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

Sia K Nicolaisen, Reimar W Thomsen, Cathrine J Lau, Henrik T Sørensen, Lars Pedersen

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 prediabetes.

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-Grey 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.
Individuals with pre-diabetes are at increased risk of later developing type 2 diabetes. 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, 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 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 and nationwide medical registries. Denmark has a tax-supported healthcare system that ensures unfettered access to medical care for all residents (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.

The Danish National Health Survey 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 and the regional Clinical Laboratory Information System Research Database at Aarhus University (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. 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; 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; 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 (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, 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,\textsuperscript{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\textsuperscript{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.\textsuperscript{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).\textsuperscript{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)\textsuperscript{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), 260,071 (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 260,071 individuals were randomly divided into a development sample (n=20,806) and a validation sample (n=5,201). 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/m² (IQR 24.1–29.8 kg/m²). 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, 2,449 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 4,026 (19.4%) individuals were followed for at least 5 years. During the same period, 1,339 (6.4%) individuals initiated a glucose-lowering treatment indicating type 2 diabetes, and a total of 2,101 (10.1%) died (online supplemental material figure S6).
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/m²), 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 poor/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/m²), 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.
Epidemiology/Health services research

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≤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 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.
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, 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/m²), 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. 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/m²), 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%), we could not make a direct comparison with our study; 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/m² in DPPOS vs median 26.7 kg/m² 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.

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. They highlighted the need for further research on factors associated with risk of progression from pre-diabetes to overt diabetes. In addition to some important and previously known predictors of developing type 2 diabetes—younger age at onset of 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).
(often associated with more obesity and a more severe pre-diabetes phenotype), 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. Mechanisms are unclear, but loneliness may associate with dysregulation in cortisol responses and heightened inflammation.

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.

In an American study 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. As individuals from higher socioeconomic groups were more likely to respond than those from lower sociodemographic groups, this may have led to an underestimation 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, and our population-based pre-diabetes cohort was rather old (median age 69.6 years) compared with other pre-diabetes cohorts. 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 and other variable selection strategies may have improved the model validity. We included ethnic origin as a potential predictor for developing diabetes, 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. 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 prediabetes 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 and substratification using new prediabetes phenotypes and genetic risk scores 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.

**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 glyceded haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a who consultation*. Geneva: World Health Organization, 2011.

3. Farmakologisk behandling av type 5 diabetes – mål og algoritmer - 2018 [article online], 2018. Available: https://vejledninger.dsam.dk/fbvy/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: 2019.

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, Achara 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;379:2579–90.

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, Leenig 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* 2015;4:441–51.

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. Jalle 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* 2017;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, Schmid SJ, 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 SJ, 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, Pyonävä K. Age of onset and type of diabetes. *Diabetes Care* 1985;8:114–7.
Epidemiology/Health services research

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;10:e0001071.

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 A1c 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, Büyükozkan 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 Ahlvqvist 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. Nicolaisen¹, Reimar W. Thomsen¹, Cathrine Lau², Henrik T. Sørensen¹, Lars Pedersen¹

¹Department of Clinical Epidemiology, Aarhus University Hospital and Aarhus University, Aarhus, Denmark
²Center 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
Supplemental Material Figure S1
Supplemental Material Figure S2
Supplemental Material Table S2
Supplemental Material Figure S3
Supplemental Material Table S3
Supplemental Material Figure S4
Supplemental Material Figure S5
Supplemental Material Table S4
Supplemental Material Figure S6
Supplemental Material Table S5
Supplemental Material Figure S7
Supplemental Material Table S6
Supplemental Material Table S7
Supplemental Material Figure S8
Supplemental Material Figure S9
Supplemental Material Figure S10
Supplemental Material Figure S11
Supplemental Material Figure S12
Supplemental Material Figure S13
Supplemental Material Figure S14
Supplemental Material Figure S15
Supplemental Material Figure S16
Supplemental Material Figure S17
Supplemental Material Figure S18
Supplemental Material Figure S19
Supplemental Material Figure S20
Supplemental Material Figure S21
Supplemental Material Figure S22
Supplemental Material Figure S23
Supplemental Material Figure S24
Supplemental Material Figure S25
Supplemental Material Figure S26
Supplemental Material Figure S27
Supplemental Material Figure S28
Supplemental Material Figure S29
Supplemental Material Figure S30
Supplemental Material Figure S31
Supplemental Material Figure S32
Supplemental Material Figure S33
Supplemental Material Figure S34
Supplemental Material Figure S35
Supplemental Material Figure S36
Supplemental Material Figure S37
Supplemental Material Figure S38
Supplemental Material Figure S39
Supplemental Material Figure S40
Supplemental Material Figure S41

