has recently been published, see van Geloven et al.[6].

The validation was performed using the R packages `riskRegression` and `timeROC`. A simplified subset of the code is provided in R code B.2.

```
###############
# Validation #
###############
#include age and sex.
valid_FG_mini_48 <- FGR(
  formula = Hist(fu_years_48,fu_status_48, cens.code='2') ~ age + sex,
  data = data,
  cause = 1 #diabetes
)

## Baseline hazards
### Main model
max(cumsum(valid_FG_main_48$crrFit$bfitj))
### Minimal model
max(cumsum(valid_FG_mini_48$crrFit$bfitj))

################################
# calculate time-dependent AUC #
################################
AUC_valid_48 <- Score(
  list(
    "Main_model" = valid_FG_main_48, #main model (all predictors)
    "Minimum_model" = valid_FG_mini_48 #minimim model (age+sex)
  ),
  formula = Hist(fu_years_48, fu_status_48, cens.code = "2") ~ 1,
  cens.model = "km", cens.method = "ipcw",
  split.method = "none",
  data = data, # use the full data
  times = c(5), # validate only after 5 years
  cause = 1,
  percent = FALSE,
  plots = c("Calibration", "ROC"),
  metrics = c("auc", "brier"),
  summary = c("risks", "IPA", "riskQuantile", "ibs")
)

### Calibration
plotCal_48 <- plotCalibration(AUC_valid_48,
                method = "quantile",
                cens.method = "local",
                times = c(5),
                auc.in.legend = TRUE,
                brier.in.legend = TRUE,
                ylim = c(0, 0.7),
                xlim = c(0, 0.7))

### ROC
plotROC_48 <- plotROC(AUC_valid_48,
                times = c(5),
                auc.in.legend = TRUE,
                brier.in.legend = TRUE)
ROC_48 <-plotROC_48$ROC$plotframe %>% filter(times == 5)
```

R code B.2: R code to perform model validation of the prediction model using the `riskRegression` and `timeROC` packages.

References

1. Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, and Pedersen L: Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. BMJ Open Diabetes Res Care 2022;**10**.
2. Andersen PK, Geskus RB, Witte T de, and Putter H: Competing risks in epidemiology: possibilities and pitfalls. Int J Epidemiol 2012;**41**:861–70.
3. Austin PC and Fine JP: Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med 2017;**36**:4391–400.
4. Wolbers M, Koller MT, Witteman JC, and Steyerberg EW: Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009;**20**:555–61.
5. Fu Z, Parikh CR, and Zhou B: Penalized variable selection in competing risks regression. Lifetime Data Anal 2017;**23**:353–76.
6. van Geloven N, Giardiello D, Bonneville EF, et al.: Validation of prediction models in the presence of competing risks: a guide through modern methods. BMJ 2022;**377**:e069249.
7. Blanche P, Kattan MW, and Gerds TA: The c-index is not proper for the evaluation of t-year predicted risks. Biostatistics 2019;**20**:347–57.
8. Blanche P, Dartigues JF, and Jacqmin-Gadda H: Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. Stat Med 2013;**32**:5381–97.
9. Kattan MW and Gerds TA: The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res 2018;**2**:7.

Appendix B

Appendix C | Latent class trajectory analyses

In the trajectory study, Study III, we used the latent class trajectory analyses to identify distinct trajectories of HbA1c in the five years prior to treatment initiation[1]. The analyses combined elements from mixed-effects modelling, spline regression, latent class analyses, and stochastic smoothing. HbA1c was modelled as a function of time (days) before treatment initiation.

Mixed-effects models

In a general linear model, the observed data are explained by explanatory variables and a set of unknown independent and identically distributed normal random errors with mean 0 and variance σ^2. This is often too restrictive. The mixed models extends this by allowing a flexible specification of the covariance matrix rather than assuming independence[2].

The mixed-effects model is written as a combination of fixed-effects and random-effects using

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{C.1}$$

where $\mathbf{y}$ is the observed data, i.e., the $M \times 1$ vector including the M observations (HbA1c measurements), with y_{ij} being the $j'th$ observation (in time), $j = 1,\ldots,m_i$, for individual i, $i = 1,\ldots,n$. $\mathbf{X}$ is the $M \times p_x$ matrix including the explanatory variables (e.g., defined using splines), $\boldsymbol{\beta}$ is the $p_x \times 1$ vector of unknown fixed-effects parameters, $\mathbf{Z}$ is the $M \times p_z n$ design matrix for the random effects, $\boldsymbol{\gamma}$ is the $p_z n \times 1$ vector of unknown random-effects parameters, and $\boldsymbol{\varepsilon}$ is the $M \times 1$ vector of error terms. In a model with random intercept and random slope, $p_z = 2$ and $\mathbf{Z}$ is specified as an $M \times 2n$ block diagonal matrix with blocks consistent with the individuals. Outside the diagonal blocks all blocks equal $\mathbf{0}$ and

$$\mathbf{z_i} = \begin{bmatrix} 1 & z_{i1} \\ \vdots & \vdots \\ 1 & z_{im_i} \end{bmatrix}$$

on the diagonal block, with 1 fitting the random intercept and $z_{ij} = x_{ij}$ the random slope. This also implies $\boldsymbol{\gamma}$ to be a $2n \times 1$ vector with

$$\boldsymbol{\gamma_i} = \begin{bmatrix} \gamma_{i0} \\ \gamma_{i1} \end{bmatrix}$$

for each individual, with γ_{i0} being the individual random intercept parameter and γ_{i1} the individual random slope parameter.

In addition, it is assumed that $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ are normally distributed with

$$\mathrm{E}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$$
$$\mathrm{Var}\begin{bmatrix}\boldsymbol{\gamma}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

with $\mathbf{G}$ being the $p_z n \times p_z n$ variance-covariance matrix for $\boldsymbol{\gamma}$, and $\mathbf{R}$ the $n \times n$ variance-covariance matrix for the error terms, $\boldsymbol{\varepsilon}$. With this structure, the variance of $\mathbf{y}$ is therefore determined as $\mathrm{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ with $'$ denoting the transposed matrix. Specifying $\mathbf{G}$ as a block diagonal matrix with $\mathbf{0}$ outside the block diagonals (to indicate that individuals are independent) and

$$diag(\mathbf{G}) = \begin{bmatrix}\sigma_0^2 & \sigma_{01}\\ \sigma_{01} & \sigma_1^2\end{bmatrix}$$

in the diagonal blocks, i.e., an unspecified variance-covariance structure, and $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I_n}$ with $\mathbf{I_n}$ denoting the $n \times n$ identity matrix and σ_ε^2 the residual standard error, allows measurements to be correlated as a function of time and accounts for repeated measurements from the same individual. The variance for the $j'th$ measurement for $i'th$ individual and the covariance between the $j'th$ and $k'th$ measurement for the $i'th$ individual will be

$$\mathrm{Var}(y_{ij}) = \sigma_0^2 + z_{ij}^2\sigma_1^2 + 2z_{ij}\sigma_{01} + \sigma_\varepsilon^2$$
$$\mathrm{Cov}(y_{ij}, y_{ik}) = \sigma_0^2 + z_{ij}\sigma_{01} + z_{ik}\sigma_{01} + z_{ij}z_{ik}\sigma_1^2 + \sigma_\varepsilon^2.$$

Using maximum likelihood estimation, the standard method is to solve the model equations and obtain the estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ using the equations

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$$
$$\hat{\boldsymbol{\gamma}} = \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}),$$

which requires the estimates for the 4 parameters $(\sigma_0^2, \sigma_1^2, \sigma_{01}, \sigma_\varepsilon)$. See for example SAS Institute Inc.[2] for further details and extensions of the mixed models.

Restricted cubic splines

To allow for flexibility in the trajectories of HbA1c, the values are fitted using restricted cubic splines. This implies that instead of including time in $\mathbf{X}$, a function of time, $f(X)$, will be included and X is then expressed as piecewise polynomials[3].

The restricted cubic spline function with k pre-specified knots given by $t_1, \ldots, t_k$ is determined by

$$f(X) = \nu_0 + \nu_1 X_1 + \cdots + \nu_{k-1} X_{k-1}$$

where $X_1 = X$, and for $j = 1, \ldots, k-2$,

$$X_{j+1} = (X - t_j)_+^3 - (X - t_{j-1})_+^3 \frac{t_k - t_j}{t_k - t_{k-1}} + (X - t_k)_+^3 \frac{t_{k-1} - t_j}{t_k - t_{k-1}}$$

where $(x-y)_+ = (x-y)\mathbb{1}_{(x>y)}$. After estimating $\nu_0, \ldots, \nu_{k-1}$, the restricted cubic spline can then be calculated using

$$f(X) = \nu_0 + \nu_1 X + \nu_2 (X - t_1)_+^2 + \cdots + \nu_{k+1}(X - t_k)_+^2$$

with

$$\nu_k = \frac{\nu_2(t_1 - t_k) + \nu_3(t_2 - t_k) + \cdots + \nu_{k-1}(t_{k-2} - t_k)}{(t_k - t_{k-1})}$$
$$\nu_{k+1} = \frac{\nu_2(t_1 - t_{k-1}) + \nu_3(t_2 - t_{k-1}) + \cdots + \nu_{k-1}(t_{k-2} - t_{k-1})}{(t_{k-1} - t_k)}$$

and $\nu_2, \ldots, \nu_{k-1}$ rescaled by $(t_k - t_1)^2$.

The number of knots, k, and the location of the k knots, $t_1, \ldots, t_k$, must be pre-specified. In our analyses, the number of knots varied from 3 to 5, and the location was defined either based on quantiles of data or to split data in equal time intervals. When including splines in the regression, it requires the parameter estimates for the k parameters $(\nu_0, \nu_1, \ldots, \nu_{k-1})$. The splines are fitted using the R package `rms`. Theoretical details can be found in Harrell[3].

Latent classes

Fitting the mixed-effects model with a restricted cubic spline allows for the model to flexibly estimate the trajectory while accounting for individual differences and the structure of data. It is still a "one-size-fits-all" model, as all individuals are bound by the same set of model parameters, assuming that the population is homogeneous[1]. By introducing latent classes, multiple sets of model parameters are included, assuming that the population is heterogeneous and represented by H latent classes of individuals characterized by H mean profiles of trajectories[1].

The latent class analyses extend the mixed-effects model in Equation (C.1) by introducing the latent classes. Rather than fitting a single set of fixed effects, the class-specific fixed effects are fitted:

$$\mathbf{y}_h = \mathbf{X}\boldsymbol{\beta}_{\boldsymbol{h}} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{C.2}$$

assuming $h = 1, \ldots, H$ latent classes and allowing covariates to vary across latent classes. Individuals are assigned to each of the classes with a certain probability, but each individual belongs to one and only one latent class. The probability is described using a multinomial logistic model

$$P(c_i = h) = \pi_{ih} = \frac{e^{\xi_{0h}}}{\sum_{l=1}^{H} e^{\xi_{0l}}}$$

with c_i being the class-membership variable for individual i that equals h if the individual belongs to class h. Class-membership probabilities are assumed not to depend on any additional covariates and, for convenience[1], $\xi_{0H} = 0$.

The standard linear mixed model (i.e., the 1-class model) with a restricted cubic spline required the estimation of the four parameters $(\sigma_0^2, \sigma_1^2, \sigma_{01}, \sigma_\varepsilon)$ from the random effects and the k parameters $(\nu_0, \nu_1, \ldots, \nu_{k-1})$ from the splines (fixed effects). Extending to include latent classes then requires estimation of four parameters $(\sigma_0^2, \sigma_1^2, \sigma_{01}, \sigma_\varepsilon)$ from the random effects, along with the kH parameters $(\nu_{h_0}, \nu_{h_1}, \ldots, \nu_{h_{k-1}})$ from the splines from each latent class, and the $H-1$ parameters $(\xi_{01}, \ldots, \xi_{0H-1})$ from the class-membership model. The latent classes are fitted using the R package `lcmm`. The R code used to fit the model is presented in R code C.1. For further details, please see Proust-Lima et al.[1].

The posterior probabilities, i.e., the probability for each individual of belonging to each of the classes, can be calculated after the model has been fitted. The calculations use the model parameters for each of the H classes and the density function for the multivariate normal distribution defined by the model. The `hlme` output already includes posterior probabilities for individuals included in data, whereas the R code in R code C.2 shows a way to calculate the posterior probabilities for new individuals. As our models were fitted based on one third of the individuals, we needed to calculate the posterior probabilities for the remaining two thirds of the individuals ourselves. Again, for further details, please see Proust-Lima et al.[1].

```
M_rcs_quant_k3_rndsi_ng4 <- gridsearch(
  rep = 50, maxiter = 50, minit = model1,
  hlme(
    fixed = hba1c ~ rcs(t, parms = c(2,17,36)),
    mixture = ~ rcs(t, parms = c(2,17,36)),
    random = ~ t,
    subject = "cpr",
    ng = 4,
    idiag = FALSE,
    cor = NULL,
    data = Mdata
  )
)
```

R code C.1: R code to fit the latent class trajectory model included in Study III. `model1` is the 1-class model (i.e., ng=1) and `Mdata` is the dataset including the variables `cpr`, `t`, and `hba1c`. The function `rcs` fits the restricted cubic spline with knots equal to 2, 17, and 36.

```
#Posterior probabilities,
#lcmm paper (Proust-Lima, 2017) equations (6), (15), and (18)

#Model
as.list(M_rcs_quant_k3_rndsi_ng4$best)
# $‘intercept class1‘
# [1] 1.559334
# $‘intercept class2‘
# [1] -0.3525771
```

```
# $‘intercept class3‘
# [1] -1.919241
# $‘intercept class1‘
# [1] 51.86207
# $‘intercept class2‘
# [1] 104.6578
# $‘intercept class3‘
# [1] 136.9756
# $‘intercept class4‘
# [1] 78.75877
# $‘rcs(t, parms = c(2, 17, 36))t class1‘
# [1] -0.4434989
# $‘rcs(t, parms = c(2, 17, 36))t class2‘
# [1] -3.988636
# $‘rcs(t, parms = c(2, 17, 36))t class3‘
# [1] -6.591044
# $‘rcs(t, parms = c(2, 17, 36))t class4‘
# [1] -2.11042
# $‘rcs(t, parms = c(2, 17, 36))t’ class1‘
# [1] 0.2945746
# $‘rcs(t, parms = c(2, 17, 36))t’ class2‘
# [1] 3.262207
# $‘rcs(t, parms = c(2, 17, 36))t’ class3‘
# [1] 5.563063
# $‘rcs(t, parms = c(2, 17, 36))t’ class4‘
# [1] 1.6657
# $‘varcov 1‘
# [1] 77.82992
# $‘varcov 2‘
# [1] -0.9839712
# $‘varcov 3‘
# [1] 0.01721758
# $stderr
# [1] -4.726917


# Class-specific estimates
class1_est <- list(
  xi = c(M_rcs_quant_k3_rndsi_ng4$best[1]),
  beta = M_rcs_quant_k3_rndsi_ng4$best[c(4, 8, 12)]
)
class2_est <- list(
  xi = c(M_rcs_quant_k3_rndsi_ng4$best[2]),
  beta = M_rcs_quant_k3_rndsi_ng4$best[c(5, 9, 13)]
)
class3_est <- list(
  xi = c(M_rcs_quant_k3_rndsi_ng4$best[3]),
  beta = M_rcs_quant_k3_rndsi_ng4$best[c(6, 10, 14)]
)
class4_est <- list(
  xi = 0,
  beta = M_rcs_quant_k3_rndsi_ng4$best[c(7, 11, 15)]
)

# Estimates from random effects
Bg <- matrix(c(
```

```
  M_rcs_quant_k3_rndsi_ng4$best[16], M_rcs_quant_k3_rndsi_ng4$best[17],
  M_rcs_quant_k3_rndsi_ng4$best[17], M_rcs_quant_k3_rndsi_ng4$best[18]
), ncol = 2)
sig_eps <- M_rcs_quant_k3_rndsi_ng4$best[19] # residual standard error

# Class-membership model
eps_0_g1 <- as.numeric(M_rcs_quant_k3_rndsi_ng4$best[1]) # class-est_xi
eps_0_g2 <- as.numeric(M_rcs_quant_k3_rndsi_ng4$best[2])
eps_0_g3 <- as.numeric(M_rcs_quant_k3_rndsi_ng4$best[3])
eps_0_g4 <- 0

# class probabilites (overall)
pi_1<-exp(eps_0_g1)/(exp(eps_0_g1)+exp(eps_0_g2)+exp(eps_0_g3)+exp(eps_0_g4))
pi_2<-exp(eps_0_g2)/(exp(eps_0_g1)+exp(eps_0_g2)+exp(eps_0_g3)+exp(eps_0_g4))
pi_3<-exp(eps_0_g3)/(exp(eps_0_g1)+exp(eps_0_g2)+exp(eps_0_g3)+exp(eps_0_g4))
pi_4<-exp(eps_0_g4)/(exp(eps_0_g1)+exp(eps_0_g2)+exp(eps_0_g3)+exp(eps_0_g4))
pi_1+pi_2+pi_3+pi_4


individuals <- r_cpr[, 1] # Full list of individuals

loop_list <- list()
for (i in 1:length(individuals)) {
  print(i)
  Idata <- Mdata %>%
    filter(cpr %in% individuals[i])

  # Xi is the design matrix
  # knots from model
  Xi <- matrix(c(rep(1, length(Idata$t)), Idata$t,
                 rcspline.eval(Idata$t, knots = c(2, 17, 36))),
                 ncol = 3)
  mui1 <- Xi %*% class1_est$beta
  mui2 <- Xi %*% class2_est$beta
  mui3 <- Xi %*% class3_est$beta
  mui4 <- Xi %*% class4_est$beta

  Zi <- matrix(c(rep(1, length(Idata$t)), Idata$t), ncol = 2)
  Vig <- Zi %*% Bg %*% t(Zi) + diag(length(Zi[, 1])) * sig_eps^2

  dmv_i1 <- dmvnorm(
    x = matrix(Idata$hba1c, ncol = length(Idata$hba1c)),
    mean = as.vector(mui1),
    sigma = Vig
  )
  dmv_i2 <- dmvnorm(
    x = matrix(Idata$hba1c, ncol = length(Idata$hba1c)),
    mean = as.vector(mui2),
    sigma = Vig
  )
  dmv_i3 <- dmvnorm(
    x = matrix(Idata$hba1c, ncol = length(Idata$hba1c)),
    mean = as.vector(mui3),
    sigma = Vig
  )
  dmv_i4 <- dmvnorm(
```

```
    x = matrix(Idata$hba1c, ncol = length(Idata$hba1c)),
    mean = as.vector(mui4),
    sigma = Vig
  )

  postprop_i1<-(pi_1*dmv_i1)/(pi_1*dmv_i1+pi_2*dmv_i2+pi_3*dmv_i3+pi_4*dmv_i4)
  postprop_i2<-(pi_2*dmv_i2)/(pi_1*dmv_i1+pi_2*dmv_i2+pi_3*dmv_i3+pi_4*dmv_i4)
  postprop_i3<-(pi_3*dmv_i3)/(pi_1*dmv_i1+pi_2*dmv_i2+pi_3*dmv_i3+pi_4*dmv_i4)
  postprop_i4<-(pi_4*dmv_i4)/(pi_1*dmv_i1+pi_2*dmv_i2+pi_3*dmv_i3+pi_4*dmv_i4)
  ppclass <- which.max(c(postprop_i1, postprop_i2, postprop_i3, postprop_i4))

  loop_list[[i]] <- c(individuals[i],
                      postprop_i1, postprop_i2, postprop_i3, postprop_i4,
                      ppclass)
}
loop_full <- data.frame(do.call(rbind, loop_list))
names(loop_full) <- c("cpr", "pp1", "pp2", "pp3", "pp4", "ppclass")
```

R code C.2: R code to calculate the posterior probabilities based on the model in Study III.

References

1. Proust-Lima C, Philipps V, and Liquet B: Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. J Stat Softw 2017;**78**:1–56.
2. SAS Institute Inc.: SAS/STAT 13.1 User's Guide The MIXED Procedure. SAS institute Incorporated, 2013.
3. Harrell F. E. J: Regression Modeling Strategies. 2nd ed. Vol. 2. Springer Series in Statistics. Springer, 2015.

Appendix C

Appendix D | Regression discontinuity design

In the regression discontinuity design (RDD) study, Study IV, we used the RDD to study the effect of being eligible for type 2 diabetes treatment on cardiovascular disease (CVD) and death[1]. Glucose-lowering treatment should be initialized when the HbA1c level exceeds 48 mmol/mol (Section 2.3) and the design uses this as a randomizing device. The design was first introduced in 1960 by Thistletwaite and Cambell[2] and is thus not a new design. In spite of this, it has only gained currency in the field of epidemiology within the last decades. In the original study, the design was used to study the effect of students' grades leading to honorary awards and scholarships[2,3], but the design can potentially be used in any situation where a continuous marker, be it grades in an exam, kilometres from a border, birth year, tax limits, or HbA1c values, is used along with a pre-specified threshold defining the choice of intervention. Individuals whose values are close to the threshold are expected to be similar and it is random variability caused by measurement error, sampling variance, and chance factors[4] that ultimately ascertains whether the individual ends below or above the threshold.

Causal effect assumptions

The aim of the study is to estimate whether treatment eligibility affects the rate of future CVD and death. As with other study designs, the RDD requires some assumptions in order to ensure the estimation of a causal effect[3,5–7]. The RDD is a special case of instrumental variable (IV) analyses and the assumptions are used to nest the RDD within the IV set-up[8].

Let X be the assignment variable the treatment guidelines rely on, i.e., the HbA1c value. This value must be measured and reported continuously[5]. Let x_0 be the threshold value, i.e., HbA1c 48 mmol/mol. Let D be the threshold indicator, i.e., $D = \mathbb{1}_{(X \geq x_0)}$. Let Y be the outcome variable. It must be observed for all individuals, independent of whether or not they were assigned the treatment[5]. The RDD can be used with continuous, binary, count, and time-to-event outcomes. In Study IV, we include death and CVD in a time-to-event analysis. Let $\mathbf{C} = \{\mathbf{O} \cup \mathbf{U}\}$ be the set of confounders on the path between treatment eligibility and CVD and death, where $\mathbf{O}$ are the observed and $\mathbf{U}$ are the unobserved (and potentially also unknown) confounders. In addition, let T be the treatment indicator denoting whether or not an individual received the treatment of interest, i.e., $T = \mathbb{1}_{(\text{"treatment"})}$. In our study, "treatment" does not refer to glucose-lowering treatment in sense of an actual pill, but rather being eligible for treatment and everything that includes, e.g., increased monitoring, cardiovascular risk factor control, glucose-lowering treatment, etc.

(Section 2.3). We are therefore estimating the effect of the treatment guidelines, not the effect of the actual treatment (the pill) itself. As a consequence, our set-up allows for the special case where $T = D$ (in the threshold) as everyone with an HbA1c value above the treatment threshold are "eligible for treatment" and no one with a value below are "eligible for treatment" (according to guidelines). This implies that $\lim_{x \downarrow x_0} P(T = 1 \mid X = x) \neq \lim_{x \uparrow x_0} P(T = 1 \mid X = x)$ and the probability of receiving treatment is therefore discontinuous in the threshold. One of the main assumptions in RDD is that the treatment is associated with the treatment guidelines and thereby the threshold indicator, i.e., $T \not\perp\!\!\!\perp D$, which is obviously true in our study. In addition, close to the threshold, the threshold indicator must not rely on anything else than the assignment variable, i.e., $D \perp\!\!\!\perp \mathbf{C} \mid X$, and individual characteristics (confounders) cannot determine whether an individual has a measurement above or below the threshold. Furthermore, the threshold indicator and the outcome must not be associated, conditional on everything else, i.e., $D \perp\!\!\!\perp Y \mid (T, X, C)$. This ensures the value of HbA1c does not itself cause the outcome. It also requires that the value of the HbA1c cannot be changed for the sole purpose of changing treatment eligibility. As HbA1c is a biomarker, this is not an issue in our studies as neither the individual nor the doctor (or anyone else for that matter) can deliberately change the HbA1c value. This is also supported by our histograms (Figure 5.1, Appendix IV). The assumption also ensures there is randomness in whether an individual's HbA1c is above or below the threshold and prompts individuals to be similar on both sides on the threshold. In order to check this, known characteristics on both sides of the threshold should be compared. The only way D can affect Y is via T and for the RDD to hold, we want any effect on the outcome to be relying on the threshold indicator — and not any other potential reason. This means that the expected outcome should be continuous around the threshold, conditional on everything else, i.e., $\mathrm{E}[Y \mid D, X = x, T, \mathbf{C}]$ is continuous in x (at x_0) for $T = 0, 1$. Continuity at the threshold may seem like a strong assumption, but it follows from the random noise introduced by the measurement error[4]. The assumption guarantees it is the threshold indicator and not anything else that causes the discontinuity in $\mathrm{E}[Y \mid X = x]$, which is exactly the discontinuity we want to estimate. The assumptions can be summarized as[3]

- $D \not\perp\!\!\!\perp T$
- $D \perp\!\!\!\perp \mathbf{C} \mid X$
- $D \perp\!\!\!\perp Y \mid (T, X, C)$
- $\mathrm{E}[Y \mid D, X = x, T, \mathbf{C}]$ is continuous in x (at x_0) for $T = 0, 1$,

and they ensure an IV set-up, where the only way having HbA1c above or below the threshold can affect CVD and death is via treatment eligibility. By design, confounding, both observed, unobserved, and unknown, will thus be eliminated (in expectation). To check the assumptions, it is recommended to plot the assignment variable (e.g., a histogram of HbA1c), to tabulate the characteristics on both sides of the threshold, and to plot the outcome as a function of the assignment variable (e.g., the outcome rate per HbA1c)[5].

Estimation

When the feasibility of the RDD has been established, the objective is to estimate the discontinuity in $\mathrm{E}[Y \mid X = x]$ in the threshold. The RDD estimates an average causal effect (ACE)[8] in the threshold[6]

$$ACE = \lim_{x \downarrow x_0} \mathrm{E}[Y \mid X = x] - \lim_{x \uparrow x_0} \mathrm{E}[Y \mid X = x],$$

which is an ITT estimate[4] and is a measure of the effect of treatment eligibility. This is exactly the aim in our study. We are interested in the causal effect on death and CVD estimated as a hazard ratio (HR) using a Cox proportional hazards model. The hazard ($h(t) = \lambda_0(t)\exp(X\beta)$ estimated as a function of time, t, but suppressed to simplify the notation) ratio is estimated as a ratio of the hazard among the *exposed*, $h^{above}(y)$, i.e., individuals whose HbA1c value was just above the threshold, and the hazard among the *unexposed*, $h^{below}(y)$, i.e., individuals whose HbA1c values was just below the threshold:

$$\begin{aligned} h^{below}(y \mid D = 0, X, x_0) &= \lambda_0^b(y)\exp\left(\beta^b(X - x_0)\right) \\ h^{above}(y \mid D = 1, X, x_0) &= \lambda_0^a(y)\exp\left(\beta^a(X - x_0)\right). \end{aligned}$$

In the threshold, i.e., when $x \to x_0$ is reached as a limit from both sides, the difference between the logarithms plays the role of the estimate of the treatment effect. The limits exist because of the continuity in the threshold:

$$\begin{aligned} ITT &= \log\left(\frac{\text{"exposed"}}{\text{"unexposed"}}\right) \\ &= \log\left(\frac{\lim_{x \downarrow x_0} h^{above}(y \mid D = 1, X, x_0)}{\lim_{x \uparrow x_0} h^{below}(y \mid D = 0, X, x_0)}\right) \\ &= \lim_{x \downarrow x_0} \log\left(\lambda_0^a(y)\exp\left(\beta^a(X - x_0)\right)\right) - \lim_{x \uparrow x_0} \log\left(\lambda_0^b(y)\exp\left(\beta^b(X - x_0)\right)\right) \\ &= \log\left(\lambda_0^a(y)\right) - \log\left(\lambda_0^b(y)\right) \\ &= \beta_2, \end{aligned}$$

where β_2 is defined to be the difference between the intercepts from the hazard functions. The overall log hazard function is calculated by dividing it into a section below and a section above the threshold, i.e.,

$$\begin{aligned} \log(h(y \mid X)) &= (1 - \mathbb{1}_{(X \geq x_0)})\log\left(h^{below}(y \mid D = 0, X, x_0)\right) + \mathbb{1}_{(X \geq x_0)}\left(h^{above}(y \mid D = 1, X, x_0)\right) \\ &= \log\left(\lambda_0^b(y)\right) - \mathbb{1}_{(X \geq x_0)}\log\left(\lambda_0^b(y)\right) + \beta^b(X - x_0) - \mathbb{1}_{(X \geq x_0)}\beta^b(X - x_0) \\ &\quad + \mathbb{1}_{(X \geq x_0)}\log\left(\lambda_0^a(y)\right) + \mathbb{1}_{(X \geq x_0)}\beta^a(X - x_0) \\ &= \log\left(\lambda_0^a(y)\right) + \beta^b(X - x_0) + \mathbb{1}_{(X < x_0)}\left(\log\left(\lambda_0^b(y)\right) - \log\left(\lambda_0^a(y)\right)\right) \\ &\quad + \mathbb{1}_{(X > x_0)}(X - x_0)(\beta^a - \beta^b) \\ &= \beta_0 + \beta_1(X - x_0) + (-\beta_2)\mathbb{1}_{(X < x_0)} + \beta_3\mathbb{1}_{(X > x_0)}(X - x_0), \end{aligned}$$

where β_0 is the logarithm of the baseline hazard above the threshold (unknown), $\beta_1 = \beta^b$ is the slope of the regression below the threshold, β_2 is the difference between the intercepts in the threshold (as defined above) and $\beta_3 = (\beta^a - \beta^b)$ ensures $\beta_3 + \beta_1 = \beta^a$ is the slope of the regression above the threshold (Figure D.1). This is the parametrization used in Equation (4.1), where β_2 is the estimate of the difference in the threshold and thus the desired estimate. The SAS code to estimate the effect is provided in SAS code D.1.

```
data ITT;
        set full_data;
        beta_1 = Hba1cValue - 48;
        beta_2 = (Hba1cvalue < 48);
        beta_3 = (Hba1cValue - 48) * (Hba1cValue > 48);
run;

proc phreg data=ITT(where=(42<=Hba1cValue<=53));
        model CensYears*Outcome(0)=beta_1 beta_2 beta_3 /risklimits ties=efron;
run;
```

SAS code D.1: SAS code to estimate the treatment effect in Study IV. The variable `CensYears` includes time to outcome in years and `Outcome` is the outcome variable with `0` denoting censoring and `1` denoting an outcome. The dataset `ITT` includes individual level information about HbA1c (`Hba1cValue`) and the outcome. In the regression, data are restricted to only include data within a reasonable small interval around the threshold in order for the assumptions to hold. In our case, data is restricted to values in the interval 42–53 mmol/mol, i.e., $42 \leq$ `Hba1cValue` ≤ 53.

Extrapolation

The RDD estimate relies on an extrapolation to the threshold, where the difference between the limits of the regression lines is the estimate of the treatment effect. In Figure D.1, the regressions are extrapolated beyond the threshold to visualize the counterfactuals[4,8]. The green dashed line illustrates what would have happened if none of the individuals whose HbA1c was above the threshold had received the treatment of interest, i.e., if the guidelines had not existed. The difference between the green dashed line and the blue line is therefore the visual effect of the treatment guidelines in the entire interval. But what would happen if the extrapolation is reverted, i.e., if the blue line is extrapolated, and we therefore act as if everyone whose HbA1c was below the threshold had actually received the treatment? The green line and the blue dashed line intersect when HbA1c= 44.68 mmol/mol. This means, for any individual whose value is below 44.7 mmol/mol, the best treatment strategy would be no treatment and similarly, for any individual whose value is above 44.7 mmol/mol, the best treatment strategy would be active treatment. If this holds, the treatment guideline threshold, 48 mmol/mol, should be lowered to 44.7 mmol/mol. It is, however, important to notice that the extrapolations are sensitive to violations of model assumptions and this should therefore only be seen as ideas for future research rather than specific results.

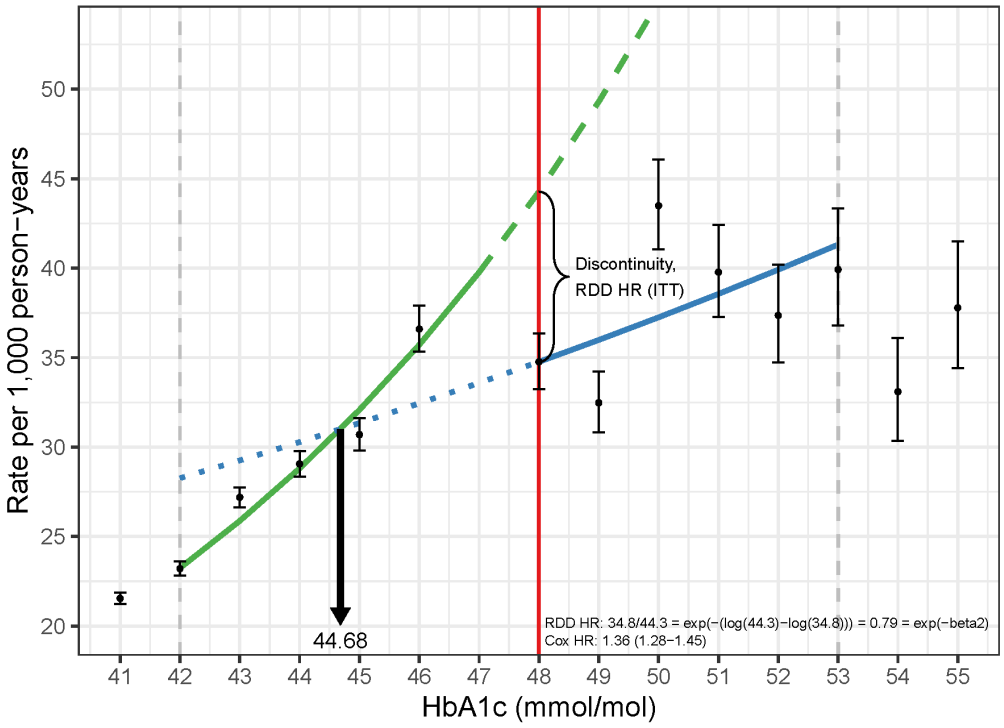


Figure D.1: Regression discontinuity from Study IV. The dots show the incidence rate per 1,000 person-years for each value of HbA1c. The red line is the HbA1c threshold, 48 mmol/mol. The range used for the RDD analysis is indicated by the grey dashed lines, 42–53 mmol/mol. The green line shows the RDD regression line (hazard rate) below the threshold and the blue line shows the RDD regression line above the threshold. In the threshold, the difference between the blue line and green line is the ITT estimate. It equals HR 0.79 (95% CI 0.69–0.90). On the other hand, a traditional (crude/unadjusted) estimate from the Cox proportional hazards model yields the result HR 1.36 (95% CI 1.28–1.45) for individuals above vs. below the threshold. The green dashed line and the blue dashed line are the extrapolations of the regressions below and above the threshold. The extrapolation of the regression estimated based on data above the threshold intersects the estimated regression based on data below the threshold when HbA1c= 44.68 mmol/mol. Adapted from Petersen et al.[1].

References

1. Petersen I, Nicolaisen SK, Ricciardi F, Sharma M, Thomsen RW, Baio G, and Pedersen L: Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study. Clin Epidemiol 2020;**12**:569–77.
2. Thistlethwaite DL and Campbell DT: Regression-discontinuity analysis: An alternative to the ex post facto experiment. Journal of Educational Psychology 1960;**51**:309–17.
3. Geneletti S, O'Keeffe AG, Sharples LD, Richardson S, and Baio G: Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. Stat Med 2015;**34**:2334–52.

4. Bor J, Moscoe E, Mutevedzi P, Newell ML, and Barnighausen T: Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology 2014;**25**:729–37.
5. Moscoe E, Bor J, and Bärnighausen T: Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol 2015;**68**:122–33.
6. Imbens GW and Lemieux T: Regression discontinuity designs: A guide to practice. Journal of Econometrics 2008;**142**:615–35.
7. Adeleke MO, Baio G, and O'Keeffe AG: Regression discontinuity designs for time-to-event outcomes: An approach using accelerated failure time models. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2022;**185**:1216–46.
8. Hernán MA and Robins JM: Causal inference: What If. Boca Raton: Chapman & Hall/CRC., 2020.

Study appendices

Appendix I
Study I: Descriptive study

HbA1c-defined prediabetes and progression to type 2 diabetes in Denmark during 2012-2018: a study based on laboratory data from routine care
Nicolaisen SK, Pedersen L, Witte DR, Sørensen HT, Thomsen RW.
Submitted.

Appendix II
Study II: Prediction study

Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark.
Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L.
BMJ Open Diabetes Research & Care 2022;10(5):e002946.

Appendix III
Study III: Trajectory study

Longitudinal HbA1c patterns before first treatment of diabetes in everyday clinical practice: A latent class trajectory analysis.
Nicolaisen SK, le Cessie S, Thomsen RW, Witte DR, Dekkers OM, Sørensen HT, Pedersen L.
In draft.

Appendix IV
Study IV: Regression discontinuity design study

Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study.
Petersen I, Nicolaisen SK, Ricciardi F, Sharma M, Thomsen RW, Baio G, Pedersen L.
Clinical Epidemiology 2020;12:569-577.

Study I | Descriptive study

HbA1c-defined prediabetes and progression to type 2 diabetes in Denmark during 2012-2018: a study based on laboratory data from routine care
Nicolaisen SK, Pedersen L, Witte DR, Sørensen HT, Thomsen RW.
Submitted.

Study I

HbA1c-defined prediabetes and progression to type 2 diabetes in Denmark during 2012-2018: a study based on laboratory data from routine care

Sia Kromann Nicolaisen[1], Lars Pedersen[1], Daniel R. Witte[2,3], Henrik Toft Sørensen[1], Reimar Wernich Thomsen[1]

[1]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

[2]Steno Diabetes Center Aarhus, Aarhus, Denmark

[3]Department of Public Health, Aarhus University, Aarhus, Denmark

Corresponding author: Sia Kromann Nicolaisen, Department of Clinical Epidemiology, Aarhus University Hospital, Olof Palmes Allé 43-45, 8200 Aarhus N, Denmark. E-mail: sknc@clin.au.dk, Telephone: +45 8716 7212.

Words: abstract 296, article 4,125

Journal: Diabetologia

Figures: 6

Tables: 1

Tweet: Prediabetes incidence, prevalence, and risk of progression to #type2Diabetes: new study based on Danish national laboratory data from @DCEAarhus @AarhusUni_Int @AUHdk @StenoAarhus

1 Abstract

Aims/hypothesis

Since glycated hemoglobin (HbA1c) was approved for diagnosing diabetes, HbA1c testing has increased substantially. Laboratory databases may now offer a tool to study prediabetes epidemiology. We examined the use of HbA1c measurements in Denmark, aiming to identify all individuals with HbA1c-defined prediabetes and examine their prevalence, incidence rates, mortality, and risk of progression to type 2 diabetes.

Methods

The study included all HbA1c measurements available in Danish laboratory databases containing test results from routine clinical care during 2012-2018. We estimated the 7-year cumulative incidence of having at least one HbA1c measurement. The population prevalence and incidence rates of prediabetes (HbA1c 42-47 mmol/mol [6.0%-6.4%]) were calculated based on all adult Danish residents. The 5-year cumulative incidence of progression to type 2 diabetes was estimated with death as competing event.

Results

Among 4,979,590 adult Danish residents, 70.8% (95% CI 70.8-70.9) had at least one HbA1c measurement recorded during 2012-2018. A total of 12,762,320 HbA1c measurements were available and 180,923 individuals were identified with incident HbA1c-defined prediabetes. This corresponded to an incidence rate of 14.2 (95% CI 14.1%-14.3%) per 1,000 person-years. The prevalence in 2018 was 7.1% (95% CI 7.1%-7.1%). Median HbA1c at prediabetes diagnosis was 43 mmol/mol (interquartile range [IQR] 42-44 mmol/mol) or 6.1% (IQR 6.0%-6.3%), median age was 66.9 years (IQR 56.7-75.7 years) and 52.0% were women. Within 5 years, 17.5% (95% CI 17.3%-17.7%) of individuals with prediabetes died. The 5-year cumulative incidence of type 2 diabetes was 21.3% (95% CI 21.1%-21.5%).

Conclusion/interpretation

The majority of the adult Danish population had HbA1c measured over a seven-year period. Out of 100 adults, 1.4 developed incident prediabetes each year and they were identified in the laboratory databases at an early stage. Within 5 years, one in five individuals with prediabetes progressed to diabetes and one in six died.

Keywords: Cohort, Haemoglobin A1c, HbA1c, Hemoglobin A1c, incidence, laboratory data, prediabetes, prevalence, progression, type 2 diabetes.

Abbreviations:

ADA	American Diabetes Association
DPP	Diabetes Prevention Program
DPPOS	Diabetes Prevention Program Outcomes Study
FPG	fasting plasma glucose
HbA1c	glycated hemoglobin
IDF	International Diabetes Foundation
IFG	impaired fasting glycaemia
IGT	impaired glucose tolerance
PY	person-years
WHO	World Health Organization

Research in context:

What is already known about this subject?

- HbA1c is now widely used to diagnose both prediabetes and type 2 diabetes.
- Current knowledge about prediabetes epidemiology is primarily based on measures other than HbA1c (*e.g.*, fasting glucose or glucose tolerance tests) and the overlap among definitions using different measures is small.
- Individuals with prediabetes will substantially burden the health care system if progression to type 2 diabetes is not prevented.

What is the key question?

- What is the prevalence, incidence rate, mortality, and risk of progression to subsequent type 2 diabetes in individuals with HbA1c-defined prediabetes in routine clinical care?

What are the new findings?

- More than 70% of the adult Danish population had at least one HbA1c measurement during 2012-2018.
- More than 1.4 out of 100 adult Danes had an incident HbA1c-defined prediabetes diagnosis each year and 7% of the adult Danes had prevalent prediabetes.
- One in five individuals with incident HbA1c-defined prediabetes progressed to type 2 diabetes within 5 years and one in six died.

How might this impact on clinical practice in the foreseeable future?

- Updated estimates of prevalence, incidence rates, and progression risks for prediabetes, reflecting patients receiving routine care who are mainly identified using HbA1c measurements, may aid in identifying at-risk individuals and help to prevent type 2 diabetes and its complications in the future.

2 Introduction

Prediabetes is defined as glucose levels above the normal level, but below the threshold for diagnosing overt type 2 diabetes, *e.g.,* glycated hemoglobin (HbA1c) levels in the range 42-47 mmol/mol (6.0%-6.4%) as defined by the World Health Organization (WHO)[1-3]. The true incidence and prevalence of prediabetes in the general population is largely unknown, as systematic population-wide screening for prediabetes or type 2 diabetes in Denmark and elsewhere is lacking. Much of our knowledge about prediabetes epidemiology is based on data from the 1990s and 2000s[3-10], when prediabetes and diabetes were defined by measures other than HbA1c (*e.g.*, fasting glucose and glucose tolerance tests). However, since the introduction of HbA1c as a tool for diagnosing type 2 diabetes (*i.e.*, HbA1c >=48 mmol/mol [6.5%]), HbA1c testing has increased greatly. HbA1c is currently one of the most frequently ordered blood tests in routine clinical care in Denmark[11, 12].

We hypothesize that the extensive use of HbA1c measurements in the Danish population makes it possible to study prevalence and incidence rates of prediabetes in routine care laboratory databases, as the frequent HbA1c testing facilitates identification of many individuals early in the course of prediabetes and that individuals can be followed from early prediabetes to potential progression to type 2 diabetes or death.

The aims of this study were threefold: first, to examine use of HbA1c measurements in Denmark after 2012; second, to exploit the Danish laboratory registry data to identify individuals with incident HbA1c-defined prediabetes and examine their prevalence, incidence rate and characteristics; and third, to follow individuals with incident HbA1c-defined prediabetes and examine their risk of progression to type 2 diabetes or death.

3 Methods

We follow the RECORD Statement throughout this paper (Electronic Supplementary Material [ESM Table S1).

3.1 Data sources

The tax-supported Danish healthcare system provides all residents (approximately 5.9 million individuals[15]) with unfettered access to medical care[13, 14] provided by general practitioners in the community and in hospitals, along with partial reimbursement for prescribed drugs[14]. The study utilized Danish registry data for all adult individuals (ages 20-100 years) living in a region of Denmark between 1 January 2012 and 31 December 2018 with available laboratory data (Figure 1).

HbA1c measurements were identified from two routine care laboratory databases: the nationwide Register of Laboratory Results for Research[11] and the regional Clinical Laboratory Information System Research Database at Aarhus University, which covers the North and Central Denmark Regions[11, 12]. Together, these registries contain virtually all laboratory measurements from both hospitals and general practitioners for the entire Danish population. At the beginning of our study period (1 January 2012), laboratory measurements had been available in the laboratory database for several years in 3 out of the 5 Danish administrative regions (North Denmark Region, Central Denmark Region, and Capital Region of Denmark). Measurements became available for the 4th region (Region Zealand) in 2014 and the 5th region (Region of Southern Denmark) in 2015 (ESM Figure S1). Thus, during the last 3 years of our study period (2016-2018), the laboratory database had full nationwide coverage.

All Danes are assigned a unique personal identification number (CPR number) at birth or immigration, making it possible to link individual information among all registries[13, 14]. We linked individuals to the following registries: the Danish National Patient Registry, containing all inpatient discharge diagnoses from all hospitals since 1977 and from emergency department and outpatient specialist clinic contacts since 1995[16]; the Danish Civil Registration System, containing vital status

and date of death for the entire Danish population; the Danish Register of Medicinal Product Statistics, containing complete prescription information from all community pharmacies since 1994[17]; and socioeconomic registries maintained by Statistics Denmark, containing data on family and household characteristics, including ethnic origin, educational level, employment status, and income[18].

3.2 Analyses

3.2.1 HbA1c measurements in the Danish registries

All HbA1c measurements available in the laboratory databases during the 2012-2018 study period were examined to determine use of HbA1c measurements, values, and availability over time. To account for regional differences in availability of laboratory data, we estimated the 7-year (2012-2018) cumulative incidence of individuals having at least one HbA1c measurement. We obtained this incidence using the non-parametric estimate of the cause-specific cumulative incidence function, with death treated as a competing event[19]. To examine annual changes in use of HbA1c measurements, all Danish residents alive on 31 December who lived in a region with available data during the entire calendar year were included in the assessment of the number of measurements.

3.2.2 Incidence of HbA1c-defined prediabetes

Criteria for inclusion in the analyses regarding risk of HbA1c-defined prediabetes included at least 5 years of permanent residence in Denmark in a region with available laboratory data (Figure 1). In addition, individuals in the study population could have no indication of pre-existing prediabetes (HbA1c 42-47 mmol/mol [6.0%-6.4%]) or diabetes (HbA1c >=48 mmol/mol [6.5%] during the previous 5 years, hospital contact with a diagnosis of diabetes during the previous 5 years, or redemption of a prescription for glucose-lowering medication during the previous 5 years). All individuals were followed from the start of follow-up (1 January 2012 or the first date when they fulfilled the inclusion criteria, whichever came last) to the date of incident HbA1c-defined prediabetes, emigration, study

end (31 December 2018) or death, whichever came first. Incident HbA1c-defined prediabetes was defined as the first HbA1c measurement within the range of 42-47 mmol/mol (6.0%-6.4%)[2, 20]. The incidence rate was calculated as the number of individuals with incident HbA1c-defined prediabetes per 1,000 person-years (PY) at risk and was calculated by age, sex, and calendar year. In addition, sex- and age-standardized incidence rates were calculated based on the WHO World 2000-2025 Standard Population Distribution[21, 22].

3.2.3 Characteristics of individuals with incident prediabetes

All individuals with incident HbA1c-defined prediabetes were characterized using the most recent data available on the date of an HbA1c measurement in the range of 42-47 mmol/mol (6.0%-6.4%) (ESM Figure S2). Baseline covariates included demographic variables, HbA1c measures, pre-existing comorbidities (based on a complete inpatient and outpatient hospital contact history during the previous 5 years), Charlson Comorbidity Index score as a measure of overall comorbidity (previous 5 years), current prescription drug use (prescriptions redeemed at community pharmacies during the previous 180 days), and variables related to socioeconomic status. Further details about the covariates are provided in ESM Table S2.

3.2.4 Prevalence of HbA1c-defined prediabetes

The prevalence of HbA1c-defined prediabetes was calculated among individuals who were alive per 31 December 2018 and had at least 5 years of residence in a Danish region with available laboratory (Figure 1). Prevalent HbA1c-defined prediabetes was defined as an HbA1c measurement in the range of 42-47 mmol/mol (6.0%-6.4%) during 2012-2018 with no indication of diabetes (with diabetes defined as HbA1c >=48 mmol/mol (6.5%), hospital contact with a diagnosis of diabetes, or redemption of a prescription for glucose-lowering medication). The prevalences were calculated by age and sex

and were also sex- and age-standardized based on the WHO World 2000-2025 Standard Population Distribution[21, 22].

3.2.5 Cumulative incidence of progression to type 2 diabetes or death

The cumulative incidences of progression from incident HbA1c-defined prediabetes to type 2 diabetes or death was estimated among individuals with 1 rather than 5 years of residence in a Danish region with available laboratory data (Figure 1). This allowed for the greatest possible amount of follow-up data. All individuals in this category with incident HbA1c-defined prediabetes (defined as above) were then followed from the date of an Hba1c measurement of 42-47 mmol/mol (6.0%-6.4%) to the time of a type 2 diabetes diagnosis, emigration, study end (31 December 2018), end of follow-up (maximum 5 years after incident HbA1c-defined prediabetes), or death, whichever came first.

Two main outcomes were studied: Type 2 diabetes and death (ESM Figure S2, ESM Table S2). Type 2 diabetes was defined as a) HbA1c-defined diabetes (first HbA1c >=48 mmol/mol [6.5%]); b) hospital contact with a diabetes diagnosis; or c) glucose-lowering treatment initiation (first redemption of a prescription for glucose-lowering medication). In addition, these three criteria were studied separately. The 5-year cumulative incidence of progression to type 2 diabetes was estimated using the non-parametric estimate of the cause-specific cumulative incidence function with death as a competing event[19]. The cumulative incidence for death was computed based on the Kaplan-Meier estimate.

All statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc, Cary, North Carolina) and R version 4.1.0 (R Core Team, 2020).

4 Results

4.1 HbA1c measurements in Denmark since 2012

A total of 4,979,590 adult Danes (ages 20-100 years) were living in Denmark in a region with available laboratory data during 2012-2018 and were thus eligible for the study (Figure 1). For these adult Danes, a total of 12,762,320 HbA1c measurements from 3,036,866 (61.0%) distinct individuals were available in the laboratory database during the 2012-2018 study period (ESM Figure S1). The median number of measurements per individual was 1 (interquartile range [IQR] 0-3); however, 18.2% (n=923,320) had at least 5 available measurements. After accounting for regional differences in coverage by the laboratory database, we estimated that 70.8% (95% confidence interval [CI] 70.8-70.9) of the entire adult Danish population had HbA1c measured at least once during 2012-2018. In the final study year, 2018, when the laboratory database achieved full nationwide coverage, 2,523,370 HbA1c measurements were available (ESM Figure S1). A majority (62.8%) of the values in the database were below the threshold for prediabetes (*i.e.,* HbA1c<42 mmol/mol [6.0%]) (Figure 2) and the median value of all recorded measurements decreased in all regions after 2012 (ESM Figure S3). A large proportion of older adults had at least one HbA1c measurement (*e.g.,* in 2012, 37.8% [95% CI 37.3-38.3] of 70-year-olds had at least one recorded HbA1c measurement and this proportion increased to 59.0% [95% CI 58.6-59.3] in 2018 (Figure 3)). The highest proportion was observed for 84-year-olds in 2018, among whom a total of 69.9% (95% CI 69.3-70.5) had at least one measurement. Older age was also a marker of having more than 1 HbA1c measurement per year (ESM Figure S4).

4.2 Incidence and characteristics of HbA1c-defined prediabetes

Overall, 3,226,748 individuals were at risk of incident HbA1c-defined prediabetes during 2012-2018, according to our inclusion criteria, i.e., they were 20-100 years old, had at least 5 years of permanent residence in a region with available laboratory data, and had no indication of pre-existing prediabetes

or diabetes (Figure 1). They contributed more than 12 million PY at risk (Figure 1, ESM Table S3). Median age on the first day at risk was 45.7 years (IQR 31.2-60.6 years) and 51.1% were women. A total of 180,923 (5.6%) individuals were identified with incident HbA1c-defined prediabetes during their time at risk (Figure 1). The overall incidence rate of HbA1c-defined prediabetes in Denmark during the 2012-2018 period was 14.2 per 1,000 PY (95% CI 14.1-14.3 per 1,000 PY) (ESM Table S3). When stratified by age, the incidence was generally lower among women than among men (Figure 4). The overall sex- and age-standardized incidence rate was 10.4 (95% CI 10.4-10.5) per 1,000 PY and the incidence rate for women was lower than that for men (10.0 [95% CI 10.0-10.1] per 1,000 PY for women and 10.8 [95% CI 10.8-10.9] per 1,000 PY for men).

Among the 180,923 individuals with incident HbA1c-defined prediabetes identified during the years 2012-2018, 93,790 (51.8%) were women, the median age was 66.9 years (IQR 56.7-75.7 years), and 10,091 (5.6%) had a Charlson Comorbidity Index score >=3 (Table 1). At the time of the prediabetes diagnosis, 64,015 (35.4%) were living alone and 81,091 (44.8%) were either unmarried, widowed, or divorced. A total of 100,024 (55.2%) individuals were in the lowest or low-to-medium income groups, 66,477 (36,7%) had no education or basic education/primary school as highest achieved education, and 118,907 (65.7%) were unemployed or not part of the workforce (including retirement). The median HbA1c level at inclusion was 43.0 mmol/mol [IQR 42.0-44.0 mmol/mol] or 6.1% [IQR 6.0%-6.2%], *i.e.,* at the lower end of the range of HbA1c levels defining prediabetes (42-47 mmol/mol [6.0%-6.4%]). While 12,875 (7.1%) of the individuals had the prediabetes-defining HbA1c test during an inpatient hospitalization (or in the emergency department), 46,270 (25.6 %) were followed in outpatient hospital specialist clinics at the time of the prediabetes diagnosis. The remaining 121,778 (67.3%) individuals were most likely tested at their general practitioner's office. A total of 153,297 (84.7%) individuals had a value in the lower half of the prediabetes range, *i.e.,* 42-44 mmol/mol (6.0%-6.2%), indicating that many were identified rather soon after their HbA1c level had progressed from normal to the range of prediabetes. Among the 83,245 (46.0%) individuals who qualified for the study based on an HbA1c of 42 mmol/mol (6.0%),

28,885 (34.7%) had had at least one HbA1c measurement in the previous year. This contrasts with only 727 (16.3%) of the 4,459 (2.5%) individuals who qualified based on an HbA1c of 47 mmol/mol (6.4%) (Table 1).

4.3 Prevalence of HbA1c-defined prediabetes

Among the 3,302,759 adult Danes (ages 20-100 years) alive in December 2018 with at least 5 years of permanent residence in a region with available laboratory data (Figure 1), a total of 234,056 individuals were identified as having prediabetes with no indication of overt diabetes during 2012-2018, yielding an overall prevalence of 7.1% (95% CI 7.1%-7.1%) (ESM Table S4). The sex- and age-standardized prevalence was 4.4% (95% CI 4.3-4.4)—4.42% (95% CI 4.4%-4.5%) for women and 4.3% (95% CI 4.3%-4.3%) for men. The prevalence increased with age.

4.4 Risk of subsequent diabetes or death in individuals with incident prediabetes

The 366,757 individuals identified with incident HbA1c-defined prediabetes based on the less restrictive definition of 1 rather than 5 years of look-back were followed for a maximum of 5 years. Median follow-up time was 2.7 years (IQR 1.6-4.5 years) and a total of 41,350 individuals (11.3%) died during follow-up (ESM Figure S5). Among the 75,907 (20.1%) individuals who were followed for all 5 years, 18,359 (24.2%) developed type 2 diabetes during follow-up. During all available follow-up time, we observed that 49,855 (13.6%) individuals developed type 2 diabetes. They were identified based mainly on elevated HbA1c: 45,109 (12.3%) had an HbA1c value >=48 mmol/mol (6.5%), 25,181 (6.9%) initiated glucose-lowering treatment, and 9,281 (2.5%) had a diabetes-related hospital contact during follow-up (ESM Figure S5, ESM Figure S6). The overall 5-year cumulative incidence of death was 17.5% (95% CI 17.3-17.7). When censoring and the competing risk of death were taken into account, the estimated overall 5-year cumulative incidence of type 2 diabetes was 21.3% (95% CI 21.1%-21.5%)

(Figure 5, Figure 6). Younger age at the time of prediabetes diagnosis and male sex were associated with increased risk. As expected, gradually higher baseline HbA1c levels in the prediabetes-defining range (42-47 mmol/mol [6.0%-6.4%]) were associated with increasingly elevated risks of developing type 2 diabetes. The 5-year cumulative incidence of diabetes increased from 11.7% (95% CI 11.5%-12.0%) for 42 mmol/mol (6.0%) to 59.4% (95% CI 57.9%-60.8%) for 47 mmol/mol (6.4%) (Figure 6). Differences in the risk of death across levels of baseline HbA1c were modest (15.3% [95% CI 15.0%-15.5%] for HbA1c 42 mmol/mol [6.0%] and 22.8% [95% CI 21.6%-24.0%] for HbA1c 47 mmol/mol [6.4%]).

5 Discussion

Our study showed that HbA1c testing in the Danish general population increased dramatically after the introduction of HbA1c as a tool for diagnosing diabetes. More than 70% of the entire Danish adult population were measured during our study period. For older people, almost two-thirds had measurements of HbA1c in any given year. More than 1.4 out of 100 adult Danes had a new HbA1c measurement in the range of prediabetes each year and the incidence increased by age. Likely due to widespread testing, individuals were identified at an early stage of prediabetes, with a median HbA1c value of 43 mmol/mol (6.1%). In 2018, 7% of the adult Danish population had prevalent HbA1c-defined prediabetes. The prevalence of HbA1c-defined prediabetes was 4% when standardized to the WHO World 2000-2025 Standard Population Distribution. Within 5 years after an HbA1c-defined prediabetes diagnosis, more than 1 in 5 progressed to type 2 diabetes and 1 in 6 died.

5.1 Comparison with other studies

Our study is among the first to examine the occurrence and prognosis of prediabetes in a population-based health care system, based on HbA1c measurements from laboratory databases with complete

population coverage. Direct comparison of our prediabetes prevalence, incidence, and progression findings with previous studies is challenging[10, 23-27], given the use of varying definitions of prediabetes (including either impaired fasting glycaemia [IFG], impaired glucose tolerance [IGT], or elevated HbA1c) and varying cut-points (*e.g.,* HbA1c values in the range of 42-47 mmol/mol [6.0%-6.4%] as in the WHO definition[2] and HbA1c in the range of 39-47 mmol/mol [5.7%-6.4%] as in the American Diabetes Association [ADA] definition[28]).

The incidence of prediabetes has been studied in different cohorts[29-34], but rarely based on HbA1c[35, 36], and virtually never based exclusively on the WHO definition used in our study (HbA1c 42-47 mmol/mol [6.0%-6.4%]). A cohort study in New Zealand based on general healthcare records of "high-need" patients found 1,276 new cases of HbA1c-defined prediabetes (HbA1c values in the range of 41-49 mmol/mol [5.9%-6.6%]) among 25,521 individuals (median age 48 years) with no prior diabetes[36], yielding an incidence of 50.0 per 1,000 PY, *i.e.,* substantially higher than our result in Denmark of 14.2 per 1,000 PY, standardized to 10.4 per 1,000 PY. A small sub-study from the Singapore Malay Eye Study included 447 normoglycemic individuals (mean age 53 years), among whom 92 had developed HbA1c-defined prediabetes (39-47 mmol/mol [5.7%-6.4%]) after 6 years, yielding an incidence of 36.9 per 1,000 PY[35]. A larger population-based Canadian study of immigrants and long-term Canadian residents (mean age 46 years) included more than 300,000 individuals with prediabetes, based on either an IFG measurement of 6.1-6.9 mmol/L, an IGT measurement of 7.8-11.0 mmol/L after a 75g oral glucose tolerance test, or an HbA1c value in the range of 42-47 mmol/mol (6.0%-6.4%)[32]. This study showed ethnic variation in incidence of prediabetes, with the lowest incidence observed among Western European immigrants, for whom an incidence of 27.9 per 1,000 PY was reported. This is close to our results despite the different definitions of prediabetes. Our prediabetes incidence findings are also consistent with those in the Danish Inter99[37] study, although the results of the latter study were not based on HbA1c. The Inter99 study surveyed 3,187 individuals (median age 46 years) who were initially free of diabetes and prediabetes, followed them for 5 years and found that 303 developed prediabetes during 16,328 years of follow-

up, *i.e.,* a prediabetes incidence of 18.6 per 1,000 PY. Our finding that men had higher prediabetes incidence rates than women across all age groups above age 55 corroborates a recent British study[38] based on electronic general practice health records during the 2009-2018 period. However, the estimated overall incidence rates in the UK were much lower than our estimates (*e.g.,* 4.5 per 1,000 PY for men and 4.7 per 1,000 PY for women, compared to 13.9 per 1,000 PY for men and 14.4 per 1,000 PY for women in our study population and 10.8 per 1,000 PY for men and 10.0 per 1,000 for women after standardization). It is possible that prediabetes is incompletely registered with Read codes in UK primary care records, due to criteria for diagnosis of non-diabetic hyperglycemia[38].

