Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Rawshani Aidin, Rawshani Araz, Franzén S, et al. Risk factors, mortality, and cardiovascular outcomes in patients with type 2 diabetes. N Engl J Med 2018;379:633-44. DOI: 10.1056/NEJMoa1800256

Risk Factors, Mortality and Cardiovascular Outcomes in Type 2 Diabetes

Supplementary Appendix

Table of Contents

SUPPLEMENTARY DISCUSSION OF THE STATISTICAL MODELS

This is a brief discussion regarding the statistical models used in this study. For this study, we applied a Cox proportional hazards model to estimate the relative risk associated with the number of risk factors at target for patients with type 2 diabetes, compared to matched controls. In addition to this, we used a developed application for the Cox model to estimate relative importance for risk factors with a statistical approach termed estimated explained relative risk (R^2) for a type 2 diabetes population with- and without prior comorbidities. We compared these results with a third model of relative importance based on a different statistical approach that estimates explainable log-likelihood that is explained by each variable (supplementary appendix Figure S2 and S3), which is only usable with the Cox regression model. We also assessed the optimal levels for three selected risk factors, by applying a prediction function to Cox models. Proportional hazards assumption and model fit as been scrutinized extensively for all models.

Risk factors assessed in the models

From the National Diabetes Registry, we included the following baseline variables: age, sex, duration of diabetes, body mass index, glycated hemoglobin levels, low-density lipoprotein cholesterol, systolic- and diastolic blood pressure, estimated glomerular filtration rate (eGFR, using the Modification of Diet in Renal Disease equation), ongoing use of lipid lowering and/or antihypertensive medications, smoking, physical activity (categories: never, less than once a week, once or twice a week, 3–5 times per week, or daily) and albuminuria, which was defined as either micro- or macroalbuminuria. Microalbuminuria was defined as two positive results for three samples obtained within one year, with positivity defined as a urinary albumin/creatinine ratio of 3-30 mg/mmol (about 30-300 mg/g) or urinary albumin clearance of 20-200 μg/min (20-300 mg/L) on two out of three consecutive tests. Macroalbuminuria was defined as a urinary albumin/creatinine ratio >30 mg/mmol (about 300 mg/g) or urinary albumin clearance >200 μg/min (>300 mg/L).

The following socioeconomic variables were included: country of birth was dichotomized as born in Sweden or immigrant; income was categorized into quintiles of annual income; marital status categorized into single, married/registered partner, divorced or widowed; and education was categorized into 9 years or less, or 10 years or more, including college and university degrees.

Imputation

Missing data for patients with type 2 diabetes was imputed to maximize power and avoid selection bias. We used multivariate imputation by chained equations (MICE) to impute five complete data sets with 10 iterations for each data set. MICE operate under the assumption that missing is missing completely at random (MCAR), meaning that the probability of a variable being missing in the study depends only on observed values.
MICE has several steps, first each variable with missing is imputed with mean values, temporarily setting all missing values equal to the mean observed value for that

variable. The imputed mean values are then set back to missing. In the next step, the model creates linear and logistic regressions for all continuous and categorical variables, using complete data, to predict missing values. All missing values are subsequently replaced with coefficients generated from the imputed models. These steps are repeated for all variables with missing values. Validation for imputed models are achieved by comparing distributions for all variables from complete data and imputed data, and by analyzing outcomes for complete cases only, to compare with imputed models. Supplemental Table S5 and Supplemental Figures S4 and Figure S5 show the frequency of missing data elements and the distribution of each parameter before and after the imputation.

Analysis for number of risk factors at target

We constructed a Cox regression model that includes number of risk factors at target ranging from 0 to 5, modeled as categorical variables. All Cox models are stratified for gender and use age as the time-scale. Attained age as time-scale is defined by study participant's age at entry into the study and age at which they experience an event or their follow-up is censored. Data was left-censored since we follow-up a patient with type 2 diabetes at inclusion in the registry, i.e. age at baseline until achieved age at event. Some literature suggest that attained age as time-scale provides most flexible models since hazard functions are estimated for several age-groups, and in cases were time doesn't accumulate in the same manner for all subjects. Also, this avoids the need to include an effect of age, since age did dominate the prediction models as the strongest predictor for all outcomes, moreover age did not satisfy the proportional hazards assumption. Socioeconomic variables such as income, education, marital and immigrant are also included in all models. Gender did not satisfy proportional hazards assumption, therefor we adjusted for gender without estimating its effect, but the models revealed that women had slightly increased risk for most outcomes.

The Cox model for the analysis of number of risk factor at target is depicted below. The analysis was performed for individuals with type 2 diabetes and controls, separately for each age-category. Results for the entire cohort are presented in the supplementary appendix Table S2.

This following model was performed for all outcomes

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{age}|\boldsymbol{age_0}, \boldsymbol{t_{death}}) \\ &= \boldsymbol{\lambda_0}(\boldsymbol{age}) \times \boldsymbol{exp}\boldsymbol{\beta_1}\boldsymbol{category_{j1}} + \boldsymbol{\beta_2}\boldsymbol{diabetesduration_{j2}} \\ &+ \boldsymbol{strata}(\boldsymbol{\beta_3}\boldsymbol{sex_{j3}}) + \boldsymbol{\beta_4}\boldsymbol{income_{j4}} + \boldsymbol{\beta_5}\boldsymbol{eduation_{j5}} \\ &+ \boldsymbol{\beta_6}\boldsymbol{marital_{j6}} + \boldsymbol{\beta_7}\boldsymbol{immigrant_{j7}} + \boldsymbol{\beta_8}\boldsymbol{pre\_chd_{j8}} \\ &+ \boldsymbol{\beta_9}\boldsymbol{pre\_hf_{j9}} + \boldsymbol{\beta_{10}}\boldsymbol{pre\_af_{j10}}\end{aligned}$$

The regression model for the main analysis does not include physical activity, body mass index and use of statins and antihypertensive medications since these data are not available for matched controls.

Analysis of relative importance for various risk factors

Estimated explained relative risk (R^2)

The coefficient of determination R^2 is a standard measure of explained risk in the normal linear model and the proportional hazards model can be specified through the semiparametric linear transformation family. The estimated explained relative risk measure by Glenn Heller is composed of entropy loss functions developed under the extreme value distribution and applied to three different models, called the full model, null model and degenerate model. The survival density for the null model is constructed to indicate the effect of ignoring the covariate in the model. The null model entropy is the explained risk lost due to ignoring the covariate (permutation).

The following equation was applied to asses the relative importance for each predictor:

$$R^2(\boldsymbol{\beta}) = \frac{\log\left[n^{-1}\sum_i \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)\right]}{0.5772 + \log\left[n^{-1}\sum_i \exp(\boldsymbol{\beta}^T \boldsymbol{x}_i)\right]}$$

For each outcome, we constructed a *Full* model that includes all predictors, and a separate Cox model for each predictor in the *Full* model. In the separate Cox models, we excluded a single covariate and the log mean of the predictors in this model is called the *Null* model for the excluded covariate.
Thus, estimating the R^2 for the *Full* model minus the R^2 for the *Null* model for each covariate, using the appended formula, generates the coefficient for determination R^2.

