EstimatioN oF mORtality risk in type2 diabetiC patiEnts (ENFORCE): an inexpensive and parsimonious prediction model.

Massimiliano Copetti, PhD, Hetal Shah, PhD, Andrea Fontana, MSc, Maria Giovanna Scarale, PhD, Claudia Menzaghi, PhD, Salvatore De Cosmo, MD, Monia Garofolo, MD, Maria Rosaria Sorrentino, MD, Olga Lamacchia, Prof, Giuseppe Penno, Prof, Alessandro Doria, Prof, Vincenzo Trischitta, Prof.

The Journal of Clinical Endocrinology & Metabolism
Endocrine Society

Submitted: January 28, 2019
Accepted: May 08, 2019
First Online: May 14, 2019

Predicting mortality in diabetic patients.

EstimatioN oF mORtality risk in type2 diabetiC patiEnts (ENFORCE): an inexpensive and parsimonious prediction model.

Massimiliano Copetti[1], PhD; Hetal Shah[2,3], PhD; Andrea Fontana[1], MSc; Maria Giovanna Scarale[4], PhD; Claudia Menzaghi[4], PhD; Salvatore De Cosmo[5], MD; Monia Garofolo[6], MD; Maria Rosaria Sorrentino[7], MD; Olga Lamacchia[7], Prof; Giuseppe Penno[6], Prof; Alessandro Doria[2,3], Prof; Vincenzo Trischitta[4,8], Prof.

[1] *Unit of Biostatistics, Fondazione IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Italy*

[2] *Research Division, Joslin Diabetes Center, Boston, MA*

[3] *Department of Medicine, Harvard Medical School, Boston, MA*

[4] *Research Unit of Diabetes and Endocrine Diseases, Fondazione IRCCS "Casa Sollievo della Sofferenza", San Giovanni Rotondo, Italy*

[5] *Department of Clinical Sciences, Fondazione IRCCS "Casa Sollievo Della Sofferenza", San Giovanni Rotondo, Italy*

[6] *Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy*

[7] *Unit of Endocrinology and Diabetology, Department of Medical and Surgical Sciences, University of Foggia, Foggia, Italy*

[8] *Department of Experimental Medicine, "Sapienza" University, Rome, Italy*

ORCiD numbers:

0000-0002-7960-5947

COPETTI

MASSIMILIANO

Received 28 January 2019. Accepted 08 May 2019.

Context. We previously developed and validated an inexpensive and parsimonious prediction model of 2-year all-cause mortality in real-life type 2 diabetic patients.
Objective. This model, now named ENFORCE, was now investigated in terms of i) prediction performance at 6 years, a more clinically useful time-horizon; ii) further validation in an independent sample; iii) performance comparison in real-life versus clinical trial setting.
Design. Observational prospective. Randomized clinical trial.
Setting. White patients with type 2 diabetes.
Patients. Gargano Mortality Study (GMS; *n*=1019), Foggia Mortality Study (FMS; *n*=1045), Pisa Mortality Study (PMS; *n*=972) as real-life samples and the standard glycemic arm of the ACCORD clinical trial (*n*=3150).
Main Outcome Measure. The endpoint was all-cause mortality. Prediction accuracy and calibration were estimated to assess model's performances.
Results. ENFORCE yielded a 6-year mortality C-statistics of 0.79, 0.78 and 0.75 in GMS, FMS and PMS, respectively (*P* heterogeneity=0.71). Pooling the three cohorts, a 6-year mortality C-statistic of 0.80 was observed. In the ACCORD trial, ENFORCE achieved a C-statistic of 0.68, a value which is significantly lower than that obtained in the pooled real-life samples (*P*<0.0001). This difference resembles that observed with other models when comparing real-life vs. clinical trial settings, thus suggesting it is a true, replicable phenomenon.

Conclusions. Time horizon of ENFORCE has been extended to 6 years and validated in three independent samples. ENFORCE is a free (http://www.operapadrepio.it/enforce/enforce.php) and user-friendly risk calculator of all-cause mortality in White type 2 diabetic patients from real-life setting.

We extended and validated an inexpensive and parsimonious prediction model of 6-year all-cause mortality in White patients with type 2 diabetes from both real-life and clinical trial settings.

Introduction

Diabetes mellitus is one of the most challenging global health problems, affecting approximately 400 million people (1) and representing a leading cause of death worldwide (2). The negative impact of diabetes on global health is projected to become even greater over the next decades given the epidemic proportions that this disease is assuming (3); it is therefore mandatory to identify the best strategies to tackle it.

For the concept of precision medicine to become a reality, the follow-up and treatment of each individual patient should be tailored to his/her individual risk profile, thereby maximizing effectiveness and minimizing costs. To pursue such ambitious goal, the availability of well-performing risk prediction models is pivotal. In the specific context of mortality, the ability to predict a high risk would allow health care providers to apply the most aggressive, most expensive, and most burdensome prevention strategies only to the most high-risk patients. It would also be important for these tools to be inexpensive, parsimonious, and simple, especially when they are to be used in health care systems with limited resources.

We have recently developed and validated a prediction model of all-cause mortality in White patients with type 2 diabetes from Central-Southern Italy enrolled in two longitudinal cohort studies (4). Our model was built based on a 2-year horizon (4). The aim of the present study was to investigate, in the same cohorts, how this risk model performs at 6 years - a time-horizon that is more useful for clinical purposes, and to validate the risk model in an additional external sample of Whites with type 2 diabetes from Italy. While we were conducting this study, a novel risk model for all-cause mortality in patients with type 2 diabetes (i.e. RECODe) was developed based on data from randomized clinical trials (5) and validated in observational longitudinal studies (5, 6). Interestingly, RECODe performed better in observational longitudinal studies (6) that in randomized clinical trials (5), thus raising the hypothesis that mortality prediction accuracy may differ between these two sets. Such a phenomenon, if confirmed, would be of great importance for comparing and interpreting epidemiological evidence derived from different datasets as well as for designing new studies. In light of this, though being well aware that beside the intrinsic study design several additional differences in genetic, environmental and clinical features characterized the two different settings, we investigated also if the all-cause mortality prediction model we set up performed differently in our observational cohort studies than in the ACCORD clinical trial.

