

# 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 <sup>1</sup>, Reimar W Thomsen,<sup>1</sup> Cathrine J Lau,<sup>2</sup> Henrik T Sørensen,<sup>1</sup> Lars Pedersen<sup>1</sup>

**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/bmjdr-2022-002946

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



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

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  $\geq$ 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.

## INTRODUCTION

Pre-diabetes is defined by glucose levels that are elevated, but below the threshold for diagnosing overt diabetes. In 2011, the WHO

## 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.

concluded that measurements of glycated hemoglobin (HbA1c)  $\geq$ 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.<sup>1 2</sup> Since then, HbA1c testing has been used for both screening and diagnosing type 2 diabetes, as well as for making treatment decisions.<sup>3–6</sup> It is currently one of the most commonly used blood tests in routine clinical care.<sup>7</sup>

Individuals with pre-diabetes are at increased risk of later developing type 2 diabetes.<sup>1–4 8–10</sup> 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,<sup>9–15</sup> 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<sup>16</sup> 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<sup>17</sup> and nationwide medical registries. Denmark has a tax-supported healthcare system that ensures unfettered access to medical care for all residents<sup>18</sup> (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.<sup>19</sup>

The Danish National Health Survey<sup>17</sup> 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<sup>7</sup> and the regional Clinical Laboratory Information System Research Database at Aarhus University<sup>7</sup> (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.<sup>7</sup> 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<sup>20</sup>; 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<sup>21</sup>; 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<sup>2</sup> (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  $\geq$ 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,<sup>22–24</sup> 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  $\geq$ 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,<sup>10 25–27</sup> 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); socio-economic 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<sup>28</sup> 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.<sup>29</sup> 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),<sup>30</sup> 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)<sup>31</sup> 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  $\geq$ 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<sup>2</sup> (IQR 24.1–29.8 kg/m<sup>2</sup>). 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  $\geq$ 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

	n (%) or median (IQR)	Missing values (%)	HR	
			HbA1c $\geq$ 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 <sup>2</sup> )	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<sup>2</sup>), 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<sup>2</sup>), 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