The Cox regression model used for this analysis is depicted below

(Example: The Full model for all-cause mortality)

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{age}|\boldsymbol{age_0}, \boldsymbol{t_{death}}) &= \boldsymbol{\lambda_0}(\boldsymbol{age}) \\ &\times \boldsymbol{exp}\left(\boldsymbol{rcs}(\boldsymbol{\beta_1 diabetesduration_{j1}}, \boldsymbol{4}) + \boldsymbol{strata}(\boldsymbol{\beta_2 sex_{j2}})\right. \\ &+ \boldsymbol{rcs}(\boldsymbol{\beta_3 cumulative_{BMI_{j3}}}, \boldsymbol{4}) \\ &+ \boldsymbol{rcs}\left(\boldsymbol{\beta_4 cumulative_{LDL_{j4}}}, \boldsymbol{4}\right) + \boldsymbol{rcs}(\boldsymbol{\beta_5 eGFR_{j5}}, \\ &\boldsymbol{4}) + \boldsymbol{rcs}(\boldsymbol{\beta_6 HbA1c_{j6}}, \boldsymbol{4}) + \boldsymbol{rcs}(\boldsymbol{\beta_7 systolicBP_{j7}}, \boldsymbol{4}) \\ &+ \boldsymbol{rcs}(\boldsymbol{\beta_8 diastolicDP_{j8}}, \boldsymbol{4}) + \boldsymbol{\beta_9 physical\_acitivity_{j9}} \\ &+ \boldsymbol{\beta_{10} albuminuria_{j10}} + \boldsymbol{\beta_{11} pre\_chd_{j11}} + \boldsymbol{\beta_{12} pre\_HF_{j12}} \\ &+ \boldsymbol{\beta_{13} pre\_AF_{j13}} + \boldsymbol{\beta_{14} income_{j14}} + \boldsymbol{\beta_{15} education_{j15}} \\ &+ \boldsymbol{\beta_{16} immigrant_{j16}} + \boldsymbol{\beta_{17} lipidmedication_{j17}} \\ &+ \boldsymbol{\beta_{18} bpmedication_{j18}} + \boldsymbol{\beta_{19} smoking_{j19}}\end{aligned}$$

Continuous variables were modeled using restricted cubic splines with 4 equally spaced knots to fulfill proportional hazards assumption, and "strata" imply stratification for the variable gender.

The ancillary analyses of relative risk factor importance does not include the "Category" variable that indicates the number of risk factors within target range, as seen in the main analysis. Those risk factors are instead modeled as continuous or categorical predictors, as appropriate. In addition to the covariables from the main analysis, these models include, physical activity, body mass index, and treatment with statins and antihypertensive medications, since these data were available for patients with diabetes.

To interpret the relative importance of individual predictors, we first identify the strongest predictors (generally top 4–5 predictors) and if the strongest predictors show a clear deviation from the other predictors. The reason for this is that the explained relative risk (R^2) model is based on permutation, therefore, it is difficult to distinguish which predictor that is strongest when there are additional predictors with almost similar R^2 value. So, a strong risk factor will contribute more to the predictive ability, compared with a less strong predictor.

In order to determine the variance explained by all the predictors included in each outcome, we subtract the sum of all R^2 values for each predictor with 1.0, this generates a percentage of the variance explained by all the predictors in each outcome. When age was included as an individual predictor in the Cox model, and adequately adjusted for with splines, age displayed a considerably higher R^2 value than any other predictor. This was one of the reasons that age was modeled as the time-scale in all Cox models.

Analysis of explained log-likelihood

Using the general Cox regression formulated above, we assessed the relative importance, i.e. partial effect, of each risk factor in terms of predicting the outcome, with a complementary method. Log likelihoods are useful for quantifying the predictive information contained in a predictor compared with the information contained in the entire set of predictors. The partial effect of each risk factor was quantified by computing the proportion of explainable log-likelihood explained by each risk factor, i.e. the Wald X^2 statistics minus the degrees of freedom (*Harrell F. Regression Modeling Strategies. Springer; 2015*).

The degrees of freedom for each covariate are depicted in the parenthesis next to each covariate in Figure S2 and S3, supplementary appendix. If a predictor interacts with another predictor the X^2 and partial R^2 measure combines the interaction with the main effects. A strong risk factor will contribute more, as compared with less a strong risk factor, to the predictive ability of the model.

Analysis of optimal levels for risk factors

We used the general Cox model and restricted cubic splines to delineate the value that was associated with the lowest risk for each outcome. The guideline target level was set as reference level for each risk factor. For glycated hemoglobin we used 53 mmol/mol (7.0%), for systolic blood pressure 140 mmHg and for low-density lipoprotein cholesterol 2.5 mmol/L. We used 4 evenly spaced knots for the restricted cubic splines and did experiment with additional splines for all continuous predictors, however between 4 to 5 knots seemed optimal, 5 and more did not contribute especially much to the model.

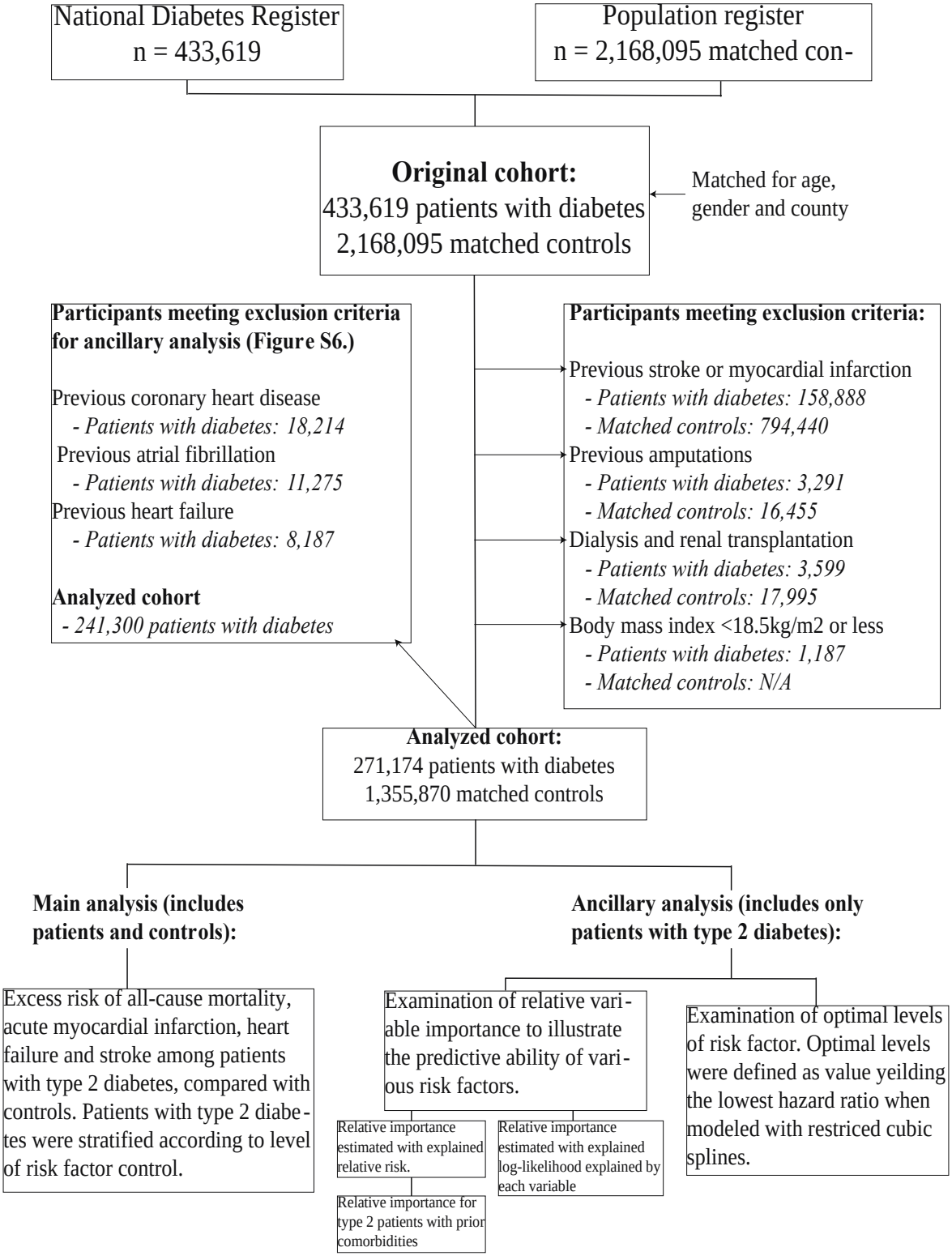
The analysis displays non-linear associations for most risk factors and outcomes, with the exception of acute myocardial infarction that displays a linear association for all risk factors. The authors believe that non-linear associations, i.e. J- and U-shaped associations are correct. In order to predict outcomes for an average type 2 diabetes population, we stratified all categorical covariables in this analysis to avoid specifying which level for each categorical covariable that we wanted to predict outcomes on.