Nicolaisen SK, et al. BMJ Open Diab Res Care 2022; 10:e002946. doi: 10.1136/bmjdrcc-2022-002946
1 Supplemental Material Table S1

2 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     | 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           | 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                | 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                                                                                                                                          |
| 5c | Give details of treatments received, if relevant. | Illustrated in the study design figure and flowchart. 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 is 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-tuning). | Fine-Gray survival models were fitted with LASSO regression for variable selection. |
| 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 | 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 | 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. |
| Model development | Specify the number of participants and outcome events in each analysis. | Details about outcome status are reported for each model. |
| | 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 | 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. |
| | Explain how to 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 | 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 | 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 | 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
| 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-Prediction-Model-Development.pdf](https://www.tripod-statement.org/wp-content/uploads/2020/01/Tripod-Checlist-Prediction-Model-Development.pdf)
2. We have adjusted the original table and added comments instead of page numbers.
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-denmark. 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.

- Temporal coverage in the laboratory database differs by region.
- 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.
- Danish National Health Survey data. See Supplemental Material Figure S3 for details about inclusion.
- Data on socioeconomic factors are from the end of the previous year. Data on employment are from the end of the previous November.
- 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.
1 **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                                      |
|-----------------------------------------|------------------|-----------------------------------------------------------------------------------------|---------------------------|-------------|----------------------------------------------------------|
| 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 |
## 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

|               | Time to event | N/A | Measurement during follow-up | LAB | NPU: NPU27300, NPU03835 | N/A |
|---------------|---------------|-----|------------------------------|-----|------------------------|-----|
| HbA1c>=48 mmol/mol (6.5%) | Time to event | N/A |                             | 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])

|               | Time to event | N/A | During follow-up | DCRS | N/A | N/A |
|---------------|---------------|-----|----------------|------|-----|-----|

## Potential predictors

### Demographic predictors

|               | Binary | Female, Male (ref.) | N/A | DCRS | N/A | 0 |
|---------------|--------|---------------------|-----|------|-----|---|