The prevalence of HbA1c-defined prediabetes in Denmark was previously estimated in a study combining registry data and population-based surveys[39]. It reported that 7% of the entire adult Danish population (ages 20-100 years) had prediabetes (HbA1c 42-47 mmol/mol [6.0%-6.4%]). The current study confirms this prevalence and additionally shows that Danish laboratory data can be used to study prediabetes occurrence and prognosis. A prevalence of 7% indicates that almost 330,000 of the 4,637,430 adults (aged 20-100 years) living in Denmark in December 2022 currently have HbA1c-defined prediabetes[15]. Our sex- and age-standardized prevalence of 4% is lower than the worldwide prevalences based on IGT (10.6%) and IFG (6.2%) reported in the International Diabetes Foundation (IDF) Diabetes Atlas[21], but reasonably close to the Europe-only estimates (7.1% based on IGT and 3.3% based on IFG measurements). However, as HbA1c was not included in the definition of prediabetes, direct comparisons are difficult.

While HbA1c levels above the lower limit for prediabetes are known to increase the risk of developing future diabetes, compared to normal levels of HbA1c[1, 2, 10, 27, 40-42], many individuals with prediabetes never progress to overt diabetes[5, 6, 43-45]. Our finding that 21% of individuals with prediabetes progressed to diabetes during 5 years of follow-up are lower than those reported in a 2018 Cochrane Review[24], in which the authors estimated a 5-year risk of 38% (95% CI 26%-51%) for progression from HbA1c-defined prediabetes (HbA1c 42-47 mmol/mol [6.0%-6.4%]) to diabetes, defined as fasting plasma glucose (FPG) >=7.0 mmol/L, HbA1c >=48 mmol/mol [6.5%], self-

reported diabetes, or glucose-lowering treatment initiation. However, only three small studies with few outcomes (1,103 individuals with 322 outcomes [46], 203 individuals with 100 outcomes [47], and 156 individuals with 58 outcomes [48]) were included in the review[46-48] and all studies were based on selected Asian populations of individuals seen at repeat health examinations. In comparison, the Diabetes Prevention Program Outcomes Study (DPPOS) [4, 40, 49] followed 2,776 of the original 3,234 Diabetes Prevention Program (DPP) [50] study participants (body mass index >=24 kg/m2 and prediabetes [FPG 5.3-6.9 mmol/L or 2-hour plasma glucose 7.8-11.0 mmol/L]) for a total of 15 years. The DPPOS study reported a 5-year cumulative incidence of diabetes of 35%, with diabetes defined as FPG >=7.0 mmol/L or 2-hour plasma glucose >=11.0 mmol/L [4, 49, 50], again considerably higher than our estimate of 21% progressing from HbA1c-defined prediabetes to diabetes within 5 years. In DPPOS, HbA1c was measured at study inclusion, but used neither as a criterion for study inclusion nor as an outcome[40]. Of interest, a meta-analysis including individual participants from 16 studies suggested that all five current prediabetes definitions identified individuals at high risk of subsequent diabetes within 5 years with similar accuracy[10]. However, while the overall progression rate from prediabetes to diabetes was 45 per 1,000 PY for prediabetes defined as HbA1c 39-47 mmol/mol (5.7%-6.4%), it was 79 per 1,000 PY according to the WHO definition (*i.e.,* HbA1c 42-47 mmol/mol [6.0%-6.4%]) used in our study[10]. This was based on 5 heterogeneous studies, with incidence rates ranging from 51.1 per 1,000 PY to 154.3 per 1,000 PY.

5.2 Study limitations

HbA1c testing relies on clinical decisions and the exact indication or reason for ordering an HbA1c test is unknown in our population. Prior to 2012, HbA1c was used mainly to monitor diabetes management. We observed that the values in those years were higher than after 2012, when median values dropped from 43-44 mmol/mol (6.1%-6.2%) to 37-40 mmol/mmol (5.5%-5.8%) and most HbA1c measurements in the 2012-2018 period were within the normal range (63% below 42 mmol/mol [6.0%]). We estimated that more than 70% of the entire adult Danish population had an HbA1c

measurement during our study period. In addition, the majority of the older Danish adults were measured at least once per year. HbA1c measurements were thus likely ordered as part of general routine health checks, because another blood test was ordered, or to rule out diabetes in individuals with non-specific symptoms, rather than because of a clear clinical suspicion of diabetes.

We aimed to include individuals with incident prediabetes by ensuring no pre-existing prediabetes nor diabetes and we included a 5-year look-back period. However, the incidence rate of prediabetes was relatively high in the first study year (2012), which may point to initial inclusion of some prevalent cases (left truncation bias). This could also be related to changes in the diagnostic criterion that occurred in 2011/2012. Some apparently new HbA1c-defined prediabetes cases may have had prior FPG- or 2-hour glucose-defined prediabetes before 2012, leading to possible overestimation of prediabetes incidence rates in early study years. Similarly, as HbA1c testing was not as common before 2012 as after 2012, some individuals might have been included with prevalent prediabetes, simply because they had no HbA1c measurements before 2012. Finally, although there is evidence that use of FPG and oral glucose tolerance testing declined rapidly in Denmark after the introduction of HbA1c in 2012[52], some individuals still might have been diagnosed with either prediabetes or diabetes based on these measures. Of note, neither plasma glucose nor oral glucose tolerance tests were included in our data, but the overlap among the different definitions of prediabetes is rather small[53-55].

In conclusion, we were able to identify individuals with HbA1c-defined prediabetes soon after their HbA1c levels progressed from the normal range to the range of prediabetes, using data from routine clinical care. This permitted estimation of updated prevalence and incidence rates for HbA1c-defined prediabetes in Denmark. As one in five individuals with prediabetes progresses to diabetes within 5 years, detailed studies of predictive factors are needed[51] as the basis for more individualized interventions to prevent type 2 diabetes and its complications in the future.

6 Additional information

Data permission

This study was reported to the Danish Data Protection Agency (Aarhus University record number 2016-051-000001/1702). The data used in this study are owned and managed by Statistics Denmark. In accordance with Danish law and data protection policies, the data used in this study were anonymized and stored and analyzed on a secured server. The data are available to researchers from research environments pre-approved by Statistics Denmark upon project approval by the Danish Data Protection Agency and Statistics Denmark. Researchers can apply for access to data after their request is approved by the Danish Data Protection Agency (https://www.datatilsynet.dk).

Ethical considerations

No ethical approval was needed. All data originate from registries and none were specifically created for this project.

Author contributions

SKN, LP, and RWT designed the study. SKN performed all analyses and had full access to the data. SKN prepared the draft of the manuscript. LP, RWT, DW, and HTS critically revised the manuscript. All authors approved the final manuscript. SKN and LP had final responsibility for the decision to submit for publication.

Conflicts of interest

The authors have no personal conflicts of interest to report. The Department of Clinical Epidemiology is involved in studies with funding from various companies as research grants to and administered by Aarhus University. None of these studies are related to the current study**.**

Role of the funding source

The study's funding source had no role in the collection, analysis, or interpretation of the data, nor in in the study design or writing of the manuscript.

7 References

[1] The International Expert Committee (2009) International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. Diabetes Care 32(7): 1327-1334. 10.2337/dc09-9033

[2] World Health Organization (2011) WHO Guidelines Approved by the Guidelines Review Committee. In: Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation. World Health Organization, Geneva

[3] Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M (2012) Prediabetes: a high-risk state for diabetes development. The Lancet 379(9833): 2279-2290. 10.1016/S0140-6736(12)60283-9

[4] Diabetes Prevention Program Research Group (2015) Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the Diabetes Prevention Program Outcomes Study. Lancet Diabetes Endocrinol 3(11): 866-875. 10.1016/S2213-8587(15)00291-0

[5] Morris DH, Khunti K, Achana F, et al. (2013) Progression rates from HbA1c 6.0-6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. Diabetologia 56(7): 1489-1493. 10.1007/s00125-013-2902-4

[6] Ligthart S, van Herpt TT, Leening MJ, et al. (2016) Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. Lancet Diabetes Endocrinol 4(1): 44-51. 10.1016/s2213-8587(15)00362-9

[7] Lindström J, Tuomilehto J (2003) The diabetes risk score: a practical tool to predict type 2 diabetes risk. Diabetes Care 26(3): 725-731. 10.2337/diacare.26.3.725

[8] Gray LJ, Taub NA, Khunti K, et al. (2010) The Leicester Risk Assessment score for detecting undiagnosed Type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. Diabet Med 27(8): 887-895. 10.1111/j.1464-5491.2010.03037.x

[9] Jølle A, Midthjell K, Holmen J, et al. (2019) Validity of the FINDRISC as a prediction tool for diabetes in a contemporary Norwegian population: a 10-year follow-up of the HUNT study. BMJ Open Diabetes Res Care 7(1): e000769. 10.1136/bmjdrc-2019-000769

[10] Lee CMY, Colagiuri S, Woodward M, et al. (2019) Comparing different definitions of prediabetes with subsequent risk of diabetes: an individual participant data meta-analysis involving 76 513 individuals and 8208 cases of incident diabetes. BMJ Open Diabetes Res Care 7(1): e000794. 10.1136/bmjdrc-2019-000794

[11] Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K (2020) Existing Data Sources in Clinical Epidemiology: Laboratory Information System Databases in Denmark. Clin Epidemiol 12: 469-475. 10.2147/clep.S245060

[12] Grann AF, Erichsen R, Nielsen AG, Froslev T, Thomsen RW (2011) Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. Clin Epidemiol 3: 133-138. 10.2147/CLEP.S17901

[13] Schmidt M, Schmidt SAJ, Adelborg K, et al. (2019) The Danish health care system and epidemiological research: from health care contacts to database records. Clin Epidemiol 11: 563-591. 10.2147/clep.S179083

[14] Schmidt M, Pedersen L, Sørensen HT (2014) The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 29(8): 541-549. 10.1007/s10654-014-9930-3

[15] Danmarks Statistik [Statistics Denmark] (2022) Folketal. Available from https://www.dst.dk/da/Statistik/emner/borgere/befolkning/befolkningstal. Accessed 05 November 2022

[16] Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT (2015) The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 7: 449-490. 10.2147/CLEP.S91125

[17] Kildemoes HW, Sørensen HT, Hallas J (2011) The Danish National Prescription Registry. Scand J Public Health 39(7 Suppl): 38-41. 10.1177/1403494810394717

[18] Thygesen LC, Daasnes C, Thaulow I, Brønnum-Hansen H (2011) Introduction to Danish (nationwide) registers on health and social issues: Structure, access, legislation, and archiving. Scandinavian Journal of Public Health 39(7_suppl): 12-16. 10.1177/1403494811399956
[19] SAS Institute Inc. (2015) SAS/STAT 14.1 User's Guide The LIFETEST Procedure. SAS institute Incorporated
[20] Videncenter For Diabetes Prædiabetes. Available from https://www.videncenterfordiabetes.dk/Type-2-diabetes/Praediabetes. Accessed 25 April 2021
[21] International Diabetes Federation (2021) IDF Diabetes Atlas. In: IDF Diabetes Atlas. International Diabetes Federation, 2021.
[22] Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJL, Lozano R, Inoue M (2000) Age standardization of rates: A new WHO standard. In:
[23] Hostalek U (2019) Global epidemiology of prediabetes - present and future perspectives. Clin Diabetes Endocrinol 5: 5. 10.1186/s40842-019-0080-0
[24] Richter B, Hemmingsen B, Metzendorf MI, Takwoingi Y (2018) Development of type 2 diabetes mellitus in people with intermediate hyperglycaemia. Cochrane Database Syst Rev 10(10): CD012661. 10.1002/14651858.CD012661.pub2
[25] Davidson MB, Kahn RA (2016) A Reappraisal of Prediabetes. The Journal of Clinical Endocrinology & Metabolism 101(7): 2628-2635. 10.1210/jc.2016-1370
[26] Makaroff LE (2017) The need for international consensus on prediabetes. Lancet Diabetes Endocrinol 5(1): 5-7. 10.1016/S2213-8587(16)30328-X
[27] Xu XY, Leung AYM, Smith R, Wong JYH, Chau PH, Fong DYT (2020) The relative risk of developing type 2 diabetes among individuals with prediabetes compared with individuals with normoglycaemia: Meta-analysis and meta-regression. J Adv Nurs 76(12): 3329-3345. 10.1111/jan.14557
[28] American Diabetes Association Professional Practice Committee (2022) 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2022. Diabetes Care 45(Suppl 1): S17-S38. 10.2337/dc22-S002
[29] Gujral UP, Narayan KMV, Kandula NR, Liu K, Kanaya AM (2020) Incidence of diabetes and prediabetes and predictors of glycemic change among South Asians in the USA: the MASALA study. BMJ Open Diabetes Res Care 8(1). 10.1136/bmjdrc-2019-001063
[30] Sadeghi M, Talaei M, Parvaresh Rizi E, Dianatkhah M, Oveisgharan S, Sarrafzadegan N (2015) Determinants of incident prediabetes and type 2 diabetes in a 7-year cohort in a developing country: The Isfahan Cohort Study. J Diabetes 7(5): 633-641. 10.1111/1753-0407.12236
[31] Latifi SM, Karandish M, Shahbazian H, Hardani Pasand L (2016) Incidence of Prediabetes and Type 2 Diabetes among People Aged over 20 Years in Ahvaz: A 5-Year Perspective Study (2009-2014). J Diabetes Res 2016: 4908647. 10.1155/2016/4908647
[32] Fazli GS, Moineddin R, Bierman AS, Booth GL (2019) Ethnic differences in prediabetes incidence among immigrants to Canada: a population-based cohort study. BMC Med 17(1): 100. 10.1186/s12916-019-1337-2
[33] Vijayakumar G, Manghat S, Vijayakumar R, et al. (2019) Incidence of type 2 diabetes mellitus and prediabetes in Kerala, India: results from a 10-year prospective cohort. BMC Public Health 19(1): 140. 10.1186/s12889-019-6445-6
[34] Hadaegh F, Derakhshan A, Zafari N, et al. (2017) Pre-diabetes tsunami: incidence rates and risk factors of pre-diabetes and its different phenotypes over 9 years of follow-up. Diabet Med 34(1): 69-78. 10.1111/dme.13034
[35] Man REK, Charumathi S, Gan ATL, et al. (2017) Cumulative incidence and risk factors of prediabetes and type 2 diabetes in a Singaporean Malay cohort. Diabetes Res Clin Pract 127: 163-171. 10.1016/j.diabres.2017.03.007
[36] Gu Y, Warren J, Kennelly J, Walker N, Harwood M (2015) Incidence Rate of Prediabetes: An Analysis of New Zealand Primary Care Data. Stud Health Technol Inform 214: 81-86

[37] Engberg S, Vistisen D, Lau C, et al. (2009) Progression to impaired glucose regulation and diabetes in the population-based Inter99 study. Diabetes Care 32(4): 606-611. 10.2337/dc08-1869
[38] Pal K, Horsfall L, Sharma M, Nazareth I, Petersen I (2021) Time trends in the incidence of clinically diagnosed type 2 diabetes and pre-diabetes in the UK 2009-2018: a retrospective cohort study. BMJ Open Diabetes Res Care 9(1). 10.1136/bmjdrc-2020-001989
[39] Jorgensen ME, Ellervik C, Ekholm O, Johansen NB, Carstensen B (2018) Estimates of prediabetes and undiagnosed type 2 diabetes in Denmark: The end of an epidemic or a diagnostic artefact? Scand J Public Health: 1403494818799606. 10.1177/1403494818799606
[40] Diabetes Prevention Program Research Group (2015) HbA1c as a Predictor of Diabetes and as an Outcome in the Diabetes Prevention Program: A Randomized Clinical Trial. Diabetes Care 38(1): 51-58. 10.2337/dc14-0886
[41] Leong A, Daya N, Porneala B, et al. (2018) Prediction of Type 2 Diabetes by Hemoglobin A(1c) in Two Community-Based Cohorts. Diabetes Care 41(1): 60-68. 10.2337/dc17-0607
[42] DeJesus RS, Breitkopf CR, Rutten LJ, Jacobson DJ, Wilson PM, Sauver JS (2017) Incidence Rate of Prediabetes Progression to Diabetes: Modeling an Optimum Target Group for Intervention. Popul Health Manag 20(3): 216-223. 10.1089/pop.2016.0067
[43] Veronese N, Noale M, Sinclair A, et al. (2022) Risk of progression to diabetes and mortality in older people with prediabetes: The English longitudinal study on ageing. Age Ageing 51(2). 10.1093/ageing/afab222
[44] Shang Y, Marseglia A, Fratiglioni L, et al. (2019) Natural history of prediabetes in older adults from a population-based longitudinal study. J Intern Med 0(0): 326-340. 10.1111/joim.12920
[45] Vistisen D, Kivimäki M, Perreault L, et al. (2019) Reversion from prediabetes to normoglycaemia and risk of cardiovascular disease and mortality: the Whitehall II cohort study. Diabetologia 62(8): 1385-1390. 10.1007/s00125-019-4895-0
[46] Kim CH, Kim HK, Kim EH, Bae SJ, Choe J, Park JY (2016) Risk of progression to diabetes from prediabetes defined by HbA1c or fasting plasma glucose criteria in Koreans. Diabetes Res Clin Pract 118: 105-111. 10.1016/j.diabres.2016.06.009
[47] Heianza Y, Arase Y, Fujihara K, et al. (2012) Screening for pre-diabetes to predict future diabetes using various cut-off points for HbA(1c) and impaired fasting glucose: the Toranomon Hospital Health Management Center Study 4 (TOPICS 4). Diabet Med 29(9): e279-285. 10.1111/j.1464-5491.2012.03686.x
[48] Nakagami T, Tanaka Y, Oya J, et al. (2016) Associations of HbA1c and fasting plasma glucose with incident diabetes: Implications for pre-diabetes thresholds in a Japanese population. Prim Care Diabetes 10(6): 407-414. 10.1016/j.pcd.2016.07.006
[49] Diabetes Prevention Program Research Group (2009) 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet 374(9702): 1677-1686. 10.1016/S0140-6736(09)61457-4
[50] Diabetes Prevention Program Research Group (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346(6): 393-403
[51] Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L (2022) Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. BMJ Open Diabetes Res Care 10(5). 10.1136/bmjdrc-2022-002946
[52] Knudsen JS, Knudsen SS, Hulman A, et al. (2022) Changes in type 2 diabetes incidence and mortality associated with introduction of HbA1c as diagnostic option: A Danish 24-year population-based study. Lancet Reg Health Eur 14: 100291. 10.1016/j.lanepe.2021.100291
[53] Barry E, Roberts S, Oke J, Vijayaraghavan S, Normansell R, Greenhalgh T (2017) Efficacy and effectiveness of screen and treat policies in prevention of type 2 diabetes: systematic review and meta-analysis of screening tests and interventions. BMJ 356: i6538. 10.1136/bmj.i6538
[54] Nielsen AA, Petersen PH, Green A, Christensen C, Christensen H, Brandslund I (2014) Changing from glucose to HbA1c for diabetes diagnosis: predictive values of one test and importance of analytical bias and imprecision. Clin Chem Lab Med 52(7): 1069-1077. 10.1515/cclm-2013-0337

[55] Vistisen D, Witte DR, Brunner EJ, et al. (2018) Risk of Cardiovascular Disease and Death in Individuals With Prediabetes Defined by Different Criteria: The Whitehall II Study. Diabetes Care 41(4): 899-906. 10.2337/dc17-2530

8 Figures and tables

Figure 1

Flowchart of individuals included in the study. All Danish adults comprised the main population and all their HbA1c measurements were assessed.

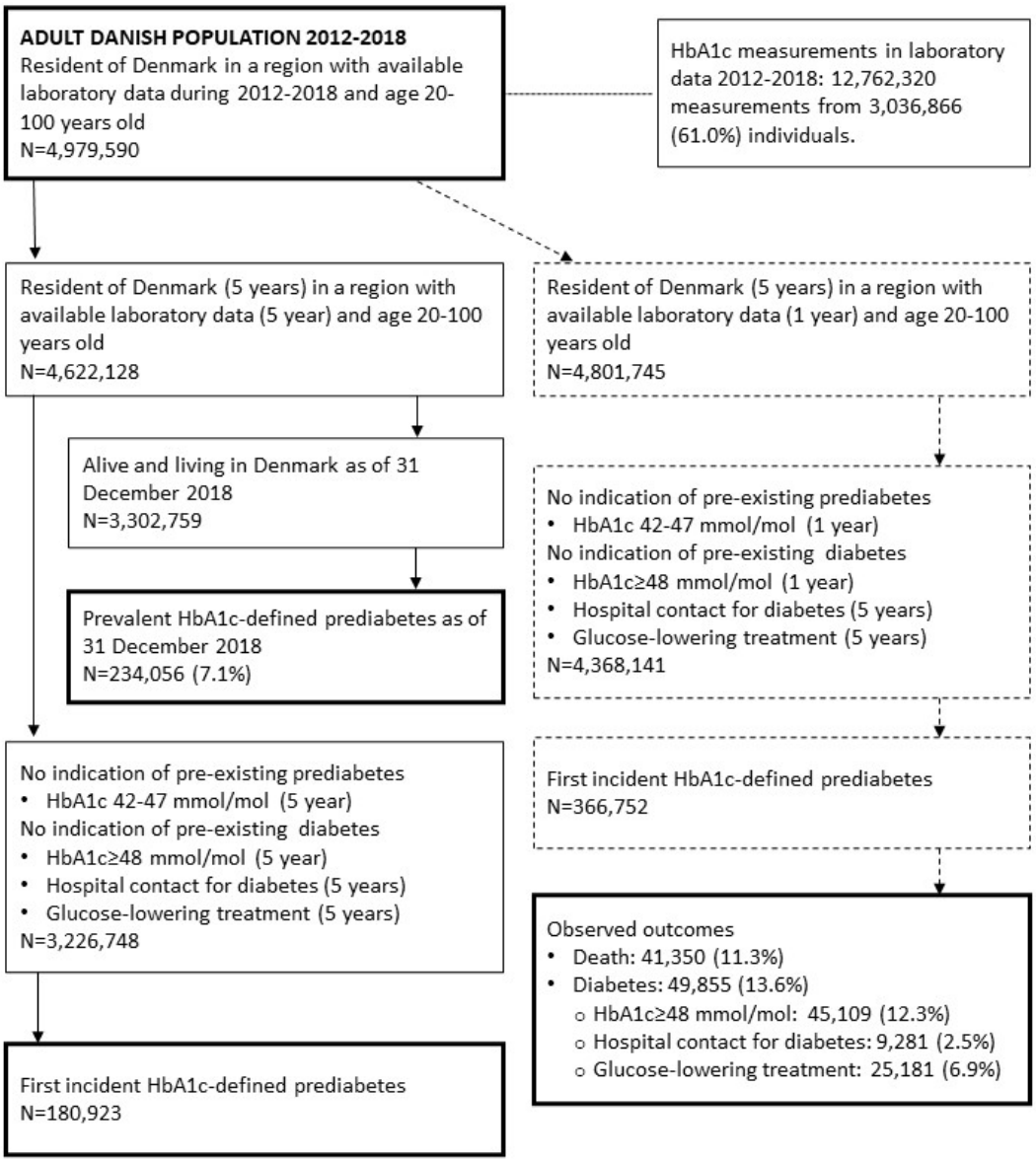


Figure 2

Histogram showing the distribution of the 12,762,320 HbA1c measurements available in the laboratory database during 2012-2018. The red dashed lines indicate the range used for prediabetes, *i.e.*, 42-47 mmol/mol (6.0%-6.4%). Measurements above 48 mmol/mol (6.5%) indicate diabetes.

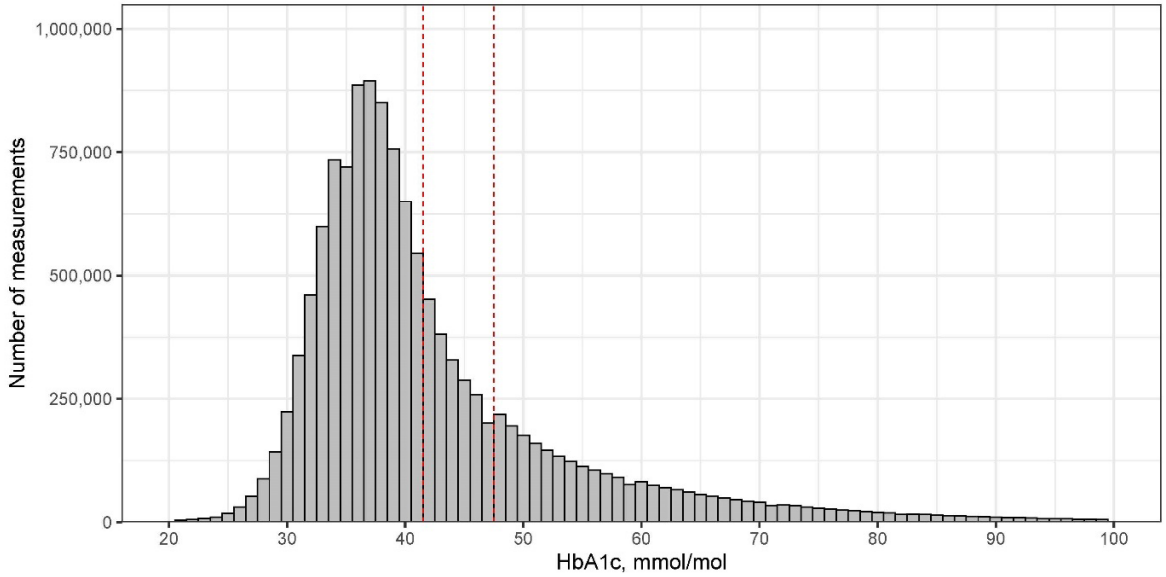


Figure 3

Proportion within each age group with at least one HbA1c measurement available during each calendar year. As an example, among individuals who turned 60 years old in 2018, 44.4% (95% CI 44.0%-44.8%) had at least one HbA1c measurement recorded in 2018, whereas 59.0% (95% CI 58.6%-59.3%) of those aged 70 years in 2018, and 68.4% (95% CI 67.9-68.9) of those aged 80 years in 2018 had at least one measurement recorded in 2018.

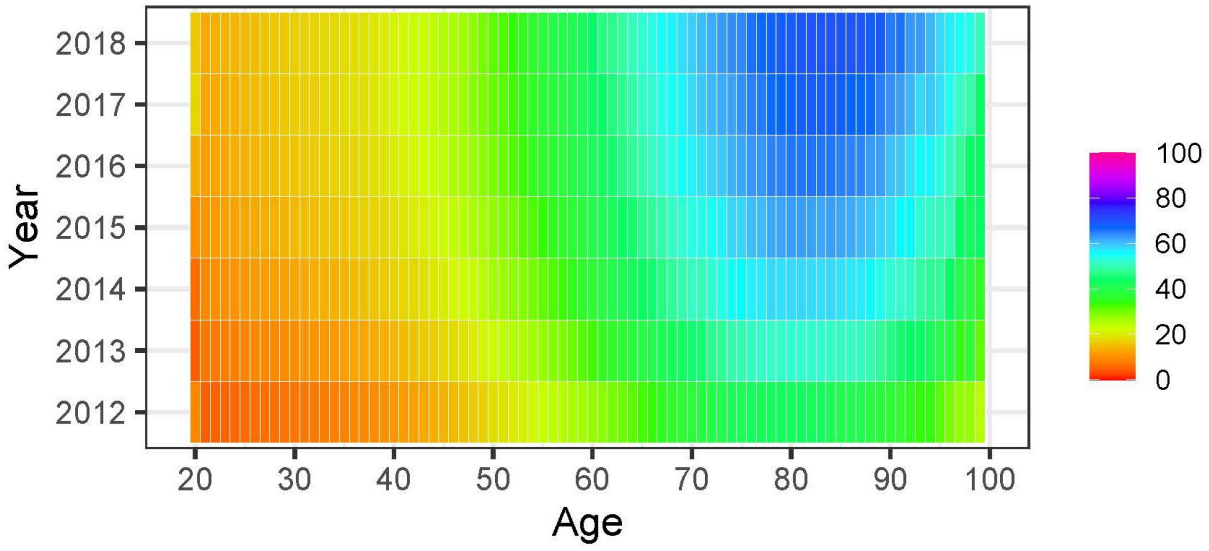


Figure 4

Incidence rates per 1,000 person-years, by age and sex, of HbA1c-defined prediabetes in the adult Danish population, 2012-2018.

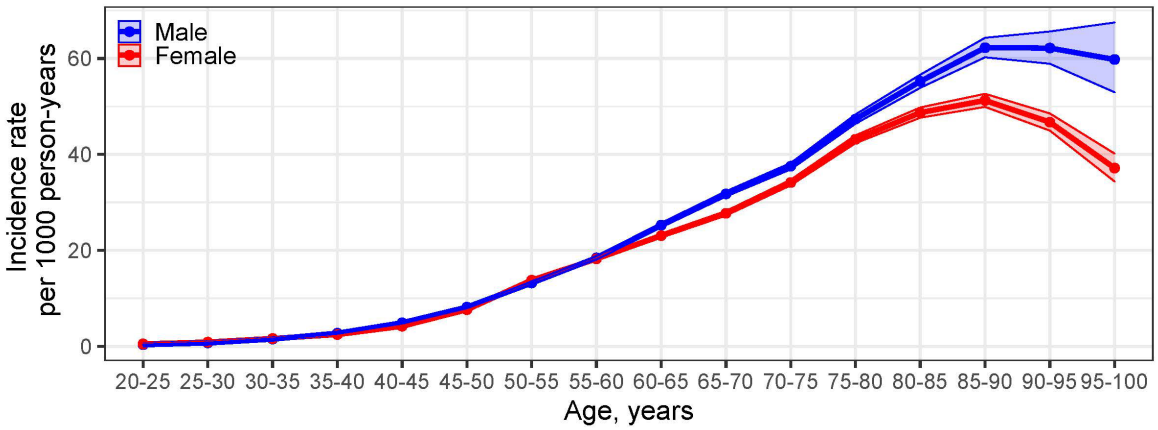


Table 1

Baseline characteristics of all individuals with incident HbA1c-defined prediabetes, overall and by baseline HbA1c level.

	N (%) or Median (IQR)	Value of prediabetes-defining HbA1c measurement					
		42 mmol/mol (6.0%)	43 mmol/mol (6.1%)	44 mmol/mol (6.2%)	45 mmol/mol (6.3%)	46 mmol/mol (6.4%)	47 mmol/mol (6.4%)
Total	180,923 (100.0%)	83,245 (100.0%)	45,340 (100.0%)	24,712 (100.0%)	14,279 (100.0%)	8,888 (100.0%)	4,459 (100.0%)
Demographic variables							
Sex							
Female	93,790 (51.8%)	44,114 (53.0%)	23,785 (52.5%)	12,458 (50.4%)	7,064 (49.5%)	4,273 (48.1%)	2,096 (47.0%)
Male	87,133 (48.2%)	39,131 (47.0%)	21,555 (47.5%)	12,254 (49.6%)	7,215 (50.5%)	4,615 (51.9%)	2,363 (53.0%)
Age at index date	66.9 (56.7-75.7)	66.9 (56.9-75.5)	67.0 (56.9-75.7)	66.9 (56.8-75.8)	66.9 (56.2-76.0)	66.6 (56.2-75.9)	65.0 (54.9-74.7)
Ethnic origin							
Danish	164,330 (90.8%)	###	###	###	###	###	###
Immigrant/descendant/unknown	16,574 (9.2%)	###	###	###	###	###	###
HbA1c measures							
Value of prediabetes-defining HbA1c measurement (mmol/mol)	43.0 (42.0-44.0)	42.0 (42.0-42.0)	43.0 (43.0-43.0)	44.0 (44.0-44.0)	45.0 (45.0-45.0)	46.0 (46.0-46.0)	47.0 (47.0-47.0)
Value of prediabetes-defining HbA1c measurement (%)	6.1 (6.0-6.2)	6.0 (6.0-6.0)	6.1 (6.1-6.1)	6.2 (6.2-6.2)	6.3 (6.3-6.3)	6.4 (6.4-6.4)	6.4 (6.4-6.4)
Place where prediabetes-defining HbA1c measurement was conducted							
Measured during inpatient hospitalization or emergency department contact	12,875 (7.1%)	5,132 (6.2%)	3,131 (6.9%)	1,939 (7.8%)	1,276 (8.9%)	919 (10.3%)	478 (10.7%)
Measured while followed at outpatient hospital specialist clinic	46,270 (25.6%)	20,691 (24.9%)	11,743 (25.9%)	6,508 (26.3%)	3,861 (27.0%)	2,356 (26.5%)	1,111 (24.9%)
Measured neither during inpatient hospitalization or emergency department contact nor while followed at outpatient hospital specialist clinic (likely during general practitioner contact)	121,778 (67.3%)	57,422 (69.0%)	30,466 (67.2%)	16,265 (65.8%)	9,142 (64.0%)	5,613 (63.2%)	2,870 (64.4%)
Presence of HbA1c measurements one year before prediabetes	53,179 (29.4%)	28,885 (34.7%)	13,316 (29.4%)	5,858 (23.7%)	2,899 (20.3%)	1,494 (16.8%)	727 (16.3%)

Prescription drug use							
Any thrombocyte-aggregation prophylaxis	41,075 (22.7%)	18,751 (22.5%)	10,491 (23.1%)	5,643 (22.8%)	3,254 (22.8%)	2,046 (23.0%)	890 (20.0%)
Any hypolipidemic treatment	58,415 (32.3%)	27,737 (33.3%)	14,827 (32.7%)	7,735 (31.3%)	4,321 (30.3%)	2,629 (29.6%)	1,166 (26.1%)
Loop-diuretics	18,540 (10.2%)	7,272 (8.7%)	4,588 (10.1%)	2,921 (11.8%)	1,881 (13.2%)	1,272 (14.3%)	606 (13.6%)
Any potential antihypertensive treatment (excl. loop-diuretics)	97,977 (54.2%)	44,887 (53.9%)	24,796 (54.7%)	13,333 (54.0%)	7,840 (54.9%)	4,865 (54.7%)	2,256 (50.6%)
Beta-blockers	40,764 (22.5%)	18,082 (21.7%)	10,292 (22.7%)	5,759 (23.3%)	3,449 (24.2%)	2,207 (24.8%)	975 (21.9%)
Calcium channel antagonists	33,503 (18.5%)	15,393 (18.5%)	8,459 (18.7%)	4,536 (18.4%)	2,688 (18.8%)	1,657 (18.6%)	770 (17.3%)
ACE inhibitors or ATII antagonists	64,111 (35.4%)	29,463 (35.4%)	16,211 (35.8%)	8,750 (35.4%)	5,138 (36.0%)	3,114 (35.0%)	1,435 (32.2%)
Oral steroids	12,703 (7.0%)	5,042 (6.1%)	3,117 (6.9%)	1,917 (7.8%)	1,267 (8.9%)	882 (9.9%)	478 (10.7%)
Any inhalation therapy for asthma/COPD	24,705 (13.7%)	10,613 (12.7%)	6,315 (13.9%)	3,518 (14.2%)	2,151 (15.1%)	1,410 (15.9%)	698 (15.7%)
Opioids	24,029 (13.3%)	10,234 (12.3%)	6,112 (13.5%)	3,453 (14.0%)	2,153 (15.1%)	1,379 (15.5%)	698 (15.7%)
Antibiotics	46,599 (25.8%)	20,258 (24.3%)	11,787 (26.0%)	6,755 (27.3%)	3,979 (27.9%)	2,559 (28.8%)	1,261 (28.3%)
Antidepressants	25,930 (14.3%)	11,609 (13.9%)	6,552 (14.5%)	3,618 (14.6%)	2,110 (14.8%)	1,379 (15.5%)	662 (14.8%)
Hypnotics/sedatives	13,541 (7.5%)	6,088 (7.3%)	3,348 (7.4%)	1,923 (7.8%)	1,080 (7.6%)	749 (8.4%)	353 (7.9%)
Comorbidities							
Charlson Comorbidity Index score							
0	128,294 (70.9%)	60,353 (72.5%)	31,970 (70.5%)	17,229 (69.7%)	9,742 (68.2%)	5,953 (67.0%)	3,047 (68.3%)
1-2	42,538 (23.5%)	18,915 (22.7%)	10,865 (24.0%)	5,944 (24.1%)	3,527 (24.7%)	2,251 (25.3%)	1,036 (23.2%)
>=3	10,091 (5.6%)	3,977 (4.8%)	2,505 (5.5%)	1,539 (6.2%)	1,010 (7.1%)	684 (7.7%)	376 (8.4%)
Cardiovascular disease	36,877 (20.4%)	16,296 (19.6%)	9,312 (20.5%)	5,295 (21.4%)	3,110 (21.8%)	1,983 (22.3%)	881 (19.8%)
Angina pectoris or CABG/PCI procedures	15,841 (8.8%)	7,200 (8.6%)	4,051 (8.9%)	2,181 (8.8%)	1,291 (9.0%)	782 (8.8%)	336 (7.5%)
Myocardial infarction	5,603 (3.1%)	2,517 (3.0%)	1,471 (3.2%)	745 (3.0%)	473 (3.3%)	277 (3.1%)	120 (2.7%)
Heart failure	6,738 (3.7%)	2,615 (3.1%)	1,688 (3.7%)	1,079 (4.4%)	685 (4.8%)	458 (5.2%)	213 (4.8%)
Atrial fibrillation/flutter	14,232 (7.9%)	5,993 (7.2%)	3,559 (7.8%)	2,125 (8.6%)	1,270 (8.9%)	880 (9.9%)	405 (9.1%)
Stroke	6,226 (3.4%)	2,775 (3.3%)	1,527 (3.4%)	923 (3.7%)	546 (3.8%)	332 (3.7%)	123 (2.8%)
Hypertension	27,513 (15.2%)	12,383 (14.9%)	6,920 (15.3%)	3,870 (15.7%)	2,300 (16.1%)	1,395 (15.7%)	645 (14.5%)
Obesity	6,087 (3.4%)	2,646 (3.2%)	1,511 (3.3%)	864 (3.5%)	515 (3.6%)	371 (4.2%)	180 (4.0%)
Chronic pulmonary disease	12,530 (6.9%)	5,209 (6.3%)	3,094 (6.8%)	1,868 (7.6%)	1,199 (8.4%)	794 (8.9%)	366 (8.2%)
Cancer (excl. non-melanoma skin cancer)	12,530 (6.9%)	5,242 (6.3%)	3,118 (6.9%)	1,846 (7.5%)	1,146 (8.0%)	766 (8.6%)	412 (9.2%)
Kidney disease	2,771 (1.5%)	1,150 (1.4%)	685 (1.5%)	418 (1.7%)	245 (1.7%)	171 (1.9%)	102 (2.3%)
Liver disease	1,153 (0.6%)	460 (0.6%)	273 (0.6%)	172 (0.7%)	119 (0.8%)	83 (0.9%)	46 (1.0%)
Dementia	2,703 (1.5%)	1,110 (1.3%)	718 (1.6%)	402 (1.6%)	225 (1.6%)	174 (2.0%)	74 (1.7%)
Thyrotoxicosis	1,781 (1.0%)	831 (1.0%)	469 (1.0%)	231 (0.9%)	131 (0.9%)	81 (0.9%)	38 (0.9%)
Hypothyroidism	2,002 (1.1%)	970 (1.2%)	516 (1.1%)	255 (1.0%)	135 (0.9%)	83 (0.9%)	43 (1.0%)
Pancreatic disease (pancreatic cancer, pancreas resection, and acute or chronic pancreatitis)	585 (0.3%)	211 (0.3%)	150 (0.3%)	84 (0.3%)	60 (0.4%)	59 (0.7%)	21 (0.5%)

Possible HbA1c-modifying conditions	14,436 (8.0%)	6,267 (7.5%)	3,730 (8.2%)	2,014 (8.1%)	1,236 (8.7%)	781 (8.8%)	408 (9.2%)
Previous ICU admission	6,401 (3.5%)	2,559 (3.1%)	1,631 (3.6%)	1,008 (4.1%)	604 (4.2%)	403 (4.5%)	196 (4.4%)
Markers of smoking	28,812 (15.9%)	12,393 (14.9%)	7,318 (16.1%)	4,159 (16.8%)	2,495 (17.5%)	1,638 (18.4%)	809 (18.1%)
Markers of alcoholism	3,800 (2.1%)	1,526 (1.8%)	983 (2.2%)	579 (2.3%)	363 (2.5%)	242 (2.7%)	107 (2.4%)
Socioeconomic variables							
Highest education achieved							
None, basic education, or primary school	66,477 (36.7%)	29,710 (35.7%)	16,727 (36.9%)	9,367 (37.9%)	5,589 (39.1%)	3,408 (38.3%)	1,676 (37.6%)
Youth education, high school, or similar	72,835 (40.3%)	33,948 (40.8%)	18,176 (40.1%)	9,777 (39.6%)	5,615 (39.3%)	3,536 (39.8%)	1,783 (40.0%)
Higher education	36,833 (20.4%)	17,507 (21.0%)	9,231 (20.4%)	4,892 (19.8%)	2,681 (18.8%)	1,670 (18.8%)	852 (19.1%)
Employment status							
Employed	62,016 (34.3%)	29,031 (34.9%)	15,335 (33.8%)	8,370 (33.9%)	4,731 (33.1%)	2,928 (32.9%)	1,621 (36.4%)
Unemployed or not part of the workforce	118,907 (65.7%)	54,214 (65.1%)	30,005 (66.2%)	16,342 (66.1%)	9,548 (66.9%)	5,960 (67.1%)	2,838 (63.6%)
Income							
Lowest income group	28,892 (16.0%)	13,179 (15.8%)	7,238 (16.0%)	3,975 (16.1%)	2,332 (16.3%)	1,432 (16.1%)	736 (16.5%)
Low-to-medium income	71,132 (39.3%)	32,233 (38.7%)	17,910 (39.5%)	9,883 (40.0%)	5,731 (40.1%)	3,630 (40.8%)	1,745 (39.1%)
Medium-to-high income	47,110 (26.0%)	21,884 (26.3%)	11,807 (26.0%)	6,320 (25.6%)	3,664 (25.7%)	2,276 (25.6%)	1,159 (26.0%)
Highest income group	33,191 (18.3%)	15,686 (18.8%)	8,234 (18.2%)	4,460 (18.0%)	2,501 (17.5%)	1,512 (17.0%)	798 (17.9%)
Type of household							
Living alone	64,015 (35.4%)	###	###	###	###	###	###
Not living alone	116,889 (64.6%)	###	###	###	###	###	###
Marital status							
Married	99,761 (55.1%)	###	###	###	###	###	###
Divorced	27,856 (15.4%)	###	###	###	###	###	###
Widow/widower	29,484 (16.3%)	###	###	###	###	###	###
Unmarried	23,751 (13.1%)	###	###	###	###	###	###

###: at least one cell (or number of missing values) contains a number < 5 and may not be reported due to Danish regulations.

Missing data: Type of household n=19 (0.0%), Ethnic origin n=19 (0.0%), Highest education achieved n=4,778 (2.6%), Income n=598 (0.3%). The rest of the variables had no missing data.

Figure 5

Overall 5-year cumulative incidences for progression from HbA1c-defined prediabetes to type 2 diabetes or death.

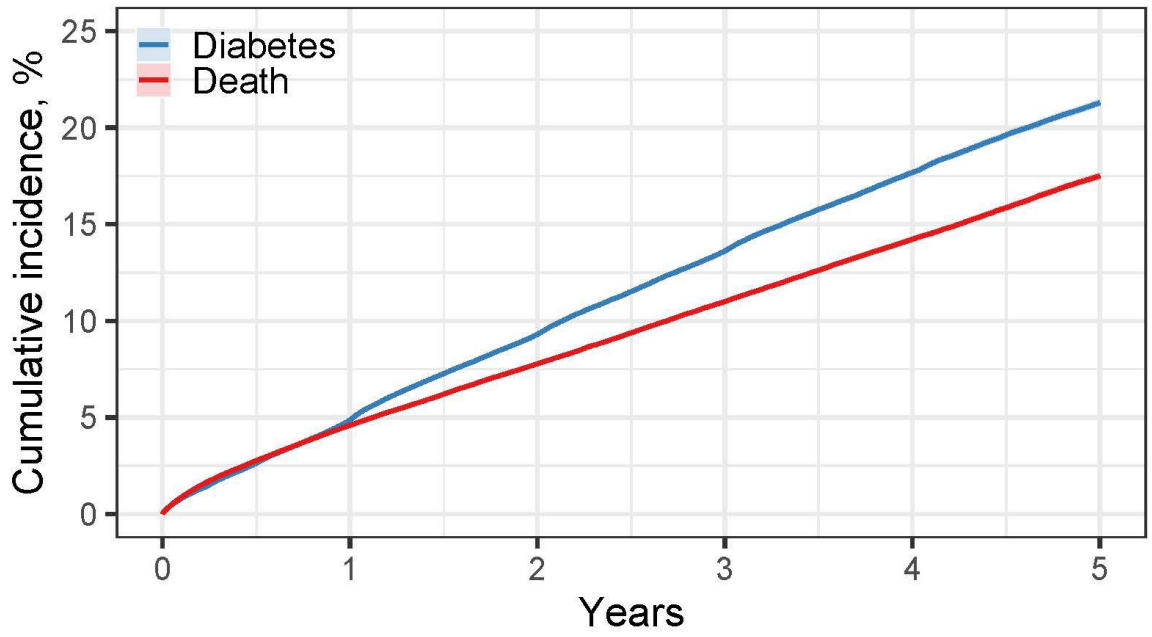
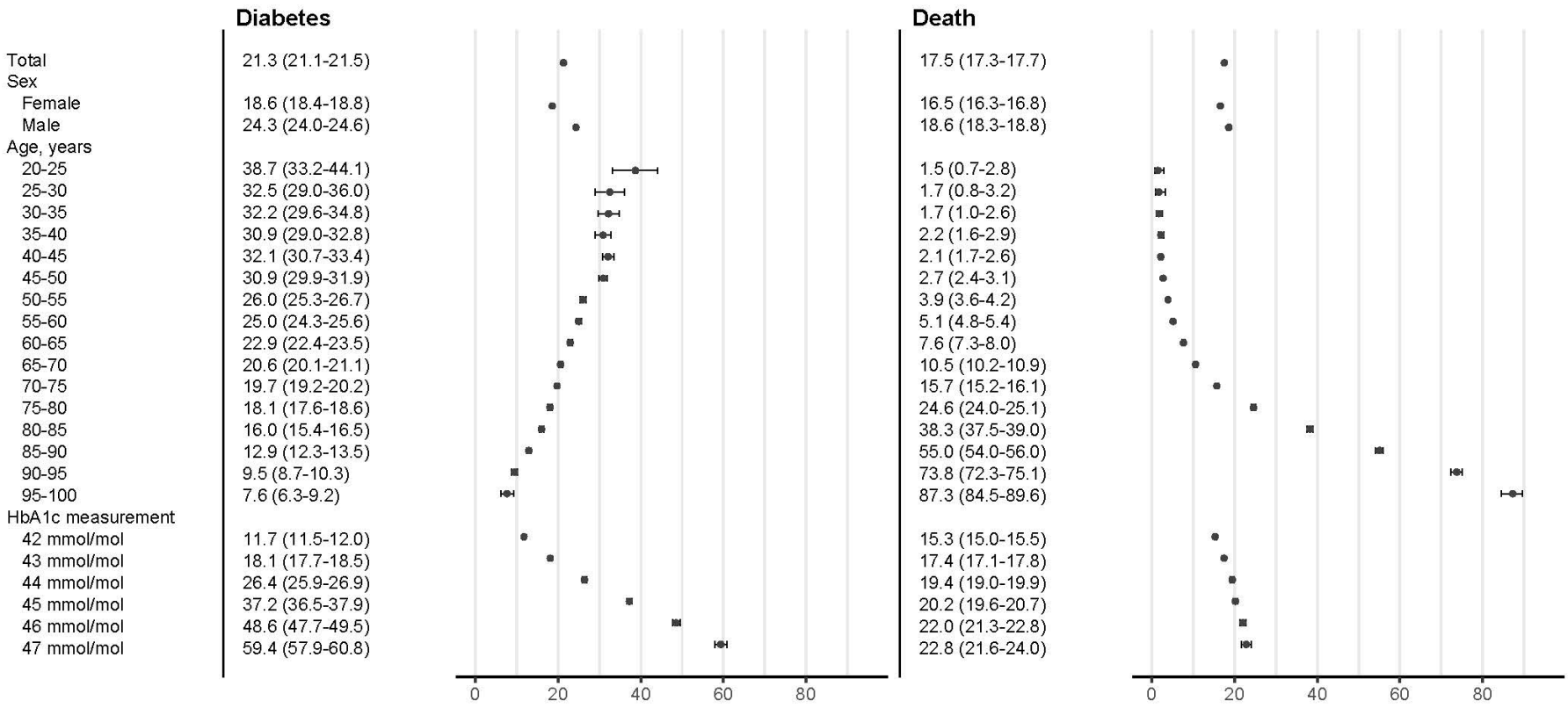


Figure 6

Estimates of the 5-year cumulative incidences for progression from HbA1c-defined prediabetes to type 2 diabetes or death stratified by sex, age, and value of the prediabetes-defining HbA1c measurement.



9 Electronic Supplementary Material

Electronic supplementary material:

HbA1c-defined prediabetes and progression to type 2 diabetes in Denmark during 2012-2018: a study based on laboratory data from routine care

Sia Kromann Nicolaisen[1], Lars Pedersen[1], Daniel R. Witte[2,3], Henrik Toft Sørensen[1], Reimar Wernich Thomsen[1]

[1]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark
[2]Steno Diabetes Center Aarhus, Aarhus, Denmark

[3]Department of Public Health, Aarhus University, Aarhus, Denmark

Corresponding author: Sia Kromann Nicolaisen, Department of Clinical Epidemiology, Aarhus University Hospital, Olof Palmes Allé 43-45, 8200 Aarhus N, Denmark. E-mail: sknc@clin.au.dk, Telephone: +45 8716 7212.

List of Supplemental Tables and Figures:

ESM Table S1

The RECORD statement: checklist of items, extended from the STROBE statement that should be reported in observational studies using routinely collected health data.

	Item No.	***STROBE items***	**Location in manuscript where items are reported**	***RECORD items***	**Location in manuscript where items are reported**
Title and abstract					
	1	*(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found*	We provide details about study design and a summary of the study in the Abstract.	*RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.* *RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.* *RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.*	In the title, it is specified that we used laboratory data from routine care in Denmark for the 2012-2018 period and the Abstract mentions use of routine clinical care databases.
Introduction					
Background rationale	*2*	*Explain the scientific background and rationale for the investigation being reported*			Use of HbA1c measurements motivates the need for updated estimates of prevalence, incidence rate, and risk of progression to diabetes. This is included in the Introduction.
Objectives	*3*	*State specific objectives, including any prespecified hypotheses*			The last part of the Introduction includes the hypothesis and aims of our study.
Methods					
Study Design	*4*	*Present key elements of study design early in the paper*			The Methods include detailed descriptions of how the study was conducted. The Supplementary Material includes additional information about the study design and variable definitions.

Setting	*5*	*Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection*			In addition to describing the Danish data sources, we have provided information about selection criteria and start/end of follow-up for our study in the Methods section.
Participants	*6*	*(a) Cohort study - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up* *Case-control study - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls* *Cross-sectional study - Give the eligibility criteria, and the sources and methods of selection of participants* *(b) Cohort study - For matched studies, give matching criteria and number of exposed and unexposed* *Case-control study - For matched studies, give matching criteria and the number of controls per case*	We have described how all populations are derived and we have provided an overview of all the populations in the Flowchart and Study Design Figure.	*RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.* *RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.* *RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.*	The Variable Description Table in the Supplementary Material provides details about all the variables used in the study (including registry codes, definitions, ranges, etc.). We justify the use of the Hba1c measurements from the laboratory data and we refer to other studies when the overlap among different definitions is discussed. We have provided information about the number of individuals in our Flowchart.
Variables	*7*	*Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.*	The Variable Description Table in the Supplemental Material provides details about all the variables used in the study.	*RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.*	The Variable Description Table in the Supplemental Material provides details about all the variables used in the study.
Data sources/ measurement	*8*	*For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group*			The Variable Description Table in the Supplemental Material provides details about all the variables used in the study.
Bias	*9*	*Describe any efforts to address potential sources of bias*			The main focus of our study was to describe the use of the HbA1c measurement and to estimate incidence,

					prevalence and risk of progression as seen in routine care. However, we have standardized our estimates of incidence and prevalence to make comparisons easier.
Study size	*10*	*Explain how the study size was arrived at*			The Flowchart provides information about the study size.
Quantitative variables	*11*	*Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why*			The Variable Description Table in the Supplemental Material provides details about all the variables used in the study.
Statistical methods	*12*	*(a) Describe all statistical methods, including those used to control for confounding* *(b) Describe any methods used to examine subgroups and interactions* *(c) Explain how missing data were addressed* *(d) Cohort study - If applicable, explain how loss to follow-up was addressed* *Case-control study - If applicable, explain how matching of cases and controls was addressed* *Cross-sectional study - If applicable, describe analytical methods taking account of sampling strategy* *(e) Describe any sensitivity analyses*			The Methods section includes information about all the statistical analyses used in the study. We have tried to make clear which population is used in each analysis by providing a detailed Flowchart. We conducted analyses accounting for loss to follow-up and risk of death as a competing event. Information about missing data is included in the Baseline Table.
Data access and cleaning methods		*..*		*RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.* *RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.*	In Additional Information, we have stated that the authors had access to the data. Information about data cleaning is included in the Variable Description Table.
Linkage		*..*		*RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.*	We had access to individual-level data and in the Methods section, we state that individual linkage is possible because all Danes are assigned a unique identification number (CPR number).
Results					

Participants	*13*	*(a) Report the numbers of individuals at each stage of the study (e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed)* *(b) Give reasons for non-participation at each stage.* *(c) Consider use of a flow diagram*	The Flowchart provides information about derivation of the study populations.	*RECORD 13.1: Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.*	The Flowchart provides information about derivation of the study populations.
Descriptive data	*14*	*(a) Give characteristics of study participants (e.g., demographic, clinical, social) and information on exposures and potential confounders* *(b) Indicate the number of participants with missing data for each variable of interest* *(c) Cohort study - summarise follow-up time (e.g., average and total amount)*			The Baseline Table includes information about the study participants and information about missing data. We also provide information about follow-up time and outcomes.
Outcome data	*15*	*Cohort study - Report numbers of outcome events or summary measures over time* *Case-control study - Report numbers in each exposure category, or summary measures of exposure* *Cross-sectional study - Report numbers of outcome events or summary measures*			We provide information about follow-up time and outcomes. In addition, we provide Venn diagrams to show the overlap of the different outcome definitions.
Main results	*16*	*(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included* *(b) Report category boundaries when continuous variables were categorized* *(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period*			Our main results comprise incidences and prevalences both in the Danish population and standardized to the WHO standard population. The risk of progression takes the competing risk of death into account. All estimates include confidence intervals.
Other analyses	*17*	*Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses*			The Methods and Results sections include presentation of all our analyses.
Discussion					
Key results	*18*	*Summarise key results with reference to study objectives*			The first part of the Discussion summarizes the key findings in our study.
Limitations	*19*	*Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias*	Our Discussion includes a section about	*RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research*	Our Discussion includes a section about the use of HbA1c data and the change of definition in 2012.

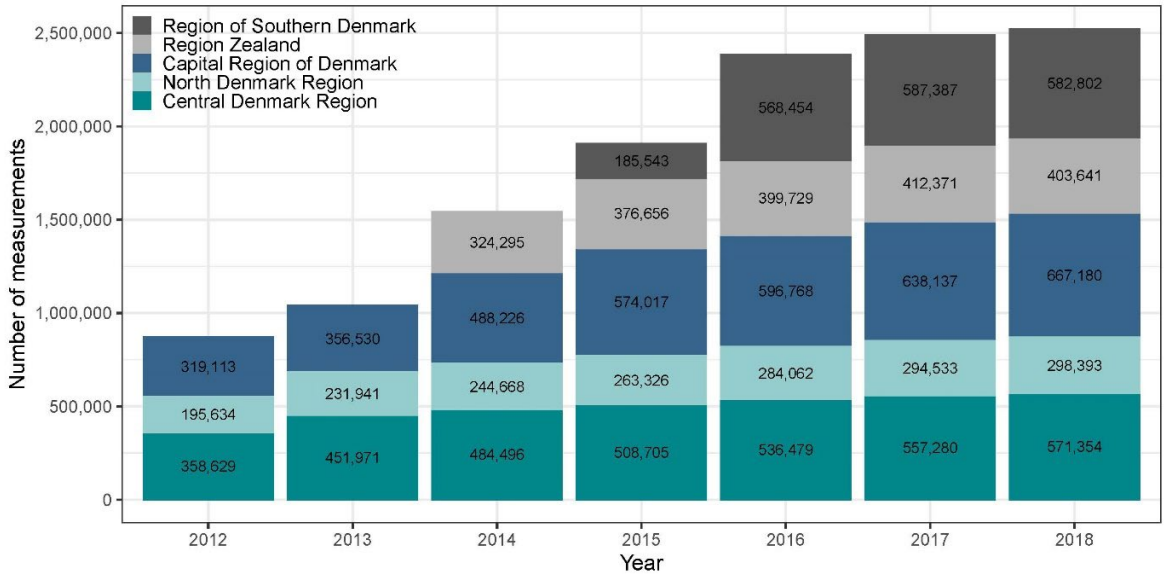
			study limitations.	*question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.*	
Interpretation	*20*	*Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence*			The Discussion provides a detailed overview of our results compared to other relevant studies.
Generalisability	*21*	*Discuss the generalisability (external validity) of the study results*			Use of nationwide registry data is discussed.
Other Information					
Funding	*22*	*Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based*			In the Additional Information, we have provided details about funding sources.
Accessibility of protocol, raw data, and programming code		*..*		*RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.*	In the Additional Information, we provide information about how to access data. In the Methods section, we provide information about the statistical software and the Variable Description Table includes information about codes. If additional information is needed, the corresponding author can be contacted.

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. PLoS Medicine 2015.

https://www.record-statement.org/checklist.php

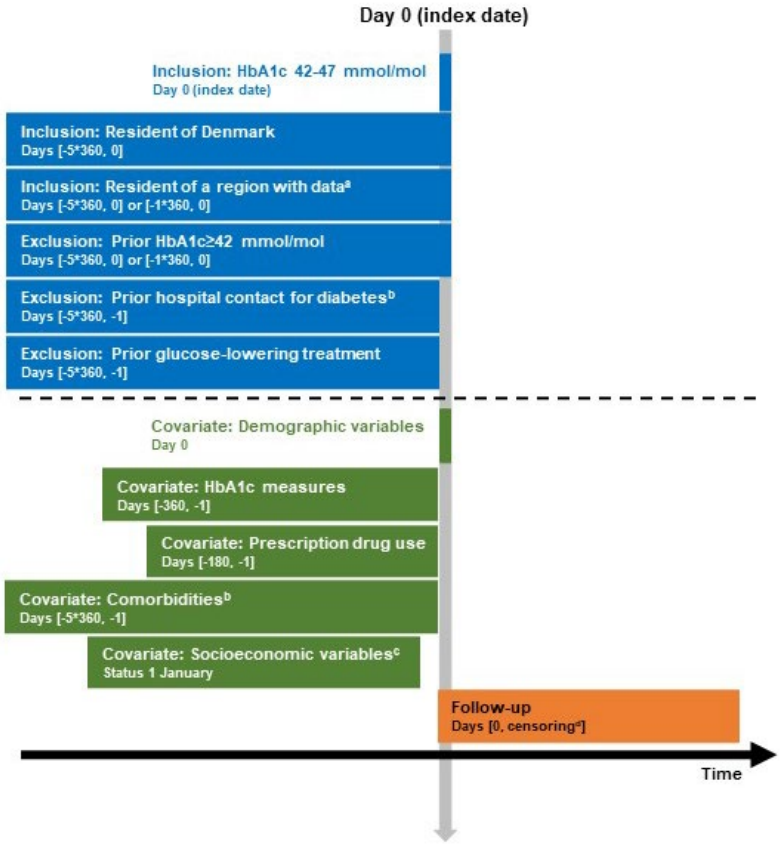
ESM Figure S1

Complete HbA1c data were available from Region of Southern Denmark since July 2015, from Region Zealand since January 2014, and from the Capital Region of Denmark since January 2010. In North Denmark Region and Central Denmark Region, data have been complete since the early 2000s. The laboratory database also contains measurements from the 1990s, but they are mostly registered with NPU codes not available in this study. A total of 12,762,320 HbA1c measurements were available in the years 2012-2018, with the Capital Region of Denmark (n=3,639,971) and the Central Denmark Region (n=3,468,914) having the most recorded measurements, followed by Region of Southern Denmark (n=1,924,186), Region Zealand (n=1,916,692) and North Denmark Region (n=1,812,557). The total number of measurements in the 3 Danish regions (Central Denmark Region, North Denmark Region, and Capital Region of Denmark) with complete HbA1c data available from the beginning of the study period in 2012 increased from 873,376 in 2012 to 1,536,927 in 2018. Data prior to 2012 were used to exclude individuals with pre-existing prediabetes/diabetes.