The effect of reverse causality makes interpretation of exact values for optimal levels of risk factors difficult, however, the analysis suggest that lower glycated hemoglobin and systolic blood pressure levels are associated with significantly lower risk for some outcomes, especially acute myocardial infarction and stroke. Moreover, to assess the influence of reverse causality, we performed additional landmark analysis were we excluded patients with type 2 diabetes with less survival time than 3 years. This did not have a significant effect on the hazard function for any of the outcomes, and patients with more than 3 years of survival time had virtually identical risk function as the entire cohort. Thus, we believe that reverse causality has a minor effect on the results for the prediction model, with the exception of particularly low values for risk factors. The increased risk that is observed for lower values is most likely an effect of reverse causality.

Statistical program and packages

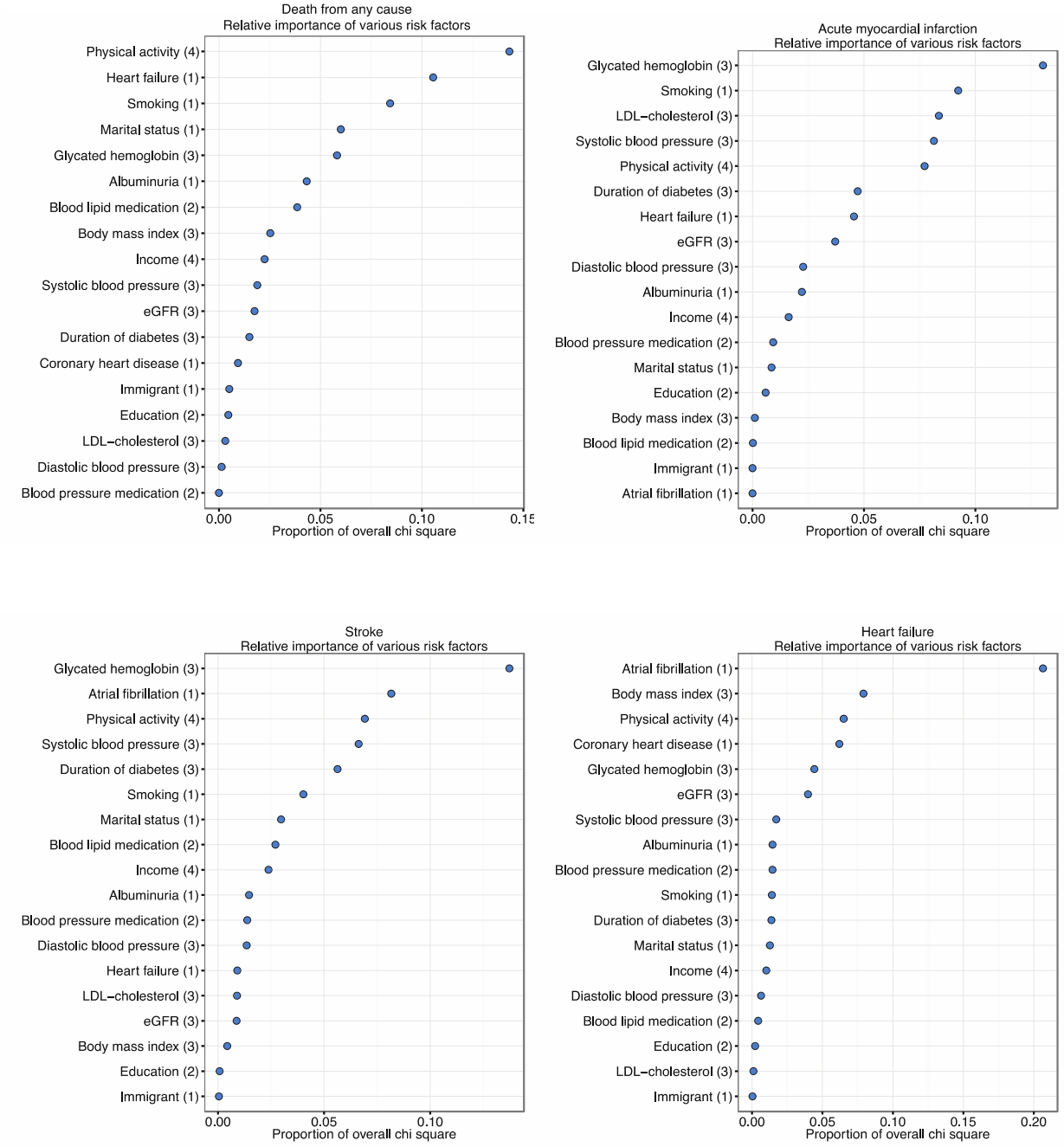
We used the following packages in RStudio: survival, rms and MICE: Multivariate Imputation by Chained Equations. R package version 2.25).