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).
| Data Type                                      | Level         | Example Value                      | Indicator |
|-----------------------------------------------|---------------|------------------------------------|-----------|
| Age on index date                             | Continuous, included linearly | 30-104 years                      | Index date |
| Ethnic origin                                 | Categorical   | Danish (ref.), Immigrant/descendant/un known | DST       |
| 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 |
| HbA1c measures                                | Categorical, only included in the hazard ratios (not in the prediction models) | 2012-2018 (ref. 2012)             | LAB       |
| Presence of HbA1c measurements before prediabetes | Binary        | No (ref.), Yes                      | LAB       |
| Prescription drug use                         |                |                                    | DRMPS     |
| Statins                                       | Binary        | No (ref.), Yes                      |           |
| Any antihypertensive drug                     | Binary        | No (ref.), Yes                      |           |
| Comorbidities                                                                 | Use | Before/After | Dates/Details                                                                 |
|------------------------------------------------------------------------------|-----|--------------|------------------------------------------------------------------------------|
| Oral steroids (C10BX13, C10BX14, C10BX15, C10BX10, C10BX07, C10BX09, C10BX11, C10BX14) | Binary | No (ref.), Yes | 180 days prior to the index date DRMP | ATC: H02AB | 0 |
| Opioid use (C10BX11, C10BX14)                                               | Binary | No (ref.), Yes | 180 days prior to the index date DRMP | 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, KFN, KFNC, KFND, KFNE, KFN, KFNG, KFNH20, DI21, DI22, DI23, DI50, DI60, DI61, DI63, DI64, DI48, DI05, DI06, DI07, DI08, DI098, DI39, DI511A, DJ22, DJ23, DJ34-DJ37, DJ26, DJ801, DJ802, DJ803 | 0 |
| Lung disease                                                                | Binary | No (ref.), Yes | Prescription redemption 180 days prior to the index date, or hospital contact day 5*360 days prior to the index date DRMP, DNPR | ATC: R03, ICD-10: DJ40, DJ684, DJ701, DJ703, DJ704, 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, etc.) | Binary | No (ref.), Yes | Prescription redemption 180 days prior to the index date DRMP, 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

| 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 |
|                                | Type                          | Categorical/Continuous/Continuous, linearly (categorical in sensitivity analysis) | 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 |
|--------------------------------|-------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|---------------------------------------------------------------------------------|------|-----|---|
| Highest education achieved     | Categorical                   | None, basic education, or primary school; Youth education, high school or similar educational level; Higher education (ref.) | Index date                                                                      | DST                                                                             | DST  | N/A | 2 |
| Employment status              | Categorical                   | Employed (ref.), Unemployed or not part of the workforce                        | End of previous November                                                        | DST                                                                             | 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                                                                             | 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                                                                             | 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 | 13-70 kg/m2 (<25 is reference in sensitivity analysis). | 2010/2013/2017                                                                | DNHS                                                                             | DNHS | N/A | 5 |
| Self-reported body mass index (BMI) | Continuous, included linearly (categorical in sensitivity analysis) | 0-1 days (ref.), 2-3 days, 4-7 days                                              | 2010/2013/2017                                                                | DNHS                                                                             | DNHS | N/A | 6 |
| Days per week with alcohol consumption | Categorical                   | 7/14 units or less per week (women/men) (ref.), More than 7/14                   | 2010/2013/2017                                                                | DNHS                                                                             | DNHS | N/A | 8 |
|                                                                 | Level  | Description                                                                 | Year(s)     | Database | Notes          | Value |
|-----------------------------------------------------------------|--------|-----------------------------------------------------------------------------|-------------|----------|----------------|-------|
| 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     |

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; DNHS, Danish National Health Survey; NPU, laboratory codes in the Nomenclature for Properties and Units; ICD-10, *International Statistical Classification of Diseases and Related Health Problems*. 
1 Classification of Diseases and Related Health Problems, Tenth Revision; ATC Anatomical Therapeutic Chemical Codes; N/A, not applicable.

2

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  |
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%).

See Christensen et al.

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.

- **Diabetes defined by HbA1c ≥48 mmol/mol**
- **Glucose-lowering treatment initiation**
- **Death**

**Cumulative incidence, %**

- 19.3% (95% CI 18.6-20.0)
- 16.3% (95% CI 15.6-16.9)
- 11.2% (95% CI 10.6-11.8)