ESM Figure S2

Study design figure, showing inclusion and exclusion criteria, covariate assessment periods, and follow-up period for all data sources for the individuals with incident HbA1c-defined prediabetes. The index date for individuals with HbA1c-defined prediabetes is the date of the initial HbA1c measurement in the range of 42-47 mmol/mol (6.0%-6.4%) (Day 0).



[a]Temporal coverage of the laboratory database depends on the region.
[b]A hospitalization was defined using primary and secondary diagnoses for both inpatient and outpatient admissions. The admission date was used as the hospital contact day.
[c]Data on socioeconomic factors were taken from the end of the previous year. Data on employment were taken from the end of the previous November.
[d]All individuals were followed to the time of an outcome, emigration, study end (31 December 2018), end of follow-up (5 years after the index date), or death, whichever came first.

ESM Table S2

Full list of all variables, definitions, codes, and data sources in the study.

A hospitalization was defined using primary and secondary diagnoses both from inpatient admissions and outpatient visits. The admission date represented the hospital contact day. Absence of a hospital contact was defined as 'no admission'. Treatment initiation was defined as a first-time redemption of a prescription. Absence of prescriptions is defined as 'no drug use'. Variables defined based on the presence of records in the healthcare registries had no missing values, as absence of records (e.g., for antihypertensive treatment, cancer, etc.) was defined as absence of the characteristic.

Variable	Type of variable/definition	Variable assessment period	Data source	Registry definition
Definitions				
Baseline HbA1c: Value of prediabetes-defining HbA1c measurement (mmol/mol)	42-47 mmol/mol (6.0%-6.4%)	This defines the prediabetes index date.	LAB	NPU: NPU27300 (mmol/mol [IFCC]), NPU03835 (% [DCCT]), (all available measurements were converted into mmol/mol and rounded to nearest integer using the formula: IFCC=(DCCT*10.93)-23.5. Data are restricted to a maximum of one measurement per day by taking the mean of possible multiple measurements. See reference for formula: Lægehåndbogen (2020) Hæmoglobin A1c (HbA1c). Available from https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/undersoegelser-og-proever/klinisk-biokemi/blodproever/haemoglobin-a1c-hba1c/. Accessed 06 February 2021) A value of 47 mmol/mol corresponds to prediabetes and is thus not reported as 6.5%, as this would indicate diabetes.
Prior hospital contact for diabetes	Binary: No, Yes	Hospital contact day 5*360 days prior to prediabetes index date	DNPR	ICD-10: DO24, DH360, DG632, DG590, DH280, DH334B, DM142, DN083, DT383, DE10-DE14
Prior glucose-lowering treatment	Binary: No, Yes	Prescription redemption 5*360 days prior to prediabetes index date	DRMPS	ATC: A10
Prior elevated HbA1c measurement (includes prediabetes and diabetes)	Binary: No, Yes. HbA1c>=42 mmol/mol (6.0%)	Elevated HbA1c 5*360 days (or 360 days, depending on the analysis) prior to prediabetes index date	LAB	NPU: NPU27300, NPU03835
HbA1c-defined diabetes	Time to event. HbA1c>=48 mmol/mol (6.5%)	Measurement during follow-up	LAB	NPU: NPU27300, NPU03835

Hospital contact for diabetes	Time to event.	Hospital contacts during follow-up.	DNPR	ICD-10: DO24, DH360, DG632, DG590, DH280, DH334B, DM142, DN083, DT383, DE10-DE14
Glucose-lowering treatment initiation	Time to event.	Prescription redemption during follow-up	DRMPS	ATC: A10
Type 2 diabetes	Time to event	HbA1c-defined diabetes, hospital contact for diabetes, or glucose-lowering treatment initiation during follow-up.	LAB, DNPR, DRMPS	NPU: NPU27300, NPU03835. ICD-10: DO24, DH360, DG632, DG590, DH280, DH334B, DM142, DN083, DT383, DE10-DE14. ATC: A10
Death	Time to event.	During follow-up	DCRS	N/A
Censoring (emigration, study end [31 December 2018], end of follow-up [5 years after index date]), or competing risk of death	Time to event.	During follow-up	DCRS	N/A
Region	Categorical: Capital Region of Denmark, Central Denmark Region, North Denmark Region, Region of Southern Denmark, Region Zealand	Prediabetes index date.	DCRS	An individual's region of residence based on the address at the time of the measurement is considered a proxy for the region in which the measurement was made.
Covariates				
Demographic variables				
Sex	Binary: Female, Male	N/A	DCRS	N/A
Age	Continuous. 20-100 years. Categorical: 20-25, …, 95-100.	Prediabetes index date	DCRS	When categorical, the lower limit is included and the upper limit is not included, *i.e.*, for example 20-25 includes >=20 and <25.
Ethnic origin	Categorical: Danish, Immigrant/descendant/unknown	N/A	DST	Danish: if at least one parent is Danish citizen and born in Denmark. Immigrant: if the individual is born abroad and no parent is both Danish citizen and born in Denmark. Descendant: if the individual is born in Denmark and no parent is both Danish citizen and born in Denmark. For further details, see https://www.dst.dk/da/Statistik/dokumentation/Times/cpr-oplysninger/ie-type
HbA1c measures				
Place where prediabetes-defining HbA1c measurement was conducted	Categorical: Measured during inpatient hospitalization or emergency department contact, Not measured during inpatient	Prediabetes index date.	DNPR	If the index date was within an inpatient admission or emergency department admission, the measurement was assumed to have been conducted during the inpatient hospitalization. If not, but the index date was during a period with an outpatient specialist clinic admission, the measurement was

	hospitalization or emergency ward visit, but while followed at outpatient hospital specialist clinic, Neither during inpatient hospitalization or at emergency ward visit nor while followed at outpatient hospital specialist clinic			assumed to be made while the individual was seen in the outpatient clinic. Otherwise, the measurement was assumed to be neither during an inpatient hospitalization nor while followed at an outpatient specialist clinic (and thus most likely made by a general practitioner).
Presence of HbA1c measurements one year before prediabetes	Binary: No, Yes	360 days prior to the prediabetes index date	LAB	NPU: NPU27300, NPU03835
Prescription drug use				
Any trombocyte-aggregation prophylaxis	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: B01AC06, N02BA01, B01AC30, B01AC07, B01AC22, B01AC04, B01AC24, B01AC25
Any hypolipidemic treatment	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: C10, A10BH51
Loop-diuretics	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: C03C, C03EB
Any potential antihypertensive treatment (excl. loop-diuretics)	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: C02, C03A, C03B, C03D, C03E (not C03EB), C07, C08, C09A, C09B, C09C, C09D, C09X, G04CA03, C10BX04, C10BX06, C10BX07, C10BX11, C10BX12, C10BX13, C10BX14, C10BX15, C10BX10, C10BX07, C10BX09, C10BX11, C10BX14
Beta-blockers	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: C07
Calcium channel antagonists	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: C08, C09BB, C09DB, C09DX01, C09DX03, C09XA53, C09XA54, C07FB, C09BX01, C09BX03, C10BX07, C10BX09, C10BX11, C10BX14
ACE inhibitors or ATII antagonists	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: C09A, C09B, C09C, C09D, C10BX04, C10BX06, C10BX07, C10BX11, C10BX12, C10BX13, C10BX14, C10BX15, C10BX10
Oral steroids	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: H02AB
Any inhalation therapy for asthma/COPD	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: R03
Opioids	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: N02A, N07BC02
Antibiotics	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: J01
Antidepressants	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: N06A
Hypnotics/sedatives	Binary: No, Yes	180 days prior to the prediabetes index date	DRMPS	ATC: N05C
Comorbidities				
Charlson Comorbidity Index score as a measure of overall comorbidity burden	Categorical: 0, 1-2, >=3	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	See reference for ICD-10 codes: Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal

				studies: development and validation. J Chronic Dis. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8. The code DQ61 (cystic kidney disease) was not included in the calculation, as it was not available in the data. Diabetes is not included in the calculation of the Charlson Comorbidity Index score as it is included in the definition of the study cohort.
Cardiovascular disease (includes stable angina pectoris [or CABG/PCI procedures], myocardial infarction, heart failure, stroke, atrial fibrillation/flutter, heart valve disease, and venous thromboembolism)	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI20 (not DI200), DI251, DI259, KFNA, KFNB, KFNC, KFND, KFNE, KFNF, KFNG, KFNH20, DI21, DI22, DI23, DI50, DI60, DI61, DI63, DI64, DI48, DI05, DI06, DI07, DI08, DI098, DI39, DI511A, DQ22, DQ23, DI34-DI37, DI26, DI801, DI802, DI803
Angina pectoris or CABG/PCI procedures	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI20 (not DI200), DI251, DI259, KFNA, KFNB, KFNC, KFND, KFNE, KFNF, KFNG, KFNH20
Myocardial infarction	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI21, DI22, DI23
Heart failure	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI50
Atrial fibrillation/flutter	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI48
Stroke	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI60, DI61, DI63, DI64
Hypertension	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI10-DI15
Obesity	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DRMPS, DNPR	ATC: A08. ICD-10: DE65-DE68.
Chronic pulmonary disease	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DJ40, DJ684, DJ701, DJ703, DJ841, DJ920, DJ961, DJ982, DJ983
Cancer (excl. non-melanoma skin cancer)	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DC00-DC99 (not DC44)
Kidney disease	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DI12, DI13, DN07, DN11, DN14, DQ61, DN08, DE102, DE112, DE142, DN00–DN05, DN18–DN19
Liver disease	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DB18, DB150, DB160, DB162, DB190, DI85, DK70, DK71, DK72, DK73, DK74, DK760, DK766
Dementia	Binary: No, Yes	Prescription redemption 180 days prior to the prediabetes index date, or hospital contact day 5*360 days prior to the prediabetes index date	DRMPS, DNPR	ATC: N06D. ICD-10: DF00, DF01, DF02, DF03, DG30, DG310B, DG311, DG318, DG319.

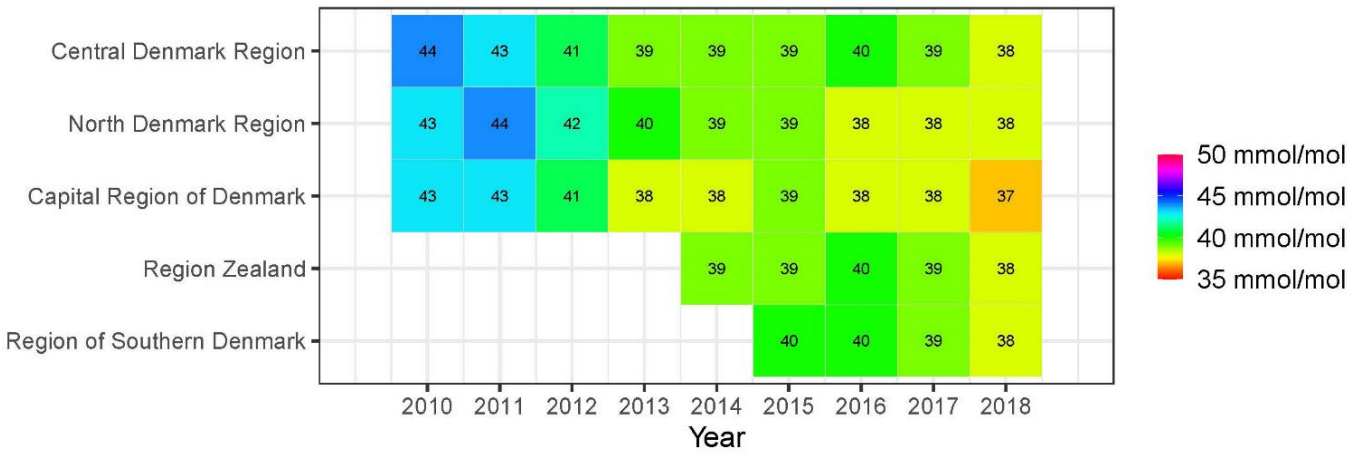
Thyrotoxicosis	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DE05
Hypothyroidism	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DE03
Pancreatic disease (includes pancreatic cancer, pancreas resection, and acute or chronic pancreatitis)	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: DC25, KJLC, DK859, DK860, DK861
Possible HbA1c-modifying conditions (includes prescription drug use of dapsone, ribavirin, antiretrovirals, trimethoprim-sulfamethoxazole, hydroxyurea, vitamin C, vitamin E, or opiates; and hospital admission or treatment with ribavirin, hemolysis, hemoglobinopathies, blood transfusion, acute blood loss/anemia, hypertriglyceridemia, chronic liver disease, pregnancy, iron deficiency, vitamin B12 deficiency, uremia, hyperbilirubinemia, end-stage renal disease [kidney transplant or dialysis], alcoholism-related diagnoses or medication, treatment with fetal hemoglobin or methemoglobin)	Binary: No, Yes	Prescription redemption 180 days prior to the prediabetes index date, or hospital contact day 5*360 days prior to the prediabetes index date	DRMPS, DNPR	ATC: J04BA02, D10AX05, J05AP01, J05, J01EE01, J04AM08, J01EA01, QJ51EA01, J01EC01, QJ01EQ11, L01XX05, A11GA, A11HA03, N02A.ICD-10: BPHM04, DD59, DD55-DD59, BOQA, DD62, BOHC, DD50, DD51, DD52, DE781, DB18, DB150, DB160, DB162, DB190, DI85, DK70, DK71, DK72, DK73, DK74, DK760, DK766, DZ321, DO00-DO99, DZ33-DZ39, DE611, DD50, DD51, BOHC2, DR39, DE804, DE806A, BJFD2, KKAS, DT861, DZ940, DF10 (not DF100), DE244, DG312, DG621, DG721, DI426, DK292, DK70, DK852, DK860, DQ860, DZ502, DZ714, DZ721, DR780, DT51, DT500A, DD564, DD74.
Previous ICU admission	Binary: No, Yes	Hospital contact day 5*360 days prior to the prediabetes index date	DNPR	ICD-10: NABE, NABB
Markers of smoking	Binary: No, Yes	Prescription redemption 180 days prior to the prediabetes index date, or hospital contact day 5*360 days prior to the prediabetes index date	DRMPS, DNPR	ATC: R03, N07BA. ICD-10: DF17, DZ716, DZ720, DJ41-DJ44.
Markers of alcoholism	Binary: No, Yes	Prescription redemption 180 days prior to the prediabetes index date, or hospital contact day 5*360 days prior to the prediabetes index date	DRMPS, DNPR	ATC: V03AA, N07BB. ICD-10: DF10 (not DF100), DE244, DG312, DG621, DG721, DI426, DK292, DK70, DK852, DK860, DQ860, DZ502, DZ714, DZ721, DR780, DT51, DT500A.
Socioeconomic variables				
Highest education achieved	Categorical: None, basic education, or primary school; Youth education, high school or similar educational level; Higher education.	Index date of prediabetes	DST	N/A

Employment status	Categorical: Employed, Unemployed, or not part of the workforce	End of previous November before prediabetes index date.	DST	N/A
Income	Categorical: Lowest income group, Low-to-medium income, Medium-to-high income, Highest income group	End of year prior to the prediabetes index year.	DST	Income group from the last calendar year. Quartiles are based on the entire Danish population per year.
Type of household	Categorical: Living alone, Not living alone	End of year prior to the prediabetes index year.	DST	N/A
Marital status	Categorical: Married, Divorced, Widow/widower, Unmarried	Index date of prediabetes	DCRS	N/A

Abbreviations: LAB, the nationwide laboratory database; DNPR, the Danish National Patient Registry; DRMPS, the Danish Register of Medicinal Product Statistics; DCRS, the Danish Civil Registration System; DST, registries maintained by Statistics Denmark.

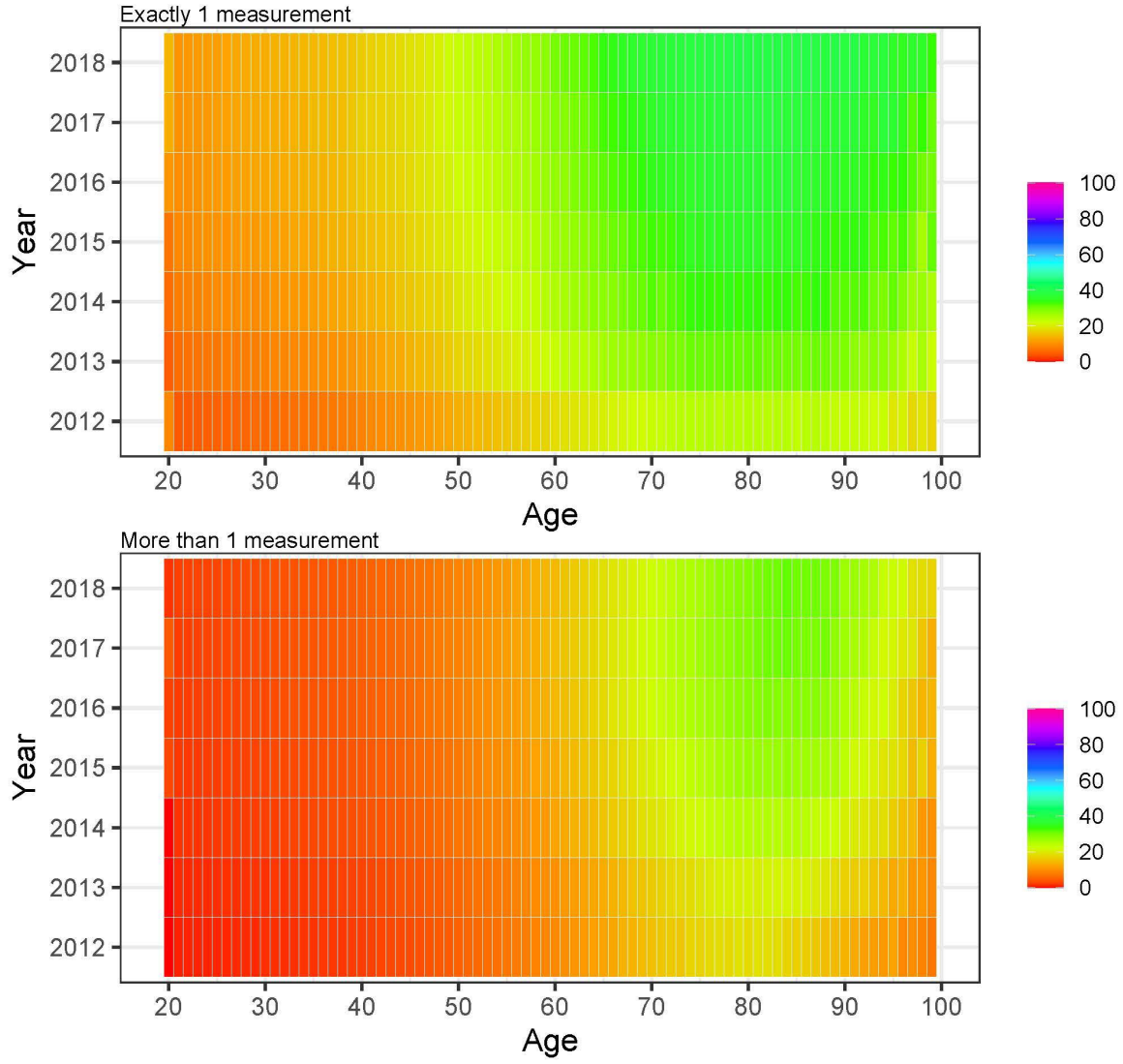
ESM Figure S3

The median value of all HbA1c measurements available in each region, for each calendar year in the study period (2012-2018) and the 2 prior calendar years (2010-2011). Prior to 2012, HbA1c was used mainly to monitor diabetes and to make decisions regarding diabetes treatment. It was used additionally as a diagnostic criterion starting in 2012.



ESM Figure S4

Proportion within each age group with only 1 and with >1 HbA1c measurement available during each calendar year.



ESM Table S3

a) Incidence rates and standardized incidence rates overall and stratified by sex, age, and calendar year.

b1), b2) Incidence rates across combinations of sex, age, and calendar year.

c) Incidence rates across combinations of sex and age.

a)	Individuals at risk		Individuals with incident HbA1c-defined prediabetes	1,000 Person-years	Incidence rate (95% CI)	Standardized incidence rate (95% CI)
Total	3,226,748		180,923	12,746.76	14.19 (14.13-14.26)	10.43 (10.38-10.48)
Sex						
Male	1,577,985		87,133	6,250.50	13.94 (13.85-14.03)	10.04 (9.97-10.10)
Female	1,648,763		93,790	6,496.26	14.44 (14.35-14.53)	10.83 (10.76-10.90)
Age (years)						
20-25	484,548		467	1,119.93	0.42 (0.38-0.46)	0.42 (0.38-0.46)
25-30	474,363		839	1,059.87	0.79 (0.74-0.85)	0.79 (0.74-0.85)
30-35	454,084		1,607	1,022.53	1.57 (1.50-1.65)	1.57 (1.50-1.65)
35-40	483,614		2,959	1,138.22	2.60 (2.51-2.70)	2.60 (2.50-2.69)
40-45	530,583		5,750	1,263.18	4.55 (4.44-4.67)	4.55 (4.43-4.67)
45-50	554,768		10,216	1,291.77	7.91 (7.76-8.06)	7.91 (7.76-8.06)
50-55	539,050		16,955	1,256.31	13.50 (13.29-13.70)	13.49 (13.29-13.70)
55-60	469,805		19,598	1,068.49	18.34 (18.09-18.60)	18.34 (18.09-18.60)
60-65	419,506		22,939	950.45	24.13 (23.82-24.45)	24.17 (23.86-24.48)
65-70	388,284		26,192	883.40	29.65 (29.29-30.01)	29.75 (29.39-30.11)
70-75	325,721		25,336	709.42	35.71 (35.28-36.16)	35.85 (35.41-36.29)
75-80	215,435		20,000	444.35	45.01 (44.39-45.64)	45.26 (44.63-45.89)
80-85	138,137		14,238	277.09	51.38 (50.55-52.24)	51.98 (51.11-52.85)
85-90	83,964		9,000	163.21	55.14 (54.02-56.30)	56.75 (55.53-57.98)
90-95	42,785		3,937	77.27	50.95 (49.38-52.57)	54.45 (52.55-56.35)
95-100	13,484		890	21.29	41.80 (39.14-44.64)	48.48 (44.57-52.38)
Calendar year						
2012	1,191,390		23,315	1,164.62	20.02 (19.76-20.28)	14.25 (14.06-14.43)
2013	1,188,322		14,718	1,162.09	12.67 (12.46-12.87)	9.02 (8.87-9.17)
2014	2,256,702		14,850	1,247.30	11.91 (11.72-12.10)	8.59 (8.44-8.73)
2015	2,307,785		37,271	2,254.83	16.53 (16.36-16.70)	12.17 (12.05-12.30)
2016	2,323,585		37,515	2,276.92	16.48 (16.31-16.64)	12.25 (12.12-12.37)
2017	2,340,053		33,274	2,287.42	14.55 (14.39-14.70)	10.87 (10.75-10.99)
2018	2,879,462		19,980	2,353.58	8.49 (8.37-8.61)	6.38 (6.29-6.47)

b1)		Incidence rate (95% CI)						
		Calendar year						
Sex	Age	2012	2013	2014	2015	2016	2017	2018
Female	20-25	0.68 (0.48-0.96)	0.39 (0.25-0.60)	0.50 (0.35-0.73)	0.65 (0.50-0.83)	0.60 (0.46-0.77)	0.45 (0.34-0.61)	0.31 (0.22-0.44)
Female	25-30	0.93 (0.67-1.29)	0.82 (0.58-1.15)	0.67 (0.47-0.96)	0.89 (0.72-1.11)	1.02 (0.83-1.24)	1.14 (0.95-1.37)	0.60 (0.47-0.77)
Female	30-35	1.86 (1.49-2.32)	1.09 (0.82-1.46)	1.11 (0.84-1.46)	1.99 (1.71-2.30)	2.14 (1.86-2.46)	1.82 (1.57-2.12)	1.03 (0.85-1.25)
Female	35-40	2.86 (2.44-3.35)	1.74 (1.42-2.14)	1.63 (1.32-2.00)	2.71 (2.41-3.04)	3.23 (2.89-3.59)	2.66 (2.36-3.00)	1.85 (1.60-2.14)
Female	40-45	4.86 (4.32-5.47)	3.02 (2.60-3.51)	3.28 (2.85-3.77)	4.81 (4.42-5.23)	5.26 (4.86-5.69)	4.53 (4.16-4.94)	2.92 (2.62-3.25)
Female	45-50	9.88 (9.13-10.69)	5.66 (5.10-6.29)	5.94 (5.38-6.56)	9.26 (8.72-9.83)	9.30 (8.76-9.88)	8.18 (7.66-8.72)	4.69 (4.31-5.10)
Female	50-55	19.01 (17.90-20.19)	10.84 (10.01-11.74)	9.99 (9.23-10.82)	16.43 (15.69-17.20)	16.64 (15.91-17.40)	14.82 (14.14-15.54)	8.42 (7.92-8.96)
Female	55-60	24.88 (23.56-26.27)	14.65 (13.64-15.73)	13.50 (12.56-14.51)	21.28 (20.37-22.23)	22.56 (21.62-23.53)	19.22 (18.36-20.12)	11.25 (10.61-11.93)
Female	60-65	31.41 (29.88-33.03)	20.34 (19.10-21.67)	17.51 (16.39-18.70)	27.08 (26.01-28.21)	27.69 (26.60-28.83)	24.23 (23.21-25.30)	13.45 (12.70-14.24)
Female	65-70	39.24 (37.45-41.10)	23.56 (22.20-25.01)	23.25 (21.94-24.65)	32.17 (30.99-33.40)	32.58 (31.37-33.84)	29.09 (27.92-30.30)	15.15 (14.32-16.03)
Female	70-75	56.99 (54.35-59.75)	32.53 (30.57-34.62)	29.20 (27.46-31.06)	40.67 (39.17-42.23)	38.75 (37.32-40.22)	34.52 (33.20-35.89)	17.76 (16.84-18.73)
Female	75-80	67.80 (64.43-71.34)	44.71 (41.98-47.62)	37.42 (35.03-39.98)	50.95 (48.88-53.12)	49.54 (47.51-51.65)	41.63 (39.79-43.55)	22.78 (21.48-24.16)
Female	80-85	75.52 (71.34-79.95)	48.67 (45.25-52.35)	42.39 (39.28-45.74)	58.11 (55.39-60.96)	53.75 (51.14-56.49)	48.51 (46.04-51.11)	24.94 (23.24-26.77)
Female	85-90	67.31 (62.63-72.34)	52.26 (48.03-56.87)	47.36 (43.40-51.68)	57.97 (54.66-61.49)	58.73 (55.34-62.32)	50.77 (47.58-54.18)	28.90 (26.52-31.49)
Female	90-95	64.68 (58.25-71.83)	42.90 (37.62-48.92)	39.10 (34.27-44.61)	53.00 (48.77-57.60)	51.62 (47.36-56.26)	48.45 (44.26-53.03)	29.54 (26.30-33.17)
Female	95-100	37.72 (29.16-48.79)	34.87 (26.50-45.89)	28.76 (21.47-38.52)	52.30 (44.86-60.98)	38.55 (32.23-46.10)	38.78 (32.43-46.38)	23.70 (18.81-29.86)

b2)		Incidence rate (95% CI)						
		Calendar year						
Sex	Age	2012	2013	2014	2015	2016	2017	2018
Male	20-25	0.23 (0.13-0.40)	0.28 (0.17-0.46)	0.15 (0.08-0.29)	0.48 (0.36-0.63)	0.43 (0.32-0.58)	0.47 (0.35-0.62)	0.18 (0.11-0.28)
Male	25-30	0.83 (0.60-1.15)	0.35 (0.22-0.57)	0.59 (0.41-0.84)	0.87 (0.70-1.08)	0.89 (0.72-1.10)	0.91 (0.74-1.11)	0.37 (0.27-0.51)
Male	30-35	1.82 (1.47-2.26)	0.94 (0.69-1.27)	1.38 (1.08-1.75)	1.78 (1.52-2.07)	1.74 (1.50-2.03)	1.67 (1.44-1.95)	1.10 (0.92-1.33)
Male	35-40	3.27 (2.83-3.79)	1.74 (1.42-2.13)	2.35 (1.98-2.78)	2.96 (2.65-3.31)	3.38 (3.05-3.76)	3.14 (2.81-3.50)	1.91 (1.66-2.19)
Male	40-45	6.67 (6.04-7.37)	3.55 (3.10-4.08)	4.23 (3.74-4.77)	5.57 (5.15-6.02)	5.64 (5.22-6.09)	5.15 (4.75-5.59)	3.39 (3.06-3.74)
Male	45-50	9.84 (9.09-10.64)	7.10 (6.47-7.79)	6.69 (6.09-7.35)	9.86 (9.30-10.45)	9.37 (8.83-9.96)	8.51 (7.98-9.07)	5.42 (5.01-5.86)
Male	50-55	17.00 (15.95-18.13)	11.26 (10.40-12.18)	10.13 (9.35-10.97)	15.95 (15.22-16.72)	15.22 (14.52-15.96)	14.17 (13.50-14.87)	8.47 (7.96-9.02)
Male	55-60	23.34 (22.05-24.71)	14.30 (13.29-15.39)	14.94 (13.94-16.02)	21.19 (20.26-22.15)	23.03 (22.07-24.03)	19.79 (18.91-20.72)	11.68 (11.01-12.38)
Male	60-65	32.92 (31.33-34.60)	21.36 (20.06-22.74)	21.84 (20.55-23.20)	28.47 (27.32-29.67)	28.96 (27.80-30.17)	26.80 (25.68-27.97)	16.75 (15.88-17.67)
Male	65-70	41.20 (39.31-43.18)	27.91 (26.38-29.54)	25.65 (24.22-27.17)	36.25 (34.92-37.63)	37.47 (36.10-38.91)	32.82 (31.50-34.20)	20.50 (19.47-21.59)
Male	70-75	55.75 (52.99-58.66)	33.24 (31.13-35.48)	32.02 (30.07-34.10)	42.99 (41.31-44.74)	44.76 (43.10-46.49)	37.90 (36.40-39.47)	22.69 (21.56-23.88)
Male	75-80	70.55 (66.73-74.59)	44.87 (41.84-48.12)	40.36 (37.60-43.33)	55.22 (52.77-57.79)	53.83 (51.43-56.33)	47.30 (45.09-49.62)	28.66 (27.01-30.40)
Male	80-85	73.71 (68.82-78.94)	55.69 (51.37-60.36)	55.95 (51.74-60.50)	64.46 (60.98-68.13)	59.82 (56.50-63.34)	55.31 (52.14-58.66)	32.39 (30.07-34.90)
Male	85-90	80.47 (73.53-88.07)	65.80 (59.43-72.85)	56.58 (50.85-62.94)	75.49 (70.36-81.00)	63.54 (58.79-68.68)	60.85 (56.16-65.92)	39.72 (36.02-43.80)
Male	90-95	71.99 (61.38-84.44)	55.87 (46.42-67.23)	55.64 (46.49-66.60)	75.18 (66.86-84.54)	65.45 (57.73-74.20)	60.02 (52.64-68.44)	49.75 (43.09-57.44)
Male	95-100	71.43 (50.24-101.6)	54.32 (35.42-83.32)	54.42 (35.84-82.66)	55.87 (41.84-74.58)	66.00 (50.55-86.17)	70.49 (53.99-92.04)	45.03 (32.02-63.35)

c)	Female				Male			
Age	Individuals at risk	Individuals with incident HbA1c-defined prediabetes	1,000 Person-years	Overall incidence rate (95% CI)	Individuals at risk	Individuals with incident HbA1c-defined prediabetes	1,000 Person-years	Overall incidence rate (95% CI)
20-25	233,876	270	536.245	0.50 (0.45-0.57)	250,672	197	583.688	0.34 (0.29-0.39)
25-30	231,541	452	513.391	0.88 (0.80-0.97)	242,822	387	546.475	0.71 (0.64-0.78)
30-35	223,194	817	500.874	1.63 (1.52-1.75)	230,890	790	521.653	1.51 (1.41-1.62)
35-40	240,251	1,390	564.508	2.46 (2.34-2.60)	243,363	1,569	573.712	2.73 (2.60-2.87)
40-45	264,802	2,647	630.341	4.20 (4.04-4.36)	265,781	3,103	632.842	4.90 (4.73-5.08)
45-50	277,831	4,947	646.717	7.65 (7.44-7.87)	276,937	5,269	645.051	8.17 (7.95-8.39)
50-55	270,723	8,715	632.235	13.78 (13.50-14.08)	268,327	8,240	624.074	13.20 (12.92-13.49)
55-60	238,020	9,901	542.871	18.24 (17.88-18.60)	231,785	9,697	525.616	18.45 (18.09-18.82)
60-65	216,078	11,317	490.452	23.07 (22.65-23.50)	203,428	11,622	460.001	25.27 (24.81-25.73)
65-70	204,156	12,886	464.626	27.73 (27.26-28.22)	184,128	13,306	418.772	31.77 (31.24-32.32)
70-75	175,151	13,080	383.022	34.15 (33.57-34.74)	150,570	12,256	326.393	37.55 (36.89-38.22)
75-80	119,698	10,734	248.725	43.16 (42.35-43.98)	95,737	9,266	195.624	47.37 (46.41-48.34)
80-85	80,756	7,975	163.702	48.72 (47.66-49.80)	57,381	6,263	113.383	55.24 (53.89-56.62)
85-90	53,433	5,407	105.488	51.26 (49.91-52.64)	30,531	3,593	57.719	62.25 (60.25-64.32)
90-95	30,429	2,623	56.137	46.73 (44.97-48.55)	12,356	1,314	21.134	62.18 (58.90-65.63)
95-100	10,538	629	16.927	37.16 (34.37-40.18)	2,946	261	4.365	59.79 (52.96-67.50)

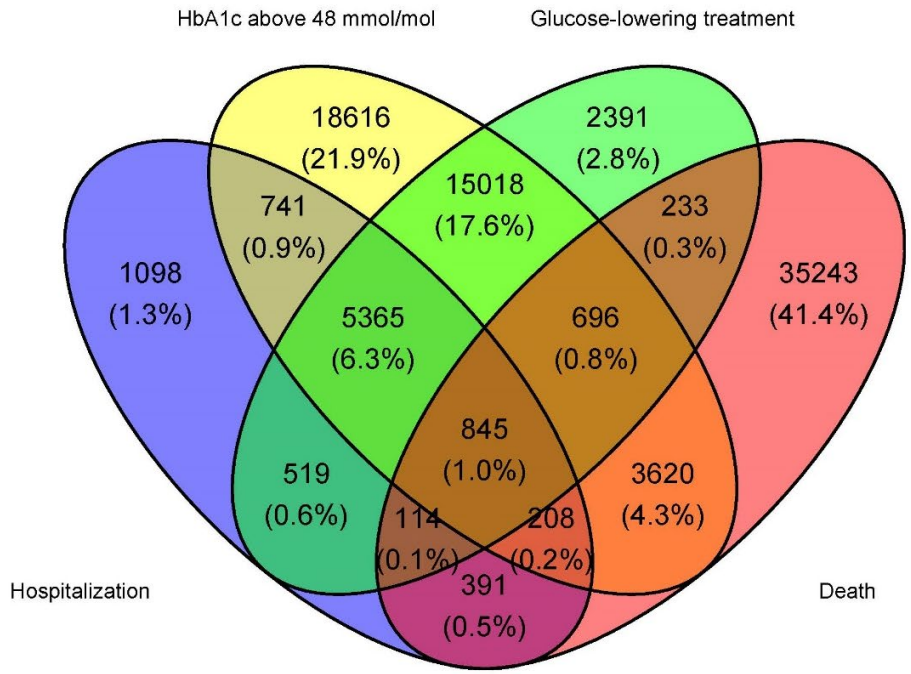
ESM Table S4

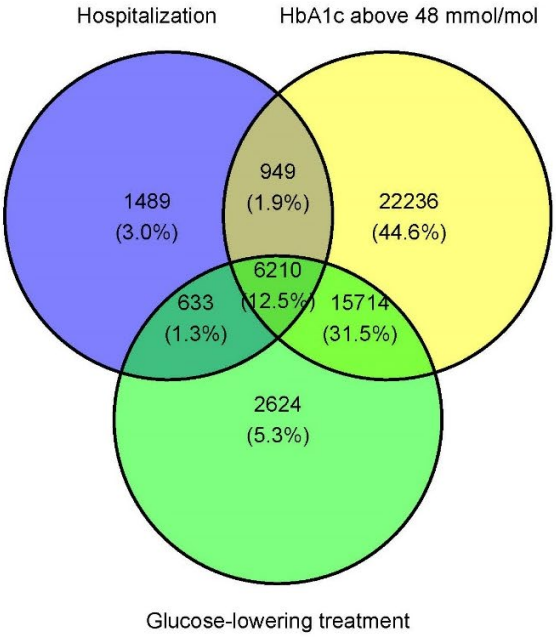
Prevalences and standardized prevalences overall and stratified by sex and age.

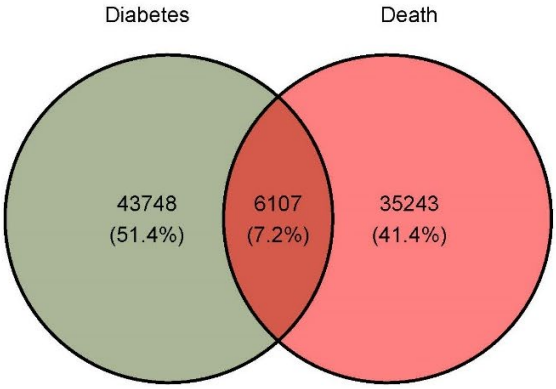
	Individuals at risk	Individuals with prevalent HbA1c-defined prediabetes	Prevalence (95% CI)	Standardized prevalence (95% CI)
Total	3,302,759	234,056	7.09 (7.06-7.11)	4.35 (4.33-4.37)
Sex				
Male	1,618,357	107,398	6.64 (6.60-6.67)	4.27 (4.25-4.30)
Female	1,684,402	126,658	7.52 (7.48-7.56)	4.42 (4.40-4.45)
Age (years)				
20-25	255,355	459	0.18 (0.16-0.20)	0.18 (0.16-0.20)
25-30	256,822	708	0.28 (0.26-0.30)	0.28 (0.26-0.30)
30-35	239,981	1,331	0.55 (0.52-0.58)	0.55 (0.52-0.58)
35-40	240,874	2,525	1.05 (1.01-1.09)	1.05 (1.01-1.09)
40-45	279,531	4,849	1.73 (1.69-1.78)	1.74 (1.69-1.78)
45-50	295,275	8,574	2.90 (2.84-2.96)	2.90 (2.84-2.96)
50-55	320,417	16,696	5.21 (5.13-5.29)	5.21 (5.13-5.29)
55-60	286,942	22,764	7.93 (7.83-8.03)	7.93 (7.83-8.03)
60-65	260,302	27,934	10.73 (10.61-10.85)	10.73 (10.61-10.84)
65-70	245,008	32,112	13.11 (12.97-13.24)	13.10 (12.97-13.24)
70-75	252,323	38,970	15.44 (15.30-15.59)	15.43 (15.29-15.58)
75-80	171,095	31,672	18.51 (18.33-18.70)	18.49 (18.30-18.67)
80-85	107,497	23,207	21.59 (21.34-21.83)	21.56 (21.31-21.80)
85-90	58,178	14,012	24.08 (23.74-24.43)	24.02 (23.67-24.38)
90-95	26,300	6,597	25.08 (24.56-25.61)	25.40 (24.82-25.97)
95-100	6,859	1,646	24.00 (22.99-25.01)	25.05 (23.77-26.32)

ESM Figure S5

Within a maximum of 5 years after diagnosis of incident HbA1c-defined prediabetes, 85,098 (23.2%) of the 366,752 individuals experienced at least one of the following outcomes: HbA1c-defined diabetes, glucose-lowering treatment initiation, hospitalization, or death.

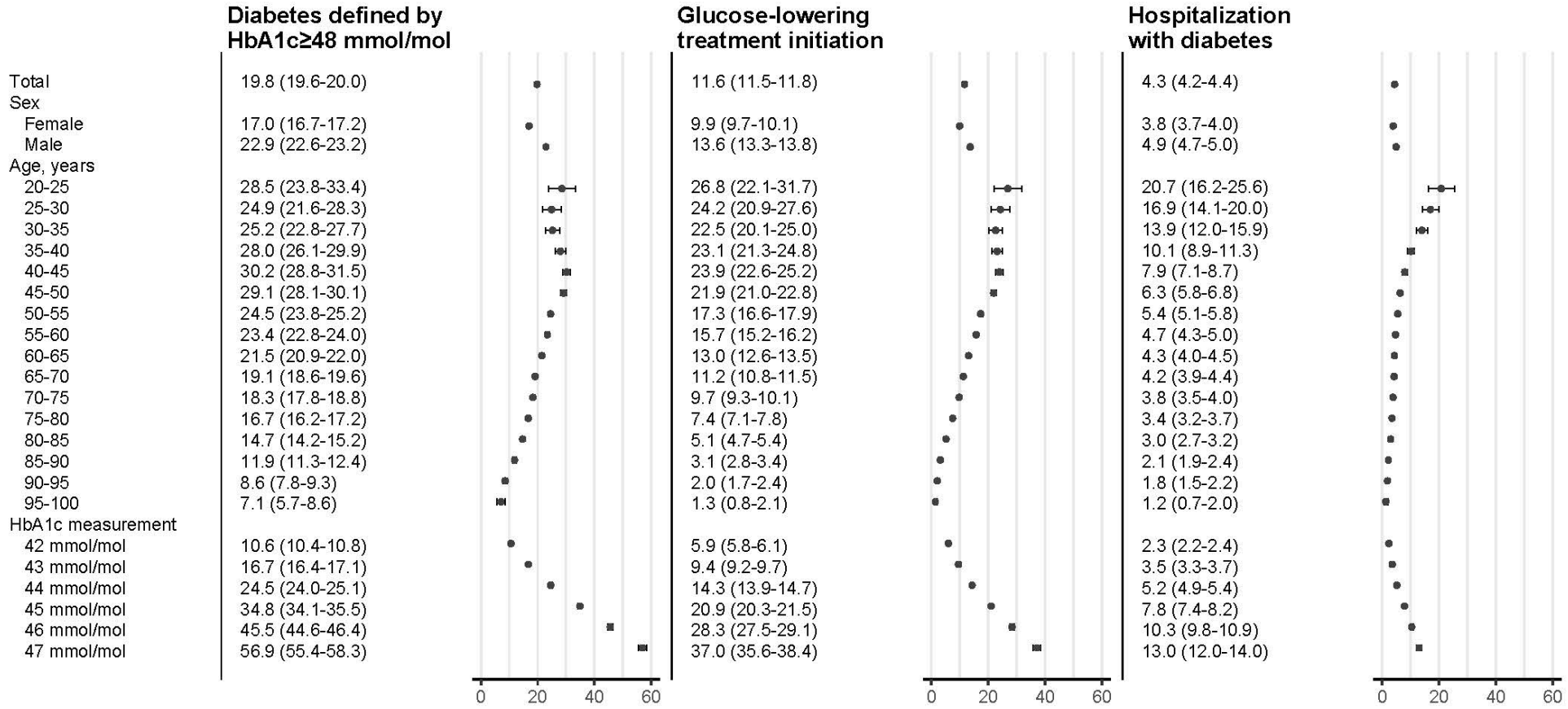






ESM Figure S6

Estimates of 5-year cumulative incidences for progression from HbA1c-defined prediabetes to HbA1c-defined diabetes, glucose-lowering treatment initiation, or hospitalization with diabetes, stratified by sex, age, and value of the baseline HbA1c measurement.



Study II | Prediction study

Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark.
Nicolaisen SK, Thomsen RW, Lau CJ, Sørensen HT, Pedersen L.
BMJ Open Diabetes Research & Care 2022;10(5):e002946.

Open access Original research

BMJ Open Diabetes Research & Care

Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark

Sia K Nicolaisen ,[1] Reimar W Thomsen,[1] Cathrine J Lau,[2] Henrik T Sørensen,[1] Lars Pedersen[1]

To cite: Nicolaisen SK, Thomsen RW, Lau CJ, *et al*. Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes in Denmark. *BMJ Open Diab Res Care* 2022;**10**:e002946. doi:10.1136/bmjdrc-2022-002946

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjdrc-2022-002946).

SKN presented preliminary results at the *37th International Conference on Pharmacoepidemiology & Therapeutic Risk Management, ICPE All Access*, August 2021. Spotlight poster SP08A.

Received 17 May 2022
Accepted 12 August 2022

For numbered affiliations see end of article.

Correspondence to
Sia K Nicolaisen;
sknc@clin.au.dk

ABSTRACT

Introduction Pre-diabetes increases the risk of type 2 diabetes, but data are sparse on predictors in a population-based clinical setting. We aimed to develop and validate prediction models for 5-year risks of progressing to type 2 diabetes among individuals with incident HbA1c-defined pre-diabetes.

Research design and methods In this population-based cohort study, we used data from the Danish National Health Survey (DNHS; n=486 495), linked to healthcare registries and nationwide laboratory data in 2012–2018. We included individuals with a first HbA1c value of 42–47 mmol/mol (6.0%–6.4%), without prior indications of diabetes. To estimate individual 5-year cumulative incidences of type 2 diabetes (HbA1c ≥48 mmol/mol (6.5%)), Fine-Gray survival models were fitted in random 80% development samples and validated in 20% validation samples. Potential predictors were HbA1c, demographics, prescriptions, comorbidities, socioeconomic factors, and self-rated lifestyle.

Results Among 335 297 (68.9%) participants in DNHS with HbA1c measurements, 26 007 had pre-diabetes and were included in the study. Median HbA1c was 43.0 mmol/mol (IQR 42.0–44.0 mmol/mol, 6.1% (IQR 6.0%–6.2%)), median age was 69.6 years (IQR 61.0–77.1 years), and 51.9% were women. During a median follow-up of 2.7 years, 11.8% progressed to type 2 diabetes and 10.1% died. The final prediction model included HbA1c, age, sex, body mass index (BMI), any antihypertensive drug use, pancreatic disease, cancer, self-reported diet, doctor's advice to lose weight or change dietary habits, having someone to talk to, and self-rated health. In the validation sample, the 5-year area under the curve was 72.7 (95% CI 71.2 to 74.3), and the model was well calibrated.

Conclusions In addition to well-known pre-diabetes predictors such as age, sex, and BMI, we found that measures of self-rated lifestyle, health, and social support are important and modifiable predictors for diabetes. Our model had an acceptable discriminative ability and was well calibrated.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Pre-diabetes increases the risk of type 2 diabetes.
⇒ HbA1c is widely used to diagnose pre-diabetes and type 2 diabetes.
⇒ Current knowledge is primarily based on pre-diabetes and diabetes defined by measures other than HbA1c (eg, fasting glucose or glucose tolerance tests).

WHAT THIS STUDY ADDS

⇒ One in five individuals with pre-diabetes will progress to HbA1c-defined diabetes within 5 years.
⇒ In addition to well-known predictors such as age, sex, and body mass index, self-rated lifestyle, health, and social support are important and modifiable predictors for type 2 diabetes.
⇒ Although we identified individuals with pre-diabetes who were at high risk, the time-dependent area under the curve was only 73 (95% CI 71 to 74) for HbA1c-defined diabetes.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The use of prognostic prediction models can aid in identifying individuals who will develop type 2 diabetes, allowing preventive interventions to be targeted more effectively.
⇒ Focus should be on physical health and on self-rated mental health and social support.

INTRODUCTION

Pre-diabetes is defined by glucose levels that are elevated, but below the threshold for diagnosing overt diabetes. In 2011, the WHO concluded that measurements of glycated hemoglobin (HbA1c) ≥48 mmol/mol (6.5%) could be used to diagnose type 2 diabetes, as a convenient alternative to existing methods based on elevated fasting blood glucose or abnormal 2-hour oral glucose tolerance tests.[1 2] Since then, HbA1c testing has been used for both screening and diagnosing type 2 diabetes, as well as for making treatment decisions.[3–6] It is currently one of the most commonly used blood tests in routine clinical care.[7]

Individuals with pre-diabetes are at increased risk of later developing type 2 diabetes.[1–4 8–10] To create risk stratification tools and effectively target preventive interventions, it is important to know the magnitude, as well as predictors, of risk for progression to type 2 diabetes. Current knowledge is based primarily on cohorts established in the 1990s and 2000s,[9–15] when pre-diabetes and type 2 diabetes were defined by measures other than HbA1c (eg, fasting glucose or glucose tolerance tests). We hypothesized that in the current era of widespread HbA1c screening in routine care, many individuals with pre-diabetes are detected early and that linked laboratory databases can aid in identifying individuals who will later develop type 2 diabetes.

We therefore examined the 5-year risk and risk predictors of type 2 diabetes in individuals with incident HbA1c-defined pre-diabetes (HbA1c 42–47 mmol/mol (6.0%–6.4%)) using the Danish National Health Survey and Danish nationwide medical registries. We restricted our analysis to data available after 2012, when identification of pre-diabetes, diagnosis of type 2 diabetes, and diabetes treatment decisions in Denmark were all based primarily on HbA1c levels.

METHODS

We follow the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis[16] reporting guidelines throughout this paper (online supplemental material table S1).

Data sources

This prognostic prediction study is a population-based cohort study based on data from the Danish National Health Survey[17] and nationwide medical registries. Denmark has a tax-supported healthcare system that ensures unfettered access to medical care for all residents[18] (approximately 5.8 million individuals in 2018), including access to general practitioners and hospitals and partial reimbursement for prescribed drugs. All Danes are assigned a unique personal identification number at birth or upon immigration, making individual linkage among registries possible.[19]

The Danish National Health Survey[17] includes self-reported information from approximately 300 000 representatively sampled Danes in each of the years 2010, 2013, and 2017. The information includes body mass index (BMI), alcohol consumption, smoking status, and dietary habits, as well as self-rated health, lifestyle, and quality of life. HbA1c measurements were obtained from the nationwide Register of Laboratory Results for Research[7] and the regional Clinical Laboratory Information System Research Database at Aarhus University[7] (online supplemental material figure S1). These registries contain virtually all laboratory measurements ordered by hospital clinicians and general practitioners for members of the Danish population.[7] Additional individual-level information was obtained from the following registries: the Danish National Patient Registry, which contains all discharge diagnoses from Danish hospitals since 1977 and from hospital emergency room and outpatient clinic contacts since 1995[20]; the Danish Civil Registration System, which contains data on vital status and date of death; the Danish Register of Medicinal Product Statistics, which contains complete prescription information from all community-based pharmacies since 1994[21]; and socioeconomic registries maintained by Statistics Denmark, which contain data on family and household socioeconomics, ethnic origin, education level, employment status, and income.

Study cohort

All individuals responding at least once to the Danish National Health Survey in the 2010, 2013 or 2017 rounds were initially eligible for this study (n=486 495). Eligibility was then restricted to individuals with at least one HbA1c measurement in the laboratory data during the 2012–2018 period. To establish a cohort of individuals with pre-diabetes, we further restricted inclusion to individuals with HbA1c measurements between 42 mmol/mol (6.0%) and 47 mmol/mol (6.4%), which is used as the definition of pre-diabetes in Denmark[2] (online supplemental material figure S2). Other eligibility criteria were at least 5 years of residency in Denmark and at least 1 year of residency in a region with available laboratory data. As our main focus was on incident pre-diabetes, a measurement was excluded if another HbA1c measurement of 42–47 mmol/mol (6.0%–6.4%) was obtained within the prior year. Measurements were also excluded if an individual had previously diagnosed or treated diabetes (ie, an HbA1c measurement ≥48 mmol/mol (6.5%) within the year prior to the measurement date, contact at any hospital with a diagnosis of diabetes within the previous 5 years, or redemption of a prescription for glucose-lowering medication within the last 5 years; (Online supplemental material figure S2 and table S2). The date of the first measurement of HbA1c-defined pre-diabetes was set as the pre-diabetes index date. Individuals aged <30 years on the index date were excluded from the analysis,[22–24] as they were likely to have type 1 diabetes. Finally, the analysis was restricted to individuals who responded to the health survey within 5 years prior to the pre-diabetes index date (online supplemental material figure S2,S3).

Study outcomes and follow-up

The primary outcome of interest was HbA1c-defined type 2 diabetes, defined as the first HbA1c measurement ≥48 mmol/mol (6.5%) during follow-up. As a secondary outcome, we examined time to glucose-lowering treatment initiation, defined as the first redemption of a prescription for a drug in the Anatomical Therapeutic Chemical Codes 'antidiabetic drug' category during follow-up (online supplemental material figure S2 and table S2). Individuals were followed from their index date to the occurrence of an outcome, emigration, study end (31 December 2018), end of follow-up (5 years after the index date), or death, whichever came first. Death was

treated as a competing risk, while emigration, study end, and end of follow-up entailed censoring in the survival models.

Potential predictors

Potential predictors of progression to type 2 diabetes were identified based on a combination of findings reported in the existing literature,[10 25–27] pathophysiological and clinical knowledge, and availability of data for our project. Online supplemental material table S2 provides information on the definitions of all potential predictors included in this study. We assessed more than 30 potential predictors on the pre-diabetes index date. These encompassed demographic variables, including sex, age, and ethnic origin; HbA1c measures, including the value of the first pre-diabetes-defining HbA1c measurement (baseline HbA1c level), as well as the presence of any HbA1c measurements during the year prior to the index date; physician-prescribed drugs purchased at pharmacies (redemption within 180 days of the index date of prescriptions for statins, any antihypertensive drugs, oral steroids, or opioids); comorbidities (hospital diagnoses within 5 years or drug use within 180 days of the index date indicating pancreatic disease, cardiovascular disease, lung disease, cancer, or possible HbA1c-modifying conditions) and the Charlson Comorbidity Index score (as a measure of overall comorbidity); socioeconomic variables, including education, employment, income, and type of household (living alone vs not living alone); and self-reported lifestyle and health indicators, including BMI, alcohol consumption, smoking status, dietary habits, and several questions on self-rated health and quality of life.

The data included only few records with missing data (a maximum of 8% missing values was recorded for alcohol consumption). We therefore deemed it appropriate to perform complete-case analyses (online supplemental material table S2).

Statistical analysis

An overall 5-year cumulative incidence curve of progression to HbA1c-defined type 2 diabetes or glucose-lowering treatment initiation was estimated using the non-parametric estimate of the cause-specific cumulative incidence function with death as a competing event. The cumulative incidence of death was estimated based on the Kaplan-Meier estimate.

Individuals were randomly split into a development sample (80%) used for model development and a validation sample (20%) used to estimate external model performance. For each potential predictor in the development sample, the hazard ratio(HR) for type 2 diabetes was estimated in a Cox model adjusted for sex, age, index year, and region of residence.

Model development

The individual risk of type 2 diabetes after 5 years was derived from cumulative incidence functions. These were estimated based on the subdistribution hazard defined by Fine and Gray[28] using the Breslow-type estimate of the underlying subdistribution hazard evaluated after 5 years.

The main model was developed in two steps. First, a Fine-Gray survival model with the least absolute shrinkage and selection operator (LASSO) was fitted to perform variable selection among all potential predictors using 1000 iterations and the Bayesian information criterion.[29] Then, a Fine-Gray survival model was refitted using the selected variables. A minimum model, a Fine-Gray survival model including only age and sex with no variable selection, was fitted for comparison purposes.

Model validity

The main and minimum models were applied to the validation sample and 5-year risks were estimated for each individual. The discrimination of the models was assessed using time-dependent receiver operating characteristic curves and the time-dependent area under the curve (AUCt),[30] both estimated after 5 years. The AUCt was estimated using inverse probability of censoring weighting with Kaplan-Meier estimated weights. Similarly, time-dependent sensitivity, specificity, positive predictive values, and negative predictive values were estimated after 5 years for prespecified risks and for the value of the maximized Youden index (sensitivity+specificity−1). Along with the Brier score, the calibration of the models was visually assessed using the calibration curves. The index of prediction accuracy (IPA, a rescaled version of the Brier score)[31] was used to consider calibration and discrimination simultaneously.

Sensitivity analyses

To ensure that model performance was not changed substantially by a possible interaction between BMI and the HbA1c level, models were fitted in which both variables were included categorically along with their interactions. The models were fitted for both outcomes and model performance was compared with the main models.

To ensure that the self-reported lifestyle and health indicators reflected the status close to the pre-diabetes index date, the cohort was restricted to individuals with data from the Danish National Health Survey 1 year prior to the index date (online supplemental material figure S3). The main model was refitted in the restricted development sample and validated in the validation sample.

To examine whether our study results were stable across middle-aged versus elderly patient groups, we reran all analyses among the individuals <60 years of age at the pre-diabetes index date and among the individuals ≥60 years of age.

To explore the impact of the limited availability of historical laboratory data (online supplemental material figure S1), we focused on the subset of individuals with at least 5 years of laboratory data and assessed the effect of this exclusion criterion.

All statistical analyses were conducted using SAS V.9.4 (SAS Institute) and R V.4.0.2 (R Core Team, 2020). For

a list of essential R packages, see online supplemental material table S3.

RESULTS

Among the 486 495 individuals with Danish National Health Survey data, 335 297 (68.9%) had at least one HbA1c measurement recorded during the 2012–2018 study period, of whom 69 303 (20.7%) had at least one HbA1c measurement in the interval of pre-diabetes at 42–47 mmol/mol (6.0%–6.4%; online supplemental material figure S4). After exclusion of individuals with previously known diabetes or pre-diabetes (1 year lookback for laboratory measurements, 5 years for hospital diagnoses and glucose-lowering treatment), 26 007 (37.5%) were identified as having incident HbA1c-defined pre-diabetes, and thus formed our study cohort for assessment of progression to type 2 diabetes. Of these, 15 737 (60.5%) individuals had at least 5 years of available laboratory data prior to inclusion (see the Sensitivity analyses section). The median follow-up time was 2.72 years (IQR 1.42–4.43 years). Overall cumulative incidence curves for type 2 diabetes with death as a competing event are shown in online supplemental material figure S5. The overall 5-year cumulative incidence was 19.3% (95% CI 18.6% to 20.0%) for type 2 diabetes defined as HbA1c ≥48 mmol/mol (6.5%) and 11.2% (95% CI 10.6% to 11.8%) for type 2 diabetes defined as initiation of glucose-lowering treatment (online supplemental material figure S5). The overall 5-year cumulative incidence of death was 16.3% (95% CI 15.6% to 16.9%).

The 26 007 individuals were randomly divided into a development sample (n=20 806) and a validation sample (n=5201). In the development sample, 10 792 (51.9%) individuals were women and the median age at pre-diabetes diagnosis was 69.6 years (IQR 61.0–77.1 years; table 1 and online supplemental material table S4). The median BMI was 26.7 kg/m^2 (IQR 24.1–29.8 kg/m^2). The median baseline HbA1c measurement was 43.0 mmol/mol (IQR 42.0–44.0 mmol/mol) or 6.1% (IQR 6.0%–6.2%), and the HR for progression to HbA1c-defined type 2 diabetes steadily increased from 1.67 (95% CI 1.47 to 1.89) for an HbA1c level of 43 mmol/mol (6.1%) vs 42 mmol/mol (6.0%; reference) to 13.69 (95% CI 11.75 to 15.94) for an HbA1c level of 47 mmol/mol (6.4%) vs 42 mmol/mol (6.0%; table 1 and online supplemental material table S4). The characteristics of individuals in the development and validation samples were nearly identical (table 1 and online supplemental material table S4).