Figure S1. Flow chart of study cohorts and analyses that were performed



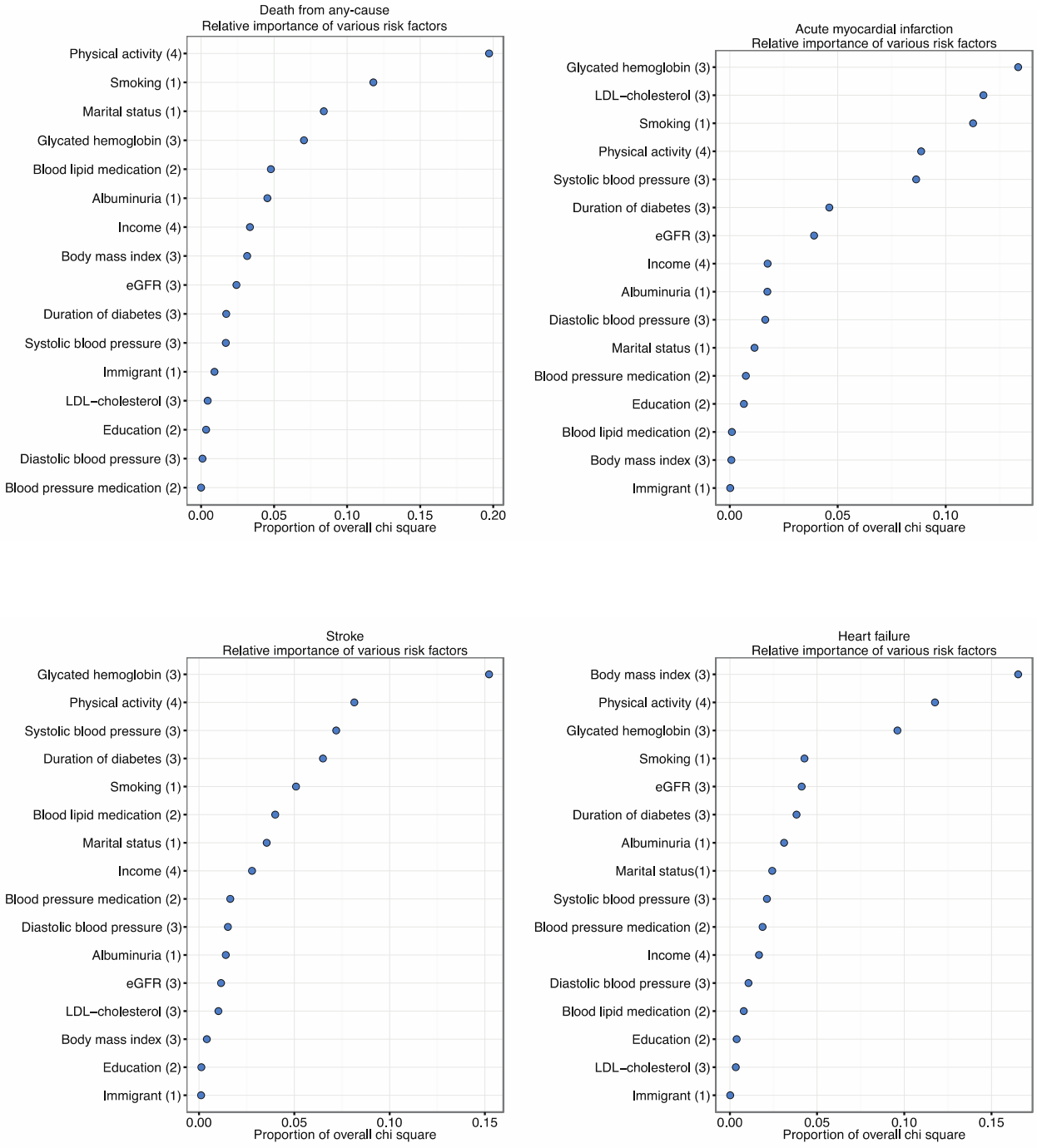
Legend: Flow chart of study cohort and analyses performed. Participants may fulfill more than one exclusion criteria

Figure S2. Relative importance of predictors for cardiovascular outcomes and all-cause mortality, by estimation of explained log-likelihood explained by each predictor, in a type 2 diabetes population with prior comorbidites



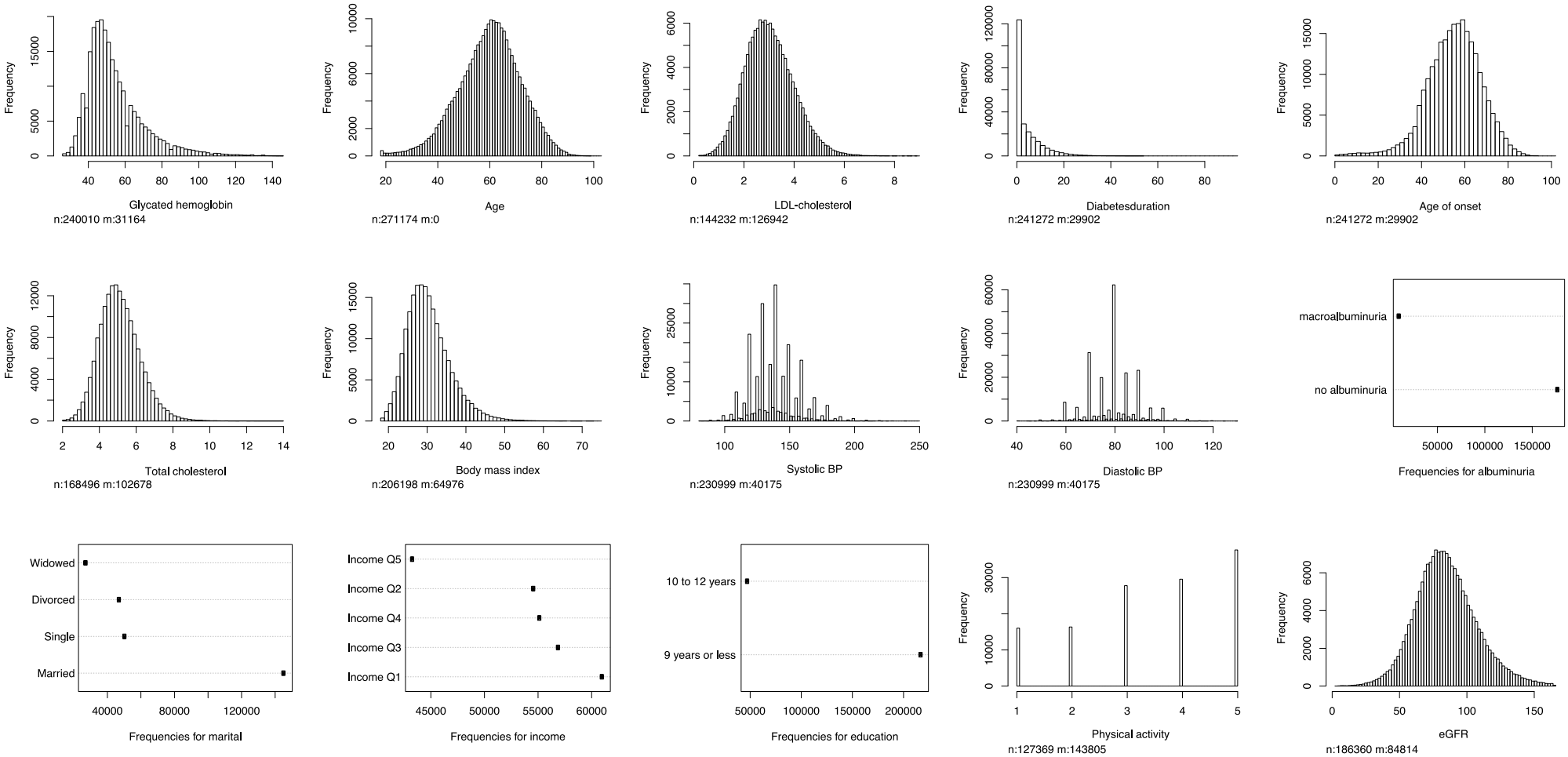
Legend: Relative variable importance was measured by means of explained log-likelihood for each pedictor. Predictors was modeled as binary or multinomial categorical variables in the Cox model. The number presented in the parenthesis next to each predictor is the number of degrees of freedom used for each predictor.

Figure S3. Relative importance of predictors for cardiovascular outcomes and all-cause mortality, by estimation of explained log-likelihood explained by each predictor, in a type 2 diabetes population without prior comorbidities.