**Years**
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.

| Demographic variables | Development sample (80%) | Validation sample (20%) |
|-----------------------|---------------------------|------------------------|
|                       | N (% or Median (IQR))     | Hazard ratio           | N (% or Median (IQR)) | Hazard ratio           |
| Total                 | 20,806 (100.0%)           |                        | 5,201 (100.0%)        |                        |
| **Demographic variables** |                          |                        |                       |                        |
| Sex                   |                           |                        |                       |                        |
| Female                | 10,792 (51.9%)            | 0.67 (0.62; 0.73)      | 2,704 (52.0%)         | 0.68 (0.61; 0.76)      |
| Male                  | 10,014 (48.1%)            | ref.                   | 2,497 (48.0%)         | ref.                   |
| Age (years)           | 69.6 (61.0-77.1)          | 0.99 (0.98; 0.99)      | 69.6 (61.6-77.1)      | 0.97 (0.96; 0.97)      |
| Age                   | 0 (0.0%)                  | ref.                   | 0 (0.0%)              | ref.                   |
| 30-39 years           | 219 (1.1%)                | ref.                   | 51 (1.0%)             | ref.                   |
| 40-49 years           | 1,082 (5.2%)              | 0.94 (0.66; 1.36)      | 280 (5.4%)            | 0.78 (0.53; 1.16)      |
| 50-59 years           | 3,390 (16.3%)             | 0.77 (0.55; 1.09)      | 792 (15.2%)           | 0.53 (0.36; 0.76)      |
| 60-69 years           | 6,023 (28.9%)             | 0.62 (0.45; 0.88)      | 1,580 (30.4%)         | 0.39 (0.27; 0.56)      |
| 70-79 years           | 6,452 (31.0%)             | 0.58 (0.41; 0.81)      | 1,621 (31.2%)         | 0.31 (0.22; 0.45)      |
| 80-89 years           | 3,132 (15.1%)             | 0.64 (0.45; 0.91)      | 748 (14.4%)           | 0.21 (0.14; 0.32)      |
| ≥90 years             | 508 (2.4%)                | 0.28 (0.16; 0.49)      | 129 (2.5%)            | 0.14 (0.06; 0.29)      |
| Ethnic origin         | 0 (0.0%)                  | ref.                   | 0 (0.0%)              | ref.                   |
| Danish                | 19,586 (94.1%)            | ref.                   | 4,899 (94.2%)         | ref.                   |
| Immigrant/descendant/ | 1,220 (5.9%)              | 1.19 (1.02; 1.39)      | 302 (5.8%)            | 1.16 (0.95; 1.42)      |
| unknown               |                           |                        |                       |                        |
| **HbA1c measures**    |                           |                        |                       |                        |
| Value of prediabetes-defining HbA1c measurement (mmol/mol) | 43.0 (42.0-44.0) | 1.69 (1.65; 1.73) | 43.0 (42.0-44.0) | 0 (0.0%) |
| Value of prediabetes-defining HbA1c measurement (%) | 6.1 (6.0-6.2) | N/A | 6.1 (6.0-6.2) | 0 (0.0%) |
| Value of prediabetes-defining HbA1c measurement | 0 (0.0%) | N/A | 0 (0.0%) | N/A |
| Development sample (80%) | Validation sample (20%) |
|--------------------------|-------------------------|
| **N (%) or Median (IQR)** | **Missing values** | **Hazard ratio** | **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%) | 0 (0.0%) |
| 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%) | 0 (0.0%) |
| 44 mmol/mol (6.2%) | 3,080 (14.8%) | 2.74 (2.41; 3.12) | 2.89 (2.43; 3.44) | 727 (14.0%) | 0 (0.0%) |
| 45 mmol/mol (6.3%) | 1,794 (8.6%) | 5.15 (4.53; 5.86) | 5.24 (4.39; 6.24) | 440 (8.5%) | 0 (0.0%) |
| 46 mmol/mol (6.4%) | 1,180 (5.7%) | 7.93 (6.93; 9.07) | 8.04 (6.71; 9.62) | 312 (6.0%) | 0 (0.0%) |
| 47 mmol/mol (6.4%) | 610 (2.9%) | 13.69 (11.75; 15.94) | 12.48 (10.15; 15.34) | 161 (3.1%) | 0 (0.0%) |