In the development sample, 2449 individuals (11.8%) had an HbA1c measurement ≥48 mmol/mol (6.5%) within 5 years. Median follow-up time was 2.73 years (IQR 1.42–4.45 years) and 4026 (19.4%) individuals were followed for at least 5 years. During the same period, 1339 (6.4%) individuals initiated a glucose-lowering treatment indicating type 2 diabetes, and a total of 2101 (10.1%) died (online supplemental material figure S6).

Table 1 Baseline characteristics of the development sample

			HR	
	n (%) or median (IQR)	Missing values (%)	HbA1c ≥48 mmol/mol (6.5%)	Glucose-lowering treatment initiation
Total	20 806 (100.0)			
Sex		0 (0.0)		
Female	10 792 (51.9)		0.67 (0.62; 0.73)	0.68 (0.61; 0.76)
Male	10 014 (48.1)		Ref	Ref
Age (years)	69.6 (61.0–77.1)	0 (0.0)	0.99 (0.98; 0.99)	0.97 (0.96; 0.97)
Body mass index (kg/m^2)	26.7 (24.1–29.8)	986 (4.7)	1.05 (1.04; 1.06)	1.07 (1.06; 1.08)
Value of pre-diabetes-defining HbA1c measurement (mmol/mol)	43.0 (42.0–44.0)	0 (0.0)	1.69 (1.65; 1.73)	1.68 (1.63; 1.74)
Value of pre-diabetes-defining HbA1c measurement (mmol/mol)		0 (0.0)		
42	9081 (43.6)		Ref	Ref
43	5061 (24.3)		1.67 (1.47; 1.89)	1.58 (1.32; 1.89)
44	3080 (14.8)		2.74 (2.41; 3.12)	2.89 (2.43; 3.44)
45	1794 (8.6)		5.15 (4.53; 5.86)	5.24 (4.39; 6.24)
46	1180 (5.7)		7.93 (6.93; 9.07)	8.04 (6.71; 9.62)
47	610 (2.9)		13.69 (11.75; 15.94)	12.48 (10.15; 15.34)

The HR is adjusted for sex, age, index year, and region of residence.

Prediction of progression to HbA1c-defined type 2 diabetes

Using LASSO, components from 11 of the potential predictors were selected for the type 2 diabetes prediction model. Within this model, a high HbA1c level at baseline was associated with increasing risk, with a subdistribution hazard ratio(SHR) of 1.64 (95% CI 1.60 to 1.69) per one-unit increase in mmol/mol (online supplemental material table S5). The prediction model also included a younger age at onset of pre-diabetes (SHR 0.99 (95% CI 0.98 to 0.99) for each 1-year increase in age), male sex (SHR 0.74 (95% CI 0.67 to 0.80) female vs male), increasing BMI (SHR 1.03 (95% CI 1.02 to 1.04) for each one-unit increase in kg/m^2), receipt of treatment for hypertension (SHR 1.17 (95% CI 1.06 to 1.28)), and presence of pre-existing pancreatic disease (SHR 2.61 (95% CI 1.49 to 4.57)). Absence of pre-existing cancer also predicted type 2 diabetes (SHR 0.76 (95% CI 0.65 to 0.90)), as cancer was a strong predictor of death, precluding later type 2 diabetes. Several self-reported health measures were also predictors of type 2 diabetes progression: self-reported unhealthy diet (SHR 1.13 (95% CI 1.01 to 1.27) for unhealthy vs average or healthy diet), having been advised by a doctor to lose weight or change dietary habits (SHR 1.40 (95% CI 1.26 to 1.56)), not having anyone to talk to when in need of support (SHR 1.29 (95% CI 1.08 to 1.55) for never/almost never vs often, mostly, or sometimes), and good self-rated health (SHR 1.13 (95% CI 1.04 to 1.23) for good vs fair/poor or excellent/very good health; online supplemental material table S5).

In the validation sample, the main model had the highest AUCt (72.7 (95% CI 71.2 to 74.3)), indicating better discriminative ability than the minimum model, which included only age and sex (AUCt 68.2 (95% CI 66.7 to 69.7); table 2 and figure 1). The main model had a lower Brier score (10.7 (95% CI 8.8 to 12.6)) and a higher IPA (18.2). This indicated better overall performance when calibration was taken into consideration (table 2). The calibration curves generally showed good calibration for both models (figure 1). Comparing the estimated probabilities in the two models, the main model assigned higher probabilities to a large subgroup of the individuals who progressed to type 2 diabetes, without overestimating the probabilities for those without the outcome (figure 2, online supplemental material figure S7 and table S6). The Youden index provided the optimal decision rule, classifying individuals with a risk >16.0% as being at high risk of type 2 diabetes, yielding a sensitivity of 68.3 (95% CI 63.9 to 72.7) and specificity of 66.3 (95% CI 65.4 to 67.1; online supplemental material table S7). The main model performed better than the minimum model for high sensitivity values (figure 1).

Prediction of progression to type 2 diabetes defined as glucose-lowering treatment initiation

The model in which type 2 diabetes was defined as initiation of glucose-lowering treatment consisted of components from only five potential predictors after using LASSO. The following variables were associated with increasing risk (online supplemental material table S5): increasing HbA1c level at baseline (SHR 1.63 (95% CI 1.58 to 1.69) per one-unit increase in mmol/mol), younger age (SHR 0.97 (95% CI 0.97 to 0.98) for each 1-year increase in age), male sex (SHR 0.76 (95% CI 0.67 to 0.85) female vs male), increasing BMI (SHR 1.05 (95% CI 1.04 to 1.06) for each one-unit increase in kg/m^2), and having been advised by a doctor to lose weight or change dietary habits (SHR 1.44 (95% CI 1.27 to 1.65)). In the validation sample, the main model for initiation of glucose-lowering treatment had an AUCt of 79.4 (95% CI 77.7 to 81.0; table 2). The main model's discriminative ability was similar to that of the minimum model (AUCt 79.8 (95% CI 78.1 to 81.4)), but it was better calibrated and had greater ability to identify individuals at high risk (table 2, online supplemental material figures S7–S9).

Sensitivity analyses

The model in which BMI and baseline HbA1c were included categorically along with the interactions improved both discriminative ability (AUCt 73.8 (95% CI 72.2 to 75.4) for HbA1c ≥48 mmol/mol (6.5%) and AUCt 80.0 (95% CI 78.4 to 81.6) for glucose-lowering treatment initiation) and calibration, but not markedly (online supplemental material figure S10).

For both outcomes, the models fitted to the restricted development sample showed similar discriminative ability

Table 2 Performance measures for the prediction models

	Definition 1: HbA1c ≥48 mmol/mol (6.5%)			Definition 2: glucose-lowering treatment initiation		
	AUCt (%)	Brier score (%)	IPA	AUCt (%)	Brier score (%)	IPA
Main model	72.7 (71.2–74.3)	10.7 (8.8–12.6)	18.2	79.4 (77.7–81.0)	7.5 (5.9–9.1)	17.1
Minimum model	68.2 (66.7–69.7)	12.8 (10.7–14.8)	2.8	79.8 (78.1–81.4)	8.6 (6.8–10.5)	4.6

The models were validated (using the validation sample) for both definitions of type 2 diabetes. High AUCt values indicate good discrimination. Low Brier scores indicate good calibration. High IPA indicates good average performance.
The main model for HbA1c ≥48 mmol/mol (6.5%) included baseline HbA1c, age, sex, body mass index (BMI), treated hypertension, pre-existing pancreatic disease, absence of cancer, unhealthy diet, doctor's advice to lose weight or change dietary habits, self-reported lack of anyone to talk to, and good self-rated health. The main model for glucose-lowering treatment initiation included baseline HbA1c, age, sex, BMI, and doctor's advice to lose weight or change dietary habits.
AUCt, time-dependent area under the curve; IPA, index of prediction accuracy.

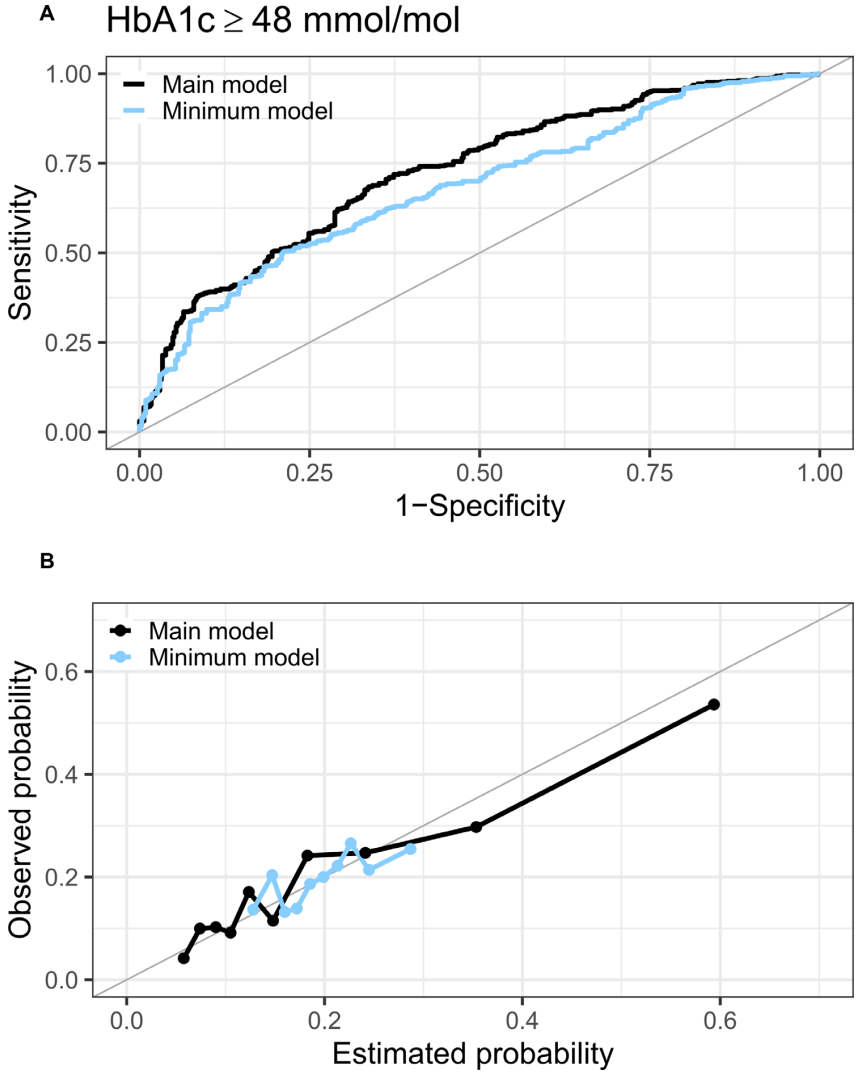


Figure 1 Comparison of the two models predicting type 2 diabetes defined as HbA1c ≥48 mmol/mol (6.5%). (A) Time-dependent receiver operating characteristic curve comparing the discriminative ability of the main model (including baseline HbA1c, age, sex, body mass index (BMI), treated hypertension, pre-existing pancreatic disease, absence of cancer, unhealthy diet, doctor's advice to lose weight or change dietary habits, self-reported lack of anyone to talk to, and good self-rated health) to the minimum model including only age and sex. (B) Calibration curve comparing the estimated and observed probabilities for the two models. The estimates for the observed probabilities were defined based on quantiles of the estimated probabilities.

(AUCt 72.9 (95% CI 71.3 to 74.4) for HbA1c ≥48 mmol/mol (6.5%) and AUCt 79.2 (95% CI 77.6 to 80.8) for glucose-lowering treatment initiation) and calibration when compared with the main models fitted to the entire development sample (online supplemental material figure S11).

When stratified by age below or above 60 years, the variable selection included fewer variables than in the main models. The coefficients in the stratified models were generally similar to the coefficients in the main models. All models showed a lower discriminative ability, and the calibration was generally impaired compared with the main models (online supplemental material figure S12).

Among the 15 737 individuals with at least 5 years of available laboratory data, we found that 2111 (13.4%) should have been excluded due to prior pre-diabetes (42≤HbA1c≤47 mmol/mol (6.0%≤HbA1c≤6.4%)), 166 (1.1%) due to prior type 2 diabetes (HbA1c≥48 mmol/mol (6.5%)), and 423 (2.7%) due to both pre-diabetes and type 2 diabetes within the past 5 years.

CONCLUSIONS

We showed that one in five individuals from our population will progress to HbA1c-defined type 2 diabetes within 5 years after their first HbA1c-defined pre-diabetes

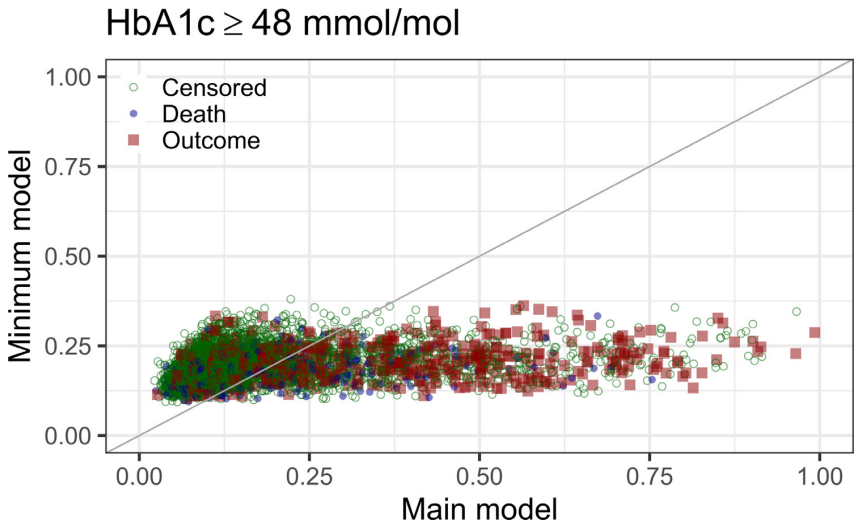


Figure 2 A comparison of the estimated probability of type 2 diabetes defined as HbA1c ≥48 mmol/mol (6.5%) from the two prediction models. The graph is colored by observed outcome: type 2 diabetes, death, or censored (ie, emigration, study end (31 December 2018), or end of follow-up (5 years after index date)). The main model includes baseline HbA1c, age, sex, body mass index (BMI), treated hypertension, pre-existing pancreatic disease, absence of cancer, unhealthy diet, doctor's advice to lose weight or change dietary habits, self-reported lack of anyone to talk to, and good self-rated health. The minimum model includes only age and sex. To avoid reporting sensitive individual-level information, random noise was added to all estimates (normal distribution, mean=0, SD=0.01).

diagnosis, and that one in nine will initiate glucose-lowering treatment within the same period. In addition to age, sex, metabolic factors and pre-existing comorbidities, we found that self-rated health, lifestyle, and existence of a social network are important predictors of the progression to type 2 diabetes. Although we could identify individuals with pre-diabetes who were at high risk, the AUCts were modest at only 73 (95% CI 71 to 74) for HbA1c-defined type 2 diabetes and 79 (95% CI 78 to 81) for glucose-lowering treatment initiation.

Comparison to other studies

HbA1c levels above the lower limit for pre-diabetes have been shown to increase the risk of future type 2 diabetes compared with normal levels of HbA1c,[1 2 27 32 33] but many individuals with pre-diabetes never progress to overt diabetes. In the Whitehall II cohort (26.4% women, mean age 61.6 years, mean HbA1c 42 mmol/mol, and mean BMI 24.6 kg/m^2[34]), an observed 14% of individuals with pre-diabetes (HbA1c 39–47 mmol/mol (5.7%–6.4%)) developed diabetes (HbA1c ≥48 mmol/mol (6.5%)) within 5 years.[34] The Whitehall II cohort was much younger than our study population (mean 61.6 years vs median 69.9 years) and included fewer women (26.4% vs 51.9% women in our study). The Whitehall II finding of 14% developing diabetes is close to the observed 12% of individuals reaching an HbA1c level ≥48 mmol/mol (6.5%) within 5 years of follow-up in our study; however, our median follow-up time was shorter (median 2.7 years of follow-up in our study vs median 6.7 years in Whitehall II). In the Diabetes Prevention Program Outcomes Study (DPPOS; 68% women, mean age 51 years, mean HbA1c 41 mmol/mol, mean BMI 34 kg/m^2),[11 32] an estimated 35% of individuals with pre-diabetes defined as elevated fasting plasma glucose (FPG; 5.3–6.9 mmol/l) or abnormal 2-hour plasma glucose (2hPG; 7.8–11.0 mmol/l) developed diabetes (FPG ≥7.0 mmol/l or 2hPG ≥11.0 mmol/l) within 5 years. As only 26% of the DPPOS participants with diabetes according to glucose criteria also had HbA1c levels ≥48 mmol/mol (6.5%),[32] we could not make a direct comparison with our study[34]; however, our estimates (19% for diabetes defined by HbA1c ≥48 mmol/mol (6.5%) and 11% for glucose-lowering treatment initiation) were markedly lower. Compared with our study population, the DPPOS included more women (68% in DPPOS vs 52% in our study) and a lower baseline HbA1c value (mean 41 mmol/mol in DPPOS vs median 43 mmol/mol in our study), with both variables predicting lower diabetes progression risk. On the other hand, DPPOS participants had a substantially higher BMI (mean BMI 34.7 kg/m^2 in DPPOS vs median 26.7 kg/m^2 in our study) and were markedly younger (mean age 51.1 years in DPPOS vs median 69.6 years in our study), with both factors increasing the risk of diabetes in our models.[11]

In a review, Jonas *et al* emphasized the current lack of evidence concerning diabetes screening and pre-diabetes interventions available from trials based on HbA1c values.[35] They highlighted the need for further research on factors associated with risk of progression from pre-diabetes to overt diabetes.[35] In addition to some important and previously known predictors of developing type 2 diabetes—younger age at onset of pre-diabetes

(often associated with more obesity and a more severe pre-diabetes phenotype[22]), male sex, high BMI, and pre-existing comorbidities—we also found self-rated health, self-reported doctor's advice regarding lifestyle problems, and measures of lack of a strong social network to be important predictors for diabetes. Mental well-being and the perception of having a supportive social network may be important factors in successful changes of poor health behavior. Moreover, perceived loneliness was recently found to be a strong independent predictor of incident type 2 diabetes, independent of living alone, socioeconomic factors, and lifestyle factors.[36] Mechanisms are unclear, but loneliness may associate with dysregulation in cortisol responses and heightened inflammation.[36]

Our models indicated that higher versus lower HbA1c at time of first pre-diabetes detection was associated with a strongly increased risk of future type 2 diabetes. This observation corroborates our current understanding of the pathophysiology of type 2 diabetes, with gradual exhaustion of beta cell capacity over time to compensate for insulin resistance, followed by an increase in blood glucose in the years immediately prior to a diabetes diagnosis.[37]

In an American study[33] assessing the performance of HbA1c in predicting long-term diabetes (glucose-lowering treatment, FPG ≥7 mmol/l, HbA1c ≥48 mmol/mol (6.5%), or self-reported diabetes), prediction models with and without HbA1c as a predictor were compared for individuals without diabetes. They reported AUCs of 66 (95% CI 63 to 68) for a model including only HbA1c, age, and sex, to 86 (95% CI 84 to 89) for a model in which fasting laboratory tests and clinical visits were added. These estimates are similar to ours (AUCt 73 (95% CI 71 to 74) for HbA1c ≥48 mmol/mol (6.5%) and AUCt 79 (95% CI 78 to 81) for glucose-lowering treatment initiation). However, our main models containing input from multiple predictors showed only slightly better discrimination than minimum models including just age and sex.

Study limitations

Ideally, individuals with pre-diabetes should be identified soon after their HbA1c levels increase to the pre-diabetes range. While we aimed to identify individuals with incident pre-diabetes, the sensitivity analysis showed that one in six might have had prior indications of pre-diabetes (more than 1 year prior to the pre-diabetes index date). Other individuals may have had undiagnosed pre-diabetes prior to study inclusion. As the median HbA1c at study inclusion was in the lower end of the pre-diabetes interval (median 43 mmol/mol (IQR 42–44 mmol/mol) or 6.1% (6.0%–6.2%)), we believe they were generally included early in the course of pre-diabetes. Still, individuals with neither HbA1c measurements nor glucose-lowering treatment or hospital-diagnosed diabetes, and individuals with type 2 diabetes based on glucose definitions who were treated only with lifestyle interventions, were not captured in our data. This could have resulted in an underestimation of type 2 diabetes risk in our study.

Another limitation is that our study cohort was based on individuals who responded to the Danish National Health Survey. The response rate for the survey was 55%–60%, and it varied along sociodemographic groups.[17] As individuals from higher sociodemographic groups were more likely to respond than those from lower sociodemographic groups, this may have led to an underestimate of the risks, and may limit the generalizability of our results. Although we aimed to include individuals as soon as they crossed the line from normal HbA1c values to pre-diabetes, increasing HbA1c levels are positively associated with increasing age on the population level,[37] and our population-based pre-diabetes cohort was rather old (median age 69.6 years) compared with other pre-diabetes cohorts.[11 34] Importantly, we corrected our estimates for the competing risk of death, and our prediction models also included age as a predictor per se; however, the high average age may have limited the comparability of our results with other cohorts.

We included a wide range of potential diabetes predictors (demographic variables, HbA1c measures, prescription drug use, comorbidities, socioeconomic variables, and self-reported lifestyle and health indicators), but data on other potential predictors[10] and other variable selection strategies may have improved the model validity. We included ethnic origin as a potential predictor for developing diabetes,[14 32 38] yet, the vast majority (95%) of our individuals were Caucasian, and model performance might not be generalizable to other ethnic groups. Unfortunately, we did not have access to other biomarkers than HbA1c in our data set, and could thus not include, for example, glucose levels, lipids, or estimates of insulin resistance and beta cell function in our models. We also missed clinical details on, for example, blood pressure, waist circumference, and family history of diabetes. These covariates are rather easily available in everyday clinical practice, and could further improve the prediction model for use in routine care.

Both HbA1c testing and the initiation of glucose-lowering treatment rely on clinical decisions influenced by potential predictors. This may have affected the variable selection and overestimated the importance of well-known risk factors. Another concern is that external model performance was estimated by split-sample validation, and this possibly overestimated the external validity. Before our models become useful for clinical work, they require additional validation along with model impact studies.[39 40] Overall, our models provide a snapshot of the current risk of progression from pre-diabetes to diabetes for a specific individual, and can thus identify individuals at high risk of progressing, thereby helping to target high-risk groups for preventive interventions in routine care. Before our models can also inform about the risk of diabetes progression under certain preventive interventions or treatment strategies, these interventions should be included in the models, and thus be part of any baseline risk assessment. We have included all relevant information in the online supplemental material and

encourage others to validate and calibrate our models in other settings.

Although we have identified individuals with pre-diabetes who are at high risk of later progression to type 2 diabetes in a real-world setting, the models' discrimination should be further improved. Additional biomarkers[41] and substratification using new pre-diabetes phenotypes and genetic risk scores[42] may lead to improved prediction models in the future. Knowing individual-level risks for progression from pre-diabetes to type 2 diabetes is crucial to effectively target preventive interventions.

Author affiliations
[1]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark
[2]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Capital Region of Denmark, Denmark

Acknowledgements We thank all the Danish National Health Survey respondents.

Contributors SKN performed all the analyses. SKN, LP, and RWT designed the study. SKN prepared the draft of the manuscript. LP, RWT, CJL, and HTS critically revised the manuscript. SKN and LP are the guarantors of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Disclaimer The study's funding source had no role in the collection, analysis, or interpretation of the data. It had no role in study design or writing of the manuscript.

Competing interests CJL holds shares in Novo Nordisk. The Department of Clinical Epidemiology is involved in studies with funding from various companies as research grants to and administered by Aarhus University. None of these studies are related to the current study.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Danish Data Protection Agency (Aarhus University record number: 2016-051-000001/1702). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The data used in this study are owned and managed by Statistics Denmark. In accordance with the Danish law and data protection policies, the data used in this study were anonymized, and stored and analyzed on a secure server. The data are available to researchers from research environments preapproved by Statistics Denmark upon project approval by the Danish Data Protection Agency and Statistics Denmark. Researchers can apply for access to data when the request is approved by the Danish Data Protection Agency (https://www.datatilsynet.dk).

ORCID iD
Sia K Nicolaisen http://orcid.org/0000-0002-6106-1945

REFERENCES

1 International Expert Committee. International expert Committee report on the role of the A1c assay in the diagnosis of diabetes. *Diabetes Care* 2009;32:1327–34.
2 World Health Organization. WHO Guidelines Approved by the Guidelines Review Committee. In: *Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a who consultation*. Geneva: World Health Organization, 2011.
3 Farmakologisk behandling af type 2 diabetes – mål og algoritmer - 2018 [article online], 2018. Available: https://vejledninger.dsam.dk/fbv-t2dm/ [Accessed 13 Apr 2022].
4 American Diabetes Association Professional Practice Committee. 2. classification and diagnosis of diabetes: standards of medical care in Diabetes-2022. *Diabetes Care* 2022;45:S17–38.
5 Internal Clinical Guidelines Team. National Institute for Health and Care Excellence: Clinical Guidelines. In: *Type 2 diabetes in adults: management*. London: National Institute for Health and Care Excellence (UK), 2015.
6 Garber AJ, Abrahamson MJ, Barzilay JI, *et al*. Consensus statement by the american association of clinical endocrinologists and american college of endocrinology on the comprehensive type 2 diabetes management algorithm - 2018 executive summary. *Endocr Pract* 2018;24:91–121.
7 Arendt JFH, Hansen AT, Ladefoged SA, *et al*. Existing data sources in clinical epidemiology: laboratory information system databases in Denmark. *Clin Epidemiol* 2020;12:469–75.
8 Makaroff LE. The need for international consensus on prediabetes. *Lancet Diabetes Endocrinol* 2017;5:5–7.
9 Morris DH, Khunti K, Achana F, *et al*. Progression rates from HbA1c 6.0-6.4% and other prediabetes definitions to type 2 diabetes: a meta-analysis. *Diabetologia* 2013;56:1489–93.
10 Tabák AG, Herder C, Rathmann W, *et al*. Prediabetes: a high-risk state for diabetes development. *Lancet* 2012;379:2279–90.
11 Diabetes Prevention Program Research Group. Long-term effects of lifestyle intervention or metformin on diabetes development and microvascular complications over 15-year follow-up: the diabetes prevention program outcomes study. *Lancet Diabetes Endocrinol* 2015;3:866–75.
12 Ligthart S, van Herpt TTW, Leening MJG, *et al*. Lifetime risk of developing impaired glucose metabolism and eventual progression from prediabetes to type 2 diabetes: a prospective cohort study. *Lancet Diabetes Endocrinol* 2016;4:44–51.
13 Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26:725–31.
14 Gray LJ, Taub NA, Khunti K, *et al*. The Leicester risk assessment score for detecting undiagnosed type 2 diabetes and impaired glucose regulation for use in a multiethnic UK setting. *Diabet Med* 2010;27:887–95.
15 Jølle A, Midthjell K, Holmen J, *et al*. Validity of the FINDRISC as a prediction tool for diabetes in a contemporary Norwegian population: a 10-year follow-up of the HUNT study. *BMJ Open Diabetes Res Care* 2019;7:e000769.
16 Moons KGM, Altman DG, Reitsma JB, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
17 Christensen AI, Lau CJ, Kristensen PL, *et al*. The Danish National health survey: study design, response rate and respondent characteristics in 2010, 2013 and 2017. *Scand J Public Health* 2022;50:180–8.
18 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.
19 Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–9.
20 Schmidt M, Schmidt SAJ, Sandegaard JL, *et al*. The Danish national patient registry: a review of content, data quality, and research potential. *Clin Epidemiol* 2015;7:449–90.
21 Kildemoes HW, Sørensen HT, Hallas J. The Danish national prescription registry. *Scand J Public Health* 2011;39:38–41.
22 Bo A, Thomsen RW, Nielsen JS, *et al*. Early-onset type 2 diabetes: age gradient in clinical and behavioural risk factors in 5115 persons with newly diagnosed type 2 diabetes-Results from the DD2 study. *Diabetes Metab Res Rev* 2018;34. doi:10.1002/dmrr.2968. [Epub ahead of print: 21 12 2017].
23 Laakso M, Pyörälä K. Age of onset and type of diabetes. *Diabetes Care* 1985;8:114–7.

24 Carstensen B, Rønn PF, Jørgensen ME. Prevalence, incidence and mortality of type 1 and type 2 diabetes in Denmark 1996-2016. *BMJ Open Diabetes Res Care* 2020;8:e001071.
25 Mor A, Berencsi K, Svensson E, *et al*. Prescribing practices and clinical predictors of glucose-lowering therapy within the first year in people with newly diagnosed type 2 diabetes. *Diabet Med* 2015;32:1546–54.
26 Rasmussen SS, Glümer C, Sandbaek A, *et al*. Determinants of progression from impaired fasting glucose and impaired glucose tolerance to diabetes in a high-risk screened population: 3 year follow-up in the addition study, Denmark. *Diabetologia* 2008;51:249–57.
27 DeJesus RS, Breitkopf CR, Rutten LJ, *et al*. Incidence rate of prediabetes progression to diabetes: modeling an optimum target group for intervention. *Popul Health Manag* 2017;20:216–23.
28 Fine JP, Gray RJ. A proportional hazards model for the Subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496–509.
29 Fu Z, Parikh CR, Zhou B. Penalized variable selection in competing risks regression. *Lifetime Data Anal* 2017;23:353–76.
30 Blanche P, Dartigues J-F, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat Med* 2013;32:5381–97.
31 Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. *Diagn Progn Res* 2018;2:7.
32 Diabetes Prevention Program Research Group. HbA1c as a predictor of diabetes and as an outcome in the diabetes prevention program: a randomized clinical trial. *Diabetes Care* 2015;38:51–8.
33 Leong A, Daya N, Porneala B, *et al*. Prediction of Type 2 Diabetes by Hemoglobin A_{1c} in Two Community-Based Cohorts. *Diabetes Care* 2018;41:60–8.
34 Vistisen D, Kivimäki M, Perreault L, *et al*. Reversion from prediabetes to normoglycaemia and risk of cardiovascular disease and mortality: the Whitehall II cohort study. *Diabetologia* 2019;62:1385–90.
35 Jonas DE, Crotty K, JDY Y. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. In: *Screening for prediabetes and type 2 diabetes mellitus: an evidence review for the US preventive services Task force Rockville (MD), agency for healthcare research and quality (US)*, 2021.
36 Hackett RA, Hudson JL, Chilcot J. Loneliness and type 2 diabetes incidence: findings from the English longitudinal study of ageing. *Diabetologia* 2020;63:2329–38.
37 Tabák AG, Jokela M, Akbaraly TN, *et al*. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *Lancet* 2009;373:2215–21.
38 Beulens J, Rutters F, Rydén L, *et al*. Risk and management of pre-diabetes. *Eur J Prev Cardiol* 2019;26:47–54.
39 Wareham NJ, Griffin SJ. Risk scores for predicting type 2 diabetes: comparing axes and spades. *Diabetologia* 2011;54:994–5.
40 Kent P, Cancelliere C, Boyle E, *et al*. A conceptual framework for prognostic research. *BMC Med Res Methodol* 2020;20:172.
41 Thorand B, Zierer A, Büyüközkan M, *et al*. A panel of 6 biomarkers significantly improves the prediction of type 2 diabetes in the MONICA/KORA study population. *J Clin Endocrinol Metab* 2021;106:1647–59.
42 Ahlqvist E, Storm P, Käräjämäki A, *et al*. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018;6:361–9.

Online-only Supplemental Material

Development of a 5-year risk prediction model for type 2 diabetes in individuals with incident HbA1c-defined prediabetes in Denmark

Short title: Progression from HbA1c-defined prediabetes

Sia K. Nicolaisen[1], Reimar W. Thomsen[1], Cathrine Lau[2], Henrik T. Sørensen[1], Lars Pedersen[1]

[1]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

[2]Center for Clinical Research and Prevention, Bispebjerg and Frederiksberg Hospital, Capital Region of Denmark, Denmark.

Corresponding author: Sia Kromann Nicolaisen, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Olof Palmes Allé 43-45, 8200 Aarhus N, Denmark. E-mail: sknc@clin.au.dk, Telephone: +45 8716 7212

List of Tables and Figures in Online-only Supplemental Material:

Supplemental Material Table S1

Checklist for the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) Statement.

Section/Topic	Item	Checklist Item	Comment
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	The title identifies the study as a development study presenting a (5-year) prognostic model for the outcome of type 2 diabetes and the target population is defined as individuals with HbA1c-defined prediabetes.
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	Background, objectives, methods (including study design, setting, participants, outcome, and statistical analyses), results (including participants and predictors), and conclusions are included in the Abstract.
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	Use of HbA1c measurements motivates the need for new prognostic prediction models for individuals with HbA1c-defined prediabetes progressing towards diabetes. Current models use other definitions of prediabetes and/or diabetes, and are primarily based on older trials, whereas the current study uses nationwide registry data after 2012. This information is included in the Introduction.
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	The study develops (and internally validates) models in a setting where HbA1c is the most often used measure. This is included in the Introduction.
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	The first subsection of the Methods section includes a description of the registries and data sources used in the study. Development and validation samples originate from the same data sources.
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	All essential dates (index date, follow-up, look-back periods, etc.) are described in the main text, and are also illustrated in the study design figure included in the Online-only Supplemental Material.
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	The study is based on nationwide registry data, which is emphasized in the main text. This is also illustrated in the study design figure and in the flowchart.
	5b	Describe eligibility criteria for participants.	The study cohort is described in detail in the main text, and is

			illustrated in the study design figure and flowchart.
	5c	Give details of treatments received, if relevant.	All treatments prior to the index date were ascertained via registry data. All codes are included in the variable description table in the Online-only Supplemental Material. No outcome-modifying interventions were included.
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	Both outcomes (diabetes defined in two different ways) are described in detail in the Methods section. Further details about codes are included in the variable description table in the Online-only Supplemental Material.
	6b	Report any actions to blind assessment of the outcome to be predicted.	Outcome assessment is included in the Discussion.
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	Predictors are listed in a subsection of the Methods section. All potential predictors are described in detail in the variable description table in the Online-only Supplemental Material. The definitions are illustrated in the study design figure, and the baseline table shows the distribution of all potential predictors in both the development sample and the validation sample. Missingness also is reported.
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	All variables were obtained from registries and population surveys. This is described in the Methods section.
Sample size	8	Explain how the study size was arrived at.	The study size is illustrated in the flowchart. The main text explains the exclusion criteria. Using Danish registry data allowed a large sample size. In combination with the statistical methods, we do not think overfitting in an issue in this study.
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	Missingness has been reported for all variables, and complete-case analyses were performed. This is emphasized in the Methods section and in the variable description table in the Online-only Supplemental Material. Missingness is described for the full cohort along with the development sample and the validation sample.
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	All potential predictors were included in the variable selection. This is described in detail in the statistical analysis section. The inclusion of each potential predictor is also described in the variable description table.
	10b	Specify type of model, all model-building procedures (including any	Fine-Gray survival models were fitted with LASSO regression for variable

		predictor selection), and method for internal validation.	selection. The validation method is split-sample validation. This is described in the Methods section.
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	Time-dependent measures handling competing risks were used. These are described in the statistical analysis section and are all referenced with relevant literature.
Risk groups	11	Provide details on how risk groups were created, if done.	Risk groups were not created *per se*, but sensitivity and specificity were calculated for various risks as part of the model performance comparison. This is described in the main text, with more detail provided in the Online-only Supplemental Material.
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	A flowchart is included. Information about follow-up is included in the main text. Details about outcome status are reported for each model.
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Baseline characteristics of participants and details about missingness are included.
Model development	14a	Specify the number of participants and outcome events in each analysis.	Details about outcome status are reported for each model.
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	Hazard ratios (adjusted for age, sex, index year, and region of residence) are included in the baseline table.
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	All coefficients, including baseline hazards, for the models are reported in the Online-only Supplemental Material.
	15b	Explain how to the use the prediction model.	A detailed description of the use of the models is included in the Online-only Supplemental Material along with a calculated example.
Model performance	16	Report performance measures (with CIs) for the prediction model.	Performance measures and figures are provided both in the main text and in the Online-only Supplemental Material.
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	The Discussion covers limitations such as possible selection bias, misclassification of outcomes, limited follow-up time, split-sample validation, and potential predictors not available in this study.
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant	The study estimates have been compared to estimates from large well-known studies/study populations and other prediction models. This is

		evidence.	described in the Discussion.
Implications	20	Discuss the potential clinical use of the model and implications for future research.	It is emphasized that our models should be used with caution. We have provided all relevant information to allow others to use and/or validate our models.
Other information			
Supplemental information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	The Online-only Supplemental Material includes additional information/tables/figures for the study. There are no resources other than what is included in the main text and the Online-only Supplemental Material.
Funding	22	Give the source of funding and the role of the funders for the present study.	Ethical considerations, Acknowledgements, Author contributions, Conflicts of interest, and Role of the funding source are all included in the main text.

1 https://www.tripod-statement.org/wp-content/uploads/2020/01/Tripod-Checlist-
2 Prediction-Model-Development.pdf

3 We have adjusted the original table and added comments instead of page numbers.

Supplemental Material Figure S1

The laboratory data on HbA1c measurements were derived from a combination of the nationwide Register of Laboratory Results for Research and the regional Clinical Laboratory Information System Research Database at Aarhus University. The data available for this study included HbA1c measurements based on NPU codes NPU03835 (HbA1c reported in % [DCCT]) and NPU27300 (HbA1c reported in mmol/mol [IFCC]).

The five Danish administrative regions had varying temporal coverage of laboratory data. For this study, the Central Denmark Region, North Denmark Region, and Capital Region of Denmark had complete and valid HbA1c data covering the entire study period from 2011 (including 1 year look-back) to 2018, whereas HbA1c data from Region Zealand and the Region of Southern Denmark were considered complete and valid only later in the study period.

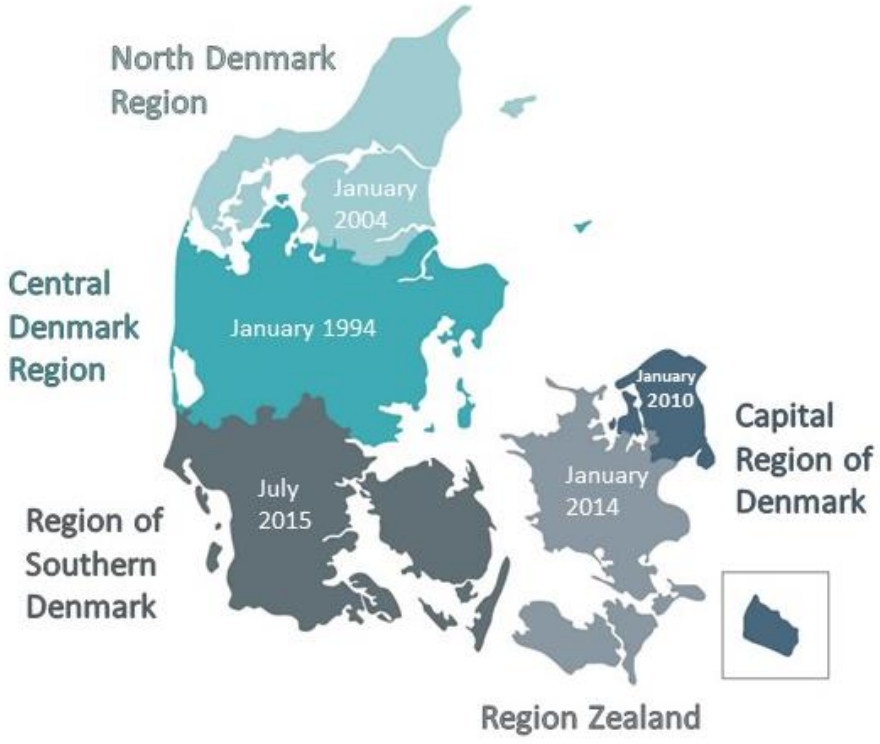
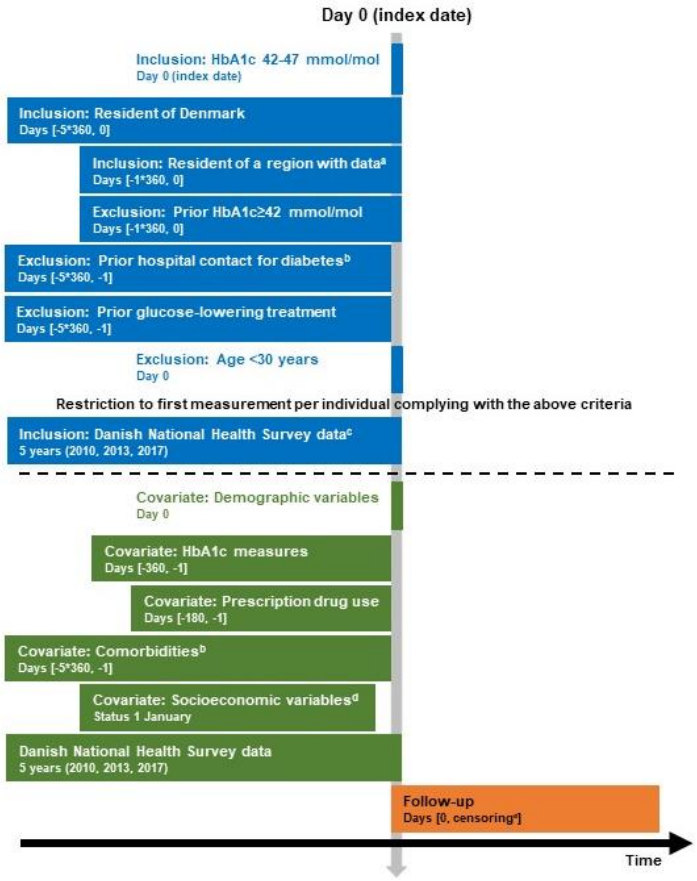


Figure adapted from https://rn.dk/genveje/fakta-om-nordjylland/regioner-i-danmark. Accessed 16 June 2021.

Supplemental Material Figure S2

Study design showing prediabetes index date, inclusion and exclusion criteria, covariate assessment periods, and follow-up periods for all data sources. See Supplemental Material Table S2 for information on all variables included in the study.



[a]Temporal coverage in the laboratory database differs by region.
[b]A hospitalization was defined using primary and secondary diagnoses for both inpatient admissions and outpatient visits. The admission date was used as the hospital contact day.
[c]Danish National Health Survey data. See Supplemental Material Figure S3 for details about inclusion.
[d]Data on socioeconomic factors are from the end of the previous year. Data on employment are from the end of the previous November.
[e]All individuals were followed to an outcome, emigration, study end (31 December 2018), end of follow-up (5 years after the index date), or death, whichever came first.

Supplemental Material Table S2

Full list of all variables, references, definitions, codes, and data sources in the study.

A hospitalization was defined using primary and secondary diagnoses both from inpatient admissions and outpatient visits. The admission date was used as the hospital contact day. Absence of a hospital contact was defined as 'no admission'. Treatment initiation was defined as a first-time redemption of a prescription. Absence of prescriptions is defined as 'no drug use'. Variables defined based on the presence of records in the healthcare registries had no missing values, as absence of records (*e.g.*, for any antihypertensive drugs, cancer, etc.) was defined as absence of the predictor. The health survey data included only a few records with missing data; a maximum of 8% missing values was noted for alcohol consumption. Therefore, we deemed it possible to perform complete-case analyses (*i.e.,* including only the individuals with no missing data in the variables included in the analyses). Among the 26,007 individuals included in the study, 20,089 (77.2%) had no missing variables. Functional forms for the continuous variables (age, BMI, and baseline HbA1c) were assessed visually and via estimated hazard ratios. All other factors were included as categorical variables.

*Percentage with missing data in total cohort (N=26,007)

Variable	Type of variable	Definition and reference	Variable assessment period	Data source	Registry definition	*
Cohort definition						
Value of prediabetes-defining HbA1c measurement (mmol/mol)	Continuous, included linearly in the prediction model (categorical in sensitivity analysis)	42-47 mmol/mol (6.0-6.4%; 42 mmol/mol [6.0%] is the reference in the sensitivity analysis)	This defines the index date of prediabetes	LAB	NPU: NPU27300 (mmol/mol [IFCC]), NPU03835 (% [DCCT]), (all available measurements were converted into mmol/mol and rounded to nearest integer using the formula: IFCC=(DCCT*10.93)-23.5. Data is restricted to a maximum of one measurement per day by taking the mean of possible multiple measurements. See reference for formula: Lægehåndbogen (2020) Hæmoglobin A1c (HbA1c). Available from https://www.sundhed.dk/sundhedsf	N/A

Appendix II

					aglig/laegehaandbogen/undersoegelser-og-proever/klinisk-biokemi/blodproever/haemoglobin-a1c-hba1c/. Accessed 06 February 2021)	
Prior hospital contact for diabetes	N/A	N/A	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DO24, DH360, DG632, DG590, DH280, DH334B, DM142, DN083, DT383, DE10-DE14	N/A
Prior glucose-lowering treatment	N/A	N/A	Prescription redemption 5*360 days prior to the index date.	DRMPS	ATC: A10	N/A
Prior elevated HbA1c measurement (includes prediabetes and diabetes)	N/A	HbA1c>=42 mmol/mol (6.0%)	Elevated HbA1c 360 days prior to the index date (5*360 in sensitivity analysis)	LAB	NPU: NPU27300, NPU03835	N/A
Outcome definition						
HbA1c>=48 mmol/mol (6.5%)	Time to event	N/A	Measurement during follow-up	LAB	NPU: NPU27300, NPU03835	N/A
Glucose-lowering treatment initiation	Time to event	N/A	Prescription redemption during follow-up	DRMPS	ATC: A10	N/A
Censoring (emigration, study end [31 December 2018], end of follow-up [5 years after index date]), or competing risk of death	Time to event	N/A	During follow-up	DCRS	N/A	N/A
Potential predictors						
Demographic variables						
Sex	Binary	Female, Male (ref.)	N/A	DCRS	N/A	0

Age on index date	Continuous, included linearly	30-104 years	Index date	DCRS	N/A	0
Ethnic origin	Categorical	Danish (ref.), Immigrant/descendant/unknown	N/A	DST	N/A	0
Region of residence	Categorical, only included in the hazard ratios (not the prediction models)	Capital Region of Denmark (ref.), Central Denmark Region, North Denmark Region, Region Zealand, Region of Southern Denmark	Index date	DCRS	N/A	0
HbA1c measures						
Year of index HbA1c measurement	Categorical, only included in the hazard ratios (not in the prediction models).	2012-2018 (ref. 2012)	Index date	LAB	N/A	0
Presence of HbA1c measurements before prediabetes	Binary	No (ref.), Yes	360 days prior to the index date	LAB	NPU: NPU27300, NPU03835	0
Prescription drug use						
Statins	Binary	No (ref.), Yes	180 days prior to the index date	DRMPS	ATC: C10AA, C10BA C10BX, A10BH51	0
Any antihypertensive drug	Binary	No (ref.), Yes	180 days prior to the index date	DRMPS	ATC: C02, C03A, C03B, C03D, C03E (not C03EB), C07, C08, C09A, C09B, C09C, C09D, C09X, G04CA03, C10BX04, C10BX06, C10BX07, C10BX11, C10BX12,	0

Appendix II

					C10BX13, C10BX14, C10BX15, C10BX10, C10BX07, C10BX09, C10BX11, C10BX14	
Oral steroids	Binary	No (ref.), Yes	180 days prior to the index date	DRMPS	ATC: H02AB	0
Opioid use	Binary	No (ref.), Yes	180 days prior to the index date	DRMPS	ATC: N02A, N07BC02	0
Comorbidities						
Pancreatic disease (includes pancreatic cancer, pancreas resection, and acute or chronic pancreatitis)	Binary	No (ref.), Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DC25, KJLC, DK859, DK860, DK861	0
Cardiovascular disease (includes stable angina pectoris [or CABG/PCI procedures], myocardial infarction, heart failure, stroke, atrial fibrillation/flutter, heart valve disease, and venous thromboembolism)	Binary	No (ref.), Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI20 (not DI200), DI251, DI259, KFNA, KFNB, KFNC, KFND, KFNE, KFNF, KFNG, KFNH20, DI21, DI22, DI23, DI50, DI60, DI61, DI63, DI64, DI48, DI05, DI06, DI07, DI08, DI098, DI39, DI511A, DQ22, DQ23, DI34-DI37, DI26, DI801, DI802, DI803	0
Lung disease	Binary	No (ref.), Yes	Prescription redemption 180 days prior to index date, or hospital contact day 5*360 days prior to the index date	DRMPS, DNPR	ATC: R03. ICD-10: DJ40, DJ684, DJ701, DJ703, DJ841, DJ920, DJ961, DJ982, DJ983, DJ41-DJ44, DJ45-DJ47, DJ60-DJ67	0
Cancer	Binary	No (ref.), Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DC00-DC99	0
Possible HbA1c-modifying conditions (includes prescription drug use of dapsone, ribavirin,	Binary	No (ref.), Yes	Prescription redemption 180 days prior to the	DRMPS, DNPR	ATC: J04BA02, D10AX05, J05AP01, J05, J01EE01, J04AM08, J01EA01, QJ51EA01, J01EC01,	0

antiretrovirals, trimethoprim-sulfamethoxazole, hydroxyurea, vitamin C, vitamin E, or opiates; and hospital admission or treatment with ribavirin, hemolysis, hemoglobinopathies, blood transfusion, acute blood loss/anemia, hypertriglyceridemia, chronic liver disease, pregnancy, iron deficiency, vitamin B12 deficiency, uremia, hyperbilirubinemia, end-stage renal disease [kidney transplant or dialysis], alcoholism-related diagnoses or medication, fetal hemoglobin, or methemoglobin)			index date, or hospital contact day 5*360 days prior to the index date		QJ01EQ11, L01XX05, A11GA, A11HA03, N02A. ICD-10: BPHM04, DD59, DD55-DD59, BOQA, DD62, BOHC, DD50, DD51, DD52, DE781, DB18, DB150, DB160, DB162, DB190, DI85, DK70, DK71, DK72, DK73, DK74, DK760, DK766, DZ321, DO00-DO99, DZ33-DZ39, DE611, DD50, DD51, BOHC2, DR39, DE804, DE806A, BJFD2, KKAS, DT861, DZ940, DF10 (not DF100), DE244, DG312, DG621, DG721, DI426, DK292, DK70, DK852, DK860, DQ860, DZ502, DZ714, DZ721, DR780, DT51, DT500A, DD564, DD74.	
Charlson Comorbidity Index score as a measure of overall comorbidity burden	Categorical	0-2 (ref.), >=3	Hospital contact day 5*360 days prior to the index date	DNPR	See reference for ICD-10 codes: Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8. The code DQ61 (cystic kidney disease) was not included in the calculation, as it was not available in the data. Diabetes is not included in the calculation of the Charlson Comorbidity Index score as it is included in the definition of the study cohort.	0
Socioeconomic variables						

Appendix II

Highest education achieved	Categorical	None, basic education, or primary school; Youth education, high school or similar educational level; Higher education (ref.)	Index date	DST	N/A	2
Employment status	Categorical	Employed (ref.), Unemployed or not part of the workforce	End of previous November	DST	N/A	0
Income	Categorical	Lowest income group, Low to medium income, Medium to high income, Highest income group (ref.)	End of previous year	DST	Income group from the last calendar year. Quartiles are based on the entire Danish population per year.	<1
Type of household	Categorical	Living alone (ref.), Not living alone	End of previous year	DST	N/A	0
Self-rated health/lifestyle/quality of life						
Years from questionnaire to first incident HbA1c-defined prediabetes	Continuous, not included in the prediction models	0-6 years (5 calendar years calculated from 1 January yields a maximum of 6 years)	The actual questionnaire date is not known, but is set to January 1 2010/2013/2017.	DNHS	N/A	0
Self-reported body mass index (BMI)	Continuous, included linearly (categorical in sensitivity analysis)	13-70 kg/m2 (<25 is reference in sensitivity analysis).	2010/2013/2017	DNHS	N/A	5
Days per week with alcohol consumption	Categorical	0-1 days (ref.), 2-3 days, 4-7 days	2010/2013/2017	DNHS	N/A	6
Alcohol consumption in relation to recommended amounts	Categorical	7/14 units or less per week (women/men) (ref.), More than 7/14	2010/2013/2017	DNHS	N/A	8

		units per week (women/men)				
Current smoking status	Categorical	Current smoker, Former smoker, Never smoker (ref.)	2010/2013/2017	DNHS	N/A	3
Overall diet	Categorical	Healthy (ref.), Average, Unhealthy	2010/2013/2017	DNHS	N/A	7
Overall self-rated health	Categorical	Fair/poor, Good, Excellent/very good (ref.)	2010/2013/2017	DNHS	N/A	2
GP advice to lose weight or change dietary habits during the past 12 months	Binary	No (ref.), Yes	2010/2013/2017	DNHS	N/A	0
Feeling stressed during the last 4 weeks	Categorical	Never/almost never (ref.), Once in a while, Often/very often	2010/2013/2017	DNHS	N/A	4
Feeling that problems were piling up during the last 4 weeks	Categorical	Never/almost never (ref.), Once in a while, Often/very often	2010/2013/2017	DNHS	N/A	3
Frequency of contact with people outside the household	Categorical	Never, Rarer than once monthly, Once or twice monthly, Once or twice weekly, Daily or almost daily (ref.),	2010/2013/2017	DNHS	People outside the household are defined as people you don't live with, *i.e.*, family, friends, colleagues, neighbors, and persons known via the internet. Contact is defined as talking by phone, writing etc.	4
Availability of someone to talk to if problems occur or support is needed	Categorical	No never or almost never, Yes mostly, Yes sometimes, Yes often (ref.)	2010/2013/2017	DNHS	N/A	3

1 Abbreviations: LAB, nationwide laboratory registry; DNPR, Danish National Patient Registry; DRMPS, Danish Register of Medicinal
2 Product Statistics; DCRS, Danish Civil Registration System; DST, registries maintained by Statistics Denmark; DNHS, Danish
3 National Health Survey; NPU, laboratory codes in the Nomenclature for Properties and Units; ICD-10, *International Statistical*

Appendix II

1 *Classification of Diseases and Related Health Problems, Tenth Revision*; ATC Anatomical Therapeutic Chemical Codes; N/A, not
2 applicable.

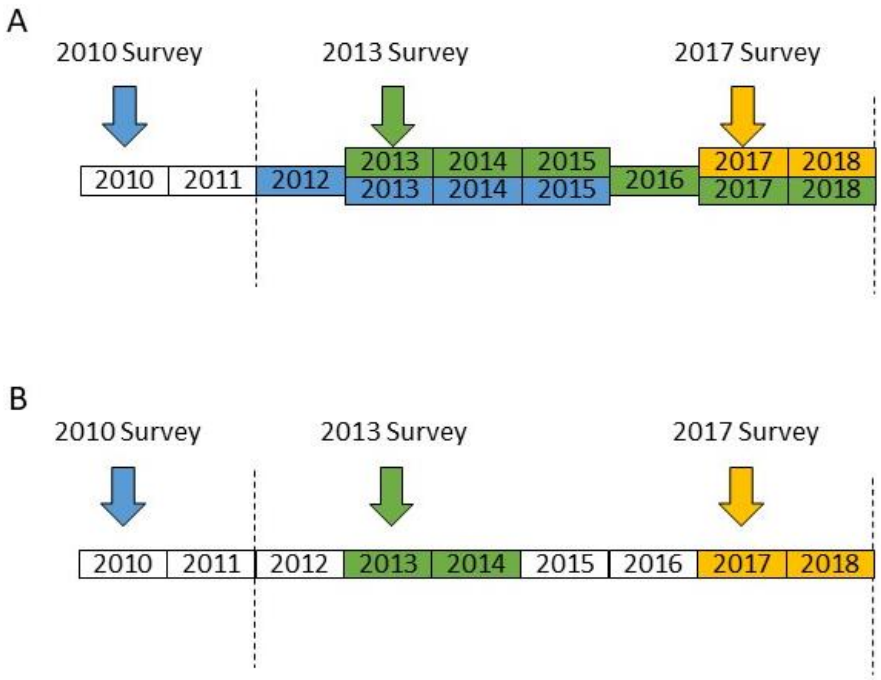
3

Supplemental Material Figure S3

The Danish National Health Survey was sent to representatively sampled Danes in 2010, 2013, and 2017 (colored arrows). The exact date of questionnaire completion is unknown.

A) An individual was included in the main analyses if a survey response was available within the last 5 calendar years prior to the year of the index date (study period 2012-2018, dashed lines). For each year, the colors represent the possible survey responses. If multiple survey responses were available within the 5 years, the most recent survey response was included.

B) In a sensitivity analysis, individuals were only included if a survey response was available for the year of the index date or the year prior to the index date. This means individuals were only included if their first incident prediabetes diagnosis was in 2013 or 2014 (included with 2013 survey response), or in 2017 or 2018 (included with 2017 survey response).



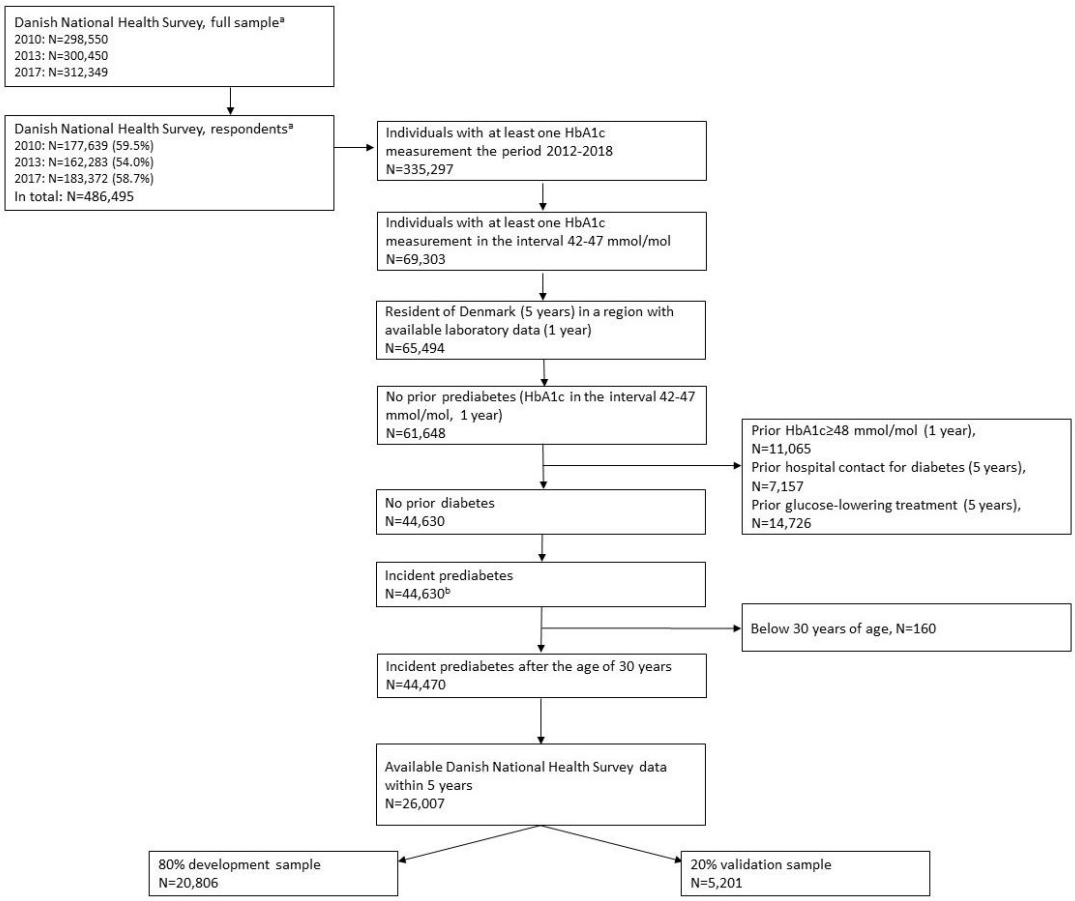
Supplemental Material Table S3

List of essential R packages (R version 4.0.2, R Core Team, 2020) and versions used in the study.