Materials and Methods

Study Design

The accuracy of the proposed model, from now on refereed to as ENFORCE (EstimatioN oF mORtality risk in type 2 diabetiC patiEnts), in predicting all-cause mortality within a 6-year horizon in patients with type 2 diabetes was investigated in the updated Gargano Mortality Study and validated in Foggia Mortality Study (GMS and FMS, respectively), in which the model was initially built and validated (4) for 2-year mortality. As a further validation step, ENFORCE's performance at 6 years was also evaluated in a new

Italian sample - the Pisa Mortality Study (PMS) (7, 8). Finally, to investigate how this model performs in a clinical trial setting, we evaluated its prediction accuracy in self-reported White individuals from the standard glycemic arm of the ACCORD clinical trial, which was carried out in patients with type 2 diabetes from the US and Canada (9).

Samples

Gargano Mortality Study (GMS)

The GMS served as training sample and includes 1028 self-reported White individuals with type 2 diabetes (diagnosed according to American Diabetes Association [ADA] 2003 criteria), who were consecutively recruited at the Scientific Institute "Casa Sollievo della Sofferenza" in San Giovanni Rotondo (Apulia, Central-Southern Italy) from November 1, 2000 to September 30, 2005 for a study aimed at identifying predictors of incident all-cause mortality. The only exclusion criterion was the presence of poor life expectancy due to malignancies. To date, this cohort has been followed-up for a median of 11.8 years (range 0.1-14.0), with the last information on vital status obtained on November 30, 2014. After excluding patients whose information on vital status at follow-up was not available (n = 9), 1019 patients (99.1% of the initial cohort) were eligible for the present analysis. Missing data rates for the nine baseline covariates included into the prognostic model varied from 0.0% to 8.2%.

Foggia Mortality Study (FMS)

The FMS served as first, external and independent validation sample and consists of 1153 self-reported White individuals with type 2 diabetes (diagnosed according to American Diabetes Association [ADA] 2003 criteria) were consecutively recruited at the Endocrine Unit of the University of Foggia (Apulia, Central-Southern Italy) from January 7, 2002 to September 30, 2008 for a study aimed at identifying predictors of incident all-cause mortality. As in the GMS, the only exclusion criterion was the presence of poor life expectancy due to malignancies. To date, this cohort has been followed-up for a median of 7.4 years (range 0.1-11.9), with the last information on vital status obtained on March 31, 2015. After excluding patients whose information on vital status at follow-up was not available (n = 108), 1 045 patients (90.1% of the initial cohort) were eligible for the present analysis. Missing data rates for the nine baseline covariates included into the prognostic model varied from 5.3% to 7.3%.

Pisa Mortality Study (PMS)

PMS served as second, external and independent validation sample. White individual (n = 972) with type 2 diabetes (diagnosed according to American Diabetes Association [ADA] 2003 criteria) were consecutively recruited at the Endocrine Unit of the University of Pisa from January 1, 2002 to February 14, 2008, for a study aimed at identifying predictors of incident all-cause mortality. As in the GMS and the FMS, the only exclusion criterion was the presence of poor life expectancy due to malignancies. To date, this cohort has been followed-up for a median of 11.2 years (range 0.3-11.7), with the last information on vital status obtained on February 28, 2015. Information on vital status at follow-up was available for all patients. Missing data rates for the nine baseline covariates included into the prognostic model varied from 0.0% to 0.5%.

Action to Control Cardiovascular Risk in Diabetes (ACCORD) study

The ACCORD clinical trial, recruited 10251 subjects with type 2 diabetes and high cardiovascular risk from 77 clinical centers across the US and Canada (9). ACCORD served as an additional external and independent validation sample with the additional scope to test the transportability of our model in a clinical trial setting. Subjects were randomized in a 1:1 ratio to intensive (targeting lowering of glycated hemoglobin, HbA1c to < 6.0%) and

standard (HbA1c 7-7.9%) glycemic treatment arms, and to blood pressure and lipid sub-trials (9). For current validation study, only self-reported White participants from the standard glycemic arm (n = 3199) were investigated. After excluding those with missing values for the nine baseline predictors at issue, 3150 participants, followed-up for a median of 5 years (range 1-7), were eligible for the present analysis.

Risk model to predict all-cause mortality

ENFORCE is based on the following nine predictors measured at baseline: age, antihypertensive and insulin therapy, body mass index (BMI), diastolic blood pressure (DBP), low density lipoprotein (LDL) cholesterol, triglyceride, high density lipoprotein cholesterol (HDL-C), and Albumin/Creatinine Ratio (ACR) levels (4). Details for selecting the above-mentioned predictors have been described in details elsewhere (4). Briefly, predictors were selected using a variables selection procedure based on the continuous Net Reclassification Improvement (10-12) within a proportional hazards Cox model (4). Continuous variables, including BMI, systolic blood pressure (SBP), DBP, LDL, triglycerides, HDL and ACR, suspected to violate the multiplicative model linearity assumption were log transformed. Important aspects including the variable selection procedure, modeling continuous prognostic factor, checking the model assumptions and complexity have been previously discussed (4).

As sensitivity analyses, in a subgroup of 1082 individuals from Italian samples for which additional clinical information were available and in self-reported White participants from the standard glycemic arm the ACCORD study, ENFORCE prediction performances were assessed after adding to the model - one by one - history of documented nonfatal myocardial infarction, stroke, retinopathy, anticoagulant therapy and eGFR.

Data analysis

Patient baseline characteristics were reported as frequency (percentage) and mean (SD) or median along with lower and upper quartiles for categorical and continuous variables, respectively. Overall and age-adjusted death incidence rates for 100 person-years were also reported and compared using a Poisson model.

Time-to-death analyses were conducted using multivariate Cox proportional hazards regression models, and risks were reported as hazard ratios (HRs) along with their 95% confidence intervals (CIs). The assumption of proportionality of the hazards was tested by using scaled Schoenfeld residuals and held for all analyses. Overall survival was defined as the time between enrollment and death. For subjects who did not experience the endpoint, survival time was censored at the time of the last available follow-up visit.