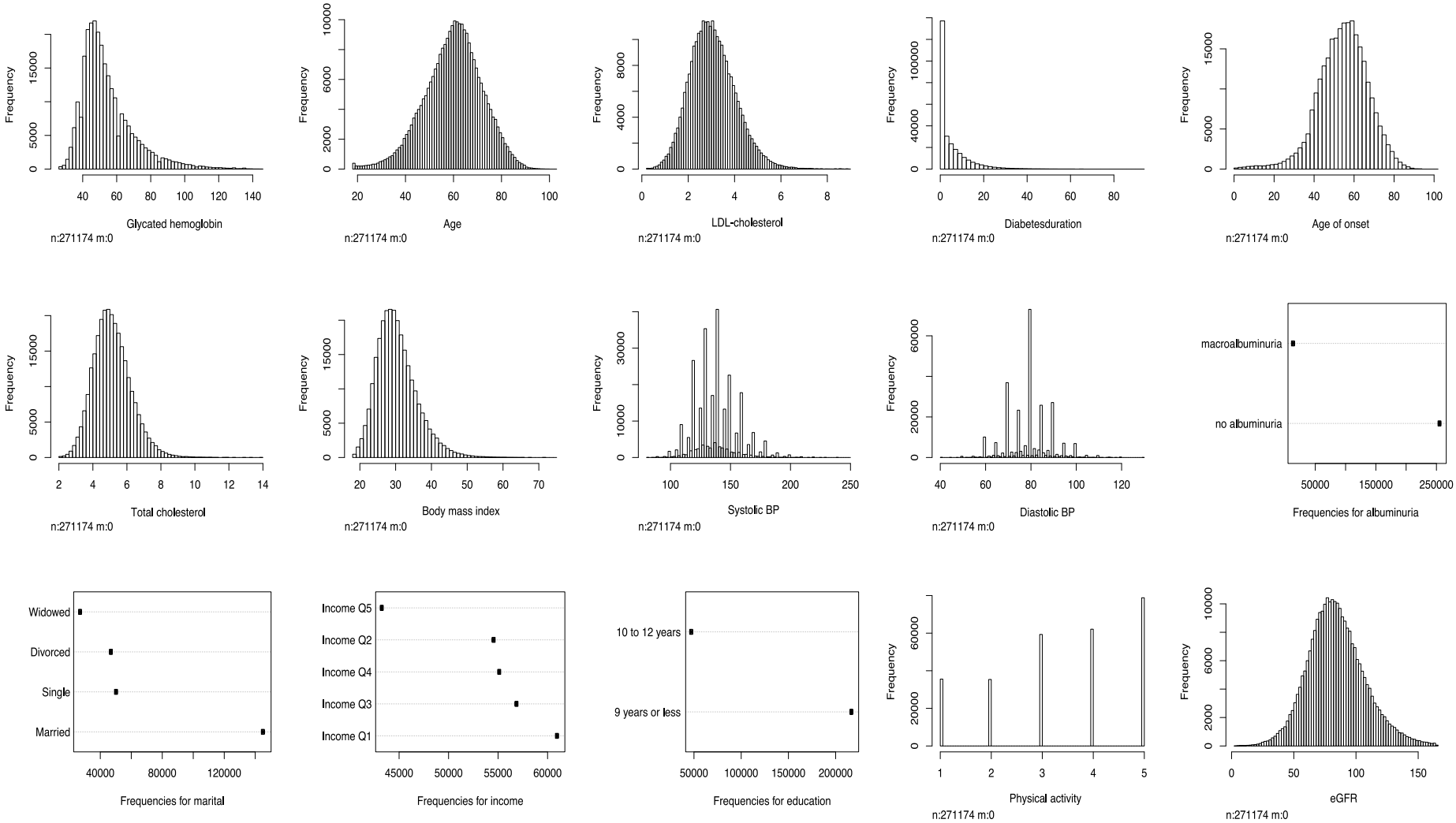
Legend: Relative variable importance was measured by means of explained log-likelihood for each pedictor. Predictors was modeled as binary or multinomial categorical variables in the Cox model. The number presented in the parenthesis next to each predictor is the number of degrees of freedom used for each predictor.

Figure S4. Distribution of variables used as predictors in the multiple imputation.



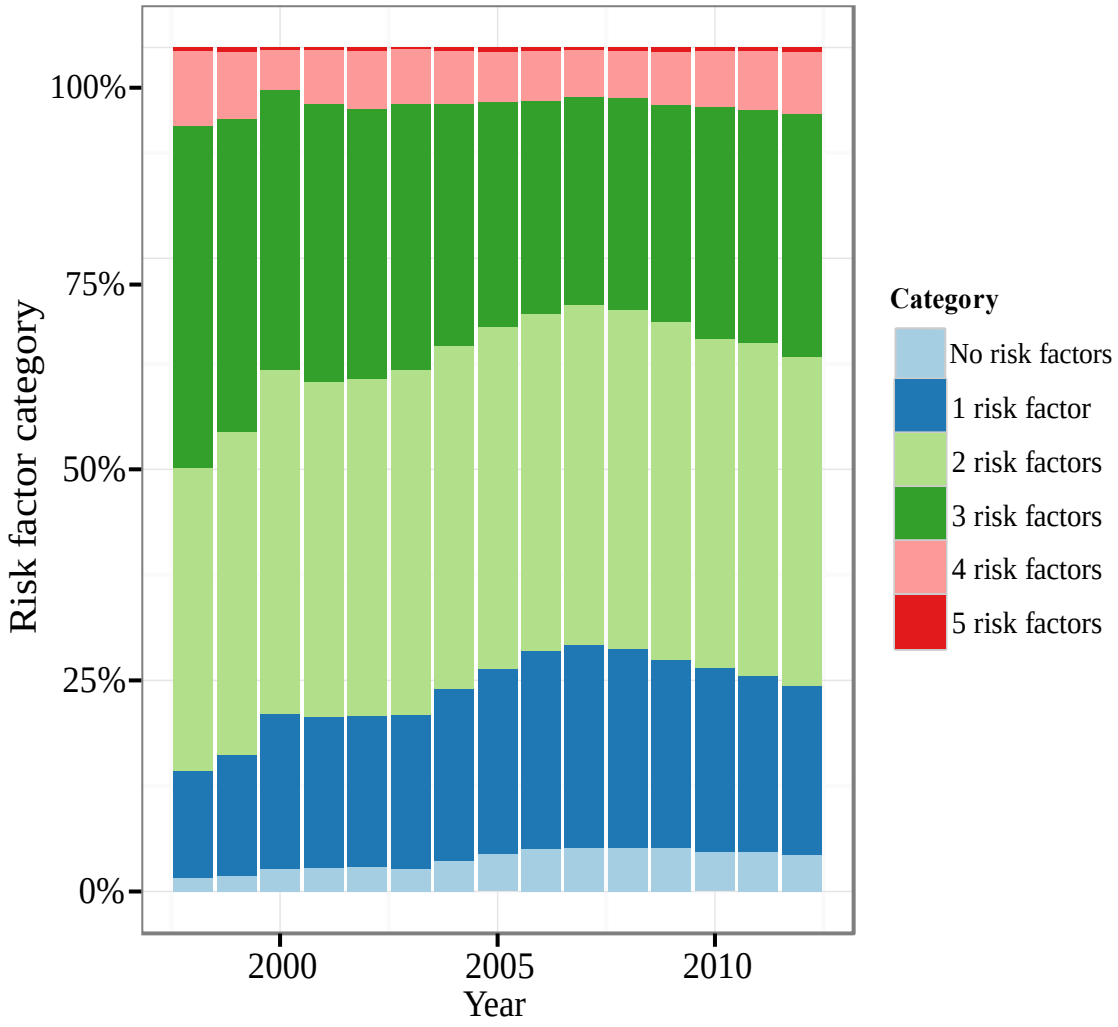
Legend: The distributions for each variable that were included in the imputation models.

Figure S5. Distributions of predictors after multiple imputations, showing the first imputed data set.