Package name	**Package version**	**Author(s)**	**CRAN**
riskRegression	2020.12.8	Thomas Alexander Gerds, Paul Blanche, Rikke Mortensen, Marvin Wright, Nikolaj Tollenaar, John Muschelli, Ulla Brasch Mogensen, Johan Sebastian Ohlendorff, Brice Ozenne	https://cran.r-project.org/package=riskRegression
prodlim	2019.11.13	Thomas A. Gerds	https://cran.r-project.org/package=prodlim
survival	3.2.3	Terry M Therneau, Thomas Lumley, Atkinson Elizabeth, Crowson Cynthia	https://cran.r-project.org/package=survival
crrp	1.0	Zhixuan Fu	https://cran.r-project.org/package=crrp
timeROC	0.4	Paul Blanche	https://cran.r-project.org/package=timeROC
cmprsk	2.2.10	Bob Gray	https://cran.r-project.org/package=cmprsk

Supplemental Material Figure S4

Flowchart of individuals included in the study. Among individuals who participated in the Danish National Health Survey, the study cohort was defined as all individuals residing in Denmark with an incident prediabetes-defining HbA1c measurement and no indication of prior diabetes. The 42-47 mmol/mol (6.0-6.4%) interval corresponds to the WHO definition of intermediate hyperglycemia, which is used as the definition of prediabetes in Denmark. Individuals were randomly split into a development sample (80%) and a validation sample (20%).

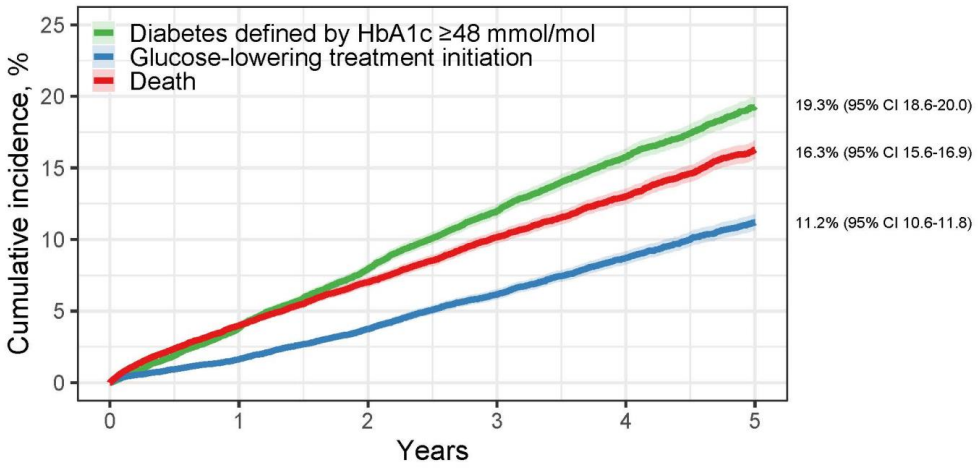


[a]See Christensen *et al*.

[b]The 44,630 individuals with prediabetes had a total of 65,499 measurements complying with the inclusion criteria. For true first incident prediabetes, only the first measurement complying with the inclusion criteria is included.

Supplemental Material Figure S5

Overall 5-year cumulative incidence curves calculated for the 26,007 individuals with incident HbA1c-defined prediabetes.



Supplemental Material Table S4

Baseline characteristics of the development sample and the validation sample. All 30 potential predictors are included in the table. The hazard ratio is adjusted for sex, age, index year, and region of residence. If no reference for the hazard ratio is provided for the categorical variables, the absence of the characteristic is the reference.

	Development sample (80%)				Validation sample (20%)	
			Hazard ratio			
	N (%) or Median (IQR)	Missing values	HbA1c ≥48 mmol/mol (6.5%)	Glucose-lowering treatment initiation	N (%) or Median (IQR)	Missing values
Total	20,806 (100.0%)				5,201 (100.0%)	
Demographic variables						
Sex		0 (0.0%)				0 (0.0%)
Female	10,792 (51.9%)		0.67 (0.62; 0.73)	0.68 (0.61; 0.76)	2,704 (52.0%)	
Male	10,014 (48.1%)		ref.	ref.	2,497 (48.0%)	
Age (years)	69.6 (61.0-77.1)	0 (0.0%)	0.99 (0.98; 0.99)	0.97 (0.96; 0.97)	69.6 (61.6-77.1)	0 (0.0%)
Age		0 (0.0%)				0 (0.0%)
30-39 years	219 (1.1%)		ref.	ref.	51 (1.0%)	
40-49 years	1,082 (5.2%)		0.94 (0.66; 1.36)	0.78 (0.53; 1.16)	280 (5.4%)	
50-59 years	3,390 (16.3%)		0.77 (0.55; 1.09)	0.53 (0.36; 0.76)	792 (15.2%)	
60-69 years	6,023 (28.9%)		0.62 (0.45; 0.88)	0.39 (0.27; 0.56)	1,580 (30.4%)	
70-79 years	6,452 (31.0%)		0.58 (0.41; 0.81)	0.31 (0.22; 0.45)	1,621 (31.2%)	
80-89 years	3,132 (15.1%)		0.64 (0.45; 0.91)	0.21 (0.14; 0.32)	748 (14.4%)	
≥90 years	508 (2.4%)		0.28 (0.16; 0.49)	0.14 (0.06; 0.29)	129 (2.5%)	
Ethnic origin		0 (0.0%)				0 (0.0%)
Danish	19,586 (94.1%)		ref.	ref.	4,899 (94.2%)	
Immigrant/descendant/ unknown	1,220 (5.9%)		1.19 (1.02; 1.39)	1.16 (0.95; 1.42)	302 (5.8%)	
HbA1c measures						
Value of prediabetes-defining HbA1c measurement (mmol/mol)	43.0 (42.0-44.0)	0 (0.0%)	1.69 (1.65; 1.73)	1.68 (1.63; 1.74)	43.0 (42.0-44.0)	0 (0.0%)
Value of prediabetes-defining HbA1c measurement (%)	6.1 (6.0-6.2)	0 (0.0%)	N/A	N/A	6.1 (6.0-6.2)	0 (0.0%)
Value of prediabetes-defining HbA1c measurement		0 (0.0%)				0 (0.0%)

	Development sample (80%)				Validation sample (20%)	
			Hazard ratio			
	N (%) or Median (IQR)	Missing values	HbA1c ≥48 mmol/mol (6.5%)	Glucose-lowering treatment initiation	N (%) or Median (IQR)	Missing values
42 mmol/mol (6.0%)	9,081 (43.6%)		ref.	ref.	2,261 (43.5%)	
43 mmol/mol (6.1%)	5,061 (24.3%)		1.67 (1.47; 1.89)	1.58 (1.32; 1.89)	1,300 (25.0%)	
44 mmol/mol (6.2%)	3,080 (14.8%)		2.74 (2.41; 3.12)	2.89 (2.43; 3.44)	727 (14.0%)	
45 mmol/mol (6.3%)	1,794 (8.6%)		5.15 (4.53; 5.86)	5.24 (4.39; 6.24)	440 (8.5%)	
46 mmol/mol (6.4%)	1,180 (5.7%)		7.93 (6.93; 9.07)	8.04 (6.71; 9.62)	312 (6.0%)	
47 mmol/mol (6.4%)	610 (2.9%)		13.69 (11.75; 15.94)	12.48 (10.15; 15.34)	161 (3.1%)	
Presence of HbA1c measurements before prediabetes		0 (0.0%)				0 (0.0%)
Yes	5,884 (28.3%)		0.73 (0.66; 0.80)	0.78 (0.68; 0.89)	1,491 (28.7%)	
No	14,922 (71.7%)		ref.	ref.	3,710 (71.3%)	
Prescription drug use						
Statins	7,706 (37.0%)	0 (0.0%)	1.03 (0.95; 1.12)	0.99 (0.88; 1.11)	1,942 (37.3%)	0 (0.0%)
Any antihypertensive drug	12,591 (60.5%)	0 (0.0%)	1.33 (1.22; 1.45)	1.31 (1.17; 1.47)	3,158 (60.7%)	0 (0.0%)
Oral steroids	1,505 (7.2%)	0 (0.0%)	1.26 (1.08; 1.46)	1.02 (0.81; 1.28)	393 (7.6%)	0 (0.0%)
Opioid use	2,601 (12.5%)	0 (0.0%)	1.21 (1.08; 1.36)	1.33 (1.14; 1.56)	670 (12.9%)	0 (0.0%)
Comorbidities						
Charlson Comorbidity Index score		0 (0.0%)				0 (0.0%)
0-2	19,544 (93.9%)		ref.	ref.	4,896 (94.1%)	
≥3	1,262 (6.1%)		1.25 (1.05; 1.48)	0.99 (0.75; 1.29)	305 (5.9%)	
Cardiovascular disease	4,530 (21.8%)	0 (0.0%)	1.03 (0.93; 1.14)	0.95 (0.83; 1.10)	1,156 (22.2%)	0 (0.0%)
Lung disease	3,096 (14.9%)	0 (0.0%)	1.16 (1.04; 1.29)	0.99 (0.84; 1.16)	810 (15.6%)	0 (0.0%)
Cancer	2,024 (9.7%)	0 (0.0%)	0.94 (0.81; 1.09)	1.01 (0.82; 1.24)	477 (9.2%)	0 (0.0%)
Pancreatitis/cancer (includes pancreatic cancer, pancreas resection, and acute or chronic pancreatitis)	67 (0.3%)	0 (0.0%)	3.08 (1.89; 5.04)	3.48 (1.92; 6.31)	16 (0.3%)	0 (0.0%)
Possible HbA1c-modifying conditions	1,499 (7.2%)	0 (0.0%)	1.11 (0.96; 1.29)	1.10 (0.90; 1.35)	390 (7.5%)	0 (0.0%)
Socioeconomic variables						
Highest education achieved		425 (2.0%)				89 (1.7%)

	Development sample (80%)				Validation sample (20%)	
			Hazard ratio			
	N (%) or Median (IQR)	Missing values	HbA1c ≥48 mmol/mol (6.5%)	Glucose-lowering treatment initiation	N (%) or Median (IQR)	Missing values
None, basic education, or primary school	7,209 (34.6%)		1.21 (1.08; 1.35)	1.19 (1.03; 1.39)	1,841 (35.4%)	
Youth education, high school, or similar educational level	8,687 (41.8%)		0.99 (0.89; 1.11)	0.97 (0.84; 1.12)	2,174 (41.8%)	
Higher education	4,485 (21.6%)		ref.	ref.	1,097 (21.1%)	
Employment status		0 (0.0%)				0 (0.0%)
Employed	6,398 (30.8%)		ref.	ref.	1,613 (31.0%)	
Unemployed or not part of the workforce	14,408 (69.2%)		1.10 (0.99; 1.21)	1.15 (1.01; 1.31)	3,588 (69.0%)	
Income		59 (0.3%)				13 (0.2%)
Lowest income group	3,698 (17.8%)		1.14 (0.99; 1.31)	1.19 (0.98; 1.44)	894 (17.2%)	
Low to medium income	8,019 (38.5%)		1.13 (1.01; 1.28)	1.19 (1.02; 1.39)	2,061 (39.6%)	
Medium to high income	5,330 (25.6%)		0.99 (0.87; 1.11)	0.98 (0.84; 1.15)	1,275 (24.5%)	
Highest income group	3,700 (17.8%)		ref.	ref.	958 (18.4%)	
Type of household		0 (0.0%)				0 (0.0%)
Living alone	6,327 (30.4%)		ref.	ref.	1,651 (31.7%)	
Not living alone	14,479 (69.6%)		0.88 (0.80; 0.96)	0.89 (0.79; 1.01)	3,550 (68.3%)	
Self-rated health/lifestyle/quality of life						
Years from questionnaire to first incident HbA1c-defined prediabetes	3.1 (1.5-4.4)	0 (0.0%)	1.00 (0.96; 1.03)	0.99 (0.97; 1.02)	3.0 (1.5-4.4)	0 (0.0%)
Body mass index (kg/m^2)	26.7 (24.1-29.8)	986 (4.7%)	1.05 (1.04; 1.06)	1.07 (1.06; 1.08)	26.8 (24.1-30.1)	239 (4.6%)
Body mass index		986 (4.7%)				239 (4.6%)
<25 kg/m^2	6,677 (32.1%)		ref.	ref.	1,654 (31.8%)	
25-30 kg/m^2	8,342 (40.1%)		1.50 (1.34; 1.67)	1.67 (1.43; 1.96)	2,045 (39.3%)	
30-35 kg/m^2	3,455 (16.6%)		2.23 (1.98; 2.51)	2.66 (2.25; 3.15)	926 (17.8%)	
≥35 kg/m^2	1,346 (6.5%)		2.39 (2.04; 2.81)	3.49 (2.85; 4.27)	337 (6.5%)	
Days per week with alcohol consumption		1,186 (5.7%)				277 (5.3%)
0-1 day	10,623 (51.1%)		ref.	ref.	2,698 (51.9%)	
2-3 days	4,306 (20.7%)		0.79 (0.71; 0.88)	0.87 (0.76; 1.01)	1,073 (20.6%)	
4-7 days	4,691 (22.5%)		0.81 (0.73; 0.90)	0.83 (0.71; 0.96)	1,153 (22.2%)	

Appendix II

	Development sample (80%)				Validation sample (20%)	
			Hazard ratio			
	N (%) or Median (IQR)	Missing values	HbA1c ≥48 mmol/mol (6.5%)	Glucose-lowering treatment initiation	N (%) or Median (IQR)	Missing values
Alcohol consumption in relation to recommended amounts		1,625 (7.8%)				385 (7.4%)
7/14 units or less per week (women/men)	15,605 (75.0%)		ref.	ref.	3,955 (76.0%)	
More than 7/14 units per week (women/men)	3,576 (17.2%)		0.98 (0.88; 1.08)	1.01 (0.88; 1.16)	861 (16.6%)	
Current smoking status		683 (3.3%)				179 (3.4%)
Current smoker	4,907 (23.6%)		1.00 (0.90; 1.12)	1.00 (0.87; 1.15)	1,222 (23.5%)	
Former smoker	8,201 (39.4%)		0.95 (0.86; 1.04)	0.93 (0.82; 1.06)	2,060 (39.6%)	
Never smoker	7,015 (33.7%)		ref.	ref.	1,740 (33.5%)	
Overall diet		1,521 (7.3%)				373 (7.2%)
Unhealthy	2,985 (14.3%)		1.19 (1.03; 1.36)	1.04 (0.86; 1.25)	720 (13.8%)	
Average	12,407 (59.6%)		1.08 (0.97; 1.20)	0.97 (0.84; 1.12)	3,153 (60.6%)	
Healthy	3,893 (18.7%)		ref.	ref.	955 (18.4%)	
Overall self-rated health		340 (1.6%)				82 (1.6%)
Excellent/very good	5,838 (28.1%)		ref.	ref.	1,448 (27.8%)	
Good	9,878 (47.5%)		1.32 (1.20; 1.46)	1.22 (1.06; 1.39)	2,434 (46.8%)	
Fair/poor	4,750 (22.8%)		1.51 (1.34; 1.69)	1.54 (1.32; 1.79)	1,237 (23.8%)	
GP advice to lose weight or change dietary habits during the past 12 months		0 (0.0%)				0 (0.0%)
Yes	3,445 (16.6%)		1.65 (1.51; 1.81)	1.80 (1.60; 2.03)	879 (16.9%)	
No	17,361 (83.4%)		ref.	ref.	4,322 (83.1%)	
Feeling stressed during the last 4 weeks		744 (3.6%)				170 (3.3%)
Never/almost never	12,259 (58.9%)		ref.	ref.	3,066 (59.0%)	
Once in a while	5,713 (27.5%)		1.09 (1.00; 1.20)	1.12 (0.99; 1.27)	1,426 (27.4%)	
Often/very often	2,090 (10.0%)		1.12 (0.97; 1.29)	1.18 (0.99; 1.41)	539 (10.4%)	
Feeling that problems were piling up during the last 4 weeks		670 (3.2%)				151 (2.9%)
Never/almost never	14,092 (67.7%)		ref.	ref.	3,525 (67.8%)	
Once in a while	4,274 (20.5%)		1.15 (1.04; 1.27)	1.17 (1.03; 1.34)	1,071 (20.6%)	

	Development sample (80%)				Validation sample (20%)	
			Hazard ratio			
	N (%) or Median (IQR)	Missing values	HbA1c ≥48 mmol/mol (6.5%)	Glucose-lowering treatment initiation	N (%) or Median (IQR)	Missing values
Often/very often	1,770 (8.5%)		1.30 (1.13; 1.49)	1.44 (1.21; 1.71)	454 (8.7%)	
Frequency of contact with people outside the household		756 (3.6%)				###
Daily or almost daily	11,197 (53.8%)		ref.	ref.	###	
Once or twice weekly	7,116 (34.2%)		0.93 (0.85; 1.02)	0.98 (0.87; 1.10)	###	
Once or twice monthly	1,351 (6.5%)		0.90 (0.76; 1.06)	0.86 (0.68; 1.07)	###	
Rarer than once monthly	362 (1.7%)		1.05 (0.79; 1.38)	1.21 (0.86; 1.70)	###	
Never	24 (0.1%)		1.92 (0.80; 4.61)	2.23 (0.72; 6.95)	###	
Availability of someone to talk to if problems occur or support is needed		673 (3.2%)				167 (3.2%)
Yes, often	10,805 (51.9%)		ref.	ref.	2,636 (50.7%)	
Yes, mostly	6,355 (30.5%)		1.10 (1.01; 1.20)	1.15 (1.02; 1.30)	1,621 (31.2%)	
Yes, sometimes	1,941 (9.3%)		1.07 (0.93; 1.23)	1.25 (1.04; 1.50)	523 (10.1%)	
No, never or almost never	1,032 (5.0%)		1.43 (1.22; 1.69)	1.36 (1.08;1.71)	254 (4.9%)	

1 ###: at least one cell contains a number < 5 and may not be reported due to Danish regulations.

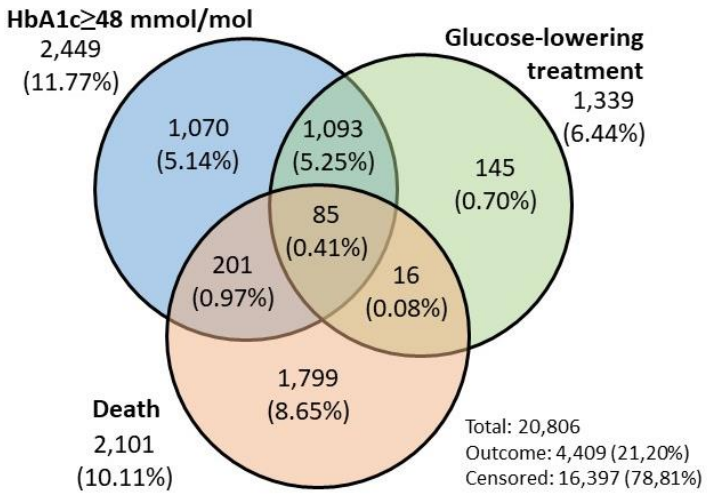
2

Appendix II

Supplemental Material Figure S6

For the primary analysis, type 2 diabetes was defined as the first HbA1c measurement ≥48 mmol/mol (6.5%) during follow-up. For the secondary analysis, type 2 diabetes was defined as glucose-lowering treatment initiation. In both analyses, death was considered a competing event. Within 5 years of follow-up in the development sample, 2,449 (11.8%) had an HbA1c measurement ≥48 mmol/mol (6.5%), and 1,339 (6.4%) initiated a glucose-lowering treatment. A total of 2,101 (10.1%) died during the follow-up period, among whom 302 (1.5%) died after having experienced at least one of the two outcomes.

In all analyses, a positive outcome at time *t* occurred when an individual had the outcome of interest before time *t*. A negative outcome occurred when there was no positive outcome, *i.e.*, when the individual was either still at risk of the outcomes of interest (uncensored and event-free) or had the competing event before time *t*.



Supplemental Material Table S5

A) Coefficients for the main model and the minimum model for type 2 diabetes defined as an HbA1c measurement ≥48 mmol/mol (6.5%).
B) Coefficients for the main model and the minimum model for type 2 diabetes defined as glucose-lowering treatment initiation.

The subdistribution hazard ratio (SHR) for an individual predictor can be obtained as $SHR = \exp(\beta)$. When SHR>1, the predictor increases the incidence of the outcome, *i.e.,* the SHR can be used to quantify the direction of the association, but it cannot be used quantify the magnitude of the association. We stress that our study is a prediction study, not a causal etiological study. Therefore, we emphasize that our models should not be used as causal models, as the estimates could be heavily confounded if interpreted causally.

A **HbA1c ≥48 mmol/mol (6.5%)**		
	Main model	**Minimum model**
Number of individuals (complete-case analyses)	18,216 (2,161 outcome, 1,434 death, 14,621 censored)	20,806 (2,449 outcome, 1,815 death, 16,542 censored)
Baseline, β_0	0.0922	0.2958
	Model coefficient, β	**Model coefficient, β**
Age (years)	-0.0124 (-0.0165; -0.0082)	-0.0163 (-0.0197; -0.0130)
Sex (female vs. male)	-0.3071 (-0.3952; -0.2190)	-0.3779 (-0.4576; -0.2982)
Prediabetes-defining HbA1c measurement (mmol/mol)	0.4973 (0.4705; 0.5242)	
Body mass index (kg/m^2)	0.0325 (0.0240; 0.0409)	
Any antihypertensive drug (Yes vs. No)	0.1529 (0.0574; 0.2484)	
Pancreatic disease (Yes vs. No)	0.9607 (0.4010; 1.5203)	
Cancer (Yes vs. No)	-0.2686 (-0.4333; -0.1040)	
Unhealthy diet (Overall diet: Unhealthy vs. Average or Healthy)	0.1221 (0.0057; 0.2385)	
Doctor's advice to lose weight or change dietary habits (GP advice to lose weight or change dietary habits during the past 12 months. Yes vs. No)	0.3380 (0.2333; 0.4426)	
Not having anyone to talk to when in need of support (Availability of someone to talk to if problems occur or support is needed? No, never or almost never vs. Yes, often, mostly, or sometimes)	0.2570 (0.0776; 0.4365)	
Good self-rated health (Overall self-rated health: Good vs. Fair/poor or Excellent/very good).	0.1215 (0.0353; 0.2077)	

For overall diet, only the questionnaire answer "Unhealthy" was included in the prediction model. Answering "Average" or "Healthy" did not increase the fit of the model and, therefore, were removed by the LASSO regression. Regarding availability of someone to talk to, "No, never or almost never" was included, whereas the answers "Yes, often", "Yes, mostly", and "Yes, sometimes" were excluded. Similarly, for self-rated health, "Good" was the only included answer, whereas both "Fair/poor" and "Excellent/very" were removed by the LASSO regression, as they did not add any precision to the model.

B **Glucose-lowering treatment initiation**		
	Main model	**Minimum model**
Number of individuals (complete-case analyses)	19,820 (1,279 outcome, 1,861 death, 16,680 censored)	20,806 (1,339 outcome, 2,000 death, 17,467 censored)
Baseline, β_0	0.0592	0.1823
	Model coefficient, β	**Model coefficient, β**
Age (years)	-0.0293 (-0.0341; -0.0245)	-0.0358 (-0.0402; -0.0314)
Sex (female vs. male)	-0.2805 (-0.3932; -0.1678)	-0.3630 (-0.4709; -0.2551)
Prediabetes-defining HbA1c measurement (mmol/mol)	0.4894 (0.4557; 0.5230)	
Body mass index (kg/m^2)	0.0473 (0.0372; 0.0575)	
Doctor's advice to lose weight or change dietary habits (GP advice to lose weight or change dietary habits during the past 12 months. Yes vs. No)	0.3679 (0.2374; 0.4984)	

The models can be used to obtain an individual's estimated probabilities based on specific characteristics as follows:

$$Prob(outcome) = 1 - \exp\left(-\beta_0 \times \exp\left(\beta_1 X_1 + \cdots + \beta_k X_k\right)\right)$$

where β_0 is the Breslow-type estimate of the underlying subdistribution hazard evaluated after 5 years, $\beta_1 \ldots \beta_k$ are the coefficients for the k predictors included in the model, and $X_1 \ldots X_k$ are the patient characteristics.

The following is a worked example of applying the prediction model to a hypothetical individual with a particular predictor profile: a 62-year-old man without prior pancreatic disease or cancer who is being treated with any antihypertensive drugs has been advised by his doctor to lose weight and change his dietary habits because he has an unhealthy diet and his BMI is 29 kg/m^2. His HbA1c was 45 mmol/mol (6.3%). He has people to talk to, and he says his overall health status is good. His risk of type 2 diabetes within 5 years when defined as HbA1c ≥48 mmol/mol (6.5%) is then calculated as:

$$Prob(outcome)$$
$$= 1 - \exp(-0.0922 \times \exp((62 - 60) \times (-0.0124) + 0 \times (-0.3071) + (45 - 42) \times (0.4973) + (29 - 25) \times (0.0325)$$
$$+ 1 \times (0.1529) + 0 \times (0.9607) + 0 \times (-0.2686) + 1 \times (0.1221) + 1 \times (0.3380) + 0(0.2 \times 570) + 1 \times (0.1215)))$$
$$= 0.6129$$

where age, BMI, and baseline HbA1c are standardized to 60 years, 25 kg/m^2, and 42 mmol/mol (6.0%).

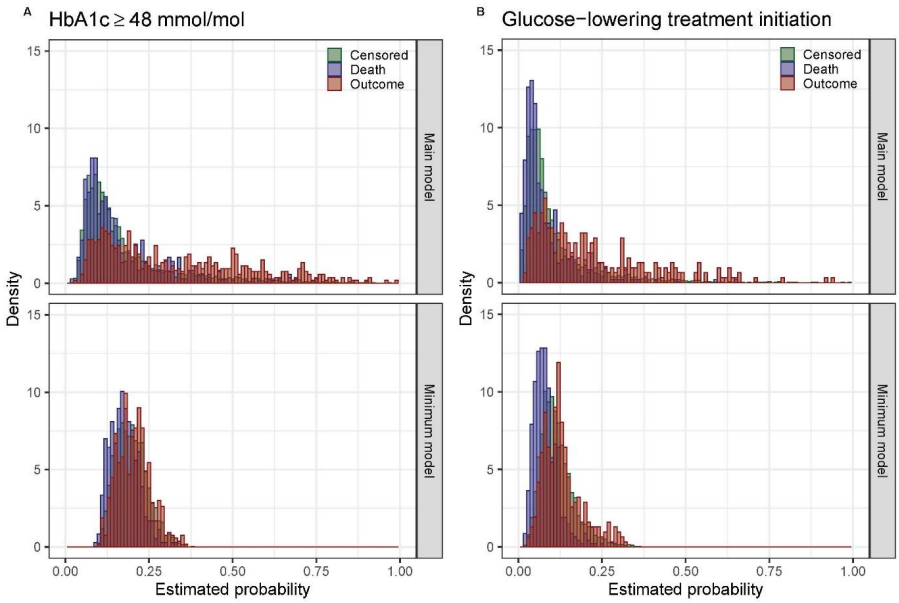
Thus, the patient's 5-year risk of HbA1c-defined type 2 diabetes is estimated to be 61%. Similarly, his 5-year risk of type 2 diabetes defined as glucose-lowering treatment initiation is 34%.

Nicolaisen SK, *et al. BMJ Open Diab Res Care* 2022; 10:e002946. doi: 10.1136/bmjdrc-2022-002946

Appendix II

Supplemental Material Figure S7

Histograms of the estimated probability of progression to type 2 diabetes defined as A) HbA1c ≥48 mmol/mol (6.5%) or B) glucose-lowering treatment initiation for all individuals in the validation sample. The graphs are colored by observed outcome: type 2 diabetes, death, or censored (*i.e.,* emigration, study end [31 December 2018], or end of follow-up [5 years after index date]).



To avoid reporting sensitive individual-level information, random noise was added to all estimates (normal distribution, mean=0, standard deviation=0.01).

Supplemental Material Table S6

For all individuals in the validation sample, the probability of type 2 diabetes was estimated based on the two models. Distributional measures of the estimated probabilities for type 2 diabetes defined as A) HbA1c ≥48 mmol/mol (6.5%) or B) glucose-lowering treatment initiation.

A **HbA1c ≥48 mmol/mol (6.5%)**										
	Min	**Max**	**Mean**	**P5**	**P10**	**Q1**	**Median**	**Q3**	**P90**	**P95**
Main model	1.94%	98.81%	19.66%	5.69%	6.59%	8.96%	13.50%	23.91%	43.17%	56.02%
Minimum model	8.63%	36.79%	19.59%	12.71%	13.84%	15.88%	19.20%	22.63%	25.99%	28.20%
B **Glucose-lowering treatment initiation**										
	Min	**Max**	**Mean**	**P5**	**P10**	**Q1**	**Median**	**Q3**	**P90**	**P95**
Main model	0.00%	97.49%	11.35%	2.09%	2.82%	4.31%	7.13%	13.32%	25.68%	35.60%
Minimum model	1.31%	38.20%	11.37%	4.90%	5.88%	7.83%	10.43%	13.85%	18.04%	21.49%

Abbreviations: P5, 5% percentile; P10, 10% percentile; P90, 90% percentile; P95, 95% percentile; Q1, first quartile; Q3, third quartile.

All values estimated from the models range from 0 to 1. To avoid reporting sensitive individual-level information, random noise was added to all estimates (normal distribution, mean=0, standard deviation=0.01) and then multiplied by 100 for reporting as a percentage.

Supplemental Material Table S7

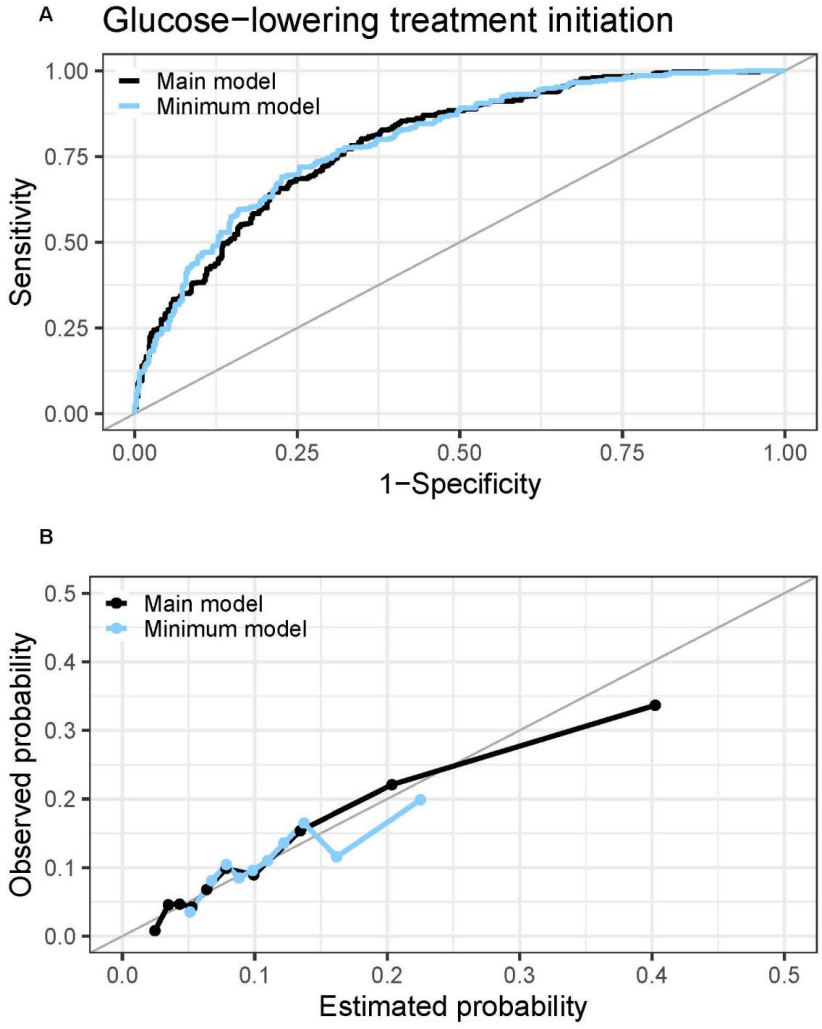
Time-dependent sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for the two models estimated based on pre-specified decision rules (10% and 20%) and on the decision rule defined by the maximized Youden index for the specific models. An individual was predicted to have the outcome according to the model if the estimated probability was larger than the decision rule and similarly predicted not to have the outcome if the estimated probability was smaller than the decision rule.

A **HbA1c ≥48 mmol/mol (6.5%)**					
Rule	**Model**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
10%	Main model	86.76 (83.50-90.01)	39.39 (38.48-40.30)	68.28 (64.47-72.09)	66.41 (59.17-73.66)
	Minimum model	100.0 (100.0-100.0)	0.00 (0.00-0.00)	60.07 (56.52-63.61)	-
20%	Main model	57.37 (52.71-62.02)	71.80 (70.99-72.62)	75.37 (70.93-79.81)	52.82 (47.75-57.90)
	Minimum model	52.05 (47.41-56.69)	75.32 (74.53-76.11)	76.03 (71.24-80.82)	51.08 (46.37-55.80)
16.04%	Main model	68.27 (63.85-72.69)	66.28 (65.43-67.14)	75.28 (71.22-79.35)	58.14 (52.68-63.60)
20.49%	Minimum model	50.06 (45.42-54.71)	78.94 (78.19-79.69)	78.15 (73.34-82.95)	51.24 (46.64-55.85)
B **Glucose-lowering treatment initiation**					
Rule	**Model**	**Sensitivity**	**Specificity**	**PPV**	**NPV**
10%	Main model	68.62 (62.88-74.35)	74.76 (74.02-75.49)	66.41 (61.00-71.82)	76.60 (72.11-81.10)
	Minimum model	69.04 (63.36-74.72)	76.62 (75.89-77.34)	68.23 (62.78-73.68)	77.28 (72.91-81.66)
20%	Main model	41.81 (35.77-47.84)	88.94 (88.38-89.51)	73.34 (66.19-80.48)	67.75 (63.54-71.96)
	Minimum model	12.29 (8.45-16.13)	98.99 (98.83-99.15)	89.83 (80.85-98.82)	60.80 (56.86-64.75)
7.39%	Main model	79.86 (74.92-84.79)	65.16 (64.34-65.98)	62.51 (57.46-67.56)	81.64 (77.12-86.16)
9.62%	Minimum model	71.58 (66.08-77.08)	74.69 (73.95-75.44)	67.30 (61.95-72.64)	78.32 (73.98-82.66)

Supplemental Material Figure S8

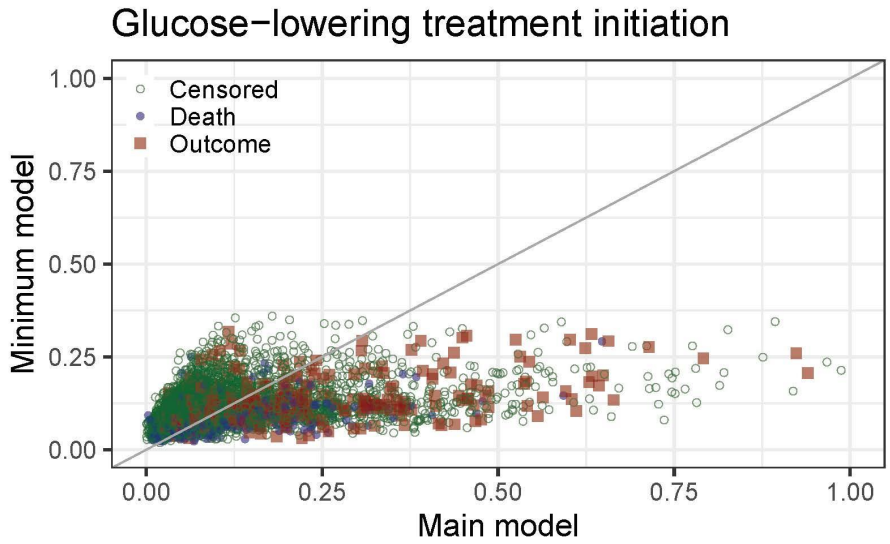
The models for predicting type 2 diabetes defined as glucose-lowering treatment initiation.

A) Time-dependent receiver operating characteristic curve comparing the discriminative ability of the main model (including baseline HbA1c level, age, sex, BMI, and doctor's advice to lose weight or change dietary habits) to that of the minimum model including only age and sex.
B) Calibration curve comparing the estimated and observed probabilities for the two models. The estimates for the observed probabilities were defined based on quantiles of the estimated probabilities.



Supplemental Material Figure S9

Comparison of the estimated probabilities from the main model (including baseline HbA1c level, age, sex, BMI, and doctor's advice to lose weight or change dietary habits) to those from the minimum model (including only age and sex for type 2 diabetes when defined as glucose-lowering treatment initiation).



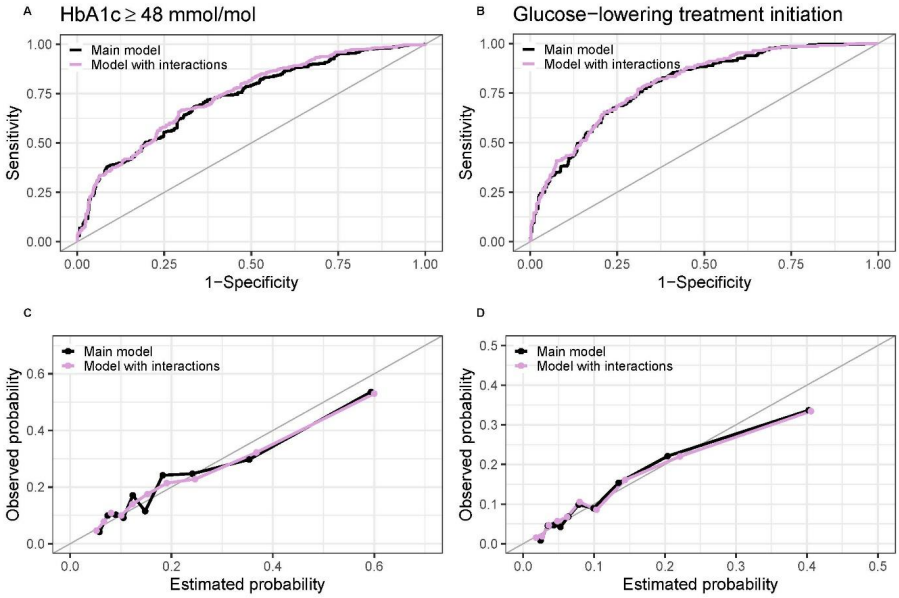
To avoid reporting sensitive individual-level information, random noise was added to all estimates (normal distribution, mean=0, standard deviation=0.01).

Supplemental Material Figure S10

Comparison of the performance of the main model and the interaction model. The interaction model included the same predictors as the main model but with BMI and baseline HbA1c levels included categorically along with their interactions. The main model for HbA1c ≥48 mmol/mol (6.5%) included baseline HbA1c, age, sex, BMI, treated hypertension, pre-existing pancreatic disease, absence of cancer, unhealthy diet, doctor's advice to lose weight or change dietary habits, self-reported lack of anyone to talk to, and good self-rated health. The main model for glucose-lowering treatment initiation included baseline HbA1c, age, sex, BMI, and doctor's advice to lose weight or change dietary habits.

A, B) Time-dependent receiver operating characteristic curves comparing the discriminative ability for type 2 diabetes defined as (A) HbA1c ≥48 mmol/mol (6.5%) and (B) glucose-lowering treatment initiation.

C, D) Calibration curves comparing the estimated and observed probabilities for the two models for type 2 diabetes defined as (C) HbA1c ≥48 mmol/mol (6.5%) and (D) glucose-lowering treatment initiation.



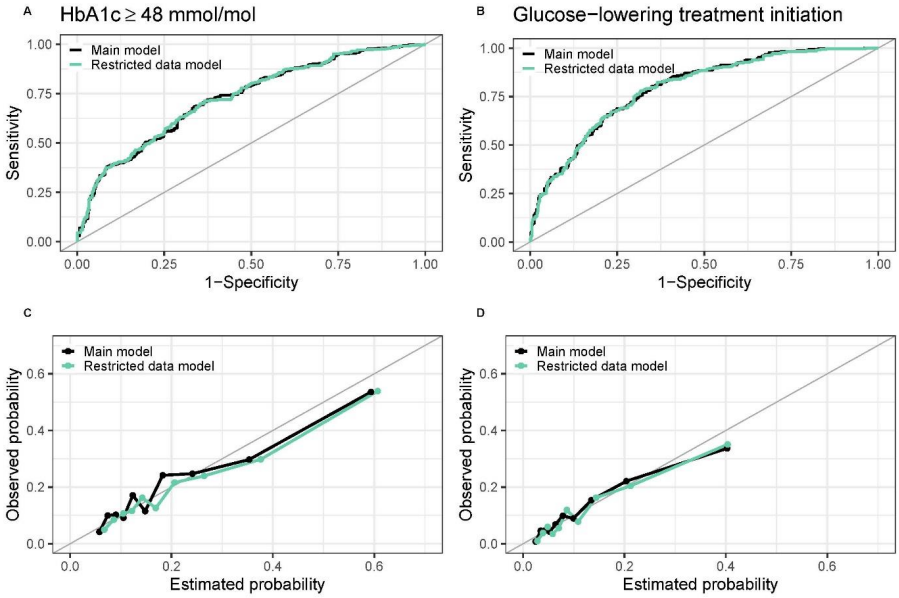
The model with the interactions had the following performance measures for HbA1c ≥48 mmol/mol (6.5%): time-dependent area under the curve (AUCt) 73.8 (95% CI 72.2-75.4); Brier score 10.6 (95% CI 8.8-12.5); index of prediction accuracy (IPA) 18.9. Similarly, performance measures for glucose-lowering treatment initiation were AUCt 80.0 (95% CI 78.4-81.6); Brier score 7.4 (95% CI 5.8-8.9); IPA 18.5.

Supplemental Material Figure S11

Comparison of the performance of the main model fitted using the entire development sample and the main model fitted using data restricted to the Danish National Health Survey in the index year or the year prior to the index year. The models for HbA1c ≥48 mmol/mol (6.5%) included baseline HbA1c, age, sex, BMI, treated hypertension, pre-existing pancreatic disease, absence of cancer, unhealthy diet, doctor's advice to lose weight or change dietary habits, self-reported lack of anyone to talk to, and good self-rated health. The models for glucose-lowering treatment initiation included baseline HbA1c, age, sex, BMI, and doctor's advice to lose weight or change dietary habits.

A, B) Time-dependent receiver operating characteristic curves comparing the discriminative ability for type 2 diabetes defined as (A) HbA1c ≥48 mmol/mol (6.5%) or (B) glucose-lowering treatment initiation.

C, D) Calibration curves comparing the estimated and observed probabilities for the two models for type 2 diabetes defined as (C) HbA1c ≥48 mmol/mol (6.5%) or (D) glucose-lowering treatment initiation.



The model based on the restricted development sample had the following performance measures for HbA1c ≥48 mmol/mol (6.5%): time-dependent area under the curve (AUCt) 72.9 (95% CI 71.3-74.4); Brier score 10.3 (95% CI 8.5-12.2); index of prediction accuracy (IPA) 21.3. Similarly, performance measures for glucose-lowering treatment initiation were AUCt 79.2 (95% CI 77.6-80.8); Brier score 7.4 (95% CI 5.8-9.0); IPA 18.0.

Supplemental Material Figure S12

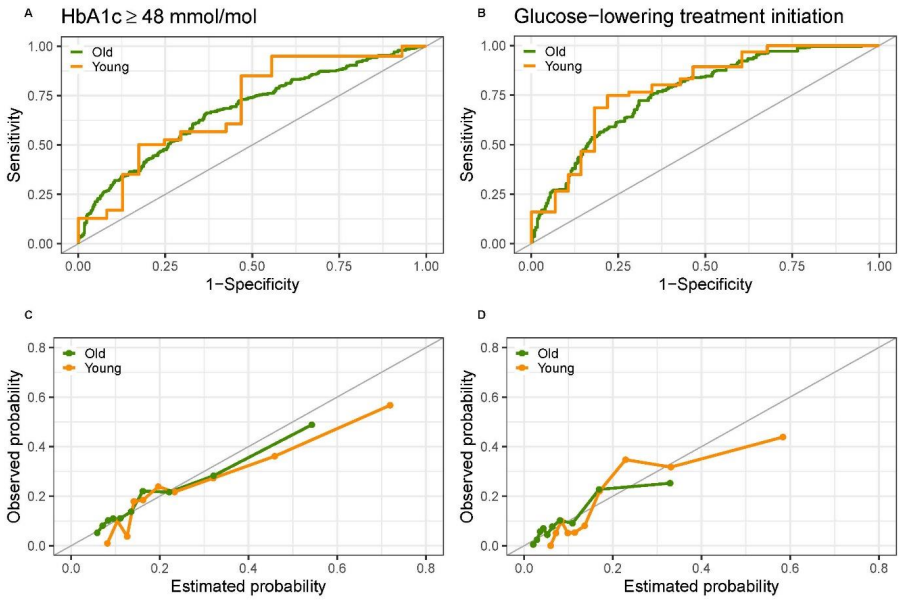
Performance of the models fitted and validated stratified by age (<60 years or ≥60 years of age on the prediabetes index date).

The model for HbA1c ≥48 mmol/mol (6.5%) for those aged below 60 years (Young) included Female sex (beta -0.39 [95% CI -0.55;-0.23]), HbA1c (beta 0.49 [95% CI 0.44-0.54]), BMI (beta 0.03 [95% CI 0.02-0.05]), and Doctor's advice to lose weight or change dietary habits (beta 0.30 [95% CI 0.13-0.47]), with $\beta_0 = 0.14$, and had the following performance measures: AUCt 69.6 (95% CI 68.4-70.9); Brier score 9.9 (95% CI 8.8-11.8); index of prediction accuracy (IPA) 26.7. For those aged above 60 years (Old), the model included Female sex (beta -0.27 [95% CI -0.37;-0.17]), HbA1c (beta 0.50 [95% CI 0.47; 0.53]), BMI (beta 0.03 [95% CI 0.02; 0.04]), and Doctor's advice to lose weight or change dietary habits (beta 0.39 [95% CI 0.26; 0.51]), with $\beta_0 = 0.09$, and the performance measures: AUCt 67.8 (95% CI 66.2-69.4); Brier score 10.8 (95% CI 9.7-11.9); index of prediction accuracy (IPA) 14.9.

The model for glucose-lowering treatment initiation for those aged below 60 years included HbA1c (beta 0.48 [95% CI 0.42-0.54]), BMI (beta 0.05 [95% CI 0.03-0.06]), and Doctor's advice to lose weight or change dietary habits (beta 0.29 [95% CI 0.09-0.50]), $\beta_0 = 0.07$, and performance measures: AUCt 78.8 (95% CI 77.4-80.1); Brier score 8.6 (95% CI 6.6-10.5); index of prediction accuracy (IPA) 26.0. For those aged above 60 years, the model included Age (beta -0.03 [95%CI -0.04;-0.03]), Female sex (beta -0.27 [95% CI -0.40;-0.13]), HbA1c (beta 0.50 [95% CI 0.46-0.54]), BMI (beta 0.05 [95% CI 0.04-0.06], and Doctor's advice to lose weight or change dietary habits (beta 0.42 [95% CI 0.25-0.59]), $\beta_0 = 0.06$, and performance measures: AUCt 76.3 (95% CI 74.6-78.0); Brier score 7.2 (95% CI 6.1-8.2); index of prediction accuracy (IPA) 12.0.

A, B) Time-dependent receiver operating characteristic curves comparing the discriminative ability for type 2 diabetes defined as (A) HbA1c ≥48 mmol/mol (6.5%) or (B) glucose-lowering treatment initiation.

C, D) Calibration curves comparing the estimated and observed probabilities for type 2 diabetes defined as (C) HbA1c ≥48 mmol/mol (6.5%) or (D) glucose-lowering treatment initiation.



Study III | Trajectory study

Longitudinal HbA1c patterns before first treatment of diabetes in everyday clinical practice: A latent class trajectory analysis.
Nicolaisen SK, le Cessie S, Thomsen RW, Witte DR, Dekkers OM, Sørensen HT, Pedersen L.
In draft.

Study III

Longitudinal HbA1c patterns before first treatment of diabetes in everyday clinical practice: A latent class trajectory analysis

Sia Kromann Nicolaisen[1], Saskia le Cessie[2], Reimar Wernich Thomsen[1], Daniel R. Witte[3,4], Olaf M. Dekkers[5] , Henrik Toft Sørensen[1], Lars Pedersen[1]

[1]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

[2]Department of Clinical Epidemiology & Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.

[3]Steno Diabetes Center Aarhus, Aarhus, Denmark

[4]Department of Public Health, Aarhus University, Aarhus, Denmark

[5]Department of Clinical Epidemiology & Department of Endocrinology and Metabolism, Leiden University Medical Center, Leiden, The Netherlands.

Corresponding author: Sia Kromann Nicolaisen, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Olof Palmes Allé 43-45, 8200 Aarhus N, Denmark. E-mail: sknc@clin.au.dk, Telephone: +45 8716 7212

Words: abstract 317, article 3,309

Tables: 1

Figures: 2

1 Abstract

Background

Since 2011, diagnosis and treatment initiation for type 2 diabetes in Denmark has mainly been based on an HbA1c threshold of 48 mmol/mol. Little is known about trajectories of HbA1c prior to diabetes treatment initiation in everyday clinical practice.

Objectives

To examine if there is heterogeneity in longitudinal patterns of HbA1c in real-world individuals who initiate glucose-lowering treatment.

Methods

In this population-based cohort study, we used HbA1c from routine laboratory databases linked to other healthcare registries. We assessed HbA1c measurements in the 5 years preceding first-ever glucose-lowering treatment. Latent class trajectory analysis was used to classify individuals based on their longitudinal HbA1c patterns.

Results

Among 21,556 individuals receiving first-ever glucose-lowering treatment in 3 Danish regions during 2017-2018, 20,733 (96%) had one or more HbA1c measurements recorded in the preceding 5 years. Four classes with distinct longitudinal HbA1c trajectories were identified. All trajectories showed slowly increasing HbA1c levels over several years, but with varyingly steep increase in HbA1c starting 9 to 17 months before diabetes treatment initiation. The largest class included 74% of the individuals and the mean trajectory had HbA1c levels above the 48 mmol/mol threshold for 9 months prior to diabetes treatment initiation, at which time the mean HbA1c had risen to 52 mmol/mol (95% CI 52-52 mmol/mol). In the remaining 3 classes, HbA1c had been above 48 mmol/mol for almost 1.5 years before treatment initiation. In these classes, HbA1c increases were steeper, reaching means of 79

mmol/mol (95% CI 78-80 mmol/mol), 105 mmol/mol (95% CI 104-106 mmol/mol), and 137 mmol/mol (95% CI 135-140 mmol/mol), respectively, before diabetes treatment initiation.

Conclusion

We identified 4 distinct longitudinal patterns of HbA1c prior to glucose-lowering treatment initiation in clinical practice, indicating distinct pathophysiological changes. Individuals in all 4 classes had mean HbA1c above the diabetes diagnostic threshold for many months before treatment was initiated, pointing towards therapeutic inertia that may potentially increase the risk of future diabetes complications.

2 Introduction

According to current guidelines, prompt multifactorial intervention with pharmacological treatment and lifestyle changes should be initiated when type 2 diabetes is diagnosed[1,2]. The overall purpose of glucose-lowering treatment intervention is good glycemic control to prevent diabetes complications[1,2]. Since 2011, the most often used diagnostic criterion has been an HbA1c measurement of 48 mmol/mol or higher[3]. In everyday clinical practice, the measured HbA1c value is often much higher when glucose-lowering treatment is initiated, with around 50% of individuals having HbA1c values above 53 mmol/mol and almost 20% having HbA1c above 75 mmol/mol in a recent real-world observational study[4].

Since 2011, HbA1c testing has increased immensely in Denmark and elsewhere, not only among individuals with prediabetes or diabetes, but also in the general population[5,6]. Population-based laboratory databases may therefore be a valuable tool to study the longitudinal history of an individual's HbA1c levels, even before the individual is diagnosed with diabetes. Type 2 diabetes is a more heterogeneous disease than previously thought[7-12] and the progression from normal values to diabetes, usually via prediabetes (*e.g.,* HbA1c 42-47 mmol/mol as defined by the World Health Organization[13-15]), may differ in different individuals. Progression to diabetes is thought to be characterized by slowly increasing glucose levels over time, followed by a more rapid increase before the onset of diabetes[16-18]. However, there might by multiple modes of progression[8,19] and evidence from population-based studies is scarce.

The aim of this study was to exploit the longitudinal HbA1c data now available in the Danish medical registries to identify distinct HbA1c trajectories prior to first-ever glucose-lowering treatment initiation and thus to study the heterogeneity in increasing levels of HbA1c and to assess any potential therapeutic inertia.

3 Methods

3.1 Data sources

This population-based cohort study is based on data from the Danish nationwide medical registries. All Danes are assigned a unique personal identification number at birth or immigration, making individual linkage between registers possible[20]. Denmark has a tax-supported health care system that ensures unfettered access to medical care for all residents[21]. This includes access to general practitioners and hospitals and partial reimbursement for prescribed drugs. Individual-level information was obtained from the following registries: the Danish Civil Registration System, which contains data on vital status and date of death; the Danish Register of Medicinal Product Statistics, which contains complete prescription information from all community-based pharmacies since 1994[22]; the Danish National Patient Registry, which contains all discharge diagnoses from Danish hospitals since 1977 and from hospital emergency room and outpatient clinics contacts since 1995[23]; and socioeconomic registries maintained by Statistics Denmark, which contain data on family and household socioeconomics, ethnic origin, educational level, employment status, and income. HbA1c measurements were obtained from the nationwide Register of Laboratory Results for Research[5] and the regional Clinical Laboratory Information System Research Database at Aarhus University[5,6]. Together, these registries contain virtually all laboratory measurements ordered both by hospital clinicians and general practitioners for all members of the Danish population[5,6]

3.2 Study cohort

All residents of Central Denmark Region, North Denmark Region, or Capital Region of Denmark (covering approximately 65% of the 5.8 million Danes in 2018[24]) in 2017-2018 were considered eligible for this study. Individuals with a first-ever redemption of a glucose-lowering drug prescription (ATC:

A10) in 2017-2018 after at least 5 years of residence and with at least one HbA1c measurement in the 5 years prior to treatment initiation were included in the study cohort.

3.3 Statistical analyses

We used a three-step approach to fit latent class models to identify distinct trajectories of HbA1c in the 5 years prior to treatment initiation[25]. HbA1c was modelled as a function of days before treatment initiation and time 0 was the day of treatment initiation.

In the first step, we fitted standard linear mixed-effects model (i.e. a 1-class model without latent classes), to find the best parametrization to model the relation between time and HbA1c. In all models, we included a random intercept and a random slope (time) with an unspecified variance-covariance structure. We fitted one model with a fixed linear, squared, and cubic time term and we additionally specified models using restricted cubic splines with 3, 4, and 5 knots[26]. The knots were either determined based on quantiles of data or were determined to split time in equally spaced intervals. Along with a visual comparison of these 7 standard linear mixed-effects models, we used the maximum of the log-likelihood function, the Bayesian information criterion (BIC), and the Akaike information criterion (AIC) to identify the best performing model parametrization.

In the second step, the best performing model parametrization was used to fit latent class models with 2, 3, 4, and 5 latent classes. To ease the computation, these models were fitted based on a random sample of one third of the individuals. Class-membership probabilities were assumed not to depend on any additional covariates and we allowed covariates to vary across latent classes.

To find the optimal latent class model, the models were compared using the BIC and the AIC. If at least one class included less than 1% of the individuals, we did not further increase the number of classes, as such small classes were not considered clinically relevant and thus only required additional parameters in the model. The individual posterior class-membership probabilities were also taken into

account, as they can be used to evaluate a model's ability to unambiguously assign individuals to a latent class. After the optimal model was chosen, the individuals were assigned to the latent class with the highest posterior class-membership probability. The posterior class-membership probabilities and assigned latent class were also calculated for the remaining two thirds of the individuals. Baseline characteristics available at the day of treatment initiation, were used to characterize the individuals in each of the 4 identified classes. Further details about the covariates are given in Supplemental Table S1.

In the third and last step, the trajectories were refitted within each class using nonparametric stochastic smoothing using all measurements from each of the 4 classes defined by the best performing model. The trajectories were smoothed using a random walk of order 2 with rounded month as time unit and individual random intercepts and they were fitted using integrated nested Laplace approximation (INLA)[27,28].

3.4 Sensitivity analyses

To ensure that the model parametrization was not changed substantially based on the number of available HbA1c measurements per individual, the analyses were repeated among individuals with at least 3 available HbA1c measurements in the 5 years preceding the first-ever glucose-lowering treatment.

All statistical analyses were conducted using SAS version 9.4 (SAS Institute Inc, Cary, North Carolina) and R version 4.0.2 (R Core Team, 2020). For a list of essential R packages, see Supplemental Table S2.

4 Results

In total, 2,964,032 individuals had lived in Central Denmark Region, North Denmark Region, and/or Capital Region of Denmark for at least 5 years in 2017-2018 and were thus eligible for our study (Figure

1). A total of 21,556 individuals initiated their first-ever glucose-lowering treatment in 2017-2018, whereof 20,733 (96.2%) had at least one HbA1c measurement in the prior 5 years and thus comprised the study population (Figure 1). The median age at treatment initiation was 60.0 years (interquartile range [IQR] 49.6-70.3) and 9,330 (45.0%) were women (Supplemental Table S3). The individuals had a median of 4 HbA1c measurements (IQR 2-7) in the 5 years prior to diabetes treatment initiation and contributed a total of 105,211 measurements. The median of all included HbA1c measurements was 46 mmol/mol (42-52 mmol/mol), *i.e.,* close to the threshold value for diabetes, 48 mmol/mol, and with the majority of the measurements being higher than the lower limit for prediabetes, 42 mmol/mol. This was also depicted in the boxplots of all measurements per month, which showed an increase in HbA1c values over time (Supplemental Figure S1).

4.1 Latent class model

Using the full data (N=20,733 with 105,211 measurements), all model types were fitted as standard linear mixed-effects models (*i.e.,* 1-class models). Based on the BIC, AIC, and the maximum log-likelihood estimate, the best model parametrization was the restricted cubic splines model with 5 knots placed based on quantiles of data (Supplemental Table S4, Supplemental Figure S2). However, as the mean curves of the models with 3 or 4 knots showed a similar fit to data (Supplemental Figure S2), the model with restricted cubic splines with 3 knots placed based on quantiles of data was chosen in order to keep the number of parameters low, as less knots substantially reduced the computing time.

The 3-knot model was fitted with 2, 3, 4, and 5 latent classes (Supplemental Table S5) using a random sample (one third) of the individuals (N=6,911 with 35,133 measurements). Based on the BIC and AIC, the 5-class model showed the best fit followed by the 4-class model. As one of the classes in the 5-class model included only 0.7% of the individuals, the 4-class model was thus preferred (Supplemental Table S6). The posterior class-membership probabilities were calculated for all

individuals and individuals were assigned to the class with the largest posterior probability. The 4 latent classes then included: Class 1, 15,283 (73.7%) individuals with 82.4% of the measurements; Class 2, 2,862 (13.8%) individuals with 11.1% of the measurements; Class 3 2,164 (10.4%) individuals with 5.6% of the measurements; and Class 4, 424 (2.1%) individuals with 1.0% of the measurements (Supplemental Table S3, Table 1).

4.2 HbA1c trajectories

The mean curves for all 4 classes are plotted in Supplemental Figure S3 and are compared to the observed data. The smoothed curves are plotted in Figure 2. The smoothed trajectory for the largest class, Class 1, had slowly increasing levels of HbA1c during the entire 5-year period, but with slightly faster increasing levels in the last year up to glucose-lowering treatment initiation (Figure 2). Individuals in Class 2 and Class 3 had slowly increasing levels of HbA1c until 1 to 1.5 years prior to treatment initiation, where the values started to increase much more rapidly than for Class 1. The smoothed trajectories for Class 1 and Class 2, had HbA1c levels above 42 mmol/mol (*i.e.,* the lower limit for prediabetes) for most of the five years prior to treatment initiation, whereas the levels were slightly lower for Class 3. The smallest class, Class 4, included only 2% of the individuals and the trajectory overall followed the same patterns as for the other classes. However, Class 4 had more unstable HbA1c values that tended to be lower than in the other classes up to 1.5 years prior to treatment initiation, followed by a very steep increase prior to treatment initiation.