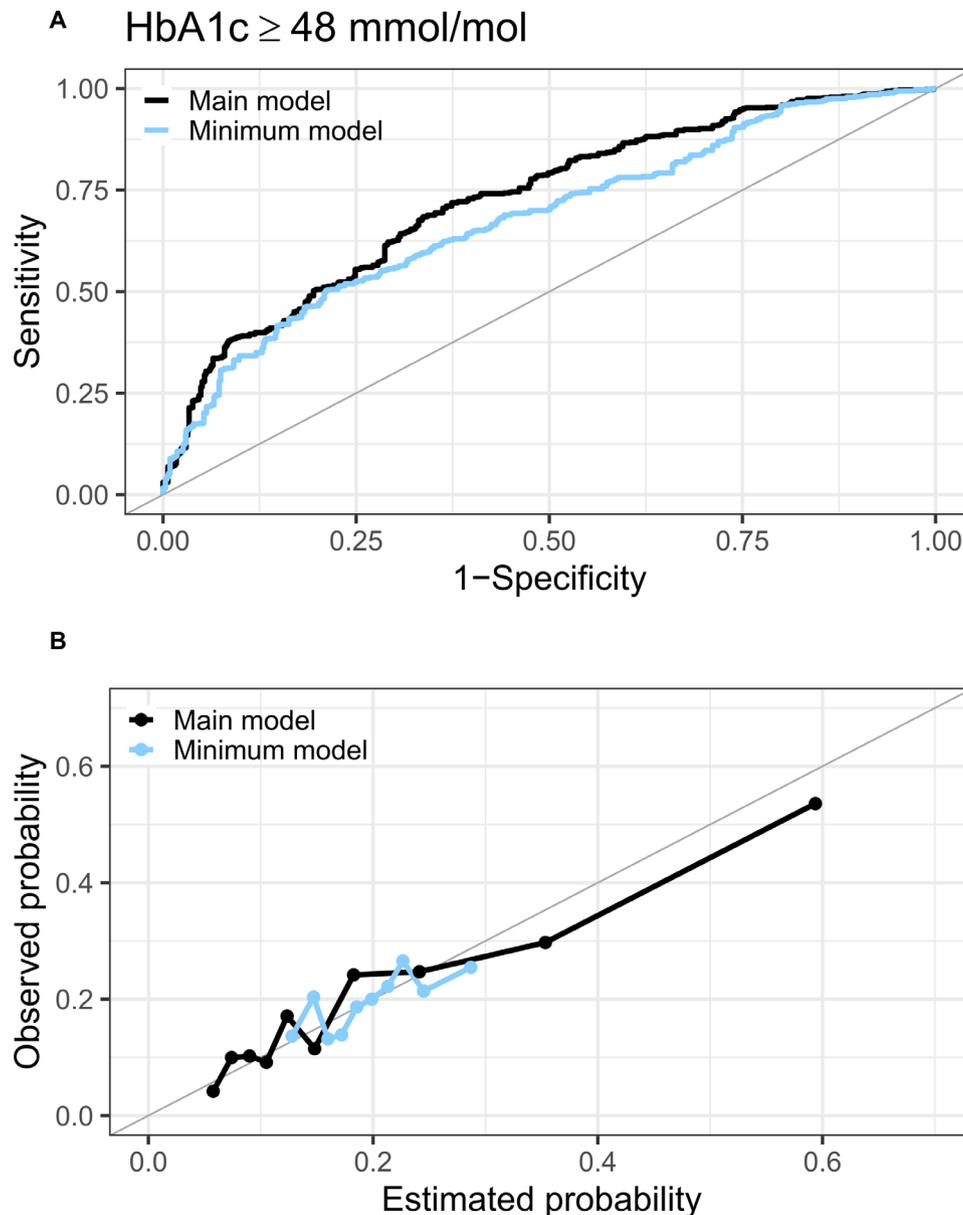
**Table 2** Performance measures for the prediction models

	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  $\geq$ 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.

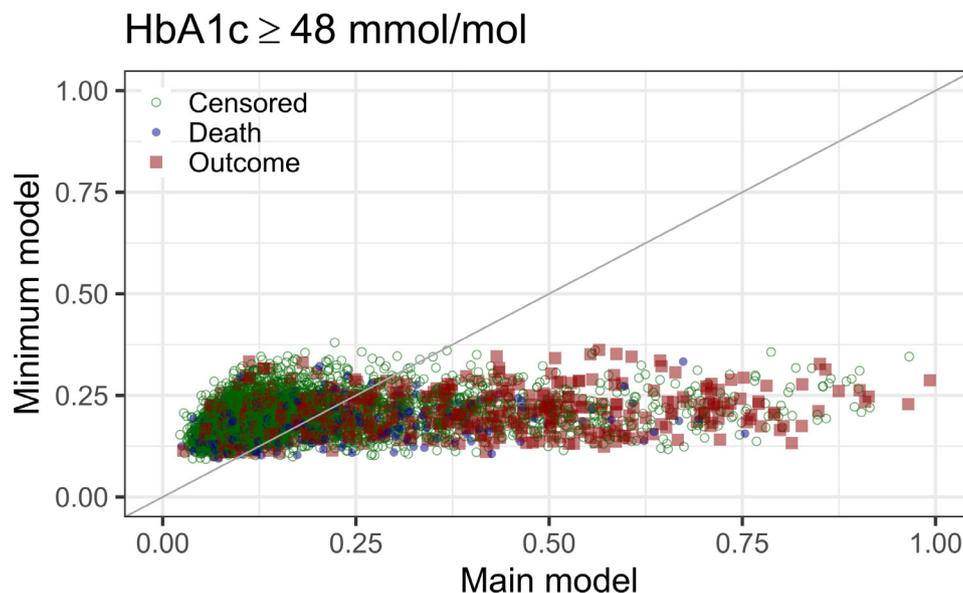
(AUC<sub>t</sub> 72.9 (95% CI 71.3 to 74.4) for HbA1c  $\geq$ 48 mmol/mol (6.5%) and AUC<sub>t</sub> 79.2 (95% CI 77.6 to 80.8) for glucose-lowering treatment initiation) and calibration when compared with the main models fitted to the entire development sample (online supplemental material figure S11).