The model discriminatory ability was assessed by estimating survival C-statistic, along with 95% CIs derived following perturbation-resampling method (13); comparisons between C-statistics were carried out according to Pencina and D'Agostino approach (14). The Greenwood-Nam-D'Agostino (GND) test (15), which measures the distance between predicted and observed Kaplan-Meier event rates over 6 years, was performed. Calibration was also reported as the slope and as the intercept of the regression line between predicted and observed Kaplan-Meier event rates over 6 years by deciles of risk. In an ideal condition, the calibration slope should be 1 and the intercept should be 0, reflecting a perfect agreement between predicted and observed event rates. Furthermore, survival conditional tree analysis (16) was performed to identify subgroups of patients with different mortality risks according to the 6-year mortality predicted probability using a conservative Bonferroni adjustment approach for the spitting rule. The free web-based calculator is available at http://www.operapadrepio.it/rcalc/rcalc.php.

As missing data rates in each of the three Italian samples were low, we performed imputations using the Random Forest framework building 100 000 trees for each sample, which has been demonstrated to be more efficient than other traditional methods (17, 18).

Two sided *P* value<0.05 was considered for statistical significance. All statistical analyses were performed using SAS Software Release 9.4 (SAS Institute, Cary, NC) and the computing environment R (R Development Core Team, version 3.3.2).

Results

All-cause mortality prediction in GMS and FMS

Baseline clinical features of patients from the GMS (*n* = 1 019) and FMS (*n* = 1 045) are reported in Table 1. Mean age was 61.1 (9.7) and 63.6 (11.8), with 41.8% and 44.9% patients being ≥ 65 years old in GMS and FMS, respectively. Mean BMI was 31.5 (5.8) and 30.3 (6.6), with 11.4% and 18.8% individuals being normal-weight, 34.3% and 31.1% over-weight and 53.3% and 49.2% obese in GMS and FMS, respectively. During follow-up, 333 (31.7%) and 309 (29.6%) patients died in the GMS and FMS, respectively. Age- and sex-adjusted mortality incidence rates were 1.5 and 3.1 events per 100 person-years in GMS and in FMS, respectively. After updating follow-up data (as well as retrieving some previously missing data or imputing them using random forest methodology), ENFORCE yielded a 6-year mortality C-statistics of 0.79 (95% CI, 0.75 to 0.82) and 0.78 (95% CI, 0.75 to 0.80) in the GMS and FMS, respectively. The previously reported prediction accuracy at 2 years (4) was confirmed both in the GMS and FMS with C-statistics equal to 0.88 (95% CI, 0.83 to 0.92) and 0.81 (95% CI, 0.75 to 0.85), respectively.

Validation in PMS

Baseline clinical features of patients from the PMS (*n* = 972) are reported in Table 1. Mean age was 59.6 (7.1), with 27.1% patients being ≥ 65 years old. Mean BMI was 29.7 (5.3), with 16.5% individuals being normal-weight, 43.1% over-weight and 40.4% obese. During follow-up, 154 (15.8%) patients died. Age- and sex-adjusted mortality incidence rate was 1.3 events per 100 person-years. Prediction accuracy of ENFORCE corresponded to a 6-year time horizon C-statistics equal to 0.75 (95% CI, 0.68 to 0.83).

Pooled Italian samples

The prediction performance of ENFORCE was similar across the three Italian cohorts, with overlapping 95% CIs for the C-statistics (*P* for heterogeneity = 0.71). When the three studies were pooled in order to increase statistical power and to obtain more robust risk estimates and prediction accuracy measures, the C-statistic for 6-year all-cause mortality was 0.80 (95% CI, 0.78 to 0.82); the calibration slope was 1.020, and the calibration intercept was -0.003 (Figure 1), while the calibration GND test *P* was 0.11. For the sake of future implementation and replication in other external samples, Cox regression coefficients for each predictor included in ENFORCE are reported in Supplemental Table 1 (19).

As shown in Table 2, ENFORCE had a better prediction accuracy in HbA1c < 8% stratum as compared to HbA1c ≥ 8% stratum.

A survival conditional tree analysis of 6-year mortality partitioned the pooled sample into four risk categories according to different levels of all-cause mortality predicted risk probabilities, namely low (i.e., predicted probability < 10%, observed 6-year mortality incidence rate = 0.7 per 100 person-years), intermediate-low (predicted probability ranging from 10% to 20%, observed 6-year mortality incidence rate = 1.2 per 100 person-years), intermediate-high (predicted probability ranging from 20% to 33%, observed 6-year mortality incidence rate = 6.1 per 100 person-years) and high (predicted probability > 33%, observed 6-year mortality incidence rate = 13.4 per 100 person-years) risk. Kaplan Meier survival curves of the four categories are shown in Figure 2. As compared with individuals with the

lower risk, those with intermediate-low, intermediate-high and high risk had HR 3.0 (95% CI, 1.2 to 4.1), HR 8.6 (95% CI, 6.3 to 11.6,) and 19.1 (95% CI, 14.6 to 25.1), respectively.

Sensitivity Analysis.
In a subgroup of the pooled sample comprising a total of 1082 individuals for whom other additional clinical information were available, ENFORCE and ENFORCE plus history of documented nonfatal myocardial infarction or stroke, gave overlapping results in terms of C-statistic for 6-year all-cause mortality, i.e. 0.76 (95% CI, 0.69 to 0.82) and 0.76 (95% CI, 0.70 to 0.82), respectively. Results did not change also if information on retinopathy or anticoagulant therapy or e-GFR CKD-EPI were added to ENFORCE (data not shown).

Validation in a clinical trial setting

The present analysis was restricted to self-reported White subjects of the standard glycaemic arm of the ACCORD trial whose baseline clinical features are reported in Table 1. Mean age was 63.3 (6.6), with 36.7% patients being $\geq$ 65 years old. As compared to Italian samples, ACCORD showed a lower proportion of females and higher BMI values, with 5.4% individuals being normal-weight, 26.3% over-weight and 68.3% obese. During follow-up, 221 (7.0%) patients died. Age- and sex-adjusted mortality incidence rate was 1.2 events per 100 person-years. For 6-year all-cause mortality, ENFORCE achieved a C-statistic of 0.68 (95% CI, 0.65 to 0.72), which was significantly lower than that obtained in the pooled Italian samples (*P* for heterogeneity < 0.0001). A very similar finding was observed after adding history of documented nonfatal myocardial infarction or stroke, with C-statistic being 0.69 (95% CI, 0.65 to 0.73). In addition, at variance with what observed in the Italian samples, C-statistics in ACCORD were not significantly different when comparing different HbA1c strata (see Table 2).