Based on the model parametrization (Supplemental Figure S3), the mean trajectory for Class 1 exceeded the limit for diabetes, *i.e.,* 48 mmol/mol, 8.8 months before treatment was initiated and Class 1 had a mean HbA1c on 51.9 mmol/mol (95% CI 51.6-52.2 mmol/mol) at the time of treatment initiation. The mean trajectory for Class 2 exceeded 48 mmol/mol 16.7 months prior to treatment initiation and reached 78.8 mmol/mol (95% CI 77.7-79.8 mmol/mol) at the time of treatment initiation. The mean trajectories for the 2 smallest classes exceeded the limit for diabetes around the

same time as the trajectory for Class 2 (16.2 months prior to treatment initiation for Class 3 and 15.1 months for Class 4). However, at the time of treatment initiation, these 2 classes had the highest mean HbA1c values; 104.7 mmol/mol (95% CI 103.5-105.9 mmol/mol) for Class 3 and 137.0 mmol/mol (95% CI 134.7-139.2 mmol/mol) for Class 4.

4.3 Baseline characteristics

The individuals in Class 1 had HbA1c measured more often than those in the other 3 classes and had a median of 5 measurements (IQR 3-8) during the 5 years before treatment. In comparison, Class 2 had median 3 measurements (IQR 2-5) and Class 3 and Class 4 both had only median 2 measurements (IQR 1-3) (Table 1, Supplemental Table S3), with 34.3% in Class 3 and 38.9% in Class 4 having only 1 measurement during the 5 years prior to diabetes treatment initiation. In all classes, the most recent HbA1c measurement was close to the time of treatment initiation (between 4 and 11 days before, Supplemental Table S3). Class 1 included more women than the other classes and the median age decreased from Class 1 to Class 4 (61.6 years [IQR 50.8-71.4] in Class 1, 57.3 years [IQR 47.8-67.6] in Class 2, 54.5 years [IQR 44.8-63.7] in Class 3, and 52.6 years [IQR 34.0-63.2] in Class 4). In particular, Class 4 included the largest share of individuals below 30 years of age (5.6% in Class 1, 4.0% in Class 2, 8.6% in Class 3, and 20.5% in Class 4). Class 4 also had higher prevalence of low education, low income, and the highest prevalence of being unmarried. Individuals in Class 3 and Class 4 had more often received a hospital diagnosis of diabetes (31.3% and 62.0% in Class 3 and 4, versus only 8.1% in Class 1) and in particular Class 4 stood out with a high use of insulin monotherapy (40.6%) or any polytherapy initiation (24.8%), whereas almost all (94.3%) in the large Class 1 started non-insulin monotherapy (typically with metformin) (Supplemental Table S7). Pre-existing cardiovascular disease or cardiovascular drug use was much more common in Class 1 than in the smaller classes, whereas Class 4 had more alcoholism-related diagnoses (Table 1, Supplemental Table S3).

4.4 Sensitivity analyses

In the analysis restricted to the 14,249 individuals with at least 3 HbA1c measurements in the 5 years prior to diabetes treatment initiation, the optimal model was the restricted cubic splines model with 3 knots placed based on quantiles and a total of 3 latent classes. The smoothed curved from the 2 largest classes in the sensitivity analysis were similar to the curves for Class 1 and Class 2 from the main analysis (Supplemental Figure S4). It was mainly individuals from Class 1 and Class 2 from the main analysis who had at least 3 measurements and the sensitivity analysis thus reproduced the trajectories for Class 1 and Class 2. The remaining individuals were included in the third class, with a trajectory reflecting an average of the curves for Class 3 and Class 4 from the main analysis.

5 Discussion

We identified 4 classes with distinct longitudinal patterns of HbA1c prior to glucose-lowering treatment initiation. The majority of individuals had HbA1c in the range of prediabetes for almost five years prior to their first treatment for diabetes. All classes had HbA1c levels that indicated diabetes (*i.e.,* HbA1c>=48 mmol/mol) for many months before diabetes treatment was initiated, but the steepness of the increase in HbA1c during the last 1 to 1.5 years before treatment initiation differed between the latent classes.

The fact that individuals in our largest class, Class 1, including three-quarters of all individuals, had HbA1c above the diabetic level for approximately 9 months before treatment was initiated, may be related to caregivers waiting for confirmatory elevated HbA1c measurements and potentially initiating non-medical interventions including advice on lifestyle changes and weight loss. Nonetheless, it is recommended that pharmacological treatment should be initiated no later than 3-6 months after the type 2 diabetes diagnosis[1,2,29]. Class 1 had more frequent measurements than those in the other classes and the mean HbA1c was 52 mmol/mol at the time when treatment (usually

metformin) was initiated. This suggests that individuals in Class 1 were detected rather early in their glycemic progression. Class 2, Class 3, and Class 4 together included only one quarter of the individuals. They were all characterized by very high HbA1c values at the time of treatment initiation and accordingly were more aggressively treated.

5.1 Comparison with other studies

The general pattern of slowly increasing HbA1c values over years observed for all 4 classes in the early phase of disease development, followed by more rapidly increasing values the last year up to diabetes diagnosis/treatment initiation is in line with previous studies[16-18]. Our analyses extend previous knowledge by providing evidence that the increase in the last phase of disease progression can differ considerably between individuals. This suggests heterogeneity in trajectories and that more focus may be appropriate on subgroups of individuals with a more rapid-onset and severe form of diabetes, to potentially avoid very high glucose levels before therapy[30]. Given the characteristics of the individuals in Class 4 (younger age, large share with diabetes hospitalizations, some with ICU admissions, rapid and steep increase of HbA1c, often initiating insulin treatment), this class may comprise many patients with type 1 rather than type 2 diabetes, while the increased share with social risk markers and e.g., alcoholism in Class 4 may suggest neglected diabetes cases in general in that group. Class 2 and Class 3 were more likely to include cases of rapid-onset type 2 diabetes, for example cases of adult-onset autoimmune diabetes associated with severe loss of insulin secretion. Unfortunately, we lacked clinical details and biomarkers to further study these potential differences.

We studied heterogeneity in prediabetes relying on HbA1c only. In the Whitehall II study, heterogeneity has also been demonstrated in other factors before treatment initiation. For example, latent class trajectory analysis was performed in the Whitehall II study to examine patterns of obesity prior to type 2 diabetes[19]. The authors identified 3 rather than 4 distinct trajectories. However, as their cohort only included confirmed type 2 diabetes, this may be in line with our results, if we assume that

our Class 4 predominantly depicts type 1 and other autoimmune types of diabetes. Based on their obesity classes, the authors also showed differences in trajectories of fasting plasma glucose and 2h plasma glucose. They found that those with a high progressive weight gain had less time before diabetes was diagnosed and had a steeper increase in their glucose measures. We unfortunately did not have data on fasting glucose and 2h plasma glucose available in our data and could thus not compare our HbA1c findings with glucose patterns in the same population[31,32].

5.2 Study limitations

One of the major difficulties using latent class trajectory analysis is to determine the optimal number of classes. Other models and number of classes might have been possible. However, we used both polynomial regressions and spline models to capture as much of the variability as possible. In addition, we fitted the best performing parametrization using multiple number of classes.

Diabetes and prediabetes can be defined based on other measures than HbA1c (*e.g.,* fasting plasma glucose or 2h fasting glucose) and we only had data available on HbA1c. Yet, HbA1c testing is now the preferred diagnostic method in clinical practice. Our study was based on HbA1c measurements from routine care laboratory databases and thus reflect real-world testing. However, the exact indication or reason for conducting an HbA1c test is unknown in our population.

Another limitation is that we did not have access to drug use during inpatient hospitalization, only prescription drugs bought at community pharmacies. Since some individuals had already experienced a hospital admission with diabetes before their first prescription, the assigned date of treatment initiation may have been misclassified in these individuals, as their treatment could already have been initiated during their hospital admission.

HbA1c measurements conducted close to index are higher than measurements conducted earlier[17] (if measurements were higher earlier, treatment would most likely also have been initiated

earlier) and measurement error can potentially "cause" the disease, as having a random high measurement can prompt treatment initiation. The HbA1c will most likely be lower after treatment initiation because of regression to the mean. It is thus important to recognize the bias introduced by having diabetes defined using a threshold value (threshold bias[17]), as the increasing values prior to treatment initiation is a result of a combination of naturally increasing values and the threshold[17].

To conclude, with more knowledge about the heterogeneity in the way diabetes manifests, treatment could possibly be improved to ensure all individuals receive the best suitable treatment. Some individuals might benefit from earlier intervention in order to avoid having very high glucose values before treatment is initiated.

6 Additional information

Data permission

This study was reported to the Danish Data Protection Agency (Aarhus University record number 2016-051-000001/1702). The data used in this study are owned and managed by Statistics Denmark. In accordance with Danish law and data protection policies, the data used in this study were anonymized and it were stored and analyzed on a secured server. The data are available to researchers from research environments pre-approved by Statistics Denmark upon project approval by the Danish Data Protection Agency and Statistics Denmark. Researchers can apply for access to data when the request is approved by the Danish Data Protection Agency (https://www.datatilsynet.dk).

Ethical considerations

No ethical approval was needed for this study. All data originated from registries and none were specifically collected for this project.

Author contributions

SKN, SLC, RWT, DW, OMD, and LP designed the study. SKN had full access to data, performed all the analyses, and prepared the draft of the manuscript. SLS, RWT, HTS, and LP critically revised the manuscript. All authors approved the final manuscript.

Conflicts of interest

The Department of Clinical Epidemiology is involved in studies with funding from various companies as research grants to and administered by Aarhus University. None of these studies are related to the current study**.** SKN received a travel grant from The Danish Diabetes Association.

Role of the funding sources

The study's funding sources had no role in the collection, analysis, or interpretation of the data. It had no role in study design or writing of the manuscript. SKN and LP had final responsibility for the decision to submit for publication.

7 References

1. American Diabetes Association Professional Practice Committee. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes—2022. *Diabetes Care.* 2021;45(Supplement_1):S125-S143.
2. Hansen KB, Kristensen JK, Balasubramaniam K, et al. Farmakologisk behandling af type 2-diabetes - Fælles retningslinje fra DES og DSAM. 2022; https://vejledninger.dsam.dk/fbv-t2dm/. Accessed 1 June 2022.
3. Knudsen JS, Knudsen SS, Hulman A, et al. Changes in type 2 diabetes incidence and mortality associated with introduction of HbA1c as diagnostic option: A Danish 24-year population-based study. *Lancet Reg Health Eur.* 2022;14:100291.
4. Knudsen JS, Hulman A, Rønn PF, et al. Trends in HbA(1c) and LDL Cholesterol in Patients With Type 2 Diabetes Receiving First-Time Treatment in Northern Denmark, 2000-2017: Population-Based Sequential Cross-Sectional Analysis. *Diabetes Care.* 2020;43(2):e17-e19.
5. Arendt JFH, Hansen AT, Ladefoged SA, Sørensen HT, Pedersen L, Adelborg K. Existing Data Sources in Clinical Epidemiology: Laboratory Information System Databases in Denmark. *Clin Epidemiol.* 2020;12:469-475.
6. Grann AF, Erichsen R, Nielsen AG, Froslev T, Thomsen RW. Existing data sources for clinical epidemiology: The clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clin Epidemiol.* 2011;3:133-138.
7. American Diabetes Association Professional Practice Committee. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2022. *Diabetes Care.* 2022;45(Suppl 1):S17-S38.
8. Færch K, Witte DR, Tabák AG, et al. Trajectories of cardiometabolic risk factors before diagnosis of three subtypes of type 2 diabetes: a post-hoc analysis of the longitudinal Whitehall II cohort study. *Lancet Diabetes Endocrinol.* 2013;1(1):43-51.
9. Ahlqvist E, Storm P, Karajamaki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol.* 2018;6(5):361-369.
10. Christensen DH, Nicolaisen SK, Ahlqvist E, et al. Type 2 diabetes classification: a data-driven cluster study of the Danish Centre for Strategic Research in Type 2 Diabetes (DD2) cohort. *BMJ Open Diabetes Res Care.* 2022;10(2).
11. Wagner R, Heni M, Tabák AG, et al. Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. *Nat Med.* 2021;27(1):49-57.
12. Walraven I, Mast MR, Hoekstra T, et al. Distinct HbA1c trajectories in a type 2 diabetes cohort. *Acta Diabetol.* 2015;52(2):267-275.
13. The International Expert Committee. International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. *Diabetes Care.* 2009;32(7):1327-1334.
14. World Health Organization. WHO Guidelines Approved by the Guidelines Review Committee. In: *Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus: Abbreviated Report of a WHO Consultation.* Geneva: World Health Organization; 2011.
15. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M. Prediabetes: a high-risk state for diabetes development. *The Lancet.* 2012;379(9833):2279-2290.
16. Tabák AG, Jokela M, Akbaraly TN, Brunner EJ, Kivimäki M, Witte DR. Trajectories of glycaemia, insulin sensitivity, and insulin secretion before diagnosis of type 2 diabetes: an analysis from the Whitehall II study. *The Lancet.* 2009;373(9682):2215-2221.
17. Mason CC, Hanson RL, Knowler WC. Progression to type 2 diabetes characterized by moderate then rapid glucose increases. *Diabetes.* 2007;56(8):2054-2061.
18. Heianza Y, Arase Y, Fujihara K, et al. Longitudinal Trajectories of HbA1c and Fasting Plasma Glucose Levels During the Development of Type 2 Diabetes: The Toranomon Hospital Health Management Center Study 7 (TOPICS 7). *Diabetes Care.* 2012;35(5):1050-1052.

19. Vistisen D, Witte DR, Tabák AG, et al. Patterns of obesity development before the diagnosis of type 2 diabetes: the Whitehall II cohort study. *PLoS Med.* 2014;11(2):e1001602.
20. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. *Eur J Epidemiol.* 2014;29(8):541-549.
21. 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;11:563-591.
22. Kildemoes HW, Sørensen HT, Hallas J. The Danish National Prescription Registry. *Scand J Public Health.* 2011;39(7 Suppl):38-41.
23. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol.* 2015;7:449-490.
24. Danmarks Statistik [Statistics Denmark]. Folketal den 1. i kvartalet efter område og tid. 2022; https://statistikbanken.dk/statbank5a/default.asp?w=1856. Accessed 23 June 2022.
25. Proust-Lima C, Philipps V, Liquet B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. *Journal of Statistical Software.* 2017;78(2):1 - 56.
26. Harrell FE, Jr. *Regression Modeling Strategies.* Vol 2. 2 ed: Springer; 2015.
27. Gómez-Rubio V. *Bayesian inference with INLA.* Chapman & Hall/CRC Press; 2020 (online version 2021-08-29).
28. Rue H, Lindgren F, van Niekerk J, Krainski E, Fattah EA. R-INLA Project. 2022; https://www.r-inla.org. Accessed 7 June 2022.
29. American Diabetes Association Professional Practice Committee. 6. Glycemic Targets: Standards of Medical Care in Diabetes-2022. *Diabetes Care.* 2022;45(Suppl 1):S83-S96.
30. Riddle MC, Gerstein HC, Home PD. Lingering Effects of Hyperglycemia in Recently Diagnosed Diabetes During Long-term Follow-up of the DCCT/EDIC and UKPDS Cohorts: More Evidence That Early Control Matters. *Diabetes Care.* 2021.
31. Faerch K, Johansen NB, Witte DR, Lauritzen T, Jorgensen ME, Vistisen D. Relationship between insulin resistance and beta-cell dysfunction in subphenotypes of prediabetes and type 2 diabetes. *J Clin Endocrinol Metab.* 2015;100(2):707-716.
32. Vistisen D, Kivimäki M, Perreault L, et al. Reversion from prediabetes to normoglycaemia and risk of cardiovascular disease and mortality: the Whitehall II cohort study. *Diabetologia.* 2019;62(8):1385-1390.

8 Figures and tables

Figure 1

Flowchart of individuals included in this study. All residents of Central Denmark Region, North Denmark Region, or Capital Region of Denmark were considered eligible. The study population included individuals with a first-ever redemption of a glucose-lowering drug prescription who had at least one HbA1c measurement in the 5 years prior to treatment initiation.

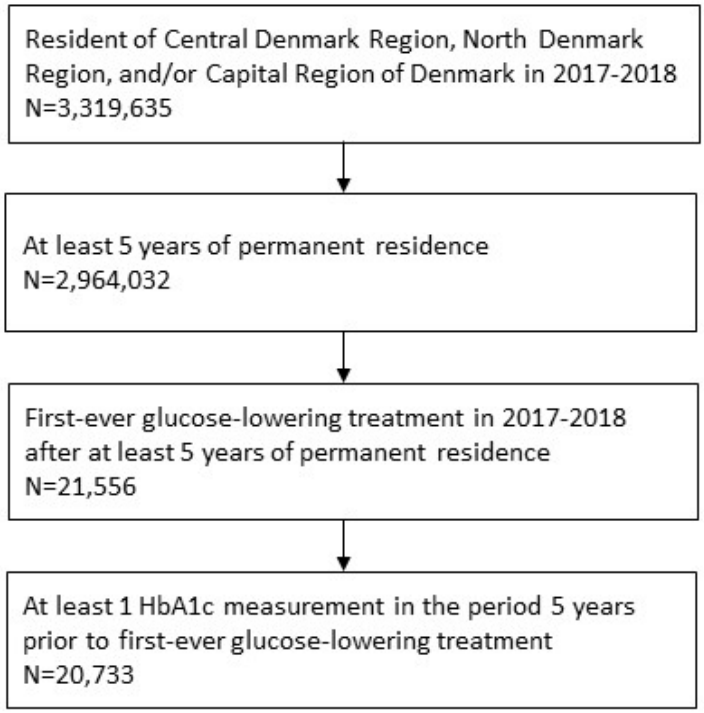


Figure 2

The smoothed trajectories for the 4 latent classes using a spline model with 3-knots placed based on quantiles of data. The red dashed lines indicate the lower limit for prediabetes, *i.e.,* HbA1c>=42 mmol/mol, and the lower limit for diabetes, *i.e.,* HbA1c>=48 mmol/mol.

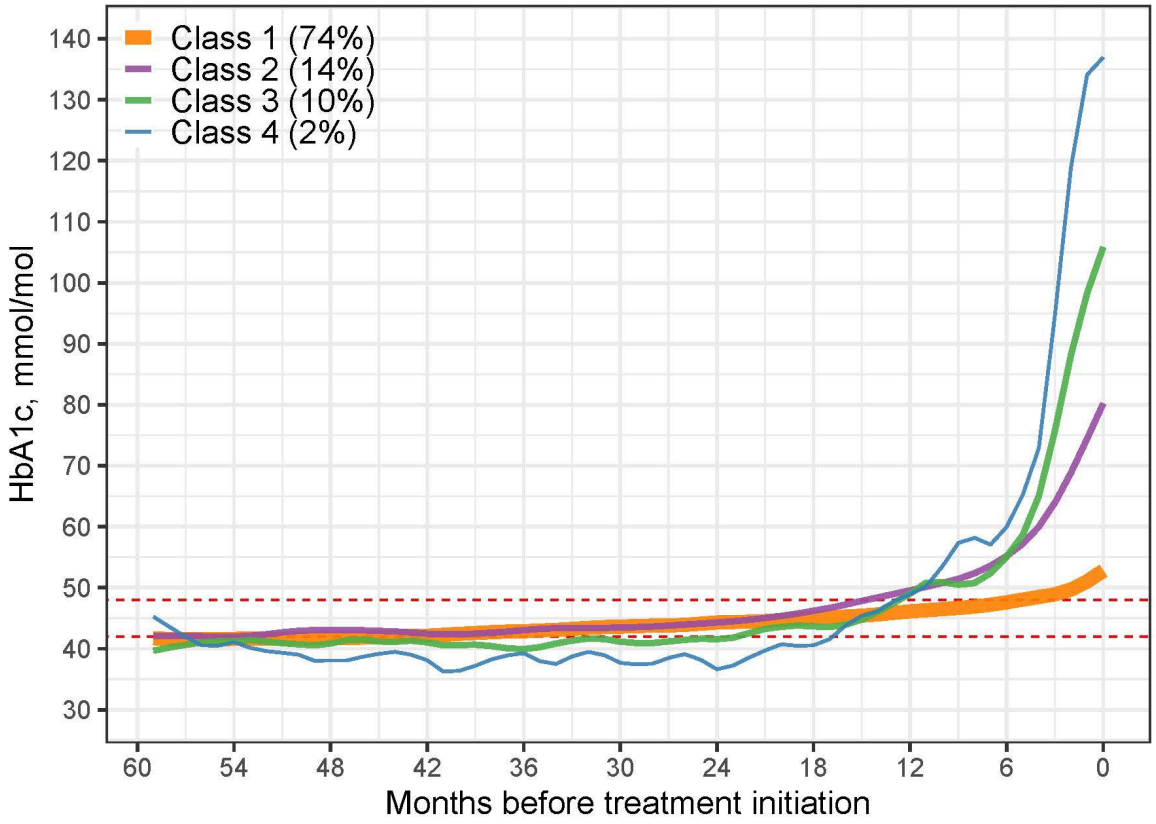


Table 1

Baseline characteristics for the 4 latent classes. Additional baseline characteristics are available in Supplemental Table S3 and additional baseline treatment regimens are available in Supplemental Table S7.

	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Total	20,733 (100.0%)	15,283 (100.0%)	2,862 (100.0%)	2,164 (100.0%)	424 (100.0%)
Sex					
Female	9,330 (45.0%)	7,561 (49.5%)	940 (32.8%)	685 (31.7%)	144 (34.0%)
Male	11,403 (55.0%)	7,722 (50.5%)	1,922 (67.2%)	1,479 (68.3%)	280 (66.0%)
Age at index (years)	60.0 (49.6-70.3)	61.6 (50.8-71.4)	57.3 (47.8-67.6)	54.5 (44.8-63.7)	52.6 (34.0-63.2)
Number of HbA1c measurements	4.0 (2.0-7.0)	5.0 (3.0-8.0)	3.0 (2.0-5.0)	2.0 (1.0-3.0)	2.0 (1.0-3.0)
Days since last HbA1c measurement	8.0 (4.0-22.0)	11.0 (6.0-32.0)	6.0 (2.0-11.0)	4.0 (1.0-9.0)	7.0 (2.0-13.0)
Socioeconomics					
Highest education achieved					
None, basic education, or primary school	6,976 (33.6%)	5,167 (33.8%)	932 (32.6%)	711 (32.9%)	166 (39.2%)
Youth education, high school, or similar educational level	8,893 (42.9%)	6,508 (42.6%)	1,256 (43.9%)	953 (44.0%)	176 (41.5%)
Higher education	4,253 (20.5%)	3,145 (20.6%)	586 (20.5%)	454 (21.0%)	68 (16.0%)
Income					
Lowest income group	3,258 (15.7%)	2,443 (16.0%)	398 (13.9%)	320 (14.8%)	97 (22.9%)
Low to medium income	7,427 (35.8%)	5,757 (37.7%)	934 (32.6%)	616 (28.5%)	120 (28.3%)
Medium to high income	5,473 (26.4%)	4,017 (26.3%)	744 (26.0%)	605 (28.0%)	107 (25.2%)
Highest income group	4,404 (21.2%)	2,991 (19.6%)	752 (26.3%)	581 (26.8%)	80 (18.9%)
Marital status					
Married	10,519 (50.7%)	7,977 (52.2%)	1,406 (49.1%)	989 (45.7%)	147 (34.7%)
Divorced	3,400 (16.4%)	2,517 (16.5%)	499 (17.4%)	315 (14.6%)	69 (16.3%)
Widow/widower	1,972 (9.5%)	1,601 (10.5%)	219 (7.7%)	136 (6.3%)	16 (3.8%)
Unmarried	4,842 (23.4%)	3,188 (20.9%)	738 (25.8%)	724 (33.5%)	192 (45.3%)
Comorbidities and prescription drug use					
Statins	6,396 (30.8%)	5,500 (36.0%)	575 (20.1%)	272 (12.6%)	49 (11.6%)
Any potential antihypertensive treatment	10,527 (50.8%)	8,559 (56.0%)	1,209 (42.2%)	637 (29.4%)	122 (28.8%)

	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Diabetes	2,709 (13.1%)	1,243 (8.1%)	526 (18.4%)	677 (31.3%)	263 (62.0%)
Cardiovascular disease	3,730 (18.0%)	2,968 (19.4%)	479 (16.7%)	241 (11.1%)	42 (9.9%)
Admission to intensive care unit	849 (4.1%)	611 (4.0%)	132 (4.6%)	68 (3.1%)	38 (9.0%)
Alcoholism-related diagnosis or medication	499 (2.4%)	330 (2.2%)	91 (3.2%)	48 (2.2%)	30 (7.1%)
Baseline glucose-lowering treatment regimen					
Overall type of treatment					
Monotherapy	19,810 (95.5%)	15,058 (98.5%)	2,641 (92.3%)	1,792 (82.8%)	319 (75.2%)
Polytherapy	923 (4.5%)	225 (1.5%)	221 (7.7%)	372 (17.2%)	105 (24.8%)
Overall type of monotherapy					
Insulin based monotherapy	1,410 (6.8%)	651 (4.3%)	234 (8.2%)	353 (16.3%)	172 (40.6%)
Non-insulin based monotherapy	18,400 (88.7%)	14,407 (94.3%)	2,407 (84.1%)	1,439 (66.5%)	147 (34.7%)

Missing data: Highest education achieved n=611 (2.9%), Income n=171 (0.8%). The rest of the variables had no missing data.

9 Supplemental Materials

Longitudinal HbA1c patterns before first treatment of diabetes in everyday clinical practice: A latent class trajectory analysis

Sia Kromann Nicolaisen[1], Saskia le Cessie[2], Reimar Wernich Thomsen[1], Daniel R. Witte[3,4], Olaf M. Dekkers[5] , Henrik Toft Sørensen[1], Lars Pedersen[1]

[1]Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark
[2]Department of Clinical Epidemiology & Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
[3]Steno Diabetes Center Aarhus, Aarhus, Denmark
[4]Department of Public Health, Aarhus University, Aarhus, Denmark
[5]Department of Clinical Epidemiology & Department of Endocrinology and Metabolism, Leiden University Medical Center, Leiden, The Netherlands.

Corresponding author: Sia Kromann Nicolaisen, Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Olof Palmes Allé 43-45, 8200 Aarhus N, Denmark. E-mail: sknc@clin.au.dk, Telephone: +45 8716 7212

List of Supplemental Tables and Figures:

Supplemental Table S1

Full list of all variables, definitions, codes, and data sources in the study.

A hospitalization was defined using primary and secondary diagnoses both from inpatient admissions and outpatient visits. The admission date was used as the hospital contact day. Absence of a hospital contact was defined as 'no admission'. Prescription drug use was defined as a redemption of a prescription. Absence of prescriptions is defined as 'no drug use'. Variables defined based on the presence of records in the healthcare registries had no missing values, as absence of records (*e.g.*, for any antihypertensive drugs, cancer, etc.) was defined as absence of the predictor.

Variable	Type of variable/definition	Variable assessment period	Data source	Registry definition
First-ever glucose-lowering treatment	N/A	Prescription redemption. This defines the index date.	DRMPS	ATC: A10. "First-ever" is defined using prescription data dated back to 1 January 1995. For definitions of combinations of treatment regimens, please see table below.
HbA1c measurements (mmol/mol)	Continuous.	All measurements 5*360 days prior to the index date	LAB	NPU: NPU27300 (mmol/mol [IFCC]), NPU03835 (% [DCCT]), (all available measurements were converted into mmol/mol and rounded to nearest integer using the formula: IFCC=(DCCT*10.93)-23.5. Data are restricted to a maximum of one measurement per day by taking the mean of possible multiple measurements. See reference for formula: Lægehåndbogen (2020) Hæmoglobin A1c (HbA1c). Available from https://www.sundhed.dk/sundhedsfaglig/laegehaandbogen/undersoegelser-og-proever/klinisk-biokemi/blodproever/haemoglobin-a1c-hba1c/. Accessed 06 February 2021)
Demographic variables				
Sex	Binary: Female, Male	N/A	DCRS	N/A
Age on index date	Continuous and categorical: <30; 30-49: 50-69; >=70 years.	Index date	DCRS	N/A
Ethnic origin	Categorical: Danish, Immigrant/descendant /unknown	N/A	DST	N/A
Prescription drug use				
Any trombocyte-aggregation prophylaxis	Binary: No, Yes	180 days prior to the index date	DRMPS	ATC: B01AC06, N02BA01, B01AC30, B01AC07, B01AC22, B01AC04, B01AC24, B01AC25
Any hypolipidemic treatment	Binary: No, Yes	180 days prior to the index date	DRMPS	ATC: C10, A10BH51
Statins	Binary: No, Yes	180 days prior to the index date	DRMPS	ATC: C10AA, C10BA C10BX, A10BH51
Any antihypertensive drug (excl. loop-diuretics)	Binary: No, Yes	180 days prior to the index date	DRMPS	ATC: C02, C03A, C03B, C03D, C03E (not C03EB), C07, C08, C09A, C09B, C09C, C09D, C09X, G04CA03, C10BX04, C10BX06, C10BX07, C10BX11, C10BX12, C10BX13, C10BX14, C10BX15, C10BX10,

				C10BX07, C10BX09, C10BX11, C10BX14
ACE inhibitors or ATII antagonists	Binary: No, Yes	180 days prior to the index date	DRMPS	ATC: C09A, C09B, C09C, C09D, C10BX04, C10BX06, C10BX07, C10BX11, C10BX12, C10BX13, C10BX14, C10BX15, C10BX10
Antidepressants	Binary: No, Yes	180 days prior to the index date	DRMPS	ATC: N06A
Comorbidities				
Charlson Comorbidity Index score as a measure of overall comorbidity burden	Categorical: 0, 1-2, >=3	Hospital contact day 5*360 days prior to the index date	DNPR	See reference for ICD-10 codes: Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8. The code DQ61 (cystic kidney disease) was not included in the calculation, as it was not available in the data. Diabetes is included in the calculation.
Diabetes	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DO24, DH360, DG632, DG590, DH280, DH334B, DM142, DN083, DT383, DE10-DE14
Cardiovascular disease (includes stable angina pectoris [or CABG/PCI procedures], myocardial infarction, heart failure, stroke, atrial fibrillation/flutter, heart valve disease, and venous thromboembolism)	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI20 (not DI200), DI251, DI259, KFNA, KFNB, KFNC, KFND, KFNE, KFNF, KFNG, KFNH20, DI21, DI22, DI23, DI50, DI60, DI61, DI63, DI64, DI48, DI05, DI06, DI07, DI08, DI098, DI39, DI511A, DQ22, DQ23, DI34-DI37, DI26, DI801, DI802, DI803
Angina pectoris or CABG/PCI procedures	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI20 (not DI200), DI251, DI259, KFNA, KFNB, KFNC, KFND, KFNE, KFNF, KFNG, KFNH20
Myocardial infarction	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI21, DI22, DI23
Heart failure	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI50
Atrial fibrillation/flutter	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI48
Stroke	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI60, DI61, DI63, DI64
Hypertension	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI10-DI15
Obesity	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DRMPS, DNPR	ATC: A08. ICD-10: DE65-DE68.
Chronic pulmonary disease	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DJ40, DJ684, DJ701, DJ703, DJ841, DJ920, DJ961, DJ982, DJ983
Cancer (excl. non-melanoma skin cancer)	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DC00-DC99 (not DC44)

Kidney disease	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DI12, DI13, DN07, DN11, DN14, DQ61, DN08, DE102, DE112, DE142, DN00–DN05, DN18–DN19
Liver disease	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: DB18, DB150, DB160, DB162, DB190, DI85, DK70, DK71, DK72, DK73, DK74, DK760, DK766
Dementia	Binary: No, Yes	Prescription redemption 180 days prior to the index date, or hospital contact day 5*360 days prior to the index date	DRMPS, DNPR	ATC: N06D. ICD-10: DF00, DF01, DF02, DF03, DG30, DG310B, DG311, DG318, DG319.
Possible HbA1c-modifying conditions (includes prescription drug use of dapsone, ribavirin, antiretrovirals, trimethoprim-sulfamethoxazole, hydroxyurea, vitamin C, vitamin E, or opiates; and hospital admission or treatment with ribavirin, hemolysis, hemoglobinopathies, blood transfusion, acute blood loss/anemia, hypertriglyceridemia, chronic liver disease, pregnancy, iron deficiency, vitamin B12 deficiency, uremia, hyperbilirubinemia, end-stage renal disease [kidney transplant or dialysis], alcoholism-related diagnoses or medication, treatment with fetal hemoglobin or methemoglobin)	Binary: No, Yes	Prescription redemption 180 days prior to the index date, or hospital contact day 5*360 days prior to the index date	DRMPS, DNPR	ATC: J04BA02, D10AX05, J05AP01, J05, J01EE01, J04AM08, J01EA01, QJ51EA01, J01EC01, QJ01EQ11, L01XX05, A11GA, A11HA03, N02A. ICD-10: BPHM04, DD59, DD55-DD59, BOQA, DD62, BOHC, DD50, DD51, DD52, DE781, DB18, DB150, DB160, DB162, DB190, DI85, DK70, DK71, DK72, DK73, DK74, DK760, DK766, DZ321, DO00-DO99, DZ33-DZ39, DE611, DD50, DD51, BOHC2, DR39, DE804, DE806A, BJFD2, KKAS, DT861, DZ940, DF10 (not DF100), DE244, DG312, DG621, DG721, DI426, DK292, DK70, DK852, DK860, DQ860, DZ502, DZ714, DZ721, DR780, DT51, DT500A, DD564, DD74.
Previous ICU admission	Binary: No, Yes	Hospital contact day 5*360 days prior to the index date	DNPR	ICD-10: NABE, NABB
Markers of smoking	Binary: No, Yes	Prescription redemption 180 days prior to the index date, or hospital contact day 5*360 days prior to the index date	DRMPS, DNPR	ATC: R03, N07BA. ICD-10: DF17, DZ716, DZ720, DJ41-DJ44.
Markers of alcoholism	Binary: No, Yes	Prescription redemption 180 days prior to the index date, or hospital contact day 5*360 days prior to the index date	DRMPS, DNPR	ATC: V03AA, N07BB. ICD-10: DF10 (not DF100), DE244, DG312, DG621, DG721, DI426, DK292, DK70, DK852, DK860, DQ860, DZ502, DZ714, DZ721, DR780, DT51, DT500A.
Socioeconomic variables				
Highest education achieved	Categorical: None, basic education, or primary school; Youth education, high school	Index date	DST	N/A

	or similar educational level; Higher education			
Employment status	Categorical: Employed; Unemployed or not part of the workforce	End of previous November	DST	N/A
Income	Categorical: Lowest income group; Low to medium income; Medium to high income; Highest income group	End of previous year	DST	Income group from the last calendar year. Quartiles are based on the entire Danish population per year.
Type of household	Categorical: Living alone, Not living alone	End of previous year	DST	N/A
Marital status	Categorical: Married, Divorced, Widow/widower, Unmarried	Index date	DST	N/A

Abbreviations: LAB, nationwide laboratory registry; DNPR, Danish National Patient Registry; DRMPS, Danish Register of Medicinal Product Statistics; DCRS, Danish Civil Registration System; DST, registries maintained by Statistics Denmark; NPU, laboratory codes in the Nomenclature for Properties and Units; ICD-10, *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*; ATC Anatomical Therapeutic Chemical Codes; N/A, not applicable.

Codes used to define baseline glucose-lowering treatment regimens (index date plus 14 days after) includes combinations of drug categories within A10. All prescriptions are from DRMPS.

Baseline glucose-lowering treatment	**Combination**	**ATC codes**
Glucose-lowering drug	All	A10
Insulin	Mono	A10A (excl. A10AE54, A10AE56)
	Combination	A10AE54, A10AE56
Metformin	Mono	A10BA02
	Combinations	A10BD01, A10BD02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15, A10BD16, A10BD17, A10BD18, A10BD20, A10BD22, A10BD23, A10BD25, A10BD26, A10BD27
Sulfonylurea	Mono	A10BB
	In combination with metformin only	A10BD01, A10BD02
	Other combinations	A10BD04, A10BD06
SGLT2	Mono	A10BK
	In combination with metformin only	A10BD15, A10BD16, A10BD20, A10BD23
	Other combinations	A10BD19, A10BD21, A10BD24, A10BD25, A10BD27
GLP1	Mono	A10BJ
	Combinations	A10AE54, A10AE56
DPP4	Mono	A10BH
	In combination with metformin only	A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD18, A10BD22
	Other combinations	A10BD09, A10BD12, A10BD19, A10BD21, A10BD24, A10BD25, A10BD27

Supplemental Table S2

List of essential R packages (R version 4.0.2, R Core Team, 2020) and versions used in the study.

Package name	Package version	Author(s)	CRAN/reference
lcmm	1.9.4	Cecile Proust-Lima, Viviane Philipps, Amadou Diakite, and Benoit Liquet	https://cran.r-project.org/web/packages/lcmm/index.html
rms	6.2.0	Frank E Harrell Jr	https://cran.r-project.org/web/packages/rms/index.html
INLA	20.3.17	Håvard Rue, Finn Lindgren, Daniel Simpson, Sara Martino, Elias Teixeira Krainski, Haakon Bakka, Andrea Riebler, Geir-Arne Fuglstad, and Martin Modrak, Christian Chiuchiolo.	https://www.r-inla.org/

Supplemental Table S3

Baseline table showing the characteristics for the entire population and stratified by latent class.

	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Total	20,733 (100.0%)	15,283 (100.0%)	2,862 (100.0%)	2,164 (100.0%)	424 (100.0%)
Demographic variables					
Sex					
Female	9,330 (45.0%)	7,561 (49.5%)	940 (32.8%)	685 (31.7%)	144 (34.0%)
Male	11,403 (55.0%)	7,722 (50.5%)	1,922 (67.2%)	1,479 (68.3%)	280 (66.0%)
Age at index (years)	60.0 (49.6-70.3)	61.6 (50.8-71.4)	57.3 (47.8-67.6)	54.5 (44.8-63.7)	52.6 (34.0-63.2)
Age at index					
<30 years	1,245 (6.0%)	856 (5.6%)	115 (4.0%)	187 (8.6%)	87 (20.5%)
30-49 years	4,123 (19.9%)	2,710 (17.7%)	737 (25.8%)	576 (26.6%)	100 (23.6%)
50-69 years	10,027 (48.4%)	7,303 (47.8%)	1,427 (49.9%)	1,112 (51.4%)	185 (43.6%)
>=70 years	5,338 (25.7%)	4,414 (28.9%)	583 (20.4%)	289 (13.4%)	52 (12.3%)
Ethnic origin					
Danish	17,733 (85.5%)	13,005 (85.1%)	2,487 (86.9%)	1,868 (86.3%)	373 (88.0%)
Immigrant/descendant/unknown	3,000 (14.5%)	2,278 (14.9%)	375 (13.1%)	296 (13.7%)	51 (12.0%)
HbA1c measures					
Number of HbA1c measurements	4.0 (2.0-7.0)	5.0 (3.0-8.0)	3.0 (2.0-5.0)	2.0 (1.0-3.0)	2.0 (1.0-3.0)
Days since first HbA1c measurement	1,212.0 (349.0-1,591	1,328.0 (699.0-1,62	999.0 (16.0-1,511.0	23.5 (6.0-1,244.0)	18.0 (7.0-783.0)
Days since last HbA1c measurement	8.0 (4.0-22.0)	11.0 (6.0-32.0)	6.0 (2.0-11.0)	4.0 (1.0-9.0)	7.0 (2.0-13.0)
Value of last HbA1c measurement	54.0 (49.0-71.0)	51.0 (48.0-56.0)	79.5 (73.0-87.0)	104.0 (97.0-113.0)	133.0 (126.0-143.0)
Only one measurement					
Yes	2,930 (14.1%)	1,480 (9.7%)	542 (18.9%)	743 (34.3%)	165 (38.9%)
No	17,803 (85.9%)	13,803 (90.3%)	2,320 (81.1%)	1,421 (65.7%)	259 (61.1%)
Prescription drug use					
Any trombocyte-aggregation prophylaxis	3,411 (16.5%)	2,836 (18.6%)	391 (13.7%)	148 (6.8%)	36 (8.5%)
Any hypolipidemic treatment	6,526 (31.5%)	5,605 (36.7%)	593 (20.7%)	278 (12.8%)	50 (11.8%)
Statins	6,396 (30.8%)	5,500 (36.0%)	575 (20.1%)	272 (12.6%)	49 (11.6%)
Any potential antihypertensive treatment (excl. loop diuretics)	10,527 (50.8%)	8,559 (56.0%)	1,209 (42.2%)	637 (29.4%)	122 (28.8%)
ACE inhibitors or ATII antagonists	7,382 (35.6%)	6,070 (39.7%)	826 (28.9%)	412 (19.0%)	74 (17.5%)
Antidepressants	2,868 (13.8%)	2,263 (14.8%)	379 (13.2%)	186 (8.6%)	40 (9.4%)

	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Comorbidities					
Charlson comorbidity index					
0	13,783 (66.5%)	10,345 (67.7%)	1,876 (65.5%)	1,387 (64.1%)	175 (41.3%)
1-2	5,389 (26.0%)	3,790 (24.8%)	729 (25.5%)	650 (30.0%)	220 (51.9%)
>=3	1,561 (7.5%)	1,148 (7.5%)	257 (9.0%)	127 (5.9%)	29 (6.8%)
Diabetes	2,709 (13.1%)	1,243 (8.1%)	526 (18.4%)	677 (31.3%)	263 (62.0%)
Cardiovascular disease	3,730 (18.0%)	2,968 (19.4%)	479 (16.7%)	241 (11.1%)	42 (9.9%)
Stable angina pectoris or CABG/PCI procedures	1,500 (7.2%)	1,225 (8.0%)	168 (5.9%)	90 (4.2%)	17 (4.0%)
Myocardial infarction	583 (2.8%)	453 (3.0%)	80 (2.8%)	43 (2.0%)	7 (1.7%)
Heart failure	687 (3.3%)	532 (3.5%)	101 (3.5%)	46 (2.1%)	8 (1.9%)
Atrial fibrillation/flutter	1,451 (7.0%)	1,163 (7.6%)	181 (6.3%)	92 (4.3%)	15 (3.5%)
Stroke	696 (3.4%)	545 (3.6%)	89 (3.1%)	52 (2.4%)	10 (2.4%)
Hypertension	3,091 (14.9%)	2,482 (16.2%)	378 (13.2%)	181 (8.4%)	50 (11.8%)
Obesity	1,584 (7.6%)	1,270 (8.3%)	198 (6.9%)	94 (4.3%)	22 (5.2%)
Chronic pulmonary disease	1,464 (7.1%)	1,191 (7.8%)	173 (6.0%)	79 (3.7%)	21 (5.0%)
Cancer (excl. non-melanoma skin cancer)	1,414 (6.8%)	1,126 (7.4%)	186 (6.5%)	86 (4.0%)	16 (3.8%)
Kidney disease diagnosis	302 (1.5%)	###	###	###	###
Liver disease	302 (1.5%)	204 (1.3%)	50 (1.7%)	36 (1.7%)	12 (2.8%)
Dementia	206 (1.0%)	###	###	###	###
Possible HbA1c-modifying conditions	2,175 (10.5%)	1,741 (11.4%)	262 (9.2%)	123 (5.7%)	49 (11.6%)
Admission to intensive care unit	849 (4.1%)	611 (4.0%)	132 (4.6%)	68 (3.1%)	38 (9.0%)
Markers of smoking	3,237 (15.6%)	2,624 (17.2%)	367 (12.8%)	203 (9.4%)	43 (10.1%)
Alcoholism-related diagnosis or medication	499 (2.4%)	330 (2.2%)	91 (3.2%)	48 (2.2%)	30 (7.1%)
Socioeconomics					
Highest education achieved					
None, basic education, or primary school	6,976 (33.6%)	5,167 (33.8%)	932 (32.6%)	711 (32.9%)	166 (39.2%)
Youth education, high school, or similar educational level	8,893 (42.9%)	6,508 (42.6%)	1,256 (43.9%)	953 (44.0%)	176 (41.5%)
Higher education	4,253 (20.5%)	3,145 (20.6%)	586 (20.5%)	454 (21.0%)	68 (16.0%)
Employment status					
In employment	9,032 (43.6%)	6,249 (40.9%)	1,369 (47.8%)	1,215 (56.1%)	199 (46.9%)
Unemployed or not part of the workforce	11,701 (56.4%)	9,034 (59.1%)	1,493 (52.2%)	949 (43.9%)	225 (53.1%)
Income					

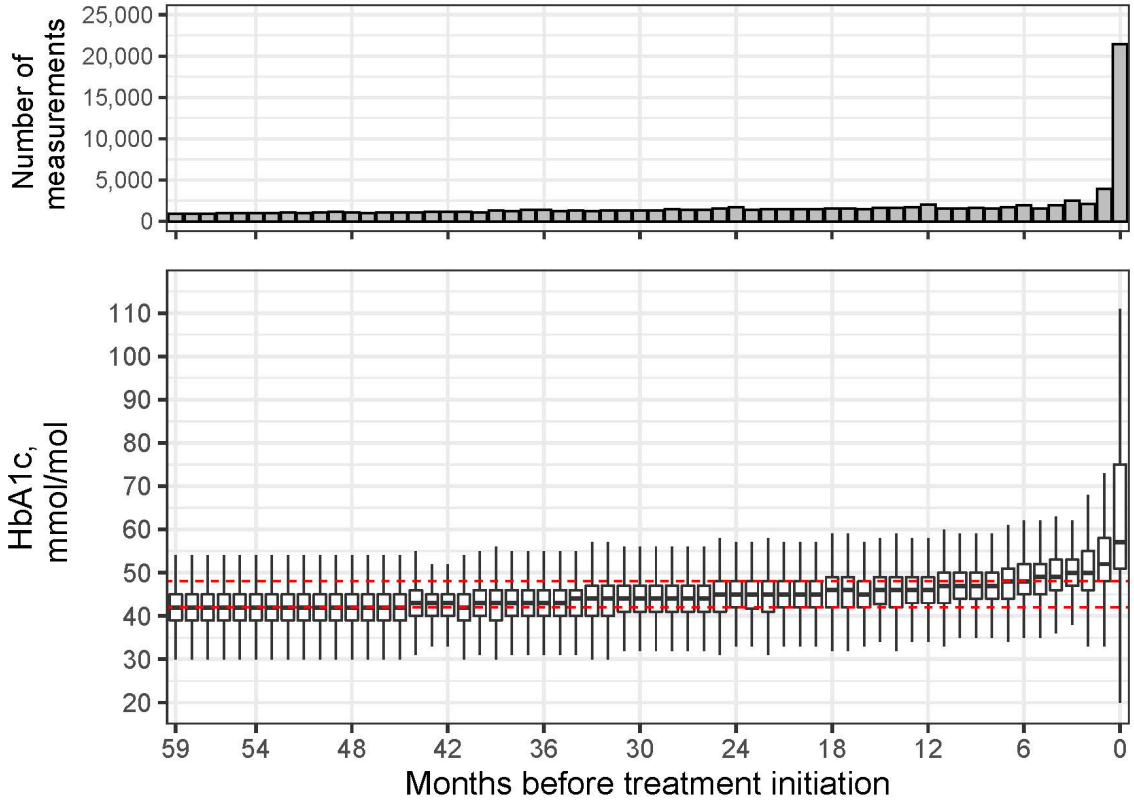
	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Lowest income group	3,258 (15.7%)	2,443 (16.0%)	398 (13.9%)	320 (14.8%)	97 (22.9%)
Low to medium income	7,427 (35.8%)	5,757 (37.7%)	934 (32.6%)	616 (28.5%)	120 (28.3%)
Medium to high income	5,473 (26.4%)	4,017 (26.3%)	744 (26.0%)	605 (28.0%)	107 (25.2%)
Highest income group	4,404 (21.2%)	2,991 (19.6%)	752 (26.3%)	581 (26.8%)	80 (18.9%)
Type of household					
Living alone	7,005 (33.8%)	5,088 (33.3%)	1,043 (36.4%)	735 (34.0%)	139 (32.8%)
Not living alone	13,728 (66.2%)	10,195 (66.7%)	1,819 (63.6%)	1,429 (66.0%)	285 (67.2%)
Marital status					
Married	10,519 (50.7%)	7,977 (52.2%)	1,406 (49.1%)	989 (45.7%)	147 (34.7%)
Divorced	3,400 (16.4%)	2,517 (16.5%)	499 (17.4%)	315 (14.6%)	69 (16.3%)
Widow/widower	1,972 (9.5%)	1,601 (10.5%)	219 (7.7%)	136 (6.3%)	16 (3.8%)
Unmarried	4,842 (23.4%)	3,188 (20.9%)	738 (25.8%)	724 (33.5%)	192 (45.3%)

means a number less than 5 is included in which cannot be reported due to Danish regulations.

Missing data: Highest education achieved n=611 (2.9%), Income n=171 (0.8%). The rest of the variables had no missing data.

Supplemental Figure S1

Histogram and boxplots showing the distributions of all the measurements for each month prior to treatment initiation. The red dotted lines in the lower panel indicate the lower limit for prediabetes (*i.e.,* HbA1c 42 mmol/mol) and diabetes (*i.e.,* HbA1c 48 mmol/mol).



Supplemental Table S4

The models were fitted based on one the full data, i.e., N=20,733 with 105,211 measurements.

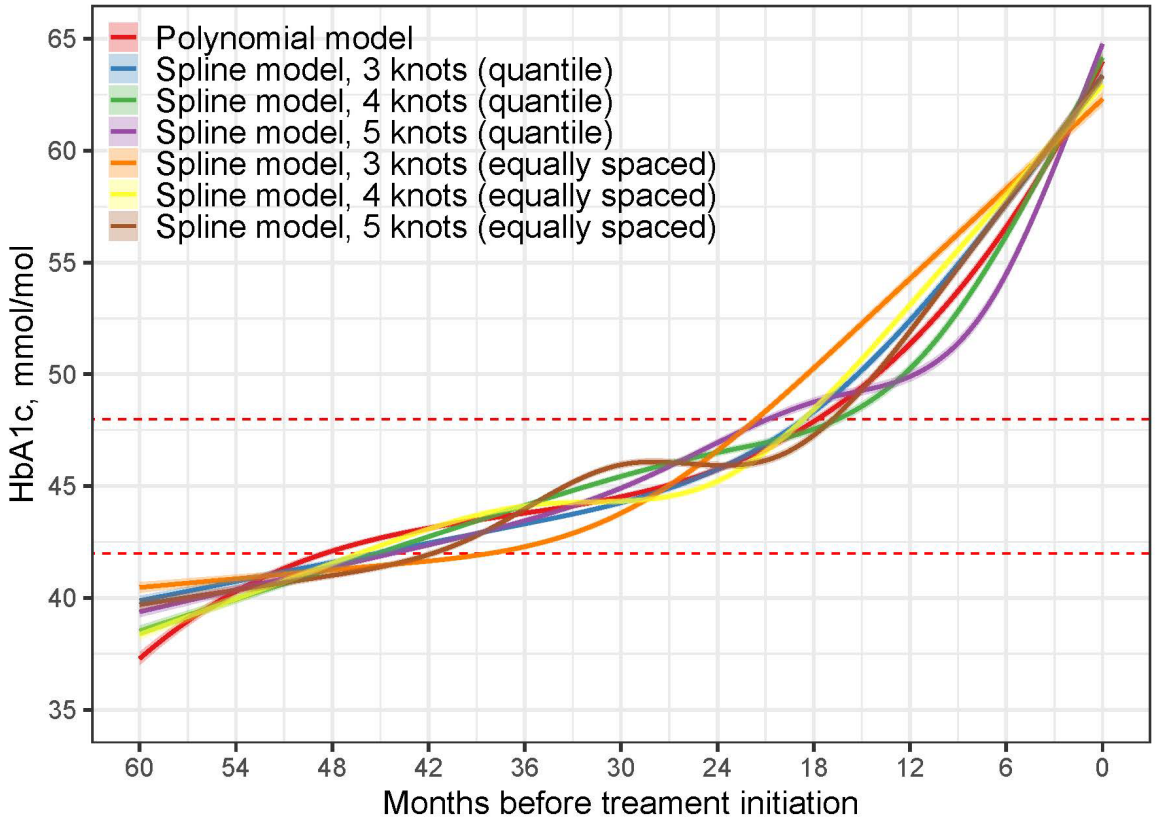
Model	Model specification/knots	Number of parameters	Maximum log-likelihood	BIC	AIC
Polynomial					
3rd degree	t+t^2+t^3	8	-376,750.2	753,579.8	753,516.3
Spline, quantile					
3 knots	Knots = (2, 17, 36)	7	-377,695.0	755,459.5	755,404.0
4 knots	Knots = (1, 11, 24, 40)	8	-376,385.1	752,850.8	752,787.3
5 knots	Knots = (1, 7, 17, 29, 43)	9	-375,697.7	751,448.9	751,377.5
Spline, equally spaced					
3 knots	Knots = (15, 30, 44)	7	-380,081.4	760,232.4	760,176.8
4 knots	Knots = (12, 24, 35, 47)	8	-378,599.9	757,279.4	757,215.9
5 knots	Knots = (10, 20, 30, 39, 49)	9	-377,644.5	755,378.4	755,306.9

The maximum log-likelihood should be as high as possible. Lower values of the Bayesian information criterion (BIC), and the Akaike information criterion (AIC) indicate a better model fit.

All models include 4 parameters from the random effects (variance-covariance matrix and residual standard error)

Supplemental Figure S2

Mean curves for all standard linear mixed models (1-class models). The models were fitted based on the full data, *i.e.,* N=20,733 with 105,211 measurements.



Supplemental Table S5

Based on the best performing standard linear mixed-effects model (the spline model with 3 knots placed based on quantiles), the number of latent classes was increased. Finding the optimal latent class model was based on number of parameters, model fit (AIC and BIC), class sizes, class membership probabilities and the clinical relevance.

The models were fitted based on one third of the individuals (N=6,911 with 35,133 measurements). To avoid a local optimum, all models were run with 50 repetitions using initial values from the 1-class model.

The class sizes are based on individuals being assigned to the class with the highest posterior class-membership probability.

Number of classes (Ng)	Number of parameters	BIC	AIC	Class sizes (percentage of individuals)	Percentage with posterior class-membership probabilities above 80%
Ng=2	11	240,818.9	240,818.6	Class 1 81.4%, Class 2 18.6%	Class 1 96.0%, Class 2 89.7%
Ng=3	15	238,067.2	237,964.6	Class 1 75.3%, Class 2 16.5%, Class 3 8.2%	Class 1 93.4%, Class 2 76.2%, Class 3 80.4%
Ng=4	19	236,927.5	236,797.6	Class 1 73.6%, Class 2 14.0%, Class 3 10.4%, Class 4 2.1%	Class 1 91.9%, Class 2 71.1%, Class 3 76.5%, Class 4 83.3%
Ng=5	23	236,164.8	236,007.4	Class 1 73.0%, Class 2 13.7%, Class 3 10.5%, Class 4 2.0%, Class 5 0.7%	Class 1 92.3%, Class 2 69.6%, Class 3 74.0%, Class 4 85.1%, Class 5 81.3%

Lower values of the BIC and the AIC indicate a better model fit.

Supplemental Table S6

Model coefficients and R code for the final model, *i.e.,* the model with 4 classes using a spline regression with 3 knots placed based on quantiles.

Coefficient	Estimate	Standard error
Class-specific fixed effects (spline)		
$v_{1_0}^{rcs}$	51.86	0.15
$v_{1_1}^{rcs}$	-0.44	0.01
$v_{1_2}^{rcs}$	0.29	0.01
$v_{2_0}^{rcs}$	104.66	0.61
$v_{2_1}^{rcs}$	-3.99	0.04
$v_{2_2}^{rcs}$	3.26	0.04
$v_{3_0}^{rcs}$	136.98	1.14
$v_{3_1}^{rcs}$	-6.59	0.09
$v_{3_2}^{rcs}$	5.56	0.10
$v_{4_0}^{rcs}$	78.76	0.54
$v_{4_1}^{rcs}$	-2.11	0.03
$v_{4_2}^{rcs}$	1.67	0.03
Random effects		
B_{11} (intercept)	77.83	N/A
$B_{12} = B_{21}$ (covariance)	-0.98	N/A
B_{22} (slope)	0.02	N/A
σ_ε (residual standard error)	4.73	0.02
Fixed effects in class-membership model		
ξ_{01}	1.56	0.04
ξ_{02}	-0.35	0.06
ξ_{03}	-1.91	0.11
$\xi_{04} = 0$ (reference)	0	N/A

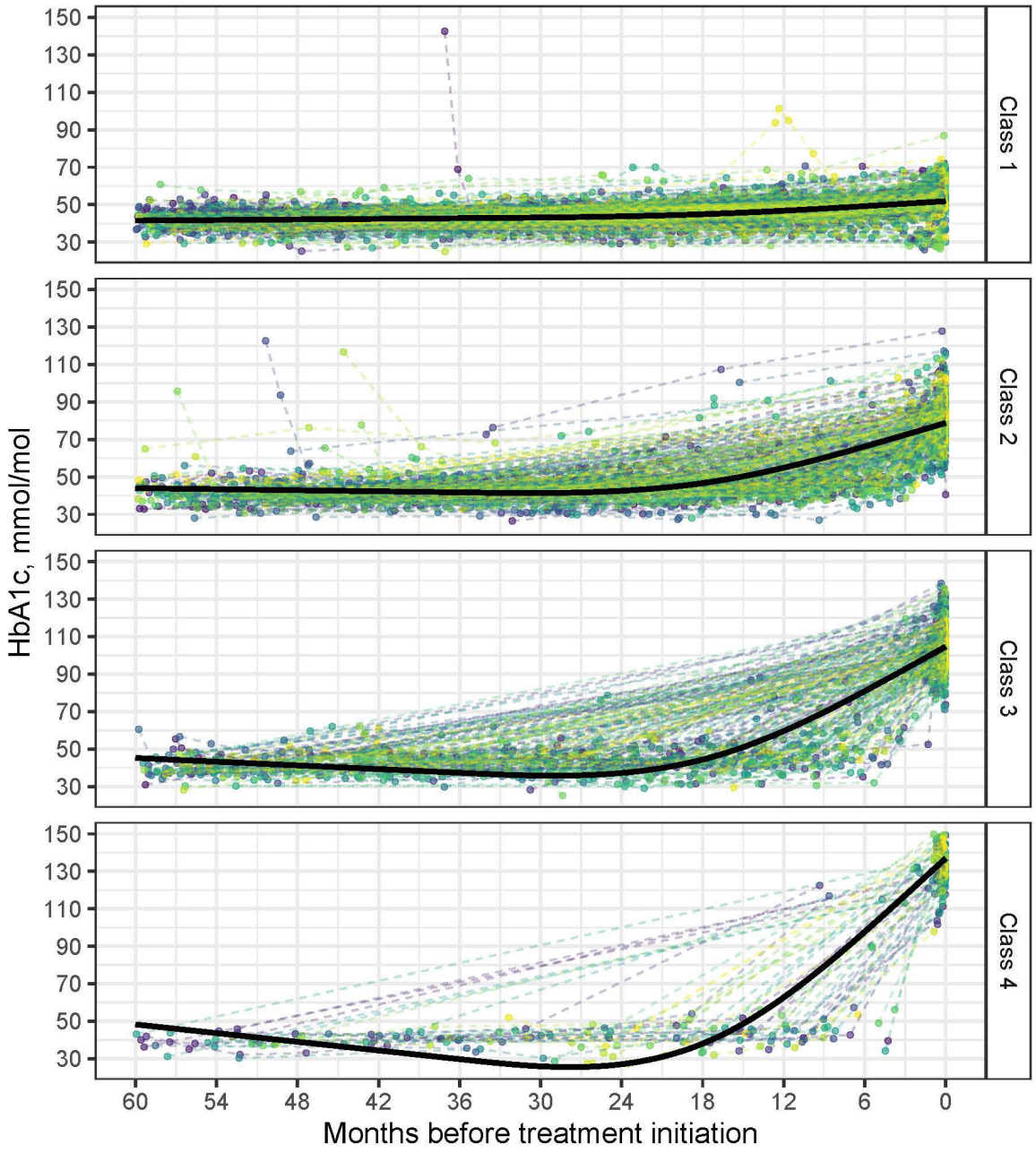
The model was fitted using the following R code:

```
gridsearch(
    rep = 50, maxiter = 50, minit = model1,
    hlme(
        fixed = hba1c ~ rcs(t, parms = c(2,17,36)),
        mixture = ~ rcs(t, parms = c(2,17,36)),
        random = ~ t,
        subject = "cpr",
        ng = 4,
        idiag = FALSE,
        cor = NULL,
        data = Mdata
    )
```

where `model1` is the 1-class model (*i.e.,* ng=1) and `Mdata` is the data including the variables `cpr`, `t` (time in days), and `hba1c`.

Supplemental Figure S3

Plot of the mean curves for the 4 latent classes together with the observed data from a maximum of 500 random individuals from each class.



To avoid reporting sensitive individual-level information, random noise was added to all HbA1c measurements (normal distribution, mean=0, standard deviation=1).

Supplemental Table S7

First-ever glucose-lowering treatment initiated at baseline. The treatment includes the treatment at index date and up to 14 days after the index date to ensure the full baseline treatment regimen is captured.