**Presence of HbA1c measurements before prediabetes**
- Yes: 5,884 (28.3%)
  - Hazard ratio: 0.73 (0.66; 0.80)
  - Glucose-lowering treatment initiation: 0.78 (0.68; 0.89)
  - N (%) or Median (IQR): 1,491 (28.7%)
- No: 14,922 (71.7%)
  - Hazard ratio: ref.
  - Glucose-lowering treatment initiation: ref.
  - N (%) or Median (IQR): 3,710 (71.3%)

**Prescription drug use**
- Statins: 7,706 (37.0%)
  - Hazard ratio: 1.03 (0.95; 1.12)
  - Glucose-lowering treatment initiation: 0.99 (0.88; 1.11)
  - N (%) or Median (IQR): 1,942 (37.3%)
- Any antihypertensive drug: 12,591 (60.5%)
  - Hazard ratio: 1.33 (1.22; 1.45)
  - Glucose-lowering treatment initiation: 1.31 (1.17; 1.47)
  - N (%) or Median (IQR): 3,158 (60.7%)
- Oral steroids: 1,505 (7.2%)
  - Hazard ratio: 1.26 (1.08; 1.46)
  - Glucose-lowering treatment initiation: 1.02 (0.81; 1.28)
  - N (%) or Median (IQR): 393 (7.6%)
- Opioid use: 2,601 (12.5%)
  - Hazard ratio: 1.21 (1.08; 1.36)
  - Glucose-lowering treatment initiation: 1.33 (1.14; 1.56)
  - N (%) or Median (IQR): 670 (12.9%)

**Comorbidities**
- Charlson Comorbidity Index score
  - 0-2: 19,544 (93.9%)
  - Hazard ratio: ref.
  - Glucose-lowering treatment initiation: ref.
  - N (%) or Median (IQR): 4,896 (94.1%)
  - ≥3: 1,262 (6.1%)
  - Hazard ratio: 1.25 (1.05; 1.48)
  - Glucose-lowering treatment initiation: 0.99 (0.75; 1.29)
  - N (%) or Median (IQR): 305 (5.9%)
- Cardiovascular disease: 4,530 (21.8%)
  - Hazard ratio: 1.03 (0.93; 1.14)
  - Glucose-lowering treatment initiation: 0.95 (0.83; 1.10)
  - N (%) or Median (IQR): 1,156 (22.2%)
- Lung disease: 3,096 (14.9%)
  - Hazard ratio: 1.16 (1.04; 1.29)
  - Glucose-lowering treatment initiation: 0.99 (0.84; 1.16)
  - N (%) or Median (IQR): 810 (15.6%)
- Cancer: 2,024 (9.7%)
  - Hazard ratio: 1.26 (1.08; 1.46)
  - Glucose-lowering treatment initiation: 1.01 (0.82; 1.24)
  - N (%) or Median (IQR): 477 (9.2%)
- Pancreatitis/cancer (includes pancreatic cancer, pancreas resection, and acute or chronic pancreatitis): 67 (0.3%)
  - Hazard ratio: 3.08 (1.89; 5.04)
  - Glucose-lowering treatment initiation: 3.48 (1.92; 6.31)
  - N (%) or Median (IQR): 16 (0.3%)
- Possible HbA1c-modifying conditions: 1,499 (7.2%)
  - Hazard ratio: 1.11 (0.96; 1.29)
  - Glucose-lowering treatment initiation: 1.10 (0.90; 1.35)
  - N (%) or Median (IQR): 390 (7.5%)