To address whether the difference in ENFORCE performance between Italian samples and ACCORD was partly due to difference in treatment intensity, C-statistics were assessed after stratifying ACCORD participants in primary as compared to secondary prevention (0.69, 95% CI, 0.64 to 0.74 vs. 0.69, 95% CI, 0.64 to 0.74) and in those included in the "blood pressure" as compared to the "lipid" sub-trial (0.69, 95% CI, 0.63 to 0.75 vs 0.69, 95% CI, 0.63 to 0.74).

Comparison between ENFORCE and RECODe

The RECODe has been recently proposed as a well-performing, validated prediction model for several diabetic complications as well as for all-cause mortality (5,6) using predictors as reported in Supplemental Table 1 (19).

In the subgroup of 1082 individuals from the pooled sample in which RECODe's variables were available, a Cox model achieved a C-statistic for 6-year all-cause mortality of 0.74 (95% CI, 0.39 to 0.82). This value is very similar to that achieved by ENFORCE (0.76, 95% CI, 0.69 to 0.82).

In the same ACCORD sub-sample used for our analysis, RECODe showed a predicted 6-year all-cause mortality C-statistic of 0.69 (95% CI, 0.65 to 0.73), again a value that is very similar to that obtained by ENFORCE (0.68, 95% CI, 0.65 to 0.72), as well as to that obtained by RECODe itself in the whole ACCORD sample (5).

Discussion

By using information commonly collected in every-day clinical practice, ENFORCE extends to a 6-year time horizon a previously reported inexpensive, parsimonious and easy-to-use prediction model of all-cause mortality in White patients with type 2 diabetes (4). ENFORCE is highly accurate and well calibrated. In addition, ENFORCE performed similarly well, and therefore was validated, in two additional independent and diverse Italian cohorts, namely FMS and PMS.

On the one hand, we emphasize that our model performed similarly well across three cohorts having somewhat different baseline clinical features and, most importantly, very different mortality rate. It is, therefore, conceivable that ENFORCE is generalizable to broader contexts. On the other hand, we acknowledge that our data were limited to Italian samples, which leaves the issue of transportability of our model to other populations still to be addressed.

Quite interestingly, the performance of ENFORCE was not as good in the ACCORD clinical trial as in the observational studies in which it was developed and validated. This finding resembles very closely the difference in performances observed between observational and interventional studies for the RECODe model (5, 6), and possibly points to a general phenomenon suggesting that predicting mortality risk in diabetic patients is more difficult in a clinical trial setting than in real-life situations. This might be the consequence of differences in intrinsic patients' motivation and cultural background, which are presumably higher in volunteers participating a clinical trial as compared to those investigated in a real-life setting. Such features characterizing volunteers can somehow flatten individual risk profiles and therefore reduce the performance of prediction models.

In addition, the different accuracy that we observed between Italian cohorts and ACCORD could at least in part due to differences in environmental and/or genetic backgrounds. Also differences in several clinical features observed between Italian and ACCORD participants (see Supplemental Table 2 (19)), could have played a role in modifying ENFORCE performance. Among these, the higher proportion of patients who were never smokers and who were on anti-hypertensive and/or anti-dyslipidemic treatments might be of particular relevance. On the contrary, as mentioned above, the relatively low mortality rate observed in the ACCORD cohort is unlikely to be responsible, given the similar, if not lower rate in the PMS, in which ENFORCE performed as well as in the other two Italian samples (i.e. GMS and FMS). In all, given the above-mentioned differences in the cohorts examined, we acknowledge that caution must be used as to the reason for differences in ENFORCE performance between the Italian vs. the ACCORD participants.

Finally, in the pooled Italian sample ENFORCE performed significantly better in individuals with relatively low HbA1c levels as compared to their counterparts with HbA1c $\geq$ 8%, thus suggesting that glycaemic control may play a role in shaping the clinical relevance of our model. However, this difference was not observed in the ACCORD study, thus again casting doubts about the possibility of transferring well performing models into real-life in the context of clinical trials setting.

Several other models have been so far described to predict all-cause mortality in patients with type 2 diabetes (5, 6, 20-30). Some of them lack formal, independent and external validation (20-24). Some other models were based on simulation studies, with the model's prognostic accuracy being, unfortunately, not reported (25-28). In addition, many models have been built based on clinical trial data (25, 28, 30), which leaves the question open as to their transportability to real-life settings. Finally, several studies (5, 6, 21-23, 25, 28-30) included patients of different ethnicity, which makes it impossible to obtain population specific models. Notably, in both Italian and ACCORD samples ENFORCE performs as well as the recently proposed RECODe (5, 6). Though this latter model has investigated individuals from different countries and settings, and used different predictors than ENFORCE (sharing only four of them; Supplemental Table 1 for details), it has been set up and validated using large and well established samples (5, 6) and represents, therefore, a useful tool for benchmarking our ENFORCE.

Limitations of our study include the lack of information on previous cardiovascular events, which are major risk factors for all-cause mortality (4). However, in a subsample representing more than one third of the whole Italian sample, no difference in the predicting

ability of ENFORCE was observed when information on previous cardiovascular events were added, thus making likely that lack of such information does not detract much from ENFORCE performance. Nonetheless, it should be underlined that despite this limitation, ENFORCE is highly accurate and well calibrated. In addition, it should be considered that, in a real-life clinical setting, information on previous cardiovascular events is obtained primarily by self-reporting, which is likely to provide unreliable data that may affect the performance of prediction models (31-33). Thus, the fact that ENFORCE is not based on previous cardiovascular history may actually be an advantage, especially in underprivileged socio-economic strata, in which the likelihood of inaccurate information on previous cardiovascular events is likely to be higher. Another limitation that should be acknowledged is that additional studies, better if in "real-life" samples, are needed to address the transportability of ENFORCE to different genetic, environmental, and cultural backgrounds.