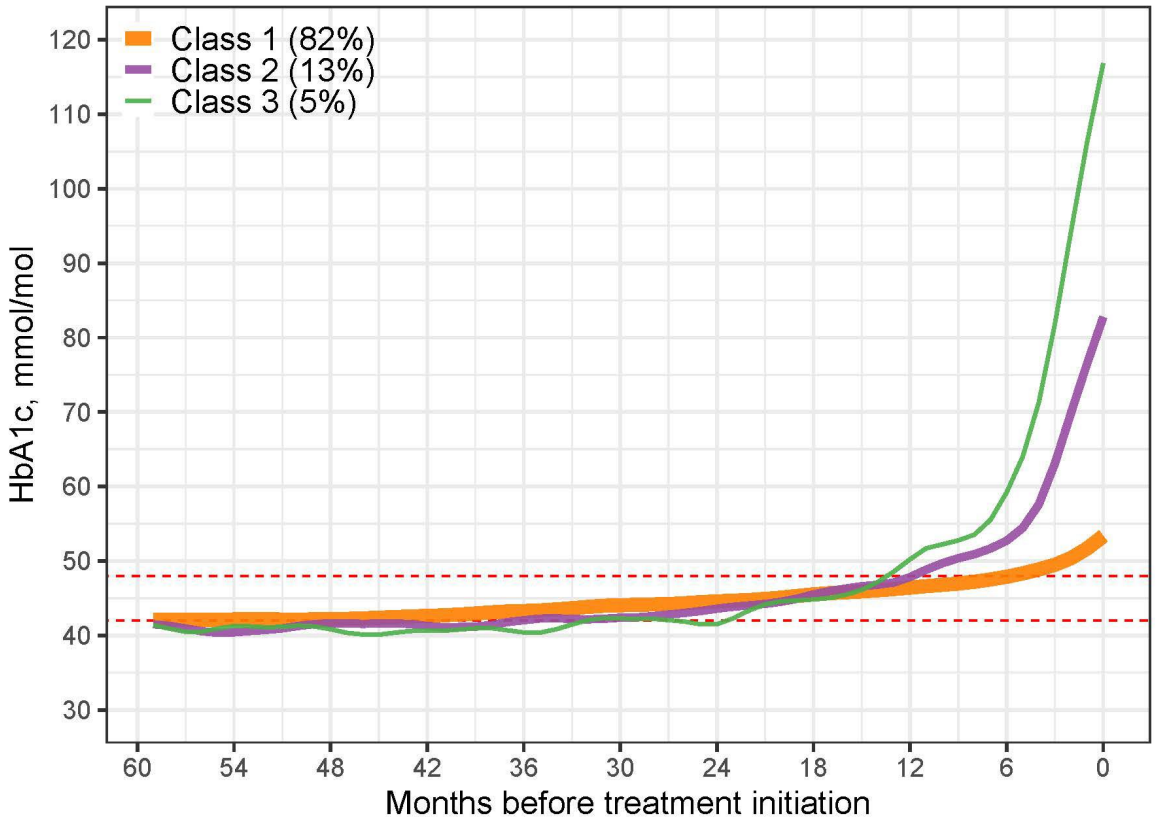
	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Total	20,733 (100.0%)	15,283 (100.0%)	2,862 (100.0%)	2,164 (100.0%)	424 (100.0%)
Baseline single glucose-lowering treatment					
Insulin	1,832 (8.8%)	714 (4.7%)	314 (11.0%)	556 (25.7%)	248 (58.5%)
Metformin	17,948 (86.6%)	13,324 (87.2%)	2,581 (90.2%)	1,794 (82.9%)	249 (58.7%)
SGLT2	205 (1.0%)	100 (0.7%)	40 (1.4%)	56 (2.6%)	9 (2.1%)
GLP1	1,065 (5.1%)	1,017 (6.7%)	13 (0.5%)	27 (1.2%)	8 (1.9%)
DPP4	519 (2.5%)	268 (1.8%)	121 (4.2%)	107 (4.9%)	23 (5.4%)
Sulfonylurea	152 (0.7%)	###	###	###	###
Baseline treatment regimen (disjoint groups)					
Overall type of treatment					
Monotherapy	19,810 (95.5%)	15,058 (98.5%)	2,641 (92.3%)	1,792 (82.8%)	319 (75.2%)
Polytherapy	923 (4.5%)	225 (1.5%)	221 (7.7%)	372 (17.2%)	105 (24.8%)
Overall type of monotherapy					
Insulin based monotherapy	1,410 (6.8%)	651 (4.3%)	234 (8.2%)	353 (16.3%)	172 (40.6%)
Non-insulin based monotherapy	18,400 (88.7%)	14,407 (94.3%)	2,407 (84.1%)	1,439 (66.5%)	147 (34.7%)
Type of monotherapy					
Insulin	1,410 (6.8%)	###	###	###	###
Metformin	17,061 (82.3%)	###	###	###	###
Other monotherapy	1,339 (6.5%)	###	###	###	###
Overall type of polytherapy					
Insulin based polytherapy	422 (2.0%)	63 (0.4%)	80 (2.8%)	203 (9.4%)	76 (17.9%)
Non-insulin based polytherapy	501 (2.4%)	162 (1.1%)	141 (4.9%)	169 (7.8%)	29 (6.8%)
Type of non-insulin based polytherapy					
Non-insulin based polytherapy excluding metformin	8 (0.0%)	###	###	###	###
Non-insulin based polytherapy including metformin	493 (2.4%)	###	###	###	###
Type of insulin based polytherapy					
Insulin+metformin	345 (1.7%)	###	###	###	###

	N (%) or Median (IQR)				
	All individuals	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Insulin+metformin+other	49 (0.2%)	###	###	###	###
Insulin+other	28 (0.1%)	###	###	###	###

means a number less than 5 is included in which cannot be reported due to Danish regulations.

Supplemental Figure S4

The smoothed mean curves for the 3 latent classes based on the spline model with 3-knots placed based on quantiles of data. Data was restricted to individuals with at least 3 measurements in the 5 years prior to first-ever glucose-lowering treatment initiation. The red dashed lines indicate the lower limit for prediabetes, *i.e.,* HbA1c>=42 mmol/mol, and the lower limit for diabetes, *i.e.,* HbA1c>=48 mmol/mol.



Study IV | Regression discontinuity design study

Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study.
Petersen I, Nicolaisen SK, Ricciardi F, Sharma M, Thomsen RW, Baio G, Pedersen L.
Clinical Epidemiology 2020;12:569-577.

Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Impact of Being Eligible for Type 2 Diabetes Treatment on All-Cause Mortality and Cardiovascular Events: Regression Discontinuity Design Study

This article was published in the following Dove Press journal:
Clinical Epidemiology

Irene Petersen[1,2]
Sia Kromann Nicolaisen[2]
Federico Ricciardi[3]
Manuj Sharma[1]
Reimar W Thomsen[2]
Gianluca Baio[3]
Lars Pedersen[2]

[1]Department of Primary Care and Population Health, University College London, London, UK; [2]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark; [3]Department of Statistical Science, University College London, London, UK

Background: Individuals with type 2 diabetes (T2D) have a twofold increased risk for cardiovascular events (CVE), and CVE is responsible for nearly 80% of the mortality. Current treatment guidelines state that individuals should immediately initiate antidiabetic treatment and cardiovascular risk-factor management from T2D diagnosis. However, the evidence base is sparse, and randomized trials are unlikely to be conducted. We examined the impact of being eligible for T2D treatment, as determined by the threshold of HbA_{1c} ≥6.5% (≥48 mmol/mol), on all-cause mortality and CVE. We hypothesised that individuals who were just above this threshold had a lower risk of CVE and all-cause mortality than individuals just below.

Methods and Findings: We used the regression discontinuity design (RDD), a quasi-experimental design, comparing rates of all-cause mortality and CVE in people just below and just above the eligibility for treatment threshold. We included Danish healthcare records from 43,070 individuals aged 40–80 years with no previous T2D record and the first record of HbA_{1c} in the range of 6.0–7.0% (42–53 mmol/mol) between 2006 and 2014. In total, 36,360 individuals had the first record of HbA_{1c} between 6.0% and 6.4% (42–47 mmol/mol), and 6710 individuals had a first record between 6.5% and 7.0% (48–53 mmol/mol). Individuals with a measurement just above 6.5% (48 mmol/mol) had a 21% lower rate of death or CVE, compared to those just below (hazard ratio: 0.79 (95% CI 0.69–0.90)). Few individuals received early metformin treatment. However, the chance of metformin treatment initiation within 3 months was substantially higher for individuals with an HbA_{1c} measurement above (14%) than below (1%) the threshold.

Conclusion: Individuals with first record of HbA_{1c} measure just above treatment threshold experienced a 21% lower rate of death or CVE than those just below. Lifestyle modifications and cardiovascular risk-factor management may contribute to this reduced rate.

Keywords: type 2 diabetes, glycated hemoglobin A1c, regression discontinuity design, cardiovascular event, mortality

Introduction

Individuals with type 2 diabetes have a twofold increased risk for cardiovascular events (CVE), and CVE is responsible for nearly 80% of the mortality in type 2 diabetes patients.[1,2] Current international treatment guidelines state that patients with type 2 diabetes should immediately initiate lifestyle modifications and be treated with metformin therapy (if tolerated) and cardiovascular risk-factor management from the time of first type 2 diabetes diagnosis, which since 2012 has been based on either plasma glucose or hemoglobin A1c (HbA_{1c}) criteria.[3,4] The rationale of early glucose-lowering

Correspondence: Irene Petersen
Department of Primary Care and Population Health, University College London, London NW3 2PF, UK
Tel +44 20 801 68032
Email i.petersen@ucl.ac.uk

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/CLEP.S251704

intervention is to hinder later micro- and macrovascular complications and mortality associated with hyperglycemia.[5] However, randomized trial evidence for the effectiveness of the early aggressive intervention in newly diagnosed type 2 diabetes patients on CVE and death is sparse.[2,6] The evidence for early metformin treatment mainly stems from a small subgroup of overweight patients in the UK Prediction Diabetes Study, in which metformin-treated patients had a 30% reduction of macrovascular complications and 36% reduced all-cause mortality.[7] Likewise, the evidence for multifactorial risk-factor management mainly stems from patients with long-standing type 2 diabetes.[8] Additional randomized controlled trials (RCTs) providing better evidence for initial type 2 diabetes treatment are unlikely to be done. To gain further insight into the effectiveness of early treatment in newly diagnosed type 2 diabetes patients, we rely on evidence based on data from clinical practice.[6] However, in clinical practice, it is often a challenge to identify patient groups that are directly comparable as in an RCT. We, therefore, used a novel method in observational medical research, the regression discontinuity design (RDD) which is a quasi-experimental design comparing rates of an event in people just below and people just above a specific treatment threshold.[9–11] The idea of the RDD, in the context of this study, is that individuals who are just below or just above the threshold for being eligible for type 2 diabetes treatment, ie, HbA_{1c} ≥6.5% (≥48 mmol/mol), are in reality very similar. Therefore, we used the RDD to estimate the effect of type 2 diabetes treatment eligibility on all-cause mortality and CVE.

Methods

Data Sources

We used population-based registry data from the Central Denmark Region and the North Denmark Region. The population is around 1.8 million, which corresponds to approximately 30% of the total Danish population. The registries include data from the clinical laboratory information system (LABKA), the Danish National Patient Registry, the Aarhus University Prescription Database (AUPD) and the Danish Civil Registration System. All data sources can be linked via a personal identifier, the CPR number.

LABKA includes biochemistry data from every blood sample analyzed in clinical chemistry department laboratories in the two regions. Data have been kept in the computerized laboratory databases since 1997. The databases cover data from hospitals as well as general practitioners and include results from HbA_{1c} measurements.[12] Every measurement is recorded according to the Nomenclature for Properties and Units (NPU) clinical laboratory terminology and includes information on the date of analysis, the NPU code, the component, the result and unit of analysis.

The Danish National Patient Registry includes information on outcomes and comorbidities.[13] This registry contains data on dates of admission and discharge from all Danish non-psychiatric hospitals since 1977 and from emergency room and outpatient clinic visits since 1995. Each hospital discharge or outpatient visit is recorded in the registry with one primary diagnosis and potentially several secondary diagnoses classified according to the *International Classification of Diseases* (ICD-10).

Since 1998, the AUPD has recorded data on reimbursed prescriptions dispensed at all community pharmacies of the Central Denmark Region and the North Denmark Region.[14] Pharmacies are equipped with electronic accounting systems primarily used to secure reimbursement from the National Health Service. The registry includes prescriber identifier, the product number, number of tablets and packages dispensed, ATC code, volume, package name, active substance, dosage, and date of dispensing.

The Danish Civil Registration System has recorded all changes in vital status and migration for the entire Danish population since 1968, with daily electronic updates.[15]

The Regression Discontinuity Design

The RDD has been widely used since the 1960s in econometrics, social sciences and politics,[16–18] but it has rarely been applied in medical and epidemiological research.[9–11] The design relies on the assumption that the threshold acts as a randomizing device for individuals close to the threshold, ie, those *just* below and those *just* above the threshold are similar. In our study, we used the RDD to examine the effect of being eligible for type 2 diabetes treatment as determined by the threshold rule HbA_{1c} ≥6.5% (≥48 mmol/mol) on all-cause mortality and CVE. According to official clinical guidelines, an individual is assigned to a treatment group (treatment vs no treatment) based on an HbA_{1c} measurement under the influence of measurement errors and various other factors that can potentially result in small changes from the actual to the measured value. Given these circumstances, one can ask if an individual whose HbA_{1c} was measured to be 47 mmol/mol is much different from an individual whose value was measured to be 48 mmol/mol? Treatment guidelines, however, will treat these two individuals very differently, which is what

the RDD utilizes as the pseudo-random treatment allocation. As the threshold only acts as a randomizing device for individuals close to the threshold, the RDD estimates the *local* effect of being eligible for treatment. It is a balancing act to only include patients close to the threshold, but also to include enough patients to ensure statistical power and to include data points for the regression model.

If the threshold is strictly adhered to when assigning treatment (treatment probability is either 0 (below) or 1 (above the threshold)), the RDD is termed *sharp* and we can easily estimate a causal effect of treatment. In most situations, however, this is not the case, ie, there are individuals above the threshold who are untreated and vice versa and the design is termed *fuzzy*.[9,11] In a fuzzy design, a simple RDD analysis is equivalent to an intention-to-treat analysis.[9,11]

For the RDD to provide valid estimates, the decision rule for treatment and cut-off value should be known.[11] In the context of this study, from 2012 and onwards, patients have been eligible for type 2 diabetes treatment, be it initiation of lifestyle modification and metformin treatment, if they had an HbA_{1c} of ≥6.5% (≥48 mmol/mol). Before 2012, a type 2 diabetes diagnosis was based on plasma glucose criteria rather than on any measured HbA_{1c} values per se. However, at least, since 2006, American Diabetes Association/European Association for the Study of Diabetes consensus recommendations have stipulated initiation of both metformin and lifestyle intervention immediately after type 2 diabetes diagnosis, and to aim for an HbA_{1c} level of <6.5% (<48 mmol/mol) in most people with newly diagnosed type 2 diabetes.[19]

Study Population

For this study, we included individuals aged 40–80 years with a first record of HbA_{1c} measurement in the range of 5.8–7.3% (40–56 mmol/mol) in the period during 2006–2014 and who had lived in the region for at least 1 year prior to the time of the measurement. As we detail below, our primary analyses were focused on those with HbA_{1c} measurements values within 6.0–7.0% (42–53 mmol/mol).

Follow-up started at the time of the initial HbA_{1c} measurement and everyone was followed until the first occurrence of CVE, death, or end of the study period (December 2014). We included information on date of birth, gender, information on hospitalization and hospital diagnoses prior to the first record of HbA_{1c} measurement, and dispensed prescriptions, including antibiotics, antihypertensives, statins and antidepressants, within 1 year prior to the first record of HbA_{1c} measurement. We excluded individuals who were lost to follow-up during the study period. Patients with a previous hospital diagnosis of type 1 diabetes or type 2 diabetes and individuals who were treated with metformin or other glucose-lowering drugs prior to the initial HbA_{1c} measurement were also excluded. In addition, we excluded individuals who experienced CVE before the initial HbA_{1c} measurement.

Outcomes

Our primary outcome was a composite of all-cause mortality and CVE (myocardial infarction or stroke). All-cause mortality and CVE were also evaluated separately. Myocardial infarction and stroke were defined as inpatient or outpatient first-time diagnoses in the Danish National Patient Registry (see Table S1 supplemental data for codes), and all-cause mortality was obtained from the Danish Civil Registration System.

Analyses

We performed a number of descriptive analyses and plots to ascertain that our data met the basic assumptions for the RDD as described by Moscoe et al.[11] 1) By plotting frequency and values of the HbA_{1c} measurements in a histogram, we assessed whether HbA_{1c} was measured and reported continuously. This is to ensure that neither patients nor doctors have been able to deliberately change the value of the measurement and hereby change treatment allocation group. 2) We tabulated the distribution of covariates to show that they were not discontinuous at the threshold and to confirm that individuals whose measurement was just below the threshold for treatment initiation and individuals just above were similar (exchangeable) at the time of the initial measurement, as is also the case in RCTs.

In our primary analyses, our focus was on those with HbA_{1c} values within 6.0–7.0% (42–53 mmol/mol) with pre-specified guideline threshold of 6.5% (48 mmol/mol). However, we also report results from the broader and the narrower ranges to evaluate the sensitivity of the results in the choice of the bandwidth. We also stratified our analyses on sex, age and calendar time (early (2006–2011) and late (2012–2014)), and finally, we adjusted our primary analyses for diuretics, antihypertensives and year.

In order to ensure our findings, using the pre-specified guideline threshold of 6.5% (48 mmol/mol), were not just due to chance, we conducted supportive analyses. We sought to evaluate whether we would observe a jump

(discontinuity) for other values of HbA_{1c} and therefore repeated our analysis for a threshold of 6.0% (42 mmol/mol) and 7.0% (53 mmol/mol), respectively.

We examined the effect of diabetes treatment *eligibility* on all-cause mortality and CVE in a Cox regression model with the following parameterization of the hazard function h:

$$\log(h(Y|Z)) = \beta_0 + \beta_1(Z-c) + \beta_2 \mathbf{1}_{(Z<c)} + \beta_3(Z-c)\mathbf{1}_{(Z\geq c)}$$

where Z denotes the vector of HbA_{1c} values, Y the vector containing the outcomes, c the threshold (6.5% (48 mmol/mol)), β_1 is the trend of the line below the threshold, $\beta_1 + \beta_3$ is the trend of the line above the threshold, and β_2 is the difference at the threshold, ie, the estimator of the intention-to-treat. We inspected the plot of the fitted hazard function from this model against the initial HbA_{1c} measurement to visually confirm the discontinuity at the threshold. A visible jump at the threshold indicates a non-zero treatment effect, whereas continuity at the threshold indicates a null effect, ie, $\beta_2 = 0$. The local effects were estimated for the three above-mentioned HbA_{1c} ranges.

All analyses were conducted using SAS version 9.4 (SAS Institute Inc., Cary, NC).

The study was approved by the Danish Data Protection Agency (record number: KEA-2015-4).

Results

In total, 717,449 individuals in the Central Denmark Region and the North Denmark Region had a first record of HbA_{1c} in the period between 2006 and 2014. Of the total population, 290,333 met the inclusion criteria. HbA_{1c} was measured and reported continuously and there was no evidence that patients nor doctors have changed treatment allocation around the threshold (Figure 1). Among these, 43,070 had a first record of HbA_{1c} within the range of 6.0–7.0% (42–53 mmol/mol), 36,360 (84%) individuals had an HbA_{1c} below the recommended threshold for treatment, 6.0–6.4% (42–47 mmol/mol) and 6710 (16%) had an HbA_{1c} above the threshold, 6.5–7.0% (48–53 mmol/mol) (Table 1). The median age was similar among individuals below and above the threshold (63 years), but the sex ratio differed slightly with 52% women below the threshold and 48% women above the threshold (Table 1). The standardized differences indicate a high degree of comparability between individuals below and above the threshold on prescribing of statins, NSAIDS, glucocorticoids, antidepressants, antibiotics and antiplatelets (Table 1). However, there were some differences for individuals receiving diuretic and antihypertensive treatment (Table 1).

The differences were larger when choosing a broader range of HbA_{1c} and became smaller when choosing a narrower range of HbA_{1c} (Table S2 supplemental data).

These patterns were similar when we stratified the sample on men and women, calendar periods and those aged below and above 60 years (Tables S3–S8 supplemental data).

The median follow-up for all-cause mortality or CVE was 4.06 (IQR 2.18–6.69) years among those with an HbA_{1c} of 6.0–6.4% (42–47 mmol/mol), and 5.18 (IQR 2.61–7.33) years for those with an HbA_{1c} of 6.5–7.0%

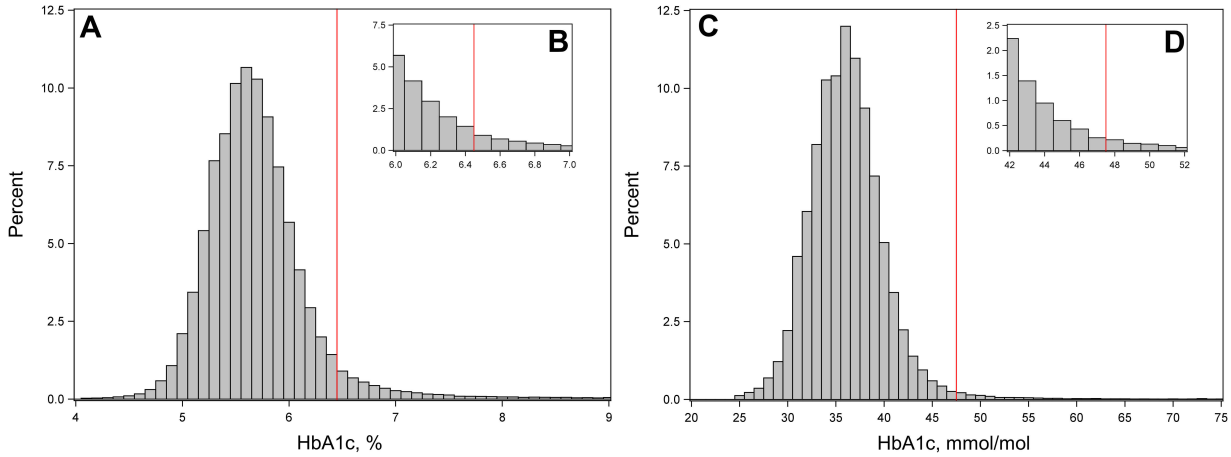


Figure 1 Hemoglobin A1c (HbA_{1c}) test results among 525,266 individuals in Central and Northern Denmark with an incident HbA_{1c} test during 2006–2012 (**A** and **B**, HbA_{1c} in %) and 2013–2014 (**C** and **D**, HbA_{1c} in mmol/mol). Histograms show continuity of HbA_{1c} test results around the threshold of 6.5% (48 mmol/mol) (vertical red lines).

Table 1 Characteristics of Individuals Within HbA_{1c} Ranges Just Below versus Just Above the 6.5% (48 mmol/mol) Threshold for Treatment Initiation, 2006–2014

	HbA_{1c} Range		
	6.0–6.4% (42–47 mmol/mol)	**6.5–7.0% (48–53 mmol/mol)**	**Standardized Difference**
Total	36,360	6710	
Treated with metformin within 3 months	484 (1.3%)	921 (13.7%)	−0.48
Female	18,916 (52.0%)	3221 (48.0%)	0.08
Median age at diagnosis (Q1;Q3)	62.9 (55.1;70.2)	63.0 (54.9;70.5)	−0.02
Comorbidities			
Chronic pulmonary disease[a]	1634 (4.5%)	422 (6.3%)	−0.08
Moderate to severe renal disease[a]	243 (0.7%)	61 (0.9%)	−0.03
Obesity[b]	1246 (3.4%)	332 (4.9%)	−0.08
Any cancer[a]	1847 (5.1%)	393 (5.9%)	−0.03
Any liver disease[a]	163 (0.4%)	54 (0.8%)	−0.05
Charlson comorbidity index score[c]			
0	30,522 (83.9%)	5372 (80.1%)	0.10
1–2	4988 (13.7%)	1112 (16.6%)	−0.08
3+	850 (2.3%)	226 (3.4%)	−0.06
Prescriptions[d]			
Statins	8847 (24.3%)	1660 (24.7%)	−0.01
NSAIDS	9618 (26.5%)	1915 (28.5%)	−0.05
Glucocorticoids	2950 (8.1%)	716 (10.7%)	−0.09
Diuretics	8844 (24.3%)	2144 (32.0%)	−0.17
Antidepressants	5135 (14.1%)	1071 (16.0%)	−0.05
Antibiotics	12,407 (34.1%)	2448 (36.5%)	−0.05
Any antihypertensives	16,981 (46.7%)	3608 (53.8%)	−0.14
Antiplatelets	6769 (18.6%)	1471 (21.9%)	−0.08

Notes: [a]Last 5 years prior to initial measurement. [b]Last 10 years prior to initial measurement. [c]Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score. [d]Last year prior to the initial measurement
Abbreviation: HbA_{1c}, hemoglobin A1c.

(48–53 mmol/mol), event rates were, in general, more frequent among individuals with a first record of HbA_{1c} measurement above the threshold than among those below the threshold. Thus, there were 27.5 (95% CI 26.7–28.3) events per 1000 person-years in those with an HbA_{1c} of 6.0–6.4% (42–47 mmol/mol) and 37.2 (95% CI 35.1–39.3) events per 1000 person-years in those with an HbA_{1c} of 6.5–7.0% (48–53 mmol/mol). However, when looking only at the threshold, there was a discontinuity (Figure 2) and individuals with a first record of HbA_{1c} measurement just above the threshold show a 21% lower rate of death or CVE than those just below the threshold (Hazard ratio (HR) 0.79 (95% CI 0.69–0.90)) (Table 2). The estimates were similar after adjustment for diuretic, antihypertensive treatment and year (Table 2).

Narrowing the overall range of HbA_{1c} to 6.2–6.8% (44–51 mmol/mol) provided similar results, but reduced the precision of the estimates as the sample was smaller (Table S9 supplemental data).

The analyses stratified on calendar time provided similar estimates as the overall results (Table 2). The analyses stratified on men and women suggested a larger positive effect for women (HR 0.73 (0.60–0.89)) than for men (HR 0.83 (0.69–0.99)) of being just above the threshold for type 2 diabetes treatment (Table 2). The analyses stratified on age below and above 60 years suggested that there was no effect in those below the age of 60, but a 22% reduction in those aged above 60 (HR 0.78 (0.68–0.91)) (Table 2).

Sensitivity Analyses

In our first sensitivity analysis, we limited the sample and only followed individuals from when they had HbA_{1c} measurements above 5.5% (37 mmol/mol) and repeated our analyses. The rate of death or CVE was similar to our original results when comparing those with an HbA_{1c} of

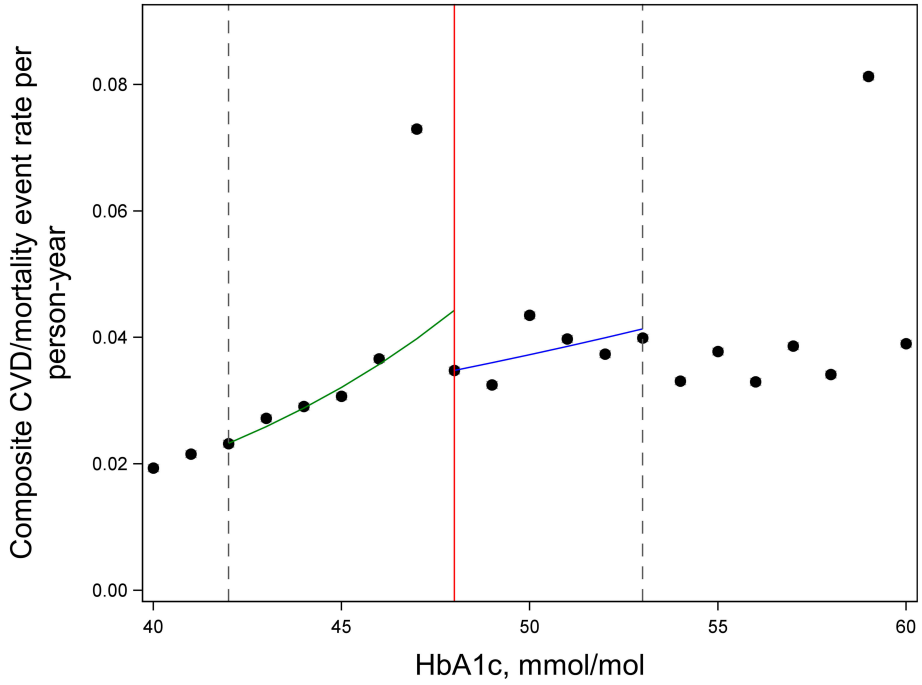


Figure 2 Rates of cardiovascular events or death according to first record of hemoglobin A1c (HbA_{1c}) measurement. HbA_{1c} threshold of 6.5% (48 mmol/mol) is shown by the red line. The range used for the primary analysis (6.0% (42 mmol/mol) to 7.0% (53 mmol/mol)) is indicated by dashed lines. The figure shows the regression discontinuity above the threshold (blue line) and below the threshold (green line).

6.0–6.4% (42–47 mmol/mol) and 6.5–7.0% (48–53 mmol/mol) (HR 0.81 (95% CI 0.71–0.92), N=44,846). As Figure 2 revealed that individuals with an HbA_{1c} of 6.4% (47 mmol/mol) had high rates of the outcome, we conducted an analysis excluding individuals (n=266) with an HbA_{1c} of 6.4% (47 mmol/mol) and found this had minimal impact on the overall results.

When moving the threshold to 6.0% (42 mmol/mol), we observed no discontinuity when comparing those just below and those above (HR 1.02 (95% CI 0.95; 1.09)) (Tables S10–S11 supplemental data) When moving the threshold to 7.0% (53 mmol/mol), we saw a slight discontinuity (HR 0.86 (0.70; 1.06)) (Tables S12–S13 supplemental data).

Metformin

Since guidelines recommend initiation of metformin treatment when the HbA_{1c} value exceeds 6.5% (48 mmol/mol), we evaluated the treatment threshold rule by plotting the initial HbA_{1c} measurement against the probability of

Table 2 Hazard Ratios of All-Cause Mortality, Cardiovascular Events and Composite Endpoint Associated with HbA_{1c} ≥6.5% (≥48 mmol/mol), Overall and Stratified Analysis. The Reference Group Is HbA_{1c} <6.5% (<48 mmol/mol)

Strata	All-Cause Mortality HR (95% CI)	Cardiovascular Event HR (95% CI)	All-Cause Mortality or Cardiovascular Event HR (95% CI)	Sample Size
All	0.73 (0.63; 0.85)	0.95 (0.76; 1.20)	0.79 (0.69; 0.90)	43,070
All, adjusted[a]	0.74 (0.64; 0.86)	0.97 (0.77; 1.22)	0.80 (0.70; 0.91)	43,070
2006–2011	0.76 (0.65; 0.89)	0.97 (0.75; 1.24)	0.81 (0.70; 0.93)	29,664
2012–2014	0.66 (0.43; 1.02)	0.93 (0.48; 1.77)	0.78 (0.54; 1.12)	13,406
Women	0.68 (0.55; 0.85)	0.79 (0.55; 1.15)	0.73 (0.60; 0.89)	22,137
Men	0.77 (0.62; 0.94)	1.08 (0.80; 1.45)	0.83 (0.69; 0.99)	20,933
Age <60y	0.72 (0.48; 1.06)	1.52 (0.95; 2.43)	0.99 (0.73; 1.35)	17,074
Age ≥60y	0.77 (0.66; 0.91)	0.86 (0.65; 1.12)	0.78 (0.68; 0.91)	25,996

Notes: [a]Adjusted for diuretic, antihypertensive treatments and year.
Abbreviation: HbA_{1c}, hemoglobin A1c.

metformin treatment initiation within 1 and 3 months, respectively, of the date of the HbA_{1c} measurement (Figure S1 supplemental data). The probability of early metformin treatment initiation was higher for patients above the threshold. However, only 921 individuals (14%) above the threshold (6.5–7.0% (48–53 mmol/mol)) and 484 individuals (1%) below the threshold (6.0–6.4% (42–47 mmol/mol)) initiated metformin treatment within 3 months after their initial HbA_{1c} record.

Discussion

In this study, we demonstrated the impact of being eligible for type 2 diabetes treatment on all-cause mortality and CVE in everyday clinical practice. We found that individuals just above the treatment threshold had a 21% lower rate of death or CVE than those just below the threshold (HR 0.79 (95% CI 0.69–0.90)). The positive effect of being eligible for treatment was slightly larger for women (HR 0.73 (0.60–0.89)) than for men (HR 0.83 (0.69–0.99)), but no effect was observed in those aged below 60 years. Relatively few individuals (14%) received metformin treatment within 3 months of their first record of HbA_{1c} measurement above the threshold, and the beneficial effect associated with being just above versus just below the treatment threshold is thus unlikely to be mediated by metformin treatment initiation alone.

For many years, type 2 diabetes guidelines have emphasized immediate lifestyle modifications, including physical activity, healthy diet and weight loss, when type 2 diabetes is first diagnosed, with metformin therapy (if tolerated) as a first-line glucose-lowering pharmacological treatment in most individuals started immediately at or soon after diagnosis. Usually, this is combined with aggressive cardiovascular risk-factor management against hypertension, dyslipidemia, and platelet aggregation.[3,5] While such multifactorial intervention has been shown to reduce CVE and mortality in patients with long-standing type 2 diabetes,[8] the effectiveness of the early intervention on subsequent cardiovascular outcomes in newly diagnosed type 2 diabetes patients has been less clear.[2,6]

It is well known that glucose levels under the current diagnostic threshold (ie, prediabetes, corresponding to HbA_{1c} 6.0–6.4% (42–47 mmol/mol)) are also associated with increased risk of CVE.[4] This is likely related to pathophysiological processes accompanying type 2 diabetes that have often been present for several years before diagnosis of manifest type 2 diabetes.[20,21] Vistisen et al recently showed that CVE and mortality rates were 37% to 54% increased with prediabetes (HbA_{1c} 6.0–6.4% or 5.7–6.4%) versus normoglycemia, and that two-thirds of this excess risk were explained by clustering of cardiometabolic risk factors such as smoking, lipids, and high blood pressure in people with prediabetes.[22] This supports our suggestions that multifactorial type 2 diabetes treatment in people reaching an HbA_{1c} of just above 6.5% (48 mmol/mol) may indeed lower the risk of CVE and death in this population. Our finding that being eligible for treatment had a slightly larger positive effect among women than men is in contrast to previous research on sex differences in type 2 diabetes.[23,24] Thus, Arnezt et al concluded that, in general, women with type 2 diabetes have higher morbidity and cardiovascular mortality than men.[23] Yet, they also emphasized a need for further research designed specifically to evaluate sex differences in the effectiveness and outcome of the available treatments.[23]

Clinical Relevance

Our study suggests that early treatment in newly diagnosed type 2 diabetes substantially reduces the risk of cardiovascular events and all-cause mortality. Since only one in seven individuals started metformin soon after becoming eligible for type 2 diabetes treatment, the prognostic benefits from "crossing the line" from prediabetes to HbA1c-defined diabetes cannot be attributed to the initiation of metformin alone, but rather multifactorial interventions. Our study thus supports current guidelines of prompt treatment intervention from the time of first type 2 diabetes diagnosis, with lifestyle counselling, aggressive cardiovascular risk-factor management, and glucose-lowering therapy. Our results also raise the possibility that individuals with prediabetes, ie, those who have an HbA1c just under the diabetes diagnostic threshold, currently receive too little clinical attention. Finally, reaching the threshold for being eligible for treatment may also have an impact on individual susceptibility to treatment and lifestyle modifications. Thus, there may be a "gosh my HbA_{1c} is too high" effect leading to the realization that changes are needed to improve health.

Strengths and Limitations

To our knowledge, this is the first time that the RDD has been applied to Danish healthcare data. One of the key strength of using data from a tax-supported, universal healthcare system is that the study is not subject to selection biases stemming from selective inclusion of specific hospital, health insurance systems, or age groups.

In this study, we examined the effect of diabetes treatment eligibility. We also considered to evaluate the causal effect of metformin treatment initiation on all-cause mortality and CVE. To estimate this effect, it is necessary to scale the effect

of diabetes treatment eligibility by the difference in the probability of metformin treatment at the threshold. However, since only one in seven of the individuals above the treatment threshold initiated metformin treatment within 3 months, we were not able to provide reliable estimates of the causal effect of metformin treatment on all-cause death and CVE. In addition, as glucose-lowering pharmacological treatment may often be combined with other cardiovascular risk-factor management,[3–5] we would not have been able to distinguish the effect of each of these components.

Our supportive analysis confirmed that there was no discontinuity in the risk of mortality or CVE when we selected thresholds of 6% (42 mmol/mol). This may help to reassure us that the observed discontinuity at 6.5% (≥48 mmol/mol) is due to clinical intervention and not just due to chance. On the other hand, our results suggest that there might also be a slight discontinuity at 7% (53 mmol/mol). Perhaps due to the fact that the 7% (53 mmol/mol) threshold has been an important, pragmatic glucose-lowering treatment target for type 2 diabetes patients in the earlier part of our study period, this threshold may also have served as a pragmatic cutoff value among many clinicians to initiate more intensive diabetes therapy.

If the test results are near the threshold for diabetes diagnosis, guidelines recommend the health care professionals to follow the patient closely and repeat the test in 3–6 months either by HbA_{1c} or alternative tests (an abnormal fasting plasma glucose or 2-hrs plasma glucose) to confirm the diabetes diagnosis.[4] However, using repeated measurements to verify a patient's diabetes status will violate the RDD assumption of random allocation of measurements around the diagnostic threshold. We, therefore, had to rely on one HbA_{1c} test in our analysis. Given the chronic progressive nature of prediabetes/type 2 diabetes, it is likely that few individuals in our study who initially were below the HbA_{1c} threshold would "cross the line" shortly after their first measurement and become eligible for treatment. Thus, our study may underestimate the actual effects of being eligible for type 2 diabetes treatment, but not changing our conclusions.

Conclusion

Individuals eligible for type 2 diabetes treatment as determined by the diagnostic threshold rule HbA_{1c} ≥6.5% (≥48 mmol/mol) experienced a 21% lower rate of death or CVE compared to those just below the threshold. Less than one in seven patients initiated metformin treatment within 3 months. Therefore, other factors, including lifestyle modifications and more aggressive cardiovascular risk-factor control following the recording of an HbA_{1c} ≥6.5% (≥48 mmol/mol), may contribute to this reduced rate of death and CVE.

Disclosure

IP, GB and FR received funding from the UK Medical Research Council for research to develop the Regression Discontinuity Design (Grant number MR/K014838/1). IP has also received funding from Novo Nordisk, but not in relation to this project. SKN, LP, and RWT did not report receiving fees, honoraria, grants or consultancies. Department of Clinical Epidemiology is, however, involved in studies with funding from various companies as research grants to (and administered by) Aarhus University. Department of Clinical Epidemiology is a member of the Danish Centre for Strategic Research in Type 2 Diabetes (DD2), supported by the Danish Agency for Science (grant nos. 09-067009 and 09-075724), the Danish Health and Medicines Authority, the Danish Diabetes Association, and an unrestricted donation from Novo Nordisk A/S. Project partners are listed on the website www.DD2.nu. Department of Clinical Epidemiology participates in the International Diabetic Neuropathy Consortium (IDNC) research program, which is supported by a Novo Nordisk Foundation Challenge program grant (Grant number NNF14SA000 6). The authors report no other conflicts of interest in this work.

References

1. Fox CS, Golden SH, Anderson C, et al. Update on prevention of cardiovascular disease in adults with type 2 diabetes mellitus in light of recent evidence: a scientific statement from the American Heart Association and the American Diabetes Association. *Diabetes Care*. 2015;38(9):1777–1803. doi:10.2337/dci15-0012
2. Abdul-Ghani M, DeFronzo RA, Del Prato S, Chilton R, Singh R, Ryder REJ. Cardiovascular disease and type 2 diabetes: has the dawn of a new era arrived? *Diabetes Care*. 2017;40(7):813–820. doi:10.2337/dc16-2736
3. National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. Guidance and guidelines. NICE. Available from: https://www.nice.org.uk/guidance/ng28. Accessed July 13, 2017.
4. Standards of Medical Care in Diabetes. American Diabetes Association. Available from: https://professional.diabetes.org/content-page/standards-medical-care-diabetes. Accessed March 17, 2018.
5. Chamberlain JJ, Rhinehart AS, Shaefer CF, Neuman A. Diagnosis and management of diabetes: synopsis of the 2016 American Diabetes Association Standards of Medical Care in Diabetes. *Ann Intern Med*. 2016;164(8):542–552. doi:10.7326/M15-3016
6. Svensson E, Baggesen LM, Johnsen SP, et al. Early glycemic control and magnitude of HbA1c reduction predict cardiovascular events and mortality: population-based cohort study of 24,752 metformin initiators. *Diabetes Care*. 2017;40(6):800–807. doi:10.2337/dc16-2271
7. King P, Peacock I, Donnelly R. The UK Prospective Diabetes Study (UKPDS): clinical and therapeutic implications for type 2 diabetes. *Br J Clin Pharmacol*. 1999;48(5):643–648. doi:10.1046/j.1365-2125.1999.00092.x

8. Gaede P, Lund-Andersen H, Parving -H-H, Pedersen O. Effect of a multifactorial intervention on mortality in type 2 diabetes. *N Engl J Med*. 2008;358(6):580–591. doi:10.1056/NEJMoa0706245
9. O'Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ*. 2014;349:g5293. doi:10.1136/bmj.g5293
10. Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology*. 2014;25(5):729–737. doi:10.1097/EDE.0000000000000138
11. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *J Clin Epidemiol*. 2015;68(2):132–143. doi:10.1016/j.jclinepi.2014.06.021
12. Grann ER, Nielsen F, Thomsen R. Existing data sources for clinical epidemiology: the clinical laboratory information system (LABKA) research database at Aarhus University, Denmark. *Clin Epidemiol*. 2011;3:133–138. doi:10.2147/CLEP.S17901
13. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449–490. doi:10.2147/CLEP.S91125
14. Ehrenstein V. Existing data sources for clinical epidemiology: aarhus university prescription database. *Clin Epidemiol*. 2010;2:273–279. doi:10.2147/CLEP.S13458
15. Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol*. 2014;29(8):541–549. doi:10.1007/s10654-014-9930-3
16. Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J Educ Psychol*. 1960;51(6):309–317. doi:10.1037/h0044319
17. Lee DS. Randomized experiments from non-random selection in U.S. House elections. *J Econom*. 2008;142(2):675–697. doi:10.1016/j.jeconom.2007.05.004
18. Imbens G, Lemieux T. Regression discontinuity designs: a guide to practice. *J Econom*. 2008;142(2):615–635. doi:10.1016/j.jeconom.2007.05.001
19. Nathan DM, Buse JB, Davidson MB, et al. Management of hyperglycaemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement from the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetologia*. 2006;49(8):1711–1721. doi:10.1007/s00125-006-0316-2
20. Hulman A, Simmons RK, Brunner EJ, et al. Trajectories of glycaemia, insulin sensitivity and insulin secretion in South Asian and white individuals before diagnosis of type 2 diabetes: a longitudinal analysis from the Whitehall II cohort study. *Diabetologia*. 2017;60(7):1252–1260. doi:10.1007/s00125-017-4275-6
21. Thrasher J. Pharmacologic management of type 2 diabetes mellitus: available therapies. *Am J Med*. 2017;130(6S):S4–S17. doi:10.1016/j.amjmed.2017.04.004
22. Vistisen D, Witte DR, Brunner EJ, et al. Risk of cardiovascular disease and death in individuals with prediabetes defined by different criteria: the Whitehall II Study. *Diabetes Care*. 2018;41(4):899–906. doi:10.2337/dc17-2530
23. Arnetz L, Ekberg NR, Alvarsson M. Sex differences in type 2 diabetes: focus on disease course and outcomes. *Diabetes Metab Syndr Obes Targets Ther*. 2014;7:409–420. doi:10.2147/DMSO.S51301
24. Roche MM, Wang PP. Sex differences in all-cause and cardiovascular mortality, hospitalization for individuals with and without diabetes, and patients with diabetes diagnosed early and late. *Diabetes Care*. 2013;36(9):2582–2590. doi:10.2337/dc12-1272

Supplemental information

Table S1 includes ICD and ATC codes used to identify conditions and treatments.

Table S2 includes characteristics of individuals from the broad range, ie, HbA_{1c} 5.8–7.3% (40–56 mmol/mol) and the narrow range, ie, HbA_{1c} 6.2–6.8% (44–51 mmol/mol)

Tables S3 and S4 include characteristics of individuals stratified by calendar year (S3: 2006–2011, S4: 2012–2014).

Tables S5 and S6 include characteristics of individuals stratified by gender (S5: Women, S6: Men)

Tables S7 and S8 include characteristics of individuals stratified by age (S7: Below 60 years, S8: above 60 years)

Table S9 includes hazard ratios of all-cause mortality, cardiovascular events and composite endpoint associated with HbA_{1c} for the broad range, ie, HbA_{1c} 5.8–7.3% (40–56 mmol/mol) and the narrow range, ie, HbA_{1c} 6.2–6.8% (44–51 mmol/mol).

Table S10 includes characteristics of individuals with HbA_{1c} ranges just below versus just above a 6.0% (42 mmol/mol) threshold for treatment initiation, 2006–2011.

Table S11 includes hazard ratios of all-cause mortality, cardiovascular events and composite endpoint associated with HbA_{1c} for patients with HbA_{1c} just above versus just below a 6.0% (42 mmol/mol) threshold for treatment initiation.

Table S12 includes characteristics of individuals with HbA_{1c} ranges just below versus just above a 7.0% (53 mmol/mol) threshold for treatment initiation, 2006–2011

Table S13 includes hazard ratios of all-cause mortality, cardiovascular events and composite endpoint associated with HbA_{1c} for patients with HbA_{1c} just above versus just below a 7.0% (53 mmol/mol) threshold for treatment initiation.

Figure S1 shows cumulative incidences of metformin treatment initiation 1 and 3 months after the first HbA_{1c} measurement.

Table S1 ICD and ATC codes used to identify conditions and treatments.

Variable	*ICD-8*	*ICD-10*
Chronic pulmonary disease	490–493, 515–518	J40–J47, J60–J67, J68.4, J70.1, J70.3, J84.1, J92.0, J96.1, J98.2, J98.3
Any liver disease	571, 573.01, 573.04, 070.00, 070.02, 070.04, 070.06, 070.08, 573.00, 456.00–456.09	B18, K70.0–K70.3, K70.9, K71, K73, K74, K76.0, B15.0, B16.0, B16.2, B19.0, K70.4, K72, K76.6, I85
Diabetes type1	249.00, 249.06, 249.07, 249.09	E10.0, E10.1, E10.9
Diabetes type2	250.00, 250.06, 250.07, 250.09	E11.0, E11.1, E11.9
Moderate to severe renal disease	403, 404, 580–583, 584, 590.09, 593.19, 753.10–753.19, 792	I12, I13, N00–N05, N07, N11, N14, N17–N19, Q61
Diabetes with end organ damage type1	249.01–249.05, 249.08	E10.2–E10.8
type2	250.01–250.05, 250.08	E11.2–E11.8
Any cancer	140–194, 204–207, 200–203, 275.59, 195–198, 199	C00–C75, C91–C95, C81–C85, C88, C90, C96, C76–C80
Obesity		E65, E66, E67, E68
Cardiovascular event (myocardial infarction or stroke)		I21, I61, I62, I64, I65

Variable	*ATC codes*
Metformin	A10BA, A10BD02, A10BD03, A10BD05, A10BD07, A10BD08, A10BD10, A10BD11, A10BD13, A10BD14, A10BD15
Other antidiabetics	A10 (excluding metformin)
Statins	C10AA, C10B, B04AB
NSAIDS	M01A
Glucocorticoids	H02AB
Diuretics	C03
Antidepressants	N06A
Antibiotics	J01
Any antihypertensives	C02, C03A, C03B, C03D, C03E, C07, C08, C09A, C09B, C09C, C09D, C09X
Antiplatelets	B01

Table S2 Characteristics of individuals from the broad range, ie, HbA_{1c} 5.8–7.3% (40–56 mmol/mol) and the narrow range, ie, HbA_{1c} 6.2–6.8% (44–51 mmol/mol)

	Broad HbA_{1c} range			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	76,031	8,011		14,216	5,416	
Treated with metformin within 3 months	565 (0.7%)	1,333 (16.6%)	-0.59	368 (2.6%)	646 (11.9%)	-0.37
Female	39,837 (52.4%)	3,802 (47.5%)	0.10	7,335 (51.6%)	2,621 (48.4%)	0.06
Median age at diagnosis (Q1;Q3)	62.0 (54.1;69.4)	62.9 (54.6;70.3)	-0.21	63.4 (55.7;70.7)	63.3 (55.1;70.7)	0.07
Comorbidities						
Chronic pulmonary disease[a]	2,916 (3.8%)	497 (6.2%)	-0.11	745 (5.2%)	348 (6.4%)	-0.05
Moderate to severe renal disease[a]	460 (0.6%)	71 (0.9%)	-0.03	100 (0.7%)	46 (0.8%)	-0.02
Obesity[b]	2,239 (2.9%)	398 (5.0%)	-0.10	552 (3.9%)	254 (4.7%)	-0.04
Any cancer[a]	3,601 (4.7%)	474 (5.9%)	-0.05	776 (5.5%)	305 (5.6%)	-0.01
Any liver disease[a]	310 (0.4%)	64 (0.8%)	-0.05	68 (0.5%)	42 (0.8%)	-0.04
Charlson comobidity index score[c]						
0	64,980 (85.5%)	6,437 (80.4%)	0.14	11,706 (82.3%)	4,344 (80.2%)	0.05
1-2	9,532 (12.5%)	1,300 (16.2%)	-0.11	2,120 (14.9%)	902 (16.7%)	-0.05
3+	1,519 (2.0%)	274 (3.4%)	-0.09	390 (2.7%)	170 (3.1%)	-0.02
Prescriptions[d]						
Statins	16,905 (22.2%)	1,942 (24.2%)	-0.05	3,615 (25.4%)	1,369 (25.3%)	0.00
NSAIDS	19,355 (25.5%)	2,323 (29.0%)	-0.08	3,873 (27.2%)	1,533 (28.3%)	-0.02
Glucocorticoids	5,355 (7.0%)	853 (10.6%)	-0.13	1,336 (9.4%)	578 (10.7%)	-0.04
Diuretics	15,797 (20.8%)	2,564 (32.0%)	-0.26	3,907 (27.5%)	1,733 (32.0%)	-0.10
Antidepressants	10,315 (13.6%)	1,259 (15.7%)	-0.06	2,085 (14.7%)	863 (15.9%)	-0.04

Cont.

	Broad HbA_{1c} range			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Antibiotics	24,564 (32.3%)	2,894 (36.1%)	-0.08	5,088 (35.8%)	1,955 (36.1%)	-0.01
Any antihypertensives	32,290 (42.5%)	4,302 (53.7%)	-0.23	7,064 (49.7%)	2,920 (53.9%)	-0.08
Antiplatelets	12,611 (16.6%)	1,740 (21.7%)	-0.13	2,873 (20.2%)	1,208 (22.3%)	-0.05

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S3 Characteristics of individuals within HbA_{1c} ranges just below versus just above the 6.5% (48 mmol/mol) threshold for treatment initiation, 2006–2011.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	47,961	6,371		24,304	5,360		9,838	4,337	
Treated with metformin within 3 months	459 (1.0%)	800 (12.6%)	-0.47	390 (1.6%)	545 (10.2%)	-0.37	286 (2.9%)	383 (8.8%)	-0.25
Female	25,070 (52.3%)	3,065 (48.1%)	0.08	12,627 (52.0%)	2,601 (48.5%)	0.07	5,117 (52.0%)	2,123 (49.0%)	0.06
Median age at diagnosis (Q1;Q3)	61.7 (53.8;69.2)	63.1 (55.0;70.6)	-0.37	62.6 (55.0;70.0)	63.2 (55.3;70.8)	-0.15	63.3 (55.8;70.7)	63.6 (55.5;71.1)	-0.03
Comorbidities									
Chronic pulmonary disease[a]	1,792 (3.7%)	404 (6.3%)	-0.12	1,049 (4.3%)	342 (6.4%)	-0.09	497 (5.1%)	275 (6.3%)	-0.06
Moderate to severe renal disease[a]	300 (0.6%)	57 (0.9%)	-0.03	166 (0.7%)	50 (0.9%)	-0.03	68 (0.7%)	40 (0.9%)	-0.03
Obesity[b]	1,386 (2.9%)	325 (5.1%)	-0.11	803 (3.3%)	271 (5.1%)	-0.09	371 (3.8%)	208 (4.8%)	-0.05
Any cancer[a]	1,977 (4.1%)	348 (5.5%)	-0.06	1,081 (4.4%)	294 (5.5%)	-0.05	469 (4.8%)	227 (5.2%)	-0.02
Any liver disease[a]	206 (0.4%)	46 (0.7%)	-0.04	113 (0.5%)	41 (0.8%)	-0.04	51 (0.5%)	33 (0.8%)	-0.03
Charlson comobidity index score[c]									
0	41,188 (85.9%)	5,116 (80.3%)	0.15	20,524 (84.4%)	4,287 (80.0%)	0.12	8,162 (83.0%)	3,470 (80.0%)	0.08
1-2	5,910 (12.3%)	1,050 (16.5%)	-0.12	3,284 (13.5%)	899 (16.8%)	-0.09	1,440 (14.6%)	738 (17.0%)	-0.07
3+	863 (1.8%)	205 (3.2%)	-0.09	496 (2.0%)	174 (3.2%)	-0.08	236 (2.4%)	129 (3.0%)	-0.04
Prescriptions[d]									
Statins	10,136 (21.1%)	1,630 (25.6%)	-0.11	5,729 (23.6%)	1,392 (26.0%)	-0.06	2,488 (25.3%)	1,148 (26.5%)	-0.03
NSAIDS	13,247 (27.6%)	1,924 (30.2%)	-0.06	6,875 (28.3%)	1,586 (29.6%)	-0.03	2,857 (29.0%)	1,269 (29.3%)	-0.00
Glucocorticoids	3,508 (7.3%)	680 (10.7%)	-0.12	1,958 (8.1%)	581 (10.8%)	-0.10	880 (8.9%)	466 (10.7%)	-0.06

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Diuretics	10,791 (22.5%)	2,164 (34.0%)	-0.26	6,304 (25.9%)	1,810 (33.8%)	-0.17	2,863 (29.1%)	1,468 (33.8%)	-0.10
Antidepressants	6,570 (13.7%)	1,006 (15.8%)	-0.06	3,496 (14.4%)	852 (15.9%)	-0.04	1,477 (15.0%)	685 (15.8%)	-0.02
Antibiotics	15,903 (33.2%)	2,340 (36.7%)	-0.07	8,433 (34.7%)	1,986 (37.1%)	-0.05	3,566 (36.2%)	1,585 (36.5%)	-0.01
Any antihypertensives	20,500 (42.7%)	3,560 (55.9%)	-0.27	11,453 (47.1%)	2,984 (55.7%)	-0.17	4,994 (50.8%)	2,416 (55.7%)	-0.10
Antiplatelets	8,143 (17.0%)	1,465 (23.0%)	-0.15	4,635 (19.1%)	1,235 (23.0%)	-0.10	2,072 (21.1%)	1,019 (23.5%)	-0.06

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S4 Characteristics of individuals within HbA_{1c} ranges just below versus just above the 6.5% (48 mmol/mol) threshold for treatment initiation, 2012–2014.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	28,070	1,640		12,056	1,350		4,378	1,079	
Treated with metformin within 3 months	106 (0.4%)	533 (32.5%)	-0.96	94 (0.8%)	376 (27.9%)	-0.84	82 (1.9%)	263 (24.4%)	-0.71
Female	14,767 (52.6%)	737 (44.9%)	0.15	6,289 (52.2%)	620 (45.9%)	0.13	2,218 (50.7%)	498 (46.2%)	0.09
Median age at diagnosis (Q1;Q3)	62.6 (54.6;69.9)	61.6 (52.8;69.3)	0.26	63.5 (55.4;70.6)	62.1 (53.2;69.5)	0.40	63.7 (55.4;70.9)	61.9 (53.3;69.5)	0.45
Comorbidities									
Chronic pulmonary disease[a]	1,124 (4.0%)	93 (5.7%)	-0.08	585 (4.9%)	80 (5.9%)	-0.05	248 (5.7%)	73 (6.8%)	-0.05
Moderate to severe renal disease[a]	160 (0.6%)	14 (0.9%)	-0.03	77 (0.6%)	11 (0.8%)	-0.02	32 (0.7%)	6 (0.6%)	0.02
Obesity[b]	853 (3.0%)	73 (4.5%)	-0.07	443 (3.7%)	61 (4.5%)	-0.04	181 (4.1%)	46 (4.3%)	-0.01
Any cancer[a]	1,624 (5.8%)	126 (7.7%)	-0.08	766 (6.4%)	99 (7.3%)	-0.04	307 (7.0%)	78 (7.2%)	-0.01
Any liver disease[a]	104 (0.4%)	18 (1.1%)	-0.09	50 (0.4%)	13 (1.0%)	-0.07	17 (0.4%)	9 (0.8%)	-0.06
Charlson comobidity index score[c]									
0	23,792 (84.8%)	1,321 (80.5%)	0.11	9,998 (82.9%)	1,085 (80.4%)	0.07	3,544 (81.0%)	874 (81.0%)	-0.00
1-2	3,622 (12.9%)	250 (15.2%)	-0.07	1,704 (14.1%)	213 (15.8%)	-0.05	680 (15.5%)	164 (15.2%)	0.01
3+	656 (2.3%)	69 (4.2%)	-0.11	354 (2.9%)	52 (3.9%)	-0.05	154 (3.5%)	41 (3.8%)	-0.02
Prescriptions[d]									
Statins	6,769 (24.1%)	312 (19.0%)	0.12	3,118 (25.9%)	268 (19.9%)	0.14	1,127 (25.7%)	221 (20.5%)	0.13
NSAIDS	6,108 (21.8%)	399 (24.3%)	-0.06	2,743 (22.8%)	329 (24.4%)	-0.04	1,016 (23.2%)	264 (24.5%)	-0.03
Glucocorticoids	1,847 (6.6%)	173 (10.5%)	-0.14	992 (8.2%)	135 (10.0%)	-0.06	456 (10.4%)	112 (10.4%)	0.00

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Diuretics	5,006 (17.8%)	400 (24.4%)	-0.16	2,540 (21.1%)	334 (24.7%)	-0.09	1,044 (23.8%)	265 (24.6%)	-0.02
Antidepressants	3,745 (13.3%)	253 (15.4%)	-0.06	1,639 (13.6%)	219 (16.2%)	-0.07	608 (13.9%)	178 (16.5%)	-0.07
Antibiotics	8,661 (30.9%)	554 (33.8%)	-0.06	3,974 (33.0%)	462 (34.2%)	-0.03	1,522 (34.8%)	370 (34.3%)	0.01
Any antihypertensives	11,790 (42.0%)	742 (45.2%)	-0.07	5,528 (45.9%)	624 (46.2%)	-0.01	2,070 (47.3%)	504 (46.7%)	0.01
Antiplatelets	4,468 (15.9%)	275 (16.8%)	-0.02	2,134 (17.7%)	236 (17.5%)	0.01	801 (18.3%)	189 (17.5%)	0.02