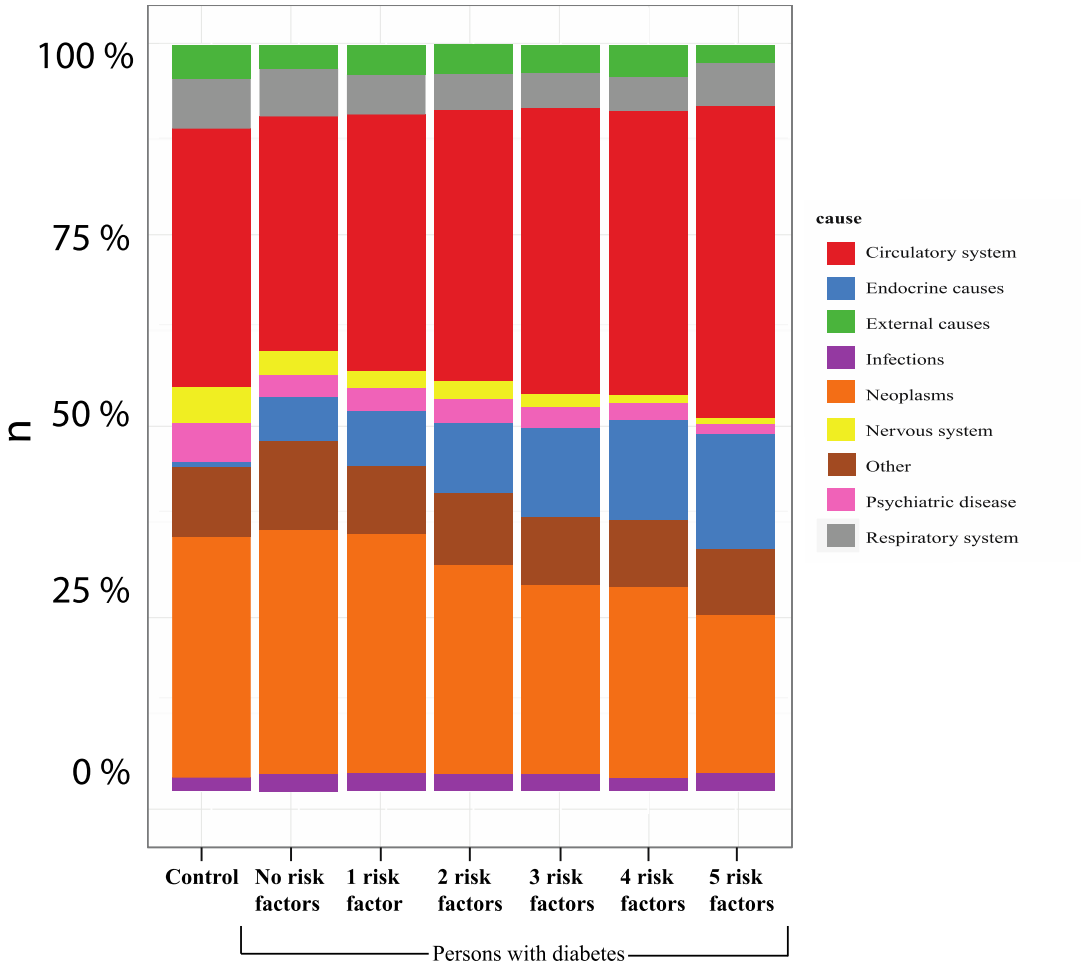
Legend: The distributions for each variable that were included in the imputation models.

Figure S6. Number of risk factors among patients with type 2 diabetes according to year



Legend: Number of risk factors among patients with type 2 diabetes according to year. The y-axis displays the proportion of patients in each risk factor category and the x-axis displays calender year.

Figure S7. Primary causes of death – all patients according to risk factors at target



Legend: Primary causes of death – all patients according to risk factors at target.

TABLE S1. ICD-9 AND ICD-10 CODES USED

Outcomes identified in hospital discharge records using the International Classification of Diseases (ICD) codes, 9th and 10th revision

Supplementary Table S1. ICD-9 and ICD-10 Codes Used to Define Baseline Conditions and Outcomes			
Outcomes of interest	**ICD-9**	**ICD-10**	**Comment**
Death			From the Cause of Death Registry
Fatal/non-fatal acute myocardial infarction	410-414	I20-I25	From the Cause of Death Registry and the inpatient registret. Non-fatal events include hospitalization for acute myocardial infarction and other ICD– codes for coronary heart disease
Fatal/non-fatal stroke	431-434, 436	I61-I64	From the Cause of Death Registry and inpatient registret. Non-fatal events include hospitalization for stroke and other ICD-codes for cerebrovascular disease
Hospitalization for heart failure	428	I50	Non-fatal hospitalization for heart failure
Non-fatal outcomes that were not a dependent variable but these variables were included in study design			
For atrial fibrillation	I427D	I48	
End-stage kidney disease	V42A, V45B, V56A, V56W	Z940, Z491, Z492, Z992	
Amputation		NHQ09, NHQ11, NHQ12, NHQ13, NHQ14, NHQ16, NHQ17, NHQ99, NGQ09, NGQ19, NGQ99, NFQ19, NFQ99	
– Includes ICD-codes as underlying diagnosis and up to 5 contributory causes for fatal outcomes. – Includes ICD-codes as principal diagnosis and up to 7 contributory causes for non-fatal outcomes. The sensitivity and specificity for these diagnoses have been validated previously (*Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. BMC Public Health. 2011;11(1):450. doi:10.1186/1471-2458-11-450.)*			

The Swedish Inpatient Registry and cause of death registry includes mandatory information on all principal and secondary hospital discharge diagnoses, death, and causes of death. The Inpatient Registry was launched in 1960s, with nationwide coverage beginning in 1987. Diagnoses in the Inpatient Registry were classified according to the International Classification of Diseases (ICD) system, and sensitivity and specificity for diagnoses of fatal/non-fatal acute myocardial infarction/coronary heart disease, hospitalization for heart failure, atrial fibrillation, and stroke have previously been validated.

Supplementary Table S2. Incidence rates and hazard ratios for mortality, acute myocardial infarction, stroke and heart failure among patients with type 2 diabetes, compared to matched controls.