In summary, we have further validated and extended the time horizon of ENFORCE - our previously described parsimonious and simple-to-use all-cause mortality prediction model for patients with type 2 diabetes. Notably, this model is inexpensive and therefore applicable also in environments where resources are limited. With the goal of helping clinicians identify individuals with type 2 diabetes at high risk of premature death, we are providing free public access to ENFORCE as a user-friendly web-based risk engine (http://www.operapadrepio.it/enforce/enforce.php). Currently, our effort is focused in further validating ENFORCE in additional samples from Italy (34) and possibly other European Countries, so to investigate its transportability in larger samples as well as in a wider geographical context.

We expect that implementation of ENFORCE as well as other predictive models (5, 6, 25, 28), according to their performance and applicability in specific real-life settings, will allow the targeting of more aggressive, expensive, and burdensome preventive strategies only to those patients who are predicted to be at very high-risk, thereby improving the cost-effectiveness of available and often limited resources.

Acknowledgments

This work was supported by the Italian Ministry of Health grants, RC2016, RC2017 and RF-2013-02356459 (to C.M.) and by Italian Ministry of University and Scientific Research "Progetti di Ricerca di Interesse Nazionale" (PRIN) 2012 grant 31 (to V.T.). M.G.S is supported by the Italian Ministry of Health grant RF-2013-02356459.

The ACCORD study (clinicaltrials.gov identifier NCT00000620) was supported by National Heart, Lung, and Blood Institute contracts N01-HC-95178, N01-HC-95179, N01-HC-95180, N01-HC-95181, N01-HC-95182, N01-HC-95183, N01-HC-95184, IAA-Y1-HC-9035, and IAA-Y1-HC-1010. Other components of the National Institutes of Health, including the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute on Aging, and the National Eye Institute, contributed funding.

Fundings

No sponsor had any role in the study design; collection, analysis, or interpretation of the data; or writing of the report. MC, AF, AD and VT had access to the raw data. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

Ministero della Salute http://dx.doi.org/10.13039/501100003196, RC2016, RC2017, RF-2013-02356459, Claudia Menzaghi; Ministero dell'Istruzione, dell'Università e della Ricerca http://dx.doi.org/10.13039/501100003407, PRIN2012 grant 31, Vincenzo Trischitta; Ministero della Salute http://dx.doi.org/10.13039/501100003196, RF-2013-02356459, Maria Giovanna Scarale; National Heart, Lung, and Blood

Institute http://dx.doi.org/10.13039/100000050, N01-HC-95178, N01-HC-95179, N01-HC-95180, N01-HC-95181, N01-HC-95182, N01-HC-95183, N01-HC-95184, IAA-Y1-HC-9035, IAA-Y1-HC-1010, Alessandro Doria

Author contributions

MC and VT participated in the study conception and design. HS, CM, SDC, MG and MRS contributed to data acquisition. MC, HS, AF and MGS performed statistical analysis. MC, HS, AF, CM AD and VT participated in data analysis and interpretation. MC, AD and VT draft the manuscript. MC, CM, SDC, OL, GP, AD and VT critically reviewed the manuscript. MC and VT are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Corresponding author and person to whom reprint requests should be addressed: Massimiliano Copetti, PhD, Unit of Biostatistics, IRCCS "Casa Sollievo della Sofferenza, Viale Padre Pio, 71013 San Giovanni Rotondo, Italy, Phone: ++390882416292, Fax: ++390882416266, E-mail: m.copetti@operapadrepio.it

and

Vincenzo Trischitta, MD, Research Unit of Diabetes and Endocrine Diseases, IRCCS "Casa Sollievo della Sofferenza, Viale Padre Pio, 71013 San Giovanni Rotondo, Italy, Phone: ++390882416276, Fax: ++390882416266, E-mail: vincenzo.trischitta@uniroma1.it

Declaration of interests
None of the authors have relevant conflicts of interest.

References

1. International Diabetes Federation. IDF Diabetes Atlas, 8th edn. Brussels, Belgium: International Diabetes Federation, 2017. http://www.diabetesatlas.org.
2. WHO Global report on diabetes, 2016. Available on the WHO website (http://www.who.int), 6 April 2016.
3. Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. *Diabetes Res Clin Pract.* 2017;128:40-50. [PMID: 28437734] doi:10.1016/j.diabres.2017.03.024.
4. De Cosmo S, Copetti M, Lamacchia O, Fontana A, Massa M, Morini E, et al. Development and validation of a predicting model of all-cause mortality in patients with type 2 diabetes. *Diabetes care.* 2013;36(9):2830-5. [PMID: 23637348] doi: 10.2337/dc12-1906.
5. Basu S, Sussman J, Berkowitz S, Hayward R, Yudkin J. Development and validation of Risk Equations for Complications of Type 2 Diabetes (RECODe) using individual participant data from randomized trials. *Lancet Diabetes Endocrinol.* 2017; 5:788-798. [PMID: 28803840] doi: 10.1016/S2213-8587(17)30221-8.
6. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Bertoni AG, Correa A, et al. Validation of Risk Equations for Complications of Type 2 Diabetes (RECODe) Using Individual Participant Data From Diverse Longitudinal Cohorts in the U.S. *Diabetes Care.* 2018;41(3):586-595. [PMID: 29269511] doi: 10.2337/dc17-2002.
7. Lamacchia O, Fontana A, Pacilli A, Copetti M, Fariello S, Garofolo M, et al. On the non-linear association between serum uric acid levels and all-cause mortality rate in patients with type 2 diabetes mellitus. *Atherosclerosis.* 2017;260:20-26. [PMID: 28334637] doi: 10.1016/j.atherosclerosis.2017.03.008.