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S5 Characteristics of women within HbA_{1c} ranges just below versus just above the 6.5% (48 mmol/mol) threshold for treatment initiation, 2006–2011.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	39,837	3,802		18,916	3,221		7,335	2,621	
Treated with metformin within 3 months	281 (0.7%)	615 (16.2%)	-0.58	235 (1.2%)	435 (13.5%)	-0.48	183 (2.5%)	308 (11.8%)	-0.37
Median age at diagnosis (Q1;Q3)	62.7 (54.9;70.1)	63.6 (55.4;71.2)	-0.19	63.6 (55.8;70.9)	63.8 (55.8;71.3)	-0.02	63.9 (56.3;71.3)	64.2 (55.9;71.5)	0.02
Comorbidities									
Chronic pulmonary disease[a]	1,620 (4.1%)	248 (6.5%)	-0.11	880 (4.7%)	212 (6.6%)	-0.08	418 (5.7%)	176 (6.7%)	-0.04
Moderate to severe renal disease[a]	185 (0.5%)	29 (0.8%)	-0.04	87 (0.5%)	25 (0.8%)	-0.04	36 (0.5%)	17 (0.6%)	-0.02
Obesity[b]	1,619 (4.1%)	247 (6.5%)	-0.11	867 (4.6%)	207 (6.4%)	-0.08	381 (5.2%)	163 (6.2%)	-0.04
Any cancer[a]	1,977 (5.0%)	229 (6.0%)	-0.05	989 (5.2%)	187 (5.8%)	-0.03	401 (5.5%)	147 (5.6%)	-0.01
Any liver disease[a]	133 (0.3%)	24 (0.6%)	-0.04	71 (0.4%)	18 (0.6%)	-0.03	27 (0.4%)	15 (0.6%)	-0.03
Charlson comobidity index score[c]									
0	34,121 (85.7%)	3,069 (80.7%)	0.13	15,952 (84.3%)	2,600 (80.7%)	0.10	6,046 (82.4%)	2,110 (80.5%)	0.05
1-2	4,946 (12.4%)	610 (16.0%)	-0.10	2,551 (13.5%)	519 (16.1%)	-0.07	1,114 (15.2%)	438 (16.7%)	-0.04
3+	770 (1.9%)	123 (3.2%)	-0.08	413 (2.2%)	102 (3.2%)	-0.06	175 (2.4%)	73 (2.8%)	-0.03
Prescriptions[d]									
Statins	9,018 (22.6%)	941 (24.8%)	-0.05	4,727 (25.0%)	808 (25.1%)	-0.00	1,936 (26.4%)	663 (25.3%)	0.03
NSAIDS	10,875 (27.3%)	1,193 (31.4%)	-0.09	5,316 (28.1%)	993 (30.8%)	-0.06	2,113 (28.8%)	808 (30.8%)	-0.04
Glucocorticoids	3,114 (7.8%)	461 (12.1%)	-0.14	1,721 (9.1%)	392 (12.2%)	-0.10	777 (10.6%)	326 (12.4%)	-0.06
Diuretics	10,072 (25.3%)	1,458 (38.3%)	-0.28	5,564 (29.4%)	1,223 (38.0%)	-0.18	2,427 (33.1%)	1,004 (38.3%)	-0.11

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Antidepressants	6,792 (17.0%)	796 (20.9%)	-0.10	3,396 (18.0%)	678 (21.0%)	-0.08	1,385 (18.9%)	560 (21.4%)	-0.06
Antibiotics	14,364 (36.1%)	1,550 (40.8%)	-0.10	7,162 (37.9%)	1,334 (41.4%)	-0.07	2,909 (39.7%)	1,078 (41.1%)	-0.03
Any antihypertensives	17,991 (45.2%)	2,207 (58.0%)	-0.26	9,403 (49.7%)	1,870 (58.1%)	-0.17	3,882 (52.9%)	1,528 (58.3%)	-0.11
Antiplatelets	5,857 (14.7%)	744 (19.6%)	-0.13	3,157 (16.7%)	635 (19.7%)	-0.08	1,364 (18.6%)	530 (20.2%)	-0.04

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S6 Characteristics of men within HbA_{1c} ranges just below versus just above the 6.5% (48 mmol/mol) threshold for treatment initiation, 2006–2011.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	36,194	4,209		17,444	3,489		6,881	2,795	
Treated with metformin within 3 months	284 (0.8%)	718 (17.1%)	-0.60	249 (1.4%)	486 (13.9%)	-0.48	185 (2.7%)	338 (12.1%)	-0.37
Median age at diagnosis (Q1;Q3)	61.2 (53.2;68.6)	62.3 (54.1;69.7)	-0.26	62.2 (54.3;69.5)	62.4 (54.2;69.9)	-0.05	62.9 (55.0;70.1)	62.7 (54.4;70.0)	0.10
Comorbidities									
Chronic pulmonary disease[a]	1,296 (3.6%)	249 (5.9%)	-0.11	754 (4.3%)	210 (6.0%)	-0.08	327 (4.8%)	172 (6.2%)	-0.06
Moderate to severe renal disease[a]	275 (0.8%)	42 (1.0%)	-0.03	156 (0.9%)	36 (1.0%)	-0.01	64 (0.9%)	29 (1.0%)	-0.01
Obesity[b]	620 (1.7%)	151 (3.6%)	-0.12	379 (2.2%)	125 (3.6%)	-0.08	171 (2.5%)	91 (3.3%)	-0.05
Any cancer[a]	1,624 (4.5%)	245 (5.8%)	-0.06	858 (4.9%)	206 (5.9%)	-0.04	375 (5.4%)	158 (5.7%)	-0.01
Any liver disease[a]	177 (0.5%)	40 (1.0%)	-0.05	92 (0.5%)	36 (1.0%)	-0.06	41 (0.6%)	27 (1.0%)	-0.04
Charlson comobidity index score[c]									
0	30,859 (85.3%)	3,368 (80.0%)	0.14	14,570 (83.5%)	2,772 (79.4%)	0.11	5,660 (82.3%)	2,234 (79.9%)	0.06
1-2	4,586 (12.7%)	690 (16.4%)	-0.11	2,437 (14.0%)	593 (17.0%)	-0.08	1,006 (14.6%)	464 (16.6%)	-0.05
3+	749 (2.1%)	151 (3.6%)	-0.09	437 (2.5%)	124 (3.6%)	-0.06	215 (3.1%)	97 (3.5%)	-0.02
Prescriptions[d]									
Statins	7,887 (21.8%)	1,001 (23.8%)	-0.05	4,120 (23.6%)	852 (24.4%)	-0.02	1,679 (24.4%)	706 (25.3%)	-0.02
NSAIDS	8,480 (23.4%)	1,130 (26.8%)	-0.08	4,302 (24.7%)	922 (26.4%)	-0.04	1,760 (25.6%)	725 (25.9%)	-0.01
Glucocorticoids	2,241 (6.2%)	392 (9.3%)	-0.12	1,229 (7.0%)	324 (9.3%)	-0.08	559 (8.1%)	252 (9.0%)	-0.03
Diuretics	5,725 (15.8%)	1,106 (26.3%)	-0.26	3,280 (18.8%)	921 (26.4%)	-0.18	1,480 (21.5%)	729 (26.1%)	-0.11

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Antidepressants	3,523 (9.7%)	463 (11.0%)	-0.04	1,739 (10.0%)	393 (11.3%)	-0.04	700 (10.2%)	303 (10.8%)	-0.02
Antibiotics	10,200 (28.2%)	1,344 (31.9%)	-0.08	5,245 (30.1%)	1,114 (31.9%)	-0.04	2,179 (31.7%)	877 (31.4%)	0.01
Any antihypertensives	14,299 (39.5%)	2,095 (49.8%)	-0.21	7,578 (43.4%)	1,738 (49.8%)	-0.13	3,182 (46.2%)	1,392 (49.8%)	-0.07
Antiplatelets	6,754 (18.7%)	996 (23.7%)	-0.12	3,612 (20.7%)	836 (24.0%)	-0.08	1,509 (21.9%)	678 (24.3%)	-0.06

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S7 Characteristics of individuals aged <60 years within HbA_{1c} ranges just below versus just above the 6.5% (48 mmol/mol) threshold for treatment initiation, 2006–2011.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	32,610	3,228		14,431	2,643		5,369	2,093	
Treated with metformin within 3 months	242 (0.7%)	642 (19.9%)	-0.66	205 (1.4%)	433 (16.4%)	-0.54	155 (2.9%)	301 (14.4%)	-0.42
Female	16,112 (49.4%)	1,463 (45.3%)	0.08	7,060 (48.9%)	1,195 (45.2%)	0.07	2,630 (49.0%)	952 (45.5%)	0.07
Median age at diagnosis (Q1;Q3)	52.7 (47.9;56.6)	52.7 (47.9;56.6)	0.02	53.1 (48.3;56.8)	52.8 (48.1;56.7)	0.10	53.3 (48.5;56.9)	52.9 (48.1;56.6)	0.14
Comorbidities									
Chronic pulmonary disease[a]	752 (2.3%)	113 (3.5%)	-0.07	382 (2.6%)	93 (3.5%)	-0.05	155 (2.9%)	72 (3.4%)	-0.03
Moderate to severe renal disease[a]	139 (0.4%)	21 (0.7%)	-0.03	64 (0.4%)	18 (0.7%)	-0.03	27 (0.5%)	14 (0.7%)	-0.02
Obesity[b]	1,381 (4.2%)	215 (6.7%)	-0.11	738 (5.1%)	178 (6.7%)	-0.07	324 (6.0%)	134 (6.4%)	-0.02
Any cancer[a]	733 (2.2%)	90 (2.8%)	-0.03	333 (2.3%)	74 (2.8%)	-0.03	137 (2.6%)	57 (2.7%)	-0.01
Any liver disease[a]	169 (0.5%)	31 (1.0%)	-0.05	85 (0.6%)	26 (1.0%)	-0.04	37 (0.7%)	21 (1.0%)	-0.03
Charlson comobidity index score[c]									
0	29,750 (91.2%)	2,846 (88.2%)	0.10	13,053 (90.5%)	2,323 (87.9%)	0.08	4,806 (89.5%)	1,846 (88.2%)	0.04
1-2	2,539 (7.8%)	336 (10.4%)	-0.09	1,206 (8.4%)	285 (10.8%)	-0.08	493 (9.2%)	221 (10.6%)	-0.05
3+	321 (1.0%)	46 (1.4%)	-0.04	172 (1.2%)	35 (1.3%)	-0.01	70 (1.3%)	26 (1.2%)	0.01
Prescriptions[d]									
Statins	4,259 (13.1%)	496 (15.4%)	-0.07	2,153 (14.9%)	417 (15.8%)	-0.02	836 (15.6%)	331 (15.8%)	-0.01
NSAIDS	8,984 (27.5%)	978 (30.3%)	-0.06	4,211 (29.2%)	795 (30.1%)	-0.02	1,590 (29.6%)	625 (29.9%)	-0.01
Glucocorticoids	1,870 (5.7%)	246 (7.6%)	-0.08	927 (6.4%)	206 (7.8%)	-0.05	383 (7.1%)	158 (7.5%)	-0.02

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Diuretics	4,154 (12.7%)	648 (20.1%)	-0.20	2,201 (15.3%)	527 (19.9%)	-0.12	955 (17.8%)	413 (19.7%)	-0.05
Antidepressants	4,934 (15.1%)	570 (17.7%)	-0.07	2,310 (16.0%)	469 (17.7%)	-0.05	876 (16.3%)	372 (17.8%)	-0.04
Antibiotics	10,201 (31.3%)	1,120 (34.7%)	-0.07	4,776 (33.1%)	921 (34.8%)	-0.04	1,860 (34.6%)	723 (34.5%)	0.00
Any antihypertensives	9,423 (28.9%)	1,277 (39.6%)	-0.23	4,749 (32.9%)	1,030 (39.0%)	-0.13	1,931 (36.0%)	799 (38.2%)	-0.05
Antiplatelets	2,283 (7.0%)	320 (9.9%)	-0.10	1,181 (8.2%)	264 (10.0%)	-0.06	500 (9.3%)	205 (9.8%)	-0.02

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S8 Characteristics of individuals aged ≥60 years within HbA_{1c} ranges just below versus just above the 6.5% (48 mmol/mol) threshold for treatment initiation, 2006–2011.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Total	43,421	4,783		21,929	4,067		8,847	3,323	
Treated with metformin within 3 months	323 (0.7%)	691 (14.4%)	-0.54	279 (1.3%)	488 (12.0%)	-0.44	213 (2.4%)	345 (10.4%)	-0.33
Female	23,725 (54.6%)	2,339 (48.9%)	0.12	11,856 (54.1%)	2,026 (49.8%)	0.09	4,705 (53.2%)	1,669 (50.2%)	0.06
Median age at diagnosis (Q1;Q3)	68.2 (64.0;73.4)	68.7 (64.4;73.9)	-0.17	68.6 (64.3;73.7)	68.8 (64.4;73.8)	-0.07	68.9 (64.3;73.9)	69.0 (64.5;73.9)	-0.03
Comorbidities									
Chronic pulmonary disease[a]	2,164 (5.0%)	384 (8.0%)	-0.12	1,252 (5.7%)	329 (8.1%)	-0.09	590 (6.7%)	276 (8.3%)	-0.06
Moderate to severe renal disease[a]	321 (0.7%)	50 (1.0%)	-0.03	179 (0.8%)	43 (1.1%)	-0.03	73 (0.8%)	32 (1.0%)	-0.01
Obesity[b]	858 (2.0%)	183 (3.8%)	-0.11	508 (2.3%)	154 (3.8%)	-0.09	228 (2.6%)	120 (3.6%)	-0.06
Any cancer[a]	2,868 (6.6%)	384 (8.0%)	-0.05	1,514 (6.9%)	319 (7.8%)	-0.04	639 (7.2%)	248 (7.5%)	-0.01
Any liver disease[a]	141 (0.3%)	33 (0.7%)	-0.05	78 (0.4%)	28 (0.7%)	-0.05	31 (0.4%)	21 (0.6%)	-0.04
Charlson comobidity index score[c]									
0	35,230 (81.1%)	3,591 (75.1%)	0.15	17,469 (79.7%)	3,049 (75.0%)	0.11	6,900 (78.0%)	2,498 (75.2%)	0.07
1-2	6,993 (16.1%)	964 (20.2%)	-0.11	3,782 (17.2%)	827 (20.3%)	-0.08	1,627 (18.4%)	681 (20.5%)	-0.05
3+	1,198 (2.8%)	228 (4.8%)	-0.11	678 (3.1%)	191 (4.7%)	-0.08	320 (3.6%)	144 (4.3%)	-0.04
Prescriptions[d]									
Statins	12,646 (29.1%)	1,446 (30.2%)	-0.02	6,694 (30.5%)	1,243 (30.6%)	-0.00	2,779 (31.4%)	1,038 (31.2%)	0.00
NSAIDS	10,371 (23.9%)	1,345 (28.1%)	-0.10	5,407 (24.7%)	1,120 (27.5%)	-0.07	2,283 (25.8%)	908 (27.3%)	-0.03
Glucocorticoids	3,485 (8.0%)	607 (12.7%)	-0.15	2,023 (9.2%)	510 (12.5%)	-0.11	953 (10.8%)	420 (12.6%)	-0.06

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.8–6.4% (40–47 mmol/mol)	*6.5–7.3% (48–56 mmol/mol)*	*Standardized difference*	*6.0–6.4% (42–47 mmol/mol)*	*6.5–7.0% (48–53 mmol/mol)*	*Standardized difference*	*6.2–6.4% (44–47 mmol/mol)*	*6.5–6.8% (48–51 mmol/mol)*	*Standardized difference*
Diuretics	11,643 (26.8%)	1,916 (40.1%)	-0.28	6,643 (30.3%)	1,617 (39.8%)	-0.20	2,952 (33.4%)	1,320 (39.7%)	-0.13
Antidepressants	5,381 (12.4%)	689 (14.4%)	-0.06	2,825 (12.9%)	602 (14.8%)	-0.06	1,209 (13.7%)	491 (14.8%)	-0.03
Antibiotics	14,363 (33.1%)	1,774 (37.1%)	-0.08	7,631 (34.8%)	1,527 (37.5%)	-0.06	3,228 (36.5%)	1,232 (37.1%)	-0.01
Any antihypertensives	22,867 (52.7%)	3,025 (63.2%)	-0.22	12,232 (55.8%)	2,578 (63.4%)	-0.16	5,133 (58.0%)	2,121 (63.8%)	-0.12
Antiplatelets	10,328 (23.8%)	1,420 (29.7%)	-0.13	5,588 (25.5%)	1,207 (29.7%)	-0.09	2,373 (26.8%)	1,003 (30.2%)	-0.07

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S9 Hazard ratios of all-cause mortality, cardiovascular events and composite endpoint associated with HbA_{1c} ≥6.5% (≥48 mmol/mol), overall and stratified analysis for the broad range, ie, HbA_{1c} 5.8–7.3% (40–56 mmol/mol) and the narrow range, ie, HbA_{1c} 6.2–6.8% (44–51 mmol/mol). The reference group is HbA_{1c} <6.5% (<48 mmol/mol).

Strata	*Range*	*All-cause mortality* *HR (95% CI)*	*Cardiovascular event* *HR (95% CI)*	*All-cause mortality or cardiovascular event* *HR (95% CI)*	*Sample size*
All	5.8–7.3% (40–56 mmol/mol)	0.77 (0.68 ; 0.87)	0.97 (0.81 ; 1.17)	0.82 (0.74 ; 0.92)	84,042
	6.2–6.8% (44–51 mmol/mol)	0.72 (0.58 ; 0.91)	0.81 (0.57 ; 1.16)	0.74 (0.61 ; 0.90)	19,632
All, adjusted[a]	5.8–7.3% (40–56 mmol/mol)	0.79 (0.70 ; 0.89)	1.00 (0.83 ; 1.20)	0.84 (0.76 ; 0.94)	84,042
	6.2–6.8% (44–51 mmol/mol)	0.74 (0.59 ; 0.93)	0.84 (0.59 ; 1.19)	0.77 (0.63 ; 0.93)	19,632
2006–2011	5.8–7.3% (40–56 mmol/mol)	0.79 (0.69 ; 0.89)	0.96 (0.79 ; 1.18)	0.83 (0.74 ; 0.93)	54,332
	6.2–6.8% (44–51 mmol/mol)	0.77 (0.60 ; 0.98)	0.82 (0.56 ; 1.21)	0.77 (0.62 ; 0.96)	14,175
2012–2014	5.8–7.3% (40–56 mmol/mol)	0.79 (0.55 ; 1.14)	0.98 (0.57 ; 1.68)	0.88 (0.65 ; 1.20)	29,710
	6.2–6.8% (44–51 mmol/mol)	0.66 (0.37 ; 1.16)	0.85 (0.35 ; 2.04)	0.72 (0.44 ; 1.18)	5,457
Women	5.8–7.3% (40–56 mmol/mol)	0.72 (0.60 ; 0.86)	0.83 (0.62 ; 1.13)	0.77 (0.65 ; 0.90)	43,639
	6.2–6.8% (44–51 mmol/mol)	0.71 (0.51 ; 0.99)	0.57 (0.33 ; 0.99)	0.69 (0.51 ; 0.93)	9,956
Men	5.8–7.3% (40–56 mmol/mol)	0.81 (0.69 ; 0.95)	1.06 (0.84 ; 1.35)	0.86 (0.75 ; 0.99)	40,403
	6.2–6.8% (44–51 mmol/mol)	0.73 (0.54 ; 1.00)	1.05 (0.66 ; 1.66)	0.79 (0.60 ; 1.02)	9,676
Age <60y	5.8–7.3% (40–56 mmol/mol)	0.77 (0.56 ; 1.05)	1.37 (0.95 ; 1.97)	0.98 (0.77 ; 1.25)	35,838
	6.2–6.8% (44–51 mmol/mol)	0.65 (0.36 ; 1.18)	0.93 (0.45 ; 1.91)	0.79 (0.49 ; 1.25)	7,462
Age ≥60y	5.8–7.3% (40–56 mmol/mol)	0.83 (0.73 ; 0.95)	0.92 (0.74 ; 1.14)	0.85 (0.75 ; 0.95)	48,204
	6.2–6.8% (44–51 mmol/mol)	0.78 (0.61 ; 0.99)	0.80 (0.54 ; 1.20)	0.77 (0.62 ; 0.95)	12,170

[a] *Adjusted for diuretic, antihypertensive treatments and year*

Table S10 Characteristics of individuals with HbA_{1c} ranges just below versus just above a 6.0% (42 mmol/mol) threshold for treatment initiation, 2006–2011.

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.4-5.9% (35-41 mmol/mol)	*6.0-6.8% (42-51 mmol/mol)*	*Standardized difference*	*5.5-5.9% (37-41 mmol/mol)*	*6.0-6.5% (42-48 mmol/mol)*	*Standardized difference*	*5.8-5.9% (40-41 mmol/mol)*	*6.0-6.2% (42-44 mmol/mol)*	*Standardized difference*
Total	165,334	41,776		126,233	38,269		39,671	28,617	
Treated with metformin within 3 months	171 (0.1%)	1,130 (2.7%)	-0.22	153 (0.1%)	632 (1.7%)	-0.16	81 (0.2%)	215 (0.8%)	-0.08
Female	86,887 (52.6%)	21,537 (51.6%)	0.02	66,297 (52.5%)	19,853 (51.9%)	0.01	20,921 (52.7%)	14,967 (52.3%)	0.01
Median age at diagnosis (Q1;Q3)	58.8 (50.5;66.8)	63.0 (55.1;70.3)	-1.12	59.6 (51.3;67.3)	62.9 (55.1;70.2)	-0.92	61.1 (53.2;68.6)	62.7 (54.9;70.0)	-0.43
Comorbidities									
Chronic pulmonary disease[a]	4,144 (2.5%)	1,982 (4.7%)	-0.12	3,431 (2.7%)	1,755 (4.6%)	-0.10	1,282 (3.2%)	1,210 (4.2%)	-0.05
Moderate to severe renal disease[a]	714 (0.4%)	289 (0.7%)	-0.03	564 (0.4%)	254 (0.7%)	-0.03	217 (0.5%)	191 (0.7%)	-0.02
Obesity[b]	3,585 (2.2%)	1,500 (3.6%)	-0.09	2,787 (2.2%)	1,326 (3.5%)	-0.08	993 (2.5%)	943 (3.3%)	-0.05
Any cancer[a]	6,358 (3.8%)	2,152 (5.2%)	-0.06	5,041 (4.0%)	1,943 (5.1%)	-0.05	1,754 (4.4%)	1,413 (4.9%)	-0.02
Any liver disease[a]	577 (0.3%)	205 (0.5%)	-0.02	439 (0.3%)	175 (0.5%)	-0.02	147 (0.4%)	126 (0.4%)	-0.01
Charlson comobidity index score[c]									
0	147,211 (89.0%)	34,866 (83.5%)	0.16	111,678 (88.5%)	32,072 (83.8%)	0.14	34,458 (86.9%)	24,204 (84.6%)	0.07
1-2	15,980 (9.7%)	5,890 (14.1%)	-0.14	12,828 (10.2%)	5,295 (13.8%)	-0.11	4,544 (11.5%)	3,782 (13.2%)	-0.05
3+	2,143 (1.3%)	1,020 (2.4%)	-0.08	1,727 (1.4%)	902 (2.4%)	-0.07	669 (1.7%)	631 (2.2%)	-0.04
Prescriptions[d]									
Statins	25,556 (15.5%)	10,216 (24.5%)	-0.23	21,075 (16.7%)	9,345 (24.4%)	-0.19	8,058 (20.3%)	6,863 (24.0%)	-0.09

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	5.4-5.9% (35-41 mmol/mol)	*6.0-6.8% (42-51 mmol/mol)*	*Standardized difference*	*5.5-5.9% (37-41 mmol/mol)*	*6.0-6.5% (42-48 mmol/mol)*	*Standardized difference*	*5.8-5.9% (40-41 mmol/mol)*	*6.0-6.2% (42-44 mmol/mol)*	*Standardized difference*
NSAIDS	37,924 (22.9%)	11,151 (26.7%)	-0.09	29,683 (23.5%)	10,165 (26.6%)	-0.07	9,737 (24.5%)	7,477 (26.1%)	-0.04
Glucocorticoids	8,424 (5.1%)	3,528 (8.4%)	-0.13	6,733 (5.3%)	3,155 (8.2%)	-0.12	2,405 (6.1%)	2,195 (7.7%)	-0.06
Diuretics	22,130 (13.4%)	10,577 (25.3%)	-0.31	18,241 (14.5%)	9,445 (24.7%)	-0.26	6,953 (17.5%)	6,629 (23.2%)	-0.14
Antidepressants	19,832 (12.0%)	5,998 (14.4%)	-0.07	15,444 (12.2%)	5,451 (14.2%)	-0.06	5,180 (13.1%)	3,974 (13.9%)	-0.02
Antibiotics	46,647 (28.2%)	14,362 (34.4%)	-0.13	36,472 (28.9%)	13,088 (34.2%)	-0.11	12,157 (30.6%)	9,610 (33.6%)	-0.06
Any antihypertensives	52,937 (32.0%)	19,901 (47.6%)	-0.32	42,570 (33.7%)	17,977 (47.0%)	-0.27	15,309 (38.6%)	13,070 (45.7%)	-0.14
Antiplatelets	18,315 (11.1%)	7,977 (19.1%)	-0.23	15,179 (12.0%)	7,175 (18.7%)	-0.19	5,842 (14.7%)	5,168 (18.1%)	-0.09

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S11 Hazard ratios of all-cause mortality, cardiovascular events and composite endpoint associated with HbA_{1c} ≥6.0% (≥42 mmol/mol), overall and stratified analysis for a broad range, ie, HbA_{1c} 5.4–6.8% (35–51 mmol/mol), medium range, ie, HbA_{1c} 5.5–6.5% (37–48 mmol/mol), and narrow range, ie, HbA_{1c} 5.8–6.2% (40–44 mmol/mol). The reference group is HbA_{1c} <6.0% (<42 mmol/mol).

Range	*All-cause mortality (95% CI)*	*Cardiovascular event (95% CI)*	*All-cause mortality or cardiovascular event (95% CI)*	*Sample size*
5.4-6.8% (35-51 mmol/mol)	1.07 (1.00 ; 1.14)	1.00 (0.91 ; 1.11)	1.06 (1.00 ; 1.12)	207,110
5.5-6.5% (37-48 mmol/mol)	1.03 (0.95 ; 1.12)	0.95 (0.85 ; 1.06)	1.02 (0.95 ; 1.09)	164,502
5.8-6.2% (40-44 mmol/mol)	0.93 (0.81 ; 1.08)	0.92 (0.76 ; 1.13)	0.98 (0.87 ; 1.10)	68,288

Table S12 Characteristics of individuals with HbA_{1c} ranges just below versus just above a 7.0% (53 mmol/mol) threshold for treatment initiation, 2006–2011

	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	6.2-6.9% (44-52 mmol/mol)	*7.0-8.0% (53-64 mmol/mol)*	*Standardized difference*	*6.5-6.9% (48-52 mmol/mol)*	*7.0-7.6% (53-60 mmol/mol)*	*Standardized difference*	*6.8-6.9% (51-52 mmol/mol)*	*7.0-7.3% (53-56 mmol/mol)*	*Standardized difference*
Total	20,351	3,420		6,135	2,702		1,638	1,876	
Treated with metformin within 3 months	1,161 (5.7%)	1,179 (34.5%)	-0.77	793 (12.9%)	855 (31.6%)	-0.46	298 (18.2%)	540 (28.8%)	-0.25
Female	10,307 (50.6%)	1,495 (43.7%)	0.14	2,972 (48.4%)	1,199 (44.4%)	0.08	784 (47.9%)	830 (44.2%)	0.07
Median age at diagnosis (Q1;Q3)	63.3 (55.4;70.7)	61.7 (53.1;69.0)	0.51	63.1 (54.8;70.6)	61.8 (53.4;69.4)	0.33	62.1 (53.5;69.9)	61.9 (53.6;69.4)	0.08
Comorbidities									
Chronic pulmonary disease[a]	1,132 (5.6%)	200 (5.8%)	-0.01	387 (6.3%)	162 (6.0%)	0.01	101 (6.2%)	110 (5.9%)	0.01
Moderate to severe renal disease[a]	157 (0.8%)	28 (0.8%)	-0.01	57 (0.9%)	20 (0.7%)	0.02	19 (1.2%)	14 (0.7%)	0.04
Obesity[b]	846 (4.2%)	177 (5.2%)	-0.05	294 (4.8%)	145 (5.4%)	-0.03	94 (5.7%)	104 (5.5%)	0.01
Any cancer[a]	1,121 (5.5%)	223 (6.5%)	-0.04	345 (5.6%)	184 (6.8%)	-0.05	91 (5.6%)	129 (6.9%)	-0.05
Any liver disease[a]	117 (0.6%)	29 (0.8%)	-0.03	49 (0.8%)	20 (0.7%)	0.01	19 (1.2%)	15 (0.8%)	0.04
Charlson comobidity index score[c]									
0	16,628 (81.7%)	2,775 (81.1%)	0.01	4,922 (80.2%)	2,177 (80.6%)	-0.01	1,319 (80.5%)	1,515 (80.8%)	-0.01
1-2	3,132 (15.4%)	519 (15.2%)	0.01	1,012 (16.5%)	423 (15.7%)	0.02	251 (15.3%)	288 (15.4%)	-0.00
3+	591 (2.9%)	126 (3.7%)	-0.04	201 (3.3%)	102 (3.8%)	-0.03	68 (4.2%)	73 (3.9%)	0.01
Prescriptions[d]									
Statins	5,148 (25.3%)	678 (19.8%)	0.13	1,533 (25.0%)	560 (20.7%)	0.10	366 (22.3%)	409 (21.8%)	0.01
NSAIDS	5,621 (27.6%)	995 (29.1%)	-0.03	1,748 (28.5%)	805 (29.8%)	-0.03	453 (27.7%)	575 (30.7%)	-0.07

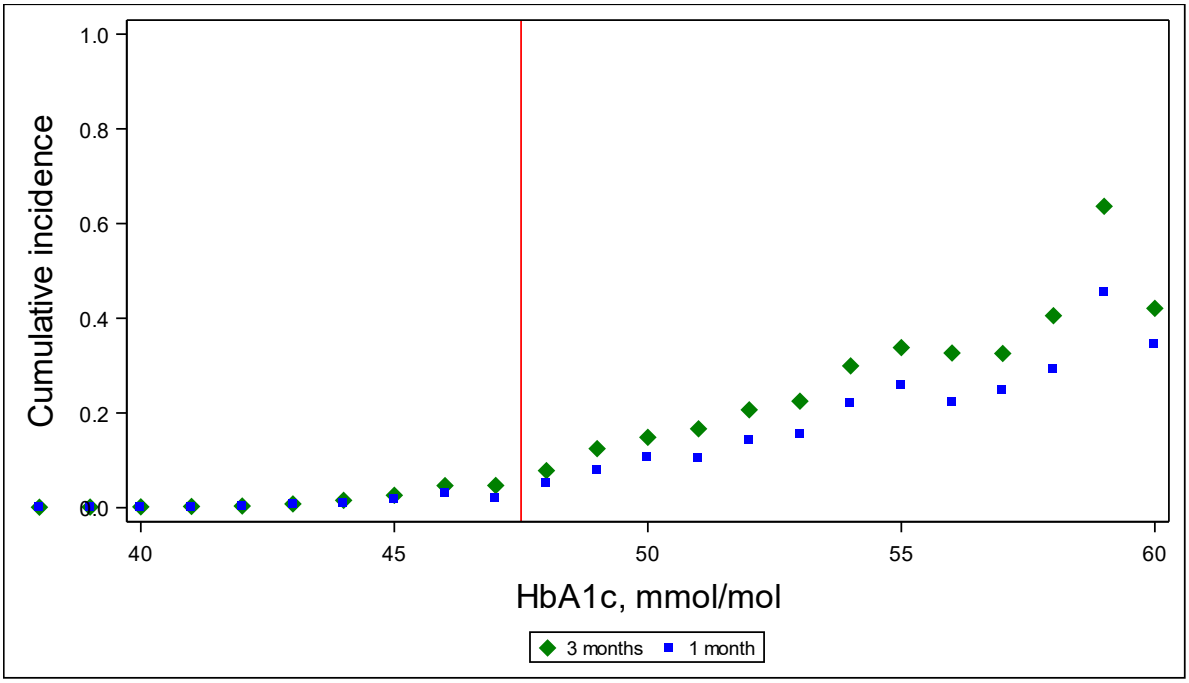
	Broad HbA_{1c} range			*Primary HbA_{1c} range*			*Narrow HbA_{1c} range*		
	6.2-6.9% (44-52 mmol/mol)	*7.0-8.0% (53-64 mmol/mol)*	*Standardized difference*	*6.5-6.9% (48-52 mmol/mol)*	*7.0-7.6% (53-60 mmol/mol)*	*Standardized difference*	*6.8-6.9% (51-52 mmol/mol)*	*7.0-7.3% (53-56 mmol/mol)*	*Standardized difference*
Glucocorticoids	1,990 (9.8%)	377 (11.0%)	-0.04	654 (10.7%)	295 (10.9%)	-0.01	177 (10.8%)	199 (10.6%)	0.01
Diuretics	5,868 (28.8%)	1,017 (29.7%)	-0.02	1,961 (32.0%)	823 (30.5%)	0.03	512 (31.3%)	603 (32.1%)	-0.02
Antidepressants	3,069 (15.1%)	485 (14.2%)	0.03	984 (16.0%)	395 (14.6%)	0.04	271 (16.5%)	275 (14.7%)	0.05
Antibiotics	7,318 (36.0%)	1,254 (36.7%)	-0.01	2,230 (36.3%)	977 (36.2%)	0.00	609 (37.2%)	664 (35.4%)	0.04
Any antihypertensives	10,372 (51.0%)	1,725 (50.4%)	0.01	3,308 (53.9%)	1,391 (51.5%)	0.05	886 (54.1%)	994 (53.0%)	0.02
Antiplatelets	4,229 (20.8%)	640 (18.7%)	0.05	1,356 (22.1%)	518 (19.2%)	0.07	326 (19.9%)	384 (20.5%)	-0.01

[a] *Last 5 years prior to initial measurement*
[b] *Last 10 years prior to initial measurement*
[c] *Last 5 years prior to initial measurement. Diabetes has been omitted from the Charlson comorbidity index score.*
[d] *Last year prior to initial measurement.*

Table S13 Hazard ratios of all-cause mortality, cardiovascular events and composite endpoint associated with HbA_{1c} ≥7.0% (≥53 mmol/mol), overall and stratified analysis for a broad range, ie, HbA_{1c} 6.2–8.0% (44–64 mmol/mol), medium range, ie, HbA_{1c} 6.5–7.6% (48–60 mmol/mol), and narrow range, ie, HbA_{1c} 6.8–7.3% (51–56 mmol/mol). The reference group is HbA_{1c} <7.0% (<53 mmol/mol).

Range	*All-cause mortality (95% CI)*	*Cardiovascular event (95% CI)*	*All-cause mortality or cardiovascular event (95% CI)*	*Sample size*
6.2-8.0% (44-64 mmol/mol)	0.92 (0.78 ; 1.09)	0.64 (0.47 ; 0.87)	0.84 (0.72 ; 0.98)	23,771
6.5-7.6% (48-60 mmol/mol)	0.94 (0.75 ; 1.19)	0.68 (0.45 ; 1.01)	0.86 (0.70 ; 1.06)	8,837
6.8-7.3% (51-56 mmol/mol)	1.18 (0.75 ; 1.85)	0.74 (0.34 ; 1.61)	1.10 (0.73 ; 1.65)	3,514

Figure S1. Cumulative incidence of metformin treatment initiation within 1 and 3 months by HbA_{1c} value for patients with a first HbA_{1c} measurement.



Reports/PhD theses from Department of Clinical Epidemiology

1. Ane Marie Thulstrup: Mortality, infections and operative risk in patients with liver cirrhosis in Denmark. Clinical epidemiological studies. PhD thesis. *2000*.

2. Nana Thrane: Prescription of systemic antibiotics for Danish children. PhD thesis. *2000*.

3. Charlotte Søndergaard. Follow-up studies of prenatal, perinatal and postnatal risk factors in infantile colic. PhD thesis. *2001*.

4. Charlotte Olesen: Use of the North Jutland Prescription Database in epidemiological studies of drug use and drug safety during pregnancy. PhD thesis. *2001*.

5. Yuan Wei: The impact of fetal growth on the subsequent risk of infectious disease and asthma in childhood. PhD thesis. *2001*.

6. Gitte Pedersen. Bacteremia: treatment and prognosis. PhD thesis. *2001*.

7. Henrik Gregersen: The prognosis of Danish patients with monoclonal gammopathy of undertermined significance: register-based studies. PhD thesis. *2002*.

8. Bente Nørgård: Colitis ulcerosa, coeliaki og graviditet; en oversigt med speciel reference til forløb og sikkerhed af medicinsk behandling. PhD thesis. *2002*.

9. Søren Paaske Johnsen: Risk factors for stroke with special reference to diet, Chlamydia pneumoniae, infection, and use of non-steroidal anti-inflammatory drugs. PhD thesis. *2002*.

10. Elise Snitker Jensen: Seasonal variation of meningococcal disease and factors associated with its outcome. PhD thesis. *2003*.

11. Andrea Floyd: Drug-associated acute pancreatitis. Clinical epidemiological studies of selected drugs. PhD thesis. *2004*.

12. Pia Wogelius: Aspects of dental health in children with asthma. Epidemiological studies of dental anxiety and caries among children in North Jutland County, Denmark. PhD thesis. *2004*.

13. Kort-og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Nordjyllands, Viborg og Århus amter 1985-2003. *2004*.

14. Reimar W. Thomsen: Diabetes mellitus and community-acquired bacteremia: risk and prognosis. PhD thesis. *2004*.

15. Kronisk obstruktiv lungesygdom i Nordjyllands, Viborg og Århus amter 1994-2004. Forekomst og prognose. Et pilotprojekt. *2005*.

16. Lungebetændelse i Nordjyllands, Viborg og Århus amter 1994-2004. Forekomst og prognose. Et pilotprojekt. *2005*.

17. Kort- og langtidsoverlevelse efter indlæggelse for nyre-, bugspytkirtel- og leverkræft i Nordjyllands, Viborg, Ringkøbing og Århus amter 1985-2004. *2005*.

18. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Nordjyllands, Viborg, Ringkøbing og Århus amter 1995-2005. *2005*.

19. Mette Nørgaard: Haematological malignancies: Risk and prognosis. PhD thesis. *2006*.

20. Alma Becic Pedersen: Studies based on the Danish Hip Arthroplastry Registry. PhD thesis. *2006*.

    Særtryk: Klinisk Epidemiologisk Afdeling - De første 5 år. *2006*.

21. Blindtarmsbetændelse i Vejle, Ringkjøbing, Viborg, Nordjyllands og Århus Amter. *2006*.

22. Andre sygdommes betydning for overlevelse efter indlæggelse for seks kræftsygdomme i Nordjyllands, Viborg, Ringkjøbing og Århus amter 1995-2005. *2006*.

23. Ambulante besøg og indlæggelser for udvalgte kroniske sygdomme på somatiske hospitaler i Århus, Ringkjøbing, Viborg, og Nordjyllands amter. *2006*.

24. Ellen M Mikkelsen: Impact of genetic counseling for hereditary breast and ovarian cancer disposition on psychosocial outcomes and risk perception: A population-based follow-up study. PhD thesis. *2006*.

25. Forbruget af lægemidler mod kroniske sygdomme i Århus, Viborg og Nordjyllands amter 2004-2005. *2006*.

26. Tilbagelægning af kolostomi og ileostomi i Vejle, Ringkjøbing, Viborg, Nordjyllands og Århus Amter. *2006*.

27. Rune Erichsen: Time trend in incidence and prognosis of primary liver cancer and liver cancer of unknown origin in a Danish region, 1985-2004. Research year report. *2007*.

28. Vivian Langagergaard: Birth outcome in Danish women with breast cancer, cutaneous malignant melanoma, and Hodgkin's disease. PhD thesis. *2007*.

29. Cynthia de Luise: The relationship between chronic obstructive pulmonary disease, comorbidity and mortality following hip fracture. PhD thesis. *2007*.

30. Kirstine Kobberøe Søgaard: Risk of venous thromboembolism in patients with liver disease: A nationwide population-based case-control study. Research year report. *2007*.

31. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Region Midtjylland og Region Nordjylland 1995-2006. *2007*.

32. Mette Skytte Tetsche: Prognosis for ovarian cancer in Denmark 1980-2005: Studies of use of hospital discharge data to monitor and study prognosis and impact of comorbidity and venous thromboembolism on survival. PhD thesis. *2007*.

33. Estrid Muff Munk: Clinical epidemiological studies in patients with unexplained chest and/or epigastric pain. PhD thesis. *2007*.

34. Sygehuskontakter og lægemiddelforbrug for udvalgte kroniske sygdomme i Region Nordjylland. *2007*.

35. Vera Ehrenstein: Association of Apgar score and postterm delivery with neurologic morbidity: Cohort studies using data from Danish population registries. PhD thesis. *2007*.

36. Annette Østergaard Jensen: Chronic diseases and non-melanoma skin cancer. The impact on risk and prognosis. PhD thesis. *2008*.

37. Use of medical databases in clinical epidemiology. *2008*.

38. Majken Karoline Jensen: Genetic variation related to high-density lipoprotein metabolism and risk of coronary heart disease. PhD thesis. *2008*.

39. Blodprop i hjertet - forekomst og prognose. En undersøgelse af førstegangsindlæggelser i Region Nordjylland og Region Midtjylland. *2008*.

40. Asbestose og kræft i lungehinderne. Danmark 1977-2005. *2008*.

41. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Region Midtjylland og Region Nordjylland 1996-2007. *2008*.

    Sandra Kruchov Thygesen. Atrial fibrillation in patients with ischemic stroke: A population-based study. Research year report. *2008*.

42. Akutte indlæggelsesforløb og skadestuebesøg på hospiter i Region Midtjylland og Region Nordjylland 2003-2007. Et pilotprojekt. *Not published*.

43. Peter Jepsen: Prognosis for Danish patients with liver cirrhosis. PhD thesis. *2009*.

44. Lars Pedersen: Use of Danish health registries to study drug-induced birth defects – A review with special reference to methodological issues and maternal use of non-steroidal anti-inflammatory drugs and Loratadine. PhD thesis. *2009*.

45. Steffen Christensen: Prognosis of Danish patients in intensive care. Clinical epidemiological studies on the impact of preadmission cardiovascular drug use on mortality. PhD thesis. *2009*.

46. Morten Schmidt: Use of selective cyclooxygenase-2 inhibitors and nonselective nonsteroidal antiinflammatory drugs and risk of cardiovascular events and death after intracoronary stenting. Research year report. *2009*.

47. Jette Bromman Kornum: Obesity, diabetes and hospitalization with pneumonia. PhD thesis. *2009*.

48. Theis Thilemann: Medication use and risk of revision after primary total hip arthroplasty. PhD thesis. *2009*.

49. Operativ fjernelse af galdeblæren. Region Midtjylland & Region Nordjylland. 1998-2008. *2009*.

50. Mette Søgaard: Diagnosis and prognosis of patients with community-acquired bacteremia. PhD thesis. *2009*.

51. Marianne Tang Severinsen. Risk factors for venous thromboembolism: Smoking, anthropometry and genetic susceptibility. PhD thesis. *2010*.

52. Henriette Thisted: Antidiabetic Treatments and ischemic cardiovascular disease in Denmark: Risk and outcome. PhD thesis. *2010*.

53. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme. Region Midtjylland og Region Nordjylland 1997-2008. *2010*.

54. Prognosen efter akut indlæggelse på Medicinsk Visitationsafsnit på Nørrebrogade, Århus Sygehus. *2010*.

55. Kaare Haurvig Palnum: Implementation of clinical guidelines regarding acute treatment and secondary medical prophylaxis among patients with acute stroke in Denmark. PhD thesis. *2010*.

56. Thomas Patrick Ahern: Estimating the impact of molecular profiles and prescription drugs on breast cancer outcomes. PhD thesis. *2010*.

57. Annette Ingeman: Medical complications in patients with stroke: Data validity, processes of care, and clinical outcome. PhD thesis. *2010*.

58. Knoglemetastaser og skeletrelaterede hændelser blandt patienter med prostatakræft i Danmark. Forekomst og prognose 1999-2007. *2010*.

59. Morten Olsen: Prognosis for Danish patients with congenital heart defects - Mortality, psychiatric morbidity, and educational achievement. PhD thesis. *2010*.

60. Knoglemetastaser og skeletrelaterede hændelser blandt kvinder med brystkræft i Danmark. Forekomst og prognose 1999-2007. *2010*.

61. Kort- og langtidsoverlevelse efter hospitalsbehandlet kræft. Region Midtjylland og Region Nordjylland 1998-2009. *2010*.

62. Anna Lei Lamberg: The use of new and existing data sources in non-melanoma skin cancer research. PhD thesis. *2011*.

63. Sigrún Alba Jóhannesdóttir: Mortality in cancer patients following a history of squamous cell skin cancer – A nationwide population-based cohort study. Research year report. *2011*.

64. Martin Majlund Mikkelsen: Risk prediction and prognosis following cardiac surgery: the EuroSCORE and new potential prognostic factors. PhD thesis. *2011*.

65. Gitte Vrelits Sørensen: Use of glucocorticoids and risk of breast cancer: a Danish population-based case-control study. Research year report. *2011*.

66. Anne-Mette Bay Bjørn: Use of corticosteroids in pregnancy. With special focus on the relation to congenital malformations in offspring and miscarriage. PhD thesis. *2012*.

67. Marie Louise Overgaard Svendsen: Early stroke care: studies on structure, process, and outcome. PhD thesis. *2012*.

68. Christian Fynbo Christiansen: Diabetes, preadmission morbidity, and intensive care: population-based Danish studies of prognosis. PhD thesis. *2012*.

69. Jennie Maria Christin Strid: Hospitalization rate and 30-day mortality of patients with status asthmaticus in Denmark – A 16-year nationwide population-based cohort study. Research year report. *2012*.

70. Alkoholisk leversygdom i Region Midtjylland og Region Nordjylland. 2007-2011. *2012*.

71. Lars Jakobsen: Treatment and prognosis after the implementation of primary percutaneous coronary intervention as the standard treatment for ST-elevation myocardial infarction. PhD thesis. *2012*.

72. Anna Maria Platon: The impact of chronic obstructive pulmonary disease on intensive care unit admission and 30-day mortality in patients undergoing colorectal cancer surgery: a Danish population-based cohort study. Research year report. *2012*.

73. Rune Erichsen: Prognosis after Colorectal Cancer - A review of the specific impact of comorbidity, interval cancer, and colonic stent treatment. PhD thesis. *2013*.

74. Anna Byrjalsen: Use of Corticosteroids during Pregnancy and in the Postnatal Period and Risk of Asthma in Offspring - A Nationwide Danish Cohort Study. Research year report. *2013*.

75. Kristina Laugesen: In utero exposure to antidepressant drugs and risk of attention deficit hyperactivity disorder (ADHD). Research year report. *2013*.

76. Malene Kærslund Hansen: Post-operative acute kidney injury and five-year risk of death, myocardial infarction, and stroke among elective cardiac surgical patients: A cohort study. Research year report. *2013*.

77. Astrid Blicher Schelde: Impact of comorbidity on the prediction of first-time myocardial infarction, stroke, or death from single-photon emission computed tomography myocardial perfusion imaging: A Danish cohort study. Research year report. *2013*.

78. Risiko for kræft blandt patienter med kronisk obstruktiv lungesygdom (KOL) i Danmark. (Online publication only). *2013*.

79. Kirurgisk fjernelse af milten og risikoen for efterfølgende infektioner, blodpropper og død. Danmark 1996-2005. (Online publication only). *2013*.

    Jens Georg Hansen: Akut rhinosinuitis (ARS) – diagnostik og behandling af voksne i almen praksis. *2013*.

80. Henrik Gammelager: Prognosis after acute kidney injury among intensive care patients. PhD thesis. *2014*.

81. Dennis Fristrup Simonsen: Patient-Related Risk Factors for Postoperative Pneumonia following Lung Cancer Surgery and Impact of Pneumonia on Survival. Research year report. *2014*.

82. Anne Ording: Breast cancer and comorbidity: Risk and prognosis. PhD thesis. *2014*.

83. Kristoffer Koch: Socioeconomic Status and Bacteremia: Risk, Prognosis, and Treatment. PhD thesis. *2014*.

84. Anne Fia Grann: Melanoma: the impact of comorbidities and postdiagnostic treatments on prognosis. PhD thesis. *2014*.

85. Michael Dalager-Pedersen: Prognosis of adults admitted to medical departments with community-acquired bacteremia. PhD thesis. *2014*.

86. Henrik Solli: Venous thromboembolism: risk factors and risk of subsequent arterial thromboembolic events. Research year report. *2014*.

87. Eva Bjerre Ostenfeld: Glucocorticoid use and colorectal cancer: risk and postoperative outcomes. PhD thesis. *2014*.

88. Tobias Pilgaard Ottosen: Trends in intracerebral haemorrhage epidemiology in Denmark between 2004 and 2012: Incidence, risk-profile and case-fatality. Research year report. *2014*.

89. Lene Rahr-Wagner: Validation and outcome studies from the Danish Knee Ligament Reconstruction Registry. A study in operatively treated anterior cruciate ligament injuries. PhD thesis. *2014*.

90. Marie Dam Lauridsen: Impact of dialysis-requiring acute kidney injury on 5-year mortality after myocardial infarction-related cardiogenic shock - A population-based nationwide cohort study. Research year report. *2014*.

91. Ane Birgitte Telén Andersen: Parental gastrointestinal diseases and risk of asthma in the offspring. A review of the specific impact of acid-suppressive drugs, inflammatory bowel disease, and celiac disease. PhD thesis. *2014*.

    Mikkel S. Andersen: Danish Criteria-based Emergency Medical Dispatch – Ensuring 112 callers the right help in due time? PhD thesis. *2014*.

92. Jonathan Montomoli: Short-term prognosis after colorectal surgery: The impact of liver disease and serum albumin. PhD thesis. *2014*.

93. Morten Schmidt: Cardiovascular risks associated with non-aspirin non-steroidal anti-inflammatory drug use: Pharmacoepidemiological studies. PhD thesis. *2014*.

94. Betina Vest Hansen: Acute admission to internal medicine departments in Denmark - studies on admission rate, diagnosis, and prognosis. PhD thesis. *2015*.

95. Jacob Gamst: Atrial Fibrillation: Risk and Prognosis in Critical Illness. PhD thesis. *2015*.

96. Søren Viborg: Lower gastrointestinal bleeding and risk of gastrointestinal cancer. Research year report. *2015*.

97. Heidi Theresa Ørum Cueto: Folic acid supplement use in Danish pregnancy planners: The impact on the menstrual cycle and fecundability. PhD thesis. *2015*.

98. Niwar Faisal Mohamad: Improving logistics for acute ischaemic stroke treatment: Reducing system delay before revascularisation therapy by reorganisation of the prehospital visitation and centralization of stroke care. Research year report. *2015*.

99. Malene Schou Nielsson: Elderly patients, bacteremia, and intensive care: Risk and prognosis. PhD thesis. *2015*.

100. Jens Tilma: Treatment Injuries in Danish Public Hospitals 2006-2012. Research year report. *2015*.

101. Thomas Lyngaa: Intensive care at the end-of-life in patients dying of cancer and non-cancer chronic diseases: A nationwide study. Research year report. *2015*.

102. Lone Winther Lietzen: Markers of immune competence and the clinical course of breast cancer. PhD thesis. *2015*.

103. Anne Høy Seemann Vestergaard: Geographical Variation in Use of Intensive Care in Denmark: A Nationwide Study. Research year report. *2015*.

104. Cathrine Wildenschild Nielsen: Fecundability among Danish pregnancy planners. Studies on birth weight, gestational age and history of miscarriage. PhD thesis. *2015*.

105. Kathrine Dyhr Lycke: Preadmission use of antidepressants and quality of care, intensive care admission and mortality of colorectal cancer surgery – a nationwide population-based cohort study. Research year report. *2015*.

106. Louise Bill: Hyponatremia in acute internal medicine patients: prevalence and prognosis. PhD thesis. *2015*.

107. Kirstine Kobberøe Søgaard: Risk and prognosis of venous thromboembolism in patients with liver disease. PhD thesis. *2015*.

108. Rikke Nørgaard Pedersen: Reoperation due to surgical bleeding in breast cancer patients and breast cancer recurrence: A Danish population-based cohort study. Research year report. *2015*.

109. Thomas Deleuran: Cirrhosis of the liver and diseases of the large joints. PhD Thesis. *2016*.

110. Anne Mette Falstie-Jensen: Hospital accreditation – what's in it for the patients? PhD thesis. *2016*.

111. Sandra Kruchov Thygesen: Respiratory distress syndrome in moderately late and late preterm infants and selected long-term neurodevelopmental outcomes. PhD thesis. *2016*.

112. Alma Bečić Pedersen: Total hip replacement surgery - occurrence and prognosis. Doctoral dissertation. *2016*.

113. Anil Mor: Type 2 Diabetes and Risk of Infections. PhD thesis. *2016*.

114. Aske Hess Rosenquist: Behavioral Development Following Early Life Organochlorine Exposure. Research year report. *2016*.

115. Simon Ramsdal Sørensen: Anti-platelet and Anti-coagulant Prescriptions and Breast Cancer Recurrence: a Danish Nationwide Prospective Cohort Study. Research year report. *2016*.

116. Regional Differences in Treatment of Patients with Inflammatory Bowel Disease in Denmark

117. Clara Reece Medici: Impact of preadmission anti-inflammatory drug use on risk of depression and anxiety after intensive care requiring mechanical ventilation. Research year report. *2016*.

118. Johan Frederik Håkonsen Arendt. Clinical implications and biochemical understanding of high plasma vitamin B12 levels. PhD thesis. *2016*.

119. Manual for using the LABKA database for research projects. *2016*.

120. Søren Christiansen: Timing of renal replacement therapy and long-term risk of chronic kidney disease and death in intensive care patients with acute kidney injury. Research year report. *2017*.

121. Ellen Hollands Steffensen: Preadmission antidepressant use and bladder cancer: a population-based cohort study of stage at diagnosis, time to surgery, and surgical outcomes. Research year report. *2017*.

122. Sara Søndergaard Schwartz: Incidence and mortality of pulmonary hypertension in adults with congenital heart disease. Research year report. *2017*.

123. Jesper Smit: Community-acquired Staphylococcus aureus bacteremia: Studies of risk and prognosis with special attention to diabetes mellitus and chronic heart failure. PhD thesis. *2017*.

124. Carina Nørskov Bagge: Risk of Dementia in Adults with Congenital Heart Disease: A Nationwide Population-Based Cohort Study. Research year report. *2017*.

125. Pia Kjær Kristensen: Hip fracture in Denmark: Quality of in-hospital care and clinical outcomes. PhD thesis. *2017*.

126. Anne Nakano Jensen: Incident heart failure in Denmark: Studies on a nationwide quality improvement initiative. PhD thesis. *2017*.

127. Kasper Adelborg: Neurological and psychiatric comorbidity in patients with heart failure − Risk and prognosis. PhD thesis. *2017*.

128. Jens Sundbøll: Depression, stroke, and dementia in patients with myocardial infarction — Studies of risk and prognosis. PhD thesis. *2017*.

129. Sigrún Alba Jóhannesdóttir Schmidt: Herpes zoster: Occurrence and risk factors. PhD thesis. *2017*.

130. Mette Glavind Bülow Pedersen: Ischemic Stroke in Adults with Congenital Heart Disease: a population-based cohort study. Research year report. *2017*.

131. Sidse Høyer: Male alcohol intake and couples' time to pregnancy - A prospective cohort study. Research year report. *2017*.

132. Cecilie Daugaard: Socioeconomic inequality in drug reimbursement during end-of-life care: A nationwide study. Research year report. *2018*.

133. Stine Bakkensen Bruun: Selective Serotonin Reuptake Inhibitor Use and Mortality, Postoperative Complications, and Quality of Care in Hip Fracture Patients - A population-based cohort study. Research year report. *2018*.

134. Maria Møller Pedersen: Prognostic role of neutrophil-lymphocyte ratio in localized and metastatic renal cell carcinoma: A population-based cohort study. Research year report. *2018*.

135. Hanne Vangsgaard Nielsen: Proton pump inhibitor prescriptions and breast cancer recurrence: A Danish nationwide prospective cohort study. Research year report. *2018*.

136. Jaimie Gradus: The Long-Term Consequences of Trauma and Stress Disorders: Evidence from the Danish Registries. Doctoral dissertation. *2018*.

137. Mette Holland-Fischer: Rheumatoid arthritis, ankylosing spondylitis and hospitalization with pneumonia. PhD thesis. *2018*.

138. Kaja Kristiane Eriksrud Kjørholt: Increasing risk of postoperative infections among hip fracture patients: A nationwide study 2005-2016. Research year report. *2018*.

139. Troels Munch: Use of opioids prior to hospital admission among the nonsurgical critically ill: Studies on risk and prognosis. PhD thesis. *2018*.

140. Charlotte Slagelse Jensen-Haarup: Colorectal Cancer Surgery and Acute Kidney Injury: A Review with Special Reference to the Risk and Prognosis of Renin-Angiotensin System Blocker Use. PhD thesis. *2018*.

141. Lisbeth Wellejus Albertsen: Development and validation of the DANish Comorbidity index for Acute Myocardial Infarction (DANCAMI): One-year mortality prediction. Research year report. *2019*.

142. Frederikke Schønfeldt Troelsen: Risk and prognosis of primary liver, gallbladder, bile duct, and pancreatic cancer a negative CT scan of the abdomen: A Danish population-based cohort study. Research year report. *2019*.

143. Emil Zâl Bjerregaard Riahi: Thromboembolic and bleeding complications in patients with liver cirrhosis and atrial fibrillation – a population-based cohort study. Research year report. *2019*.

144. Kristine Fogh Andersen: Hematuria and subsequent long-term risk of chronic kidney disease. A Danish population-based cohort study. Research year report. *2019*.

145. Diana Hedevang Christensen: Diabetic polyneuropathy in type 2 diabetes. Prevalence, risk factors, mental health, and diagnostic validity. PhD thesis. *2019*.

146. Kristina Laugesen: Glucocorticoid treatment: Population-based studies on utilization and selected adverse effects. PhD thesis. *2020*.

147. Frederik Pagh Kristensen: Statin therapy and risk of diabetic polyneuropathy: A population-based cohort study. Research year report. *2020*.

148. Jeppe Damgren Vesterager: Hospital variation in the risk of infection after hip fracture surgery: A population-based cohort study. Research year report. *2020*.

149. Katrine Hjuler Lund: Perceived stress and in-home assessed semen quality - A cross-sectional study. Research year report. *2020*.

150. Jakob Schöllhammer Knudsen: Type 2 diabetes in real-life: Population-based Danish studies of incidence, prognosis, and treatment effectiveness. PhD thesis. *2020*.

151. Anne Gedebjerg: Complications of type 2 diabetes: Prevalence and association with mannose-binding lectin. PhD thesis. *2020*.

152. Philip Munch: Is the risk of cardiovascular disease increased in living kidney donors? A Danish population-based cohort study. Research year report. *2020*.

    Henrik Toft Sørensen: Studies on the Associations among Venous Thromboembolism, Cancer, Arterial Cardiovascular Events, and Mortality. Doctoral Dissertation, Higher Doctorate (Doctor of Science). *2021*.

153. Anne Staub Rasmussen: Surgery during pregnancy – risk and prognosis. PhD thesis. *2021*.

154. Søren Viborg Vestergaard: Epidemiology of Nephrotic Syndrome - Occurrence and Prognosis in a Nationwide Cohort of Adults. PhD thesis. *2021*.

155. Rikke Nørgaard: Late breast cancer recurrence - Risk, prognosis and prevention. PhD thesis. *2021*.

156. Nina McKinnon Edwards: The impact of socioeconomic status on total hip arthroplasty - Utilization and postoperative complications. PhD thesis. *2021*.

157. Maria Daniella Bergholt: Impact of first-time hospital accreditation on recommended care, patient experiences and clinical outcomes. PhD thesis. *2021*.

158. Morten Schmidt: Multimorbidity and cardiovascular disease: Population-based studies of risk, prognosis, and prediction. Doctoral dissertation. *2022*.

159. Cathrine Fonnesbech Hjorth: Breast cancer survivorship after taxane-based chemotherapy - Studies in premenopausal women. PhD thesis. *2022*.

160. Thomas Johannesson Hjelholt: Associations of Hip Fracture and Stroke in Older Adults - Occurrence, Risk, and Prognosis. PhD thesis. *2022*.

161. Anne Høy Seemann Vestergaard: Care at the end of life in Denmark. The role of underlying disease and socioeconomic position. PhD thesis. *2022*.

162. Nils Skajaa: Epidemiology of Stroke in Denmark – Studies of occurrence and prognosis with an emphasis on younger adults. PhD thesis. *2022*.

163. Mette Kielsholm Thomsen: Participation, risk-selection and effect on mortality in the Danish screening program for colorectal cancer. PhD thesis. *2022*.

164. Buket Öztürk Esen: Studies of antidepressant use in pregnancy: Biases and how to address them. PhD thesis. *2022*.

165. Elena Dudukina: Threatened abortion and the long-term health of children and mothers. PhD thesis. *2022*.

166. Lise Skovgaard Svingel: Epidemiology of acute pyelonephritis: Occurrence, microbiology, and prognosis. PhD thesis. *2023*.