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 15737 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 ( $42 \leq \text{HbA1c} \leq 47$  mmol/mol ( $6.0\% \leq \text{HbA1c} \leq 6.4\%$ )), 166 (1.1%) due to prior type 2 diabetes (HbA1c  $\geq$ 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  $\geq 48$  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 AUCs 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,<sup>12 27 32 33</sup> 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<sup>2</sup><sup>34</sup>), an observed 14% of individuals with pre-diabetes (HbA1c 39–47 mmol/mol (5.7%–6.4%)) developed diabetes (HbA1c  $\geq 48$  mmol/mol (6.5%)) within 5 years.<sup>34</sup> 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  $\geq 48$  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<sup>2</sup>),<sup>11 32</sup> 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  $\geq 7.0$  mmol/l or 2hPG  $\geq 11.0$  mmol/l) within 5 years. As only 26% of the DPPOS participants with diabetes according to glucose criteria also had HbA1c levels  $\geq 48$  mmol/mol (6.5%),<sup>32</sup> we could not make a direct comparison with our study<sup>34</sup>; however, our estimates (19% for diabetes defined by HbA1c  $\geq 48$  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<sup>2</sup> in DPPOS vs median 26.7 kg/m<sup>2</sup> 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.<sup>11</sup>

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.<sup>35</sup> They highlighted the need for further research on factors associated with risk of progression from pre-diabetes to overt diabetes.<sup>35</sup> 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<sup>22</sup>), 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.<sup>36</sup> Mechanisms are unclear, but loneliness may associate with dysregulation in cortisol responses and heightened inflammation.<sup>36</sup>

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.<sup>37</sup>

In an American study<sup>33</sup> assessing the performance of HbA1c in predicting long-term diabetes (glucose-lowering treatment, FPG  $\geq 7$  mmol/l, HbA1c  $\geq 48$  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  $\geq 48$  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.<sup>17</sup> 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,<sup>37</sup> and our population-based pre-diabetes cohort was rather old (median age 69.6 years) compared with other pre-diabetes cohorts.<sup>11 34</sup> 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<sup>10</sup> and other variable selection strategies may have improved the model validity. We included ethnic origin as a potential predictor for developing diabetes,<sup>14 32 38</sup> 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.<sup>39 40</sup> 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<sup>41</sup> and substratification using new pre-diabetes phenotypes and genetic risk scores<sup>42</sup> 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

<sup>1</sup>Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark

<sup>2</sup>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>).

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Sia K Nicolaisen <http://orcid.org/0000-0002-6106-1945>

#### REFERENCES

- 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.
- 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.
- Farmakologisk behandling af type 2 diabetes – mål og algoritmer - 2018 [article online], 2018. Available: <https://vejledning.dsam.dk/fbv-t2dm/> [Accessed 13 Apr 2022].
- 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.
- 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.
- 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.
- 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.
- Makaroff LE. The need for international consensus on prediabetes. *Lancet Diabetes Endocrinol* 2017;5:5–7.
- 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.
- Tabák AG, Herder C, Rathmann W, *et al*. Prediabetes: a high-risk state for diabetes development. *Lancet* 2012;379:2279–90.
- 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.
- 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.
- Lindström J, Tuomilehto J. The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* 2003;26:725–31.
- 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.
- 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.
- 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.
- 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.
- 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.
- Schmidt M, Pedersen L, Sørensen HT. The Danish civil registration system as a tool in epidemiology. *Eur J Epidemiol* 2014;29:541–9.
- 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.
- Kildemoes HW, Sørensen HT, Hallas J. The Danish national prescription registry. *Scand J Public Health* 2011;39:38–41.
- 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].
- 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<sub>1c</sub> 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, Büyükközkcan 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.