8. Prudente S, Di Paola R, Pezzilli S, Garofolo M, Lamacchia O, Filardi T, et al. Pharmacogenetics of oral antidiabetes drugs: evidence for diverse signals at the IRS1 locus. *Pharmacogenomics J*. 2018;18(3):431-435. [PMID: 28696414] doi: 10.1038/tpj.2017.32.
9. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC Jr, Bigger JT, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-59. [PMID: 18539917] doi: 10.1056/NEJMoa0802743.
10. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008;27:157-172. [PMID: 17569110] doi: 10.1002/sim.2929.
11. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med*. 2011;30:11-21. [PMID: 21204120] doi: 10.1002/sim.4085.
12. Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med*. 2012;31:101-113. [PMID: 22147389] doi: 10.1002/sim.4348.
13. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med*. 2011;30(10):1105-17. [PMID: 21484848] doi: 10.1002/sim.4154.
14. Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med* 2004;23:2109-2123. [PMID: 15211606] doi: 10.1002/sim.1802.
15. Demler OV, Paynter NP, Cook NR. Tests of calibration and goodness-of-fit in the survival setting. *Stat Med*. 2015;34:1659-1680. [PMID: 25684707] doi: 10.1002/sim.6428.
16. Hothorn T, Zeileis A. partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*. 2015;16:3905-3909.
17. Tang F, Ishwaran H. Random forest missing data algorithms. *Stat Anal Data Min: The ASA Data Sci Journal*. 2017;10:363-377. [PMID: 29403567] doi: 10.1002/sam.11348.
18. Shah A, Bartlett J, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: A CALIBER study. *American Journal of Epidemiology*. 2014;179:764-74. [PMID: 24589914] doi: 10.1093/aje/kwt312.
19. Copetti M, Shah H, Fontana A, Scarale MG, Menzaghi C, De Cosmo S, Garofolo M, Sorrentino MR, Lamacchia O, Penno G, Doria A, Trischitta V. EstimatioN oF mORtality risk in type2 diabetiC patiEnts (ENFORCE): an inexpensive and parsimonious prediction model. Figsharecom Deposited April 24, 2019. https://doi.org/10.6084/m9.figshare.8035547.v1
20. Hong Kong Diabetes Registry, Yang X, So WY, Tong PC, Ma RC, Kong AP, et al. Development and validation of an all-cause mortality risk score in type 2 diabetes. *Archives of Internal Medicine*. 2008;168:451-457. [PMID: 18332288] doi: 10.1001/archinte.168.5.451.
21. Wells BJ, Jain A, Arrigain S, Yu C, Rosenkrans WA Jr, Kattan MW. Predicting 6-year mortality risk in patients with type 2 diabetes. *Diabetes Care*. 2008; 31:2301-2306. [PMID: 18809629] doi: 10.2337/dc08-1047.
22. McEwen LN, Karter AJ, Waitzfelder BE, Crosson JC, Marrero DG, Mangione CM, et al. Predictors of mortality over 8 years in type 2 diabetic patients: Translating Research Into Action for Diabetes (TRIAD). *Diabetes Care*. 2012;35:1301-1309. [PMID: 19228862] doi: 10.2337/dc08-1911.
23. Wells BJ, Roth R, Nowacki AS, Arrigain S, Yu C, Rosenkrans WA Jr, et al. Prediction of morbidity and mortality in patients with type 2 diabetes. *Peer J*. 2013;1:e87. [PMID: 23781409] doi: 10.7717/peerj.87.

24. Raffield LM, Hsu FC, Cox AJ, Carr JJ, Freedman BI, Bowden DW. Predictors of all-cause and cardiovascular disease mortality in type 2 diabetes: Diabetes Heart Study. *Diabetology & Metabolic Syndrome*. 2015;7:58. [PMID: 26146522] doi: 10.1186/s13098-015-0055-y.
25. Hayes AJ, Leal J, Gray AM, Holman RR, Clarke PM. UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia*. 2013;56:1925-1933. [PMID: 23793713] doi: 10.1007/s00125-013-2940-y.
26. Palmer AJ, Roze S, Valentine WJ, Minshall ME, Foos V, Lurati FM, et al. The CORE Diabetes Model: Projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making. *Current Medical Research and Opinion*. 2004;20 Suppl 1:S5-26. [PMID: 15324513] doi: 10.1185/030079904X1980.
27. Palmer AJ1, Roze S, Valentine WJ, Minshall ME, Foos V, Lurati FM, et al. Validation of the CORE Diabetes Model against epidemiological and clinical studies. *Current Medical Research and Opinion*. 2004;20 Suppl 1:S27-40. [PMID: 15324514] doi: 10.1185/030079904X2006.
28. McEwan P, Foos V, Palmer JL, Lamotte M, Lloyd A, Grant D. Validation of the IMS CORE Diabetes Model. *Value in Health*. 2014;17:714-724. [PMID: 25236995] doi: 10.1016/j.jval.2014.07.007.
29. Robinson TE, Elley CR, Kenealy T, Drury PL. Development and validation of a predictive risk model for all-cause mortality in type 2 diabetes. *Diabetes Research and Clinical Practice*. 2015;108:482-488. doi: 10.1016/j.diabres.2015.02.015.
30. Shao H, Fonseca V, Stoecker C, Liu S, Shi L. Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO). *PharmacoEconomics*. 2018;36:1125-1134. [PMID: 29725871] doi: 10.1007/s40273-018-0662-1.
31. Rosella LC, Corey P, Stukel TA, Mustard C, Hux J, Manuel DG. The influence of measurement error on calibration, discrimination, and overall estimation of a risk prediction model. *Popul Health Metr*. 2012;10(1):20. [PMID: 23113916] doi: 10.1186/1478-7954-10-20.
32. Molenaar EA, Van Ameijden EJ, Grobbee DE, Numans ME. Comparison of routine care self-reported and biometrical data on hypertension and diabetes: results of the Utrecht Health Project. *Eur J Public Health*. 2007;17(2):199-205. [PMID: 16887835] doi: 10.1093/eurpub/ckl113.
33. Bays HE, Chapman RH, Fox KM, Grandy S, SHIELD Study Group. Comparison of self-reported survey (SHIELD) versus NHANES data in estimating prevalence of dyslipidemia. *Curr Med Res Opin*. 2008;24(4):1179-86. [PMID: 18346311] doi: 10.1185/030079908X280527.
34. Barchetta I, Capoccia D, Baroni MG, Buzzetti R, Cavallo MG, De Cosmo S, Leonetti F, Leotta S, Morano S, Morviducci L, Prudente S, Pugliese G, Trischitta V; SUMMER Study in Diabetes Group. The "Sapienza University Mortality and Morbidity Event Rate (SUMMER) study in diabetes": Study. Nutr Metab Cardiovasc Dis. 2016 Feb; 26(2):103-8. https://www.summerstudyindiabetes.it

Figure 1. Calibration plot. Expected Kaplan-Meier mortality rate, based on the proposed 6-year all-cause mortality prediction model in the Italian pooled sample, versus observed Kaplan-Meier mortality rate.