**Socioeconomic variables**
- Highest education achieved: 425 (2.0%)
  - Hazard ratio: 1.67 (1.47; 1.89)
  - Glucose-lowering treatment initiation: 1.58 (1.32; 1.89)
  - N (%) or Median (IQR): 1,300 (25.0%)
- Socioeconomic status: 89 (1.7%)
| Development sample (80%) | Validation sample (20%) |
|--------------------------|-------------------------|
| **N (%) or Median (IQR)** | **Hazard ratio** | **N (%) or Median (IQR)** | **Missing values** |
| **HbA1c ≥48 mmol/mol (6.5%)** | **Glucose-lowering treatment initiation** | **Missing values** | **Median (IQR)** |
| 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 | 6,398 (30.8%) | 0 (0.0%) | 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%) | 0.88 (0.80; 0.96) | 0.89 (0.79; 1.01) | 3,550 (68.3%) |
| Not living alone | 14,479 (69.6%) | ref. | ref. | 1,651 (31.7%) |
| **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/m²) | 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/m² | 6,677 (32.1%) | ref. | ref. | 1,654 (31.8%) |
| 25-30 kg/m² | 8,342 (40.1%) | 1.50 (1.34; 1.67) | 1.67 (1.43; 1.96) | 2,045 (39.3%) |
| 30-35 kg/m² | 3,455 (16.6%) | 2.23 (1.98; 2.51) | 2.66 (2.25; 3.15) | 926 (17.8%) |
| ≥35 kg/m² | 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%) |
|                          | Development sample (80%) | Validation sample (20%) |
|--------------------------|--------------------------|-------------------------|
|                          | N (%) or Median (IQR)    | hazard ratio            | N (%) or Median (IQR)    | Missing values |
| Alcohol consumption in relation to recommended amounts |                          |                         |                         | 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                                  |                          |                         |                         | 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%)       |
| 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%)  |
| Frequency of contact with people outside the household | Development sample (80%) | Validation sample (20%) |
|-------------------------------------------------------|---------------------------|-------------------------|
|                                                       | N (%) or Median (IQR)     | Missing values          | N (%) or Median (IQR) | Missing values |
| Often/very often                                       | 1,770 (8.5%)              |                         | 454 (8.7%)             |
| Frequency of contact with people outside the household | 756 (3.6%)                | 1.30 (1.13; 1.49)       | 1.44 (1.21; 1.71)      |
| 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%)                | 647 (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%)     |

###: at least one cell contains a number < 5 and may not be reported due to Danish regulations.
For the primary analysis, type 2 diabetes was defined as the first HbA1c measurement $\geq 48$ 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 $\geq 48$ 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 $t$. 

![Venn Diagram](image)
Supplemental Material Table S5

A) Coefficients for the main model and the minimum model for type 2 diabetes defined as an HbA1c measurement $\geq 48$ 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 $\text{SHR} = \exp(\beta)$. 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 to 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.

|                                                                 | 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, $\beta_0$                                            | 0.0922                                                                      | 0.2958                                                                       |
| 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, \( \beta_0 \) | 0.0592 | 0.1823 |
| 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) | 0.3679 (0.2374; 0.4984) |
| Body mass index (kg/m\(^2\)) | 0.0473 (0.0372; 0.0575) | 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:

\[
\text{Prob(outcome)} = 1 - \exp(-\beta_0 \times \exp(\beta_1 X_1 + \cdots + \beta_k X_k))
\]

where \( \beta_0 \) is the Breslow-type estimate of the underlying subdistribution hazard evaluated after 5 years, \( \beta_1 \) … \( \beta_k \) are the coefficients for the \( k \) predictors included in the model, and \( X_1 \) … \( X_k \) 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 \( \geq \) 48 mmol/mol (6.5%) is then calculated as:
Prob(outcome) = 1 − exp(−0.0922 × exp((62 − 60) × (−0.0124) + (45 − 42) × (0.4973) + (29 − 25) × (0.0325) + 1 × (0.1529) + 0 × (0.9607) + 0 × (−0.2686) + 1 × (0.1221) + 1 × (0.3380) + 0(0.2 × 570) + 1 × (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%.
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 (58.52-63.61) | -          |
| 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 (63.85-74.69) | 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.
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 \( \geq 48 \) 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 \( \geq 48 \) 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 \( \geq 48 \) 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 \( \geq 48 \) 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.