OUTCOMES	GROUP	EVENTS	PERSON−YEARS	INCIDENCE RATE	HAZARD RATIO
Mortality, overall	Control	137,520	8,609,491	15.97 (15.89 to 16.06)	Reference
	No risk factors	1,220	66,606	18.31 (17.12 to 19.50)	1.06 (1.00 to 1.12)
	1 risk factor	6,679	344,511	19.39 (18.85 to 19.93)	1.09 (1.06 to 1.12)
	2 risk factors	14,188	670,362	21.17 (20.80 to 21.53)	1.19 (1.17 to 1.22)
	3 risk factors	12,322	488,145	25.24 (24.73 to 25.75)	1.50 (1.47 to 1.54)
	4 risk factors	3,119	102,306	30.49 (29.30 to 31.67)	2.29 (2.20 to 2.38)
	5 risk factors	297	6,397	46.48 (38.94 to 54.03)	3.55 (2.92 to 4.32)
Mortality, women	Control	73,878	4,293,026	17.21 (17.08 to 17.33)	Reference
	No risk factors	606	34,602	17.52 (15.80 to 19.24)	1.05 (0.96 to 1.15)
	1 risk factor	3,553	176,970	20.08 (19.32 to 20.83)	1.10 (1.06 to 1.14)
	2 risk factors	7,727	341,056	22.66 (22.14 to 23.17)	1.18 (1.14 to 1.21)
	3 risk factors	6,523	238,641	27.33 (26.62 to 28.05)	1.47 (1.43 to 1.52)
	4 risk factors	1,400	43,847	31.93 (30.08 to 33.77)	2.43 (2.27 to 2.60)
	5 risk factors	121	2369	51.00 (39.63 to 62.36)	3.98 (2.81 to 5.64)
Mortality, men	Control	63,642	4,316,465	14.74 (14.63 to 14.86)	Reference
	No risk factors	614	32,004	19.17 (17.42 to 20.92)	1.06 (0.97 to 1.16)
	1 risk factor	3,126	167,541	18.66 (17.89 to 19.43)	1.08 (1.04 to 1.13)
	2 risk factors	6,461	329,306	19.62 (19.11 to 20.13)	1.21 (1.18 to 1.25)
	3 risk factors	5,798	249,504	23.24 (22.54 to 23.94)	1.54 (1.49 to 1.59)
	4 risk factors	1,719	58,459	29.41 (27.66 to 31.16)	2.18 (2.05 to 2.32)
	5 risk factors	177	4,029	43.83 (35.07 to 52.59)	3.31 (2.76 to 3.96)
Acute myocardial infarction	Control	46,055	8,486,081	5.43 (5.38 to 5.48)	Reference
	No risk factors	352	65,794	5.35 (4.73 to 5.96)	0.84 (0.75 to 0.93)
	1 risk factor	2361	338,600	6.97 (6.59 to 7.36)	1.11 (1.05 to 1.18)
	2 risk factors	6,107	653,718	9.34 (9.06 to 9.63)	1.53 (1.48 to 1.58)
	3 risk factors	6,117	471,193	12.98 (12.62 to 13.34)	2.21 (2.14 to 2.28)
	4 risk factors	1,665	97,879	17.01 (15.98 to 18.04)	3.29 (3.09 to 3.50)
	5 risk factors	169	6,052	27.86 (22.73 to 33.00)	5.28 (4.48 to 6.21)
Stroke	Control	46,618	8,479,199	5.50 (5.45 to 5.55)	Reference
	No risk factors	372	65,701	5.67 (4.92 to 6.41)	0.95 (0.84 to 1.07)
	1 risk factor	2,382	338,138	7.05 (6.71 to 7.38)	1.16 (1.11 to 1.22)
	2 risk factors	5,506	654,729	8.41 (8.15 to 8.67)	1.40 (1.36 to 1.44)
	3 risk factors	4,987	472,493	10.56 (10.23 to 10.88)	1.87 (1.80 to 1.93)
	4 risk factors	1,131	98,743	11.45 (10.47 to 12.44)	2.49 (2.26 to 2.73)
	5 risk factors	96	6127	15.60 (11.28 to 19.91)	3.58 (2.70 to 4.76)
Heart failure	Control	48,712	8,490,223	5.74 (5.69 to 5.79)	Reference
	No risk factors	769	64,607	11.90 (10.92 to 12.88)	1.45 (1.34 to 1.57)
	1 risk factor	3,813	334,487	11.40 (10.98 to 11.82)	1.51 (1.44 to 1.58)
	2 risk factors	7,688	649,782	11.83 (11.54 to 12.13)	1.69 (1.65 to 1.74)
	3 risk factors	6,621	470,263	14.08 (13.71 to 14.45)	2.19 (2.12 to 2.26)
	4 risk factors	1,644	98,029	16.77 (15.90 to 17.65)	3.28 (3.06 to 3.52)
	5 risk factors	160	6,023	26.64 (21.74 to 31.54)	5.64 (4.68 to 6.80)

Age and sex standardized incidence rate per 10,000 person-years.

Supplementary Table S3. Variables Used in the Imputation Algorithm
Age, sex, year of visit, duration of diabetes, age at onset of diabetes, smoking status, use of antihypertensive medication, use of lipid lowering medications, macroalbuminuria, LDL-cholesterol, total cholesterol, body mass index, glycated hemoglobin, systolic blood pressure, diastolic blood pressure, marital status, immigrant status, income, education, physical activity, eGFR, county, history of coronary heart disease, history of heart failure, history of atrial fibrillation, survival time for fatal/nonfatal acute myocardial infarction, outcome for fatal/nonfatal acute myocardial infarction, survival time for fatal/nonfatal stroke, outcome for fatal/nonfatal stroke, survival time for total mortality, outcome for total mortality.

Table S4. Complete baseline Characteristics of Patients with Type 2 Diabetes and Matched Controls

	Matched controls	Persons with diabetes with complete data on all five risk factor						
	Overall	Overall	Number of risk factors beyond therapeutic target					
Number of participants			No risk factors	1 risk factors	2 risk factors	3 risk factors	4 risk factors	5 risk factors
n, complete case data set	483,365	96,673	4,852	22,584	39,673	24,341	4927	296
n, imputed data set 1	1,355,870	271,174	11,612	57,000	107,840	76,325	17,227	1,170
n, imputed data set 2	1,355,870	271,174	11,569	57,033	107709	76,455	17,213	1,195
n, imputed data set 3	1,355,870	271,174	11,685	57,164	107,521	76,517	17,102	1,185
n, imputed data set 4	1,355,870	271,174	11,669	57,217	107,551	76,431	17,137	1,169
n, imputed data set 5	1,355,870	271,174	11,562	57,246	107,551	76,392	17,219	1,204
Women – no. (%)	238,885 (49.4)	47,777 (49.4)	2,525 (52.0)	11,528 (51.0)	19,799 (49.9)	11,713 (48.1)	2,085 (42.3)	127 (42.9)
Age (years) – mean (SD)§	60.58 (10.89)	60.58 (10.89)	60.96 (12.06)	61.04 (11.23)	60.93 (10.79)	60.00 (10.54)	58.32 (10.06)	57.27 (10.21)
Marital status – no. (%)								
Divorced	82,197 (17.0)	16,540 (17.1)	683 (14.1)	3,517 (15.6)	6,550 (16.5)	4,524 (18.6)	1,186 (24.1)	80 (27.0)
Married	277,842 (57.5)	53,754 (55.6)	2,930 (60.4)	13,131 (58.1)	22,482 (56.7)	12,790 (52.5)	2,296 (46.6)	125 (42.2)
Single	83,295 (17.2)	17,472 (18.1)	804 (16.6)	3,754 (16.6)	6,885 (17.4)	4,837 (19.9)	1,118 (22.7)	74 (25.0)
Widowed	40,001 (8.3)	8,907 (9.2)	435 (9.0)	2,182 (9.7)	3,756 (9.5)	2,190 (9.0)	327 (6.6)	17 (5.7)
Education – no. (%)								
9 years or less (%)	348,899 (73.0)	77,778 (81.7)	3,755 (78.5)	17,617 (79.1)	31,781 (81.3)	20,142 (84.1)	4,223 (87.2)	260 (89.0)
Income quintile – no. (%)								
Income quintile 1 (lowest)	84,831 (17.6)	2,0443 (21.1)	1,030 (21.2)	4,667 (20.7)	8,345 (21.0)	5,294 (21.7)	1,038 (21.1)	69 (23.3)
Income quintile 2	75,027 (15.5)	18,453 (19.1)	957 (19.7)	4,320 (19.1)	7,384 (18.6)	4,698 (19.3)	1,023 (20.8)	71 (24.0)
Income quintile 3	92,811 (19.2)	20,610 (21.3)	1,075 (22.2)	4,782 (21.2)	8,396 (21.2)	5,169 (21.2)	1,126 (22.9)	62 (20.9)
Income quintile 4	114,919 (23.8)	20,808 (21.5)	976 (20.1)	4,785 (21.2)	8,606 (21.7)	5,298 (21.8)	1,080 (21.9)	63 (21.3)
Income quintile 5 (highest)	115,747 (23.9)	16,359 (16.9)	814 (16.8)	4,030 (17.8)	6,942 (17.5)	3,882 (15.9)	660 (13.4)	31 (10.5)
Immigrants – no. (%)	60,147 (12.4)	17,310 (17.9)	869 (17.9)	3,909 (17.3)	6,800 (17.1)	4,591 (18.9)	1,065 (21.6)	76 (25.7)