Figure 2. **Kaplan Meier survival curves.** Kaplan Meier survival curves of the four risk categories as determined by the survival conditional tree analysis.

Table 1. Baseline clinical, demographical and laboratory characteristics of diabetic patients enrolled in the four samples

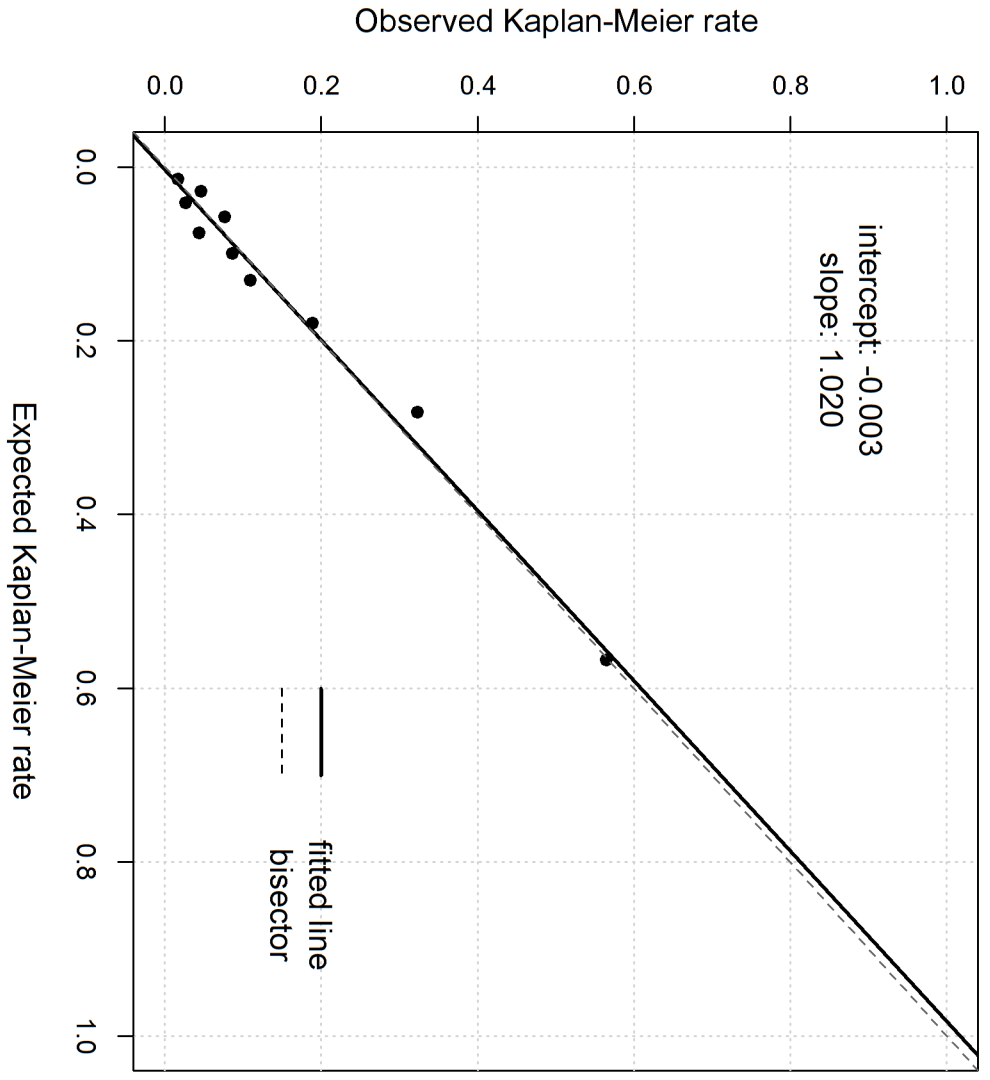
	Category	GMS (*n* = 1 019)	FMS (*n* = 1 045)	PMS (*n* = 972)	ACCORD (*n* = 3 150)
Age (years)	mean (SD)	61.1 (9.7)	63.6 (11.8)	59.6 (7.1)	63.3 (6.6)
Sex, n (%)	Males	512 (50.2)	510 (48.8)	580 (59.7)	1079 (34.3)
Smoking habits (%)	Never smokers	789 (77.6)	531 (55.2)	768 (79.1)	2815 (89.4)
	Ex-smokers	86 (8.5)	264 (27.4)	0 (0.0)	0 (0.0)
	Smokers	142 (14.0)	167 (17.4)	203 (20.9)	335 (10.6)
BMI (kg/m^2)	mean (SD)	31.0 (5.8)	30.3 (6.6)	29.7 (5.3)	31.9 (5.2)
Duration of diabetes (years)	mean (SD)	10.9 (9.0)	13.1 (10.0)	10.1 (8.5)	10.6 (7.5)
Dyslipidemia, n (%)*	Yes	875 (86.8)	848 (86.1)	815 (83.8)	2,730 (86.7)
Glycated hemoglobin (%)	mean (SD)	8.7 (1.0)	9.0 (1.1)	7.6 (1.2)	8.2 (0.9)
SBP (mmHg)	mean (SD)	134.5 (16.6)	130.4 (15.5)	142.8 (19.2)	135.1 (16.3)
DBP (mmHg)	mean (SD)	78.3 (8.9)	76.6 (9.0)	82.4 (10.1)	74.0 (10.2)
HDL-C (mg/dl)	Median [IQR]	42 [35, 52]	45 [37, 54]	48 [41, 57]	39 [33, 46]
LDL (mg/dl)	Median [IQR]	119 [93, 143]	101 [77, 127]	130 [109, 149]	98 [79, 123]
Total cholesterol (mg/dl)	Median [IQR]	192 [166, 223]	182 [151, 211]	201 [178, 224]	177 [154, 207]
Triglycerides (mg/dl)	Median [IQR]	131 [94, 187]	141 [99, 202]	133 [98, 199]	170 [120, 248]
Uric acid (mg/dl)	Median [IQR]	5.1 [4.2, 6.2]	5.4 [4.5, 7.1]	5.2 [4.3, 6.2]	NA
Creatinine (mg/dl)	Median [IQR]	0.9 [0.8,1.1]	0.8 [0.7,1.0]	0.8 [0.7,1.0]	0.9 [0.8,1.0]
ACR (mg/mmol)	Median [IQR]	1.3 [0.6, 4.1]	1.8 [0.7, 8.2]	0.7 [0.4, 1.8]	1.3 [0.7, 3.9]
Albuminuria (mg/g)	Median [IQR]	10.7 [4.3, 29.4]	16.3 [6.1, 64.2]	4.1 [3.0, 15.4]	1.50 [0.7, 4.2]
e-GFR CKD-EPI (mL/min per 1.73m^2)	Median [IQR]	76.2 [63.2, 88.5]	85.7 [63.2, 98.9]	89.9 [76.6, 97.6]	85.8 [71.3, 95.5]
Anti-hypertensive TX, n (%)	Yes	540 (57.8)	679 (68.6)	516 (53.1)	2713 (86.1)
Insulin TX, n (%)	Yes	424 (41.8)	363 (36.7)	249 (25.6)	1098 (34.9)
Anti-dyslipidemic TX, n (%)	No	695 (68.4)	652 (66.0)	653 (67.2)	987 (31.5)
	Statins	295 (29.0)	302 (30.6)	269 (27.7)	1933 (61.6)
	Fibrates	26 (1.6)	34 (3.4)	50 (5.1)	216 (6.9)**
Follow-up (years)	Median [IQR]	11.8 [9.4, 13.0]	7.4 [6.5, 8.3]	11.2 [10.9, 11.7]	5.0 [4.1, 5.7]
Vital status n (%)	Deaths	333 (31.7)	309 (29.6)	154 (15.8)	221 (7.0)