Coexisting conditions – no. (%)†								
Atrial fibrillation – no. (%)	11,294 (2.3)	3,672 (3.8)	292 (6.0)	1,047 (4.6)	1,423 (3.6)	777 (3.2)	125 (2.5)	8 (2.7)
Coronary heart disease – no. (%)	14,127 (2.9)	6,433 (6.7)	636 (13.1)	1,905 (8.4)	2,403 (6.1)	1,220 (5.0)	253 (5.1)	16 (5.4)
Heart failure – no. (%)	4,619 (1.0)	2,320 (2.4)	214 (4.4)	609 (2.7)	880 (2.2)	500 (2.1)	110 (2.2)	7 (2.4)
Information in the National Diabetes Register								
Duration of diabetes – mean (SD)		4.53 (5.75)	4.09 (5.28)	4.08 (5.35)	4.29 (5.61)	5.19 (6.18)	5.51 (6.41)	6.09 (6.60)
Age at onset of diabetes – mean (SD)		56.09 (11.11)	56.84 (12.20)	56.97 (11.44)	56.69 (11.02)	54.87 (10.64)	52.76 (10.20)	51.34 (11.11)
Glycated hemoglobin – mean (SD)								
Millimoles per mole‡		53.22 (13.96)	44.31 (5.14)	46.75 (8.67)	51.10 (12.32)	61.58 (15.25)	66.28 (15.01)	70.51 (15.45)
Percent*		7.02 (± 1.28)	6.21 (± 0.47)	6.43 (± 0.79)	6.83 (± 1.13)	7.79 (± 1.39)	8.22 (± 1.37)	8.60 (± 1.41)
LDL cholesterol – mmol/l (SD)		3.00 (0.95)	1.98 (0.39)	2.55 (0.87)	3.09 (0.92)	3.38 (0.85)	3.50 (0.84)	3.65 (0.81)
LDL cholesterol – mg/dl (SD)		116.0 (36.7)	76.5 (15.1)	98.6 (33.6)	119.5 (35.5)	130.7 (32.8)	135.3 (32.4)	141.1 (31.3)
Total cholesterol – mean (SD)		5.11 (1.05)	4.04 (0.58)	4.64 (0.95)	5.20 (1.01)	5.51 (0.97)	5.67 (0.98)	5.88 (0.99)
Smoker – no. (%)		16486 (17.1)	0 (0.0)	886 (3.9)	4268 (10.8)	7125 (29.3)	3911 (79.4)	296 (100.0)
Body Mass Index – mean (SD)		30.36 (5.53)	29.53 (5.48)	29.85 (5.37)	30.40 (5.48)	30.85 (5.68)	30.74 (5.67)	30.85 (5.86)
Systolic blood pressure – mean (SD)		137.94 (17.01)	123.02 (9.38)	131.58 (15.38)	139.37 (16.63)	142.96 (16.63)	144.84 (17.17)	147.79 (18.71)
Diastolic blood pressure – mean (SD)		79.16 (9.57)	69.54 (5.88)	75.08 (8.89)	80.08 (9.17)	82.35 (8.88)	83.89 (8.92)	84.53 (9.88)
Macroalbuminuria – no. (%)		4695 (4.9)	0 (0.0)	177 (0.8)	908 (2.3)	1802 (7.4)	1512 (30.7)	296 (100.0)
eGFR – mean (SD)**		84.19 (21.52)	82.29 (20.71)	83.01 (20.59)	83.43 (20.94)	85.90 (22.39)	88.76 (24.72)	91.93 (29.16)
Treatment – no. (%)								
Statin – no (%)		57,945 (61.5)	2,907 (61.5)	13,312 (60.4)	24,206 (62.6)	14,449 (61.0)	2,862 (59.8)	209 (73.1)
Antihypertensive – no (%)		40,553 (42.4)	2,934 (61.0)	11,035 (49.3)	15,839 (40.4)	8,809 (36.7)	1,823 (37.6)	113 (38.4)

§ Plus–minus values are means ± SD. Controls were matched for age, sex, and county.
† Diagnostic codes for the conditions listed are from the International Classification of Diseases, 9th Revision and 10th Revision.
‡ Concentrations of glycated hemoglobin were based on values from the International Federation of Clinical Chemistry and Laboratory Medicine.
* Percentages for the glycated hemoglobin level were based on values from the National Glycohemoglobin Standardization Program.
** The GFR was estimated with the use of the Modification of Diet in Renal Disease equation.

Supplementary Table S5. Baseline Characteristics of Patients with Type 2 Diabetes With Missing Data.		
	Persons with diabetes with atleast one missing risk factor	**Persons with diabetes with no missing data**
n	174,501	96,673
Women – no. (%)	87,286 (50.0)	47777 (49.4)
Age (years) – mean (SD)§	60.28 (12.46)	60.58 (10.89)
Marital status – no. (%)		
Divorced	30,886 (17.7)	16540 (17.1)
Married	91,743 (52.6)	53754 (55.6)
Single	33,259 (19.1)	17472 (18.1)
Widowed	18,613 (10.7)	8907 (9.2)
Education – no. (%)		
9 years or less	139,794 (82.1)	77778 (81.7)
Income quintile – no. (%)		
Income quintile 1 (lowest)	40,596 (23.3)	20443 (21.1)
Income quintile 2	36,187 (20.7)	18453 (19.1)
Income quintile 3	36,340 (20.8)	20610 (21.3)
Income quintile 4	34,398 (19.7)	20808 (21.5)
Income quintile 5	26,980 (15.5)	16359 (16.9)
Immigrants – no. (%)	38,520 (22.1)	17310 (17.9)
Coexisting conditions – no. (%)†		
Atrial fibrillation	7,603 (4.4)	3672 (3.8)
Coronary heart disease	11,781 (6.8)	6433 (6.7
Heart failure	5,867 (3.4)	2320 (2.4)
Information in the National Diabetes Register		
Duration of diabetes – mean (SD)	5.14 (6.93)	4.53 (5.75)
Age at onset of diabetes – mean (SD)	55.17 (13.12)	56.09 (11.11)
Glycated haemoglobin – mean (SD)		
Millimoles per mole‡	55.95 (16.43)	53.22 (13.96)
Percent*	7.02 (1.28	7.27 (1.50)
LDL cholesterol – mean (SD) (mmol/mol)	3.09 (0.99)	3.00 (0.95)
LDL cholesterol – mean (SD) (mg/dl)	119.5 (38.3)	116.0 (36.7)
Total cholesterol – mean (SD)	5.23 (1.17)	5.11 (1.05)
Smoker – no. (%)	23,012 (19.1)	16,486 (17.1)
Body Mass Index – mean (SD)	30.32 (5.65)	30.36 (5.53)
Systolic blood pressure – mean (SD)	139.79 (18.46)	137.94 (17.01)
Diastolic blood pressure – mean (SD)	79.98 (9.85)	79.16 (9.57)
Albuminuria – n (%)	5,389 (5.9)	4,695 (4.9)
eGFR – mean (SD)**	85.14 (23.53)	84.19 (21.52)
Treatment – no. (%)		
Statin – no (%)	51,592 (32.7)	40,553 (42.4)
Antihypertensive – no (%)	88,461 (55.4)	57,945 (61.5)

§ Plus–minus values are means ± SD. Controls were matched for age, sex, and county. In the incidence analysis, each patient with type 2 diabetes was matched with five controls.
† Diagnostic codes for the conditions listed are from the *International Classification of Diseases, 9th Revision and 10th Revision.*
‡ Concentrations of glycated hemoglobin were based on values from the International Federation of Clinical Chemistry and Laboratory Medicine.
* Percentages for the glycated hemoglobin level were based on values from the National Glycohemoglobin Standardization Program.
** The GFR was estimated with the use of the Modification of Diet in Renal Disease equation.