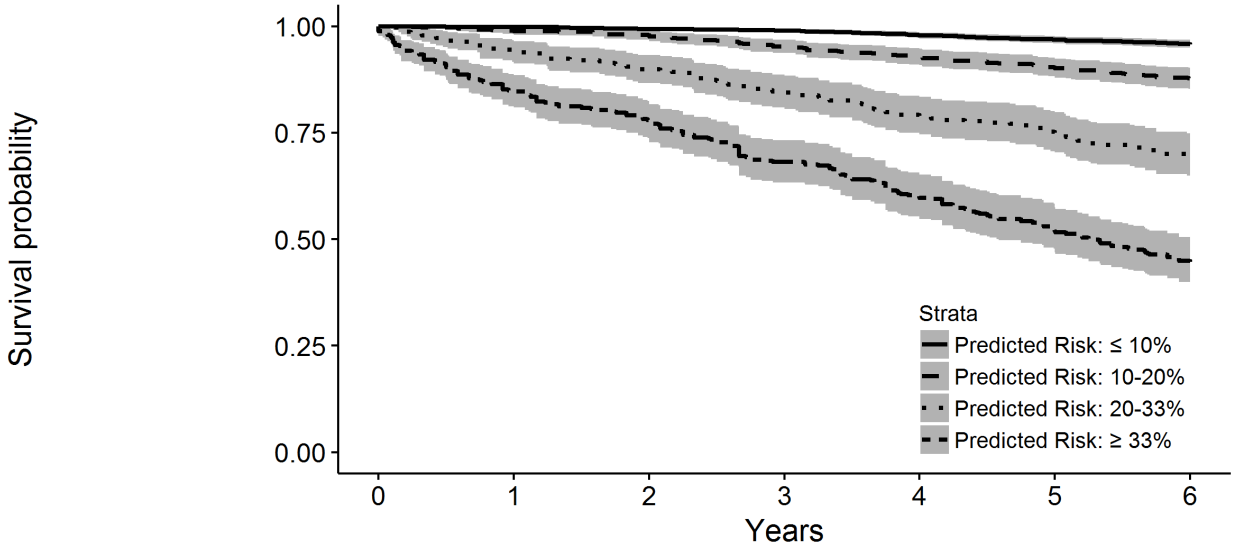
GMS = Gargano Mortality Study; FMS = Foggia Mortality Study; PMS = Pisa Mortality Study; ACCORD = Action to Control Cardiovascular Risk in Diabetes. BMI= body mass index, SBP=systolic blood pressure, DBP=diastolic blood pressure, HDL-C = high-density lipoprotein-cholesterol; ACR = urinary albumin-to-creatinine ratio; LDL = low-density lipoprotein, TX=therapy.
* Dyslipidemia: in GMS, FMS and PMS defined as: XXX LDL > 100 mg/dl or HDL-C < 4.0 mg/dl or Triglycerides ≥ 150 mg/dl in ACCORD defined as: LDL > 100 mg/dl or HDL-C < 4.0 mg/dl or Triglycerides ≥ 150 mg/dl.
** Patient treated with fibrates only or with both fibrates and statins.

Table 2. ENFORCE 6-years prediction accuracies according to HbA1c stratum in the Italian pooled sample and in the ACCORD study

Study	Subsample	N participants	N events	C-statistic (95% CI)	P value
Italian samples*	HbA1c ≥ 8% stratum	1539	267	0.78 (0.75-0.80)	0.003
	HbA1c < 8% stratum	1414	167	0.84 (0.81-0.87)	
ACCORD#	HbA1c ≥ 8% stratum	1 693	119	0.72 (0.67-0.77)	0.179
	HbA1c < 8% stratum	1 439	81	0.67 (0.61-0.72)	

*83 patients were excluded because of missing HbA1c information
#18 patients were excluded because of missing HbA1c information





Number at risk

Predicted Risk: ≤ 10%	1686	1683	1675	1669	1651	1632	1615
Predicted Risk: 10-20%	678	671	664	646	630	612	593
Predicted Risk: 20-33%	327	308	294	277	259	246	228
Predicted Risk: ≥ 33%	345	293	269	235	208	182	155

Cumulative number of censoring

Predicted Risk: ≤ 10%	0	0	0	0	0	1	1614
Predicted Risk: 10-20%	0	0	0	0	0	1	594
Predicted Risk: 20-33%	0	0	0	0	0	0	226
Predicted Risk: ≥ 33%	0	0	0	0	0	0	154