**Prediction of mortality and major cardiovascular complications in type 2 diabetes: External validation of UK Prospective Diabetes Study outcomes model version 2 in two European observational cohorts**

Full title

**Prediction of mortality and complications in type 2 diabetes: external validation of UKPDS outcomes model version 2 in the Casale Monferrato Survey and Hoorn Diabetes Care System cohorts.**

Short running title

Validation of UKPDS-OM2 in two European cohorts

Authors:

Eva Pagano[1], MSc; Stefan RA Konings[2], MSc; Daniela Di Cuonzo, PhD[1]; Rosalba Rosato[3], Professor; Graziella Bruno[4], Professor; Amber A van der Heijden[5,6], PhD; Joline Beulens, PhD[6], Roderick Slieker[6,7], PhD, Jose Leal[8], DPhil; Talitha L Feenstra[9,10], PhD.

1. Unit of Clinical Epidemiology, "Città della Salute e della Scienza" Hospital, Turin, Italy and CPO Piemonte, Turin, Italy.
2. University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), Groningen, The Netherlands.
3. Department of Psychology, University of Turin, Italy.
4. Laboratory of Diabetic Nephropathy, Department of Medical Sciences, University of Turin, Turin, Italy. University of Turin, Italy.
5. Department of General Practice, Amsterdam Public Health Institute, Amsterdam UMC, location VUMC, Amsterdam, the Netherlands
6. Department of Epidemiology and Biostatistics, Amsterdam Public Health Institute, Amsterdam UMC, location VUMC, Amsterdam, the Netherlands
7. Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands
8. Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, UK
9. University of Groningen, Faculty of Science and Engineering, Groningen Research Institute of Pharmacy, Groningen, The Netherlands .
10. RIVM, Bilthoven, The Netherlands.

Corresponding author:

Eva Pagano

Unit of Clinical Epidemiology

“Città della Salute e della Scienza” Hospital

Via Santena 5, 10126, Torino

Tel. 0039 011 633 68 55

Abstract word count: 249

Main text word count: 4007

Number of tables and figures: 2 tables and 2 figures

## Abstract

**Aims:** The United Kingdom Prospective Diabetes Study Outcomes Model version 2 (UKPDS-OM2) is a health economic decision model supporting long-term evaluation of treatment options in diabetes. The aim of the study was to externally validate the model by comparing the predicted and observed outcomes in two European population-based cohorts of people with type 2 diabetes.

**Materials and methods:** We used data from the Casale Monferrato Survey (CMS) [n=1931] and a subgroup of the Hoorn Diabetes Care System (DCS) cohorts [N=5188]. The following outcomes were analysed: all-cause mortality, myocardial infarction (MI), ischaemic heart disease (IHD), stroke, and congestive heart failure (CHF). Model performance was assessed by comparing predictions with observed cumulative incidences in each cohort during follow-up.

**Results:** All-cause mortality was overestimated by the UKPDS-OM2 in both the cohorts, with a bias of 0.05 in CMS and 0.12 in DCS at 10 years of follow-up. For MI, predictions were consistently higher than observed incidence over the entire follow-up in both cohorts (10 years bias 0.07 for CMS and 0.10 for DCS). The model performed well for stroke and IHD outcomes in both cohorts. Heart failure incidence was predicted well for DCS (5 years bias -0.001) but underestimated for CMS.

**Conclusions**: UKPDS-OM2 consistently overpredicted the risk of mortality and MI in both cohorts during follow-up. Period effects may partially explain the differences. Results indicate that transferability is not satisfactory for all outcomes and new or adjusted risk equations may be needed before applying the model to the Italian or Dutch settings.

## Introduction

Several diabetes outcomes simulation models have been developed and validated for type 2 diabetes (T2D) mellitus populations to support health care policy making. They are mainly used to extrapolate long-term clinical outcomes and costs from clinical trials, to assess cost-effectiveness and to estimate expected total budget impact of introducing new types of treatment. These analyses aid decision makers as part of a Health Technology Assessment (HTA) to support health care priority setting. Models that enable simulation at the level of individual patients (microsimulation models) in particular are useful in tracking multiple comorbidities and patient outcomes to properly reflect heterogeneity and facilitate evaluations of precision medicine (1).

The United Kingdom Prospective Diabetes Study Outcomes Model version 2 (UKPDS-OM2) is an individual-level microsimulation tool used to predict health outcomes of individuals diagnosed with T2D (2). The model predicts several types of complications common to people with T2D, using Monte Carlo simulation and risk equations fitted on the UKPDS data that account for a range of patient characteristics, such as age, sex, pre-existing complications, lab values and lifestyle habits. The model, as described by Hayes et al., (2) is transparent and has been validated internally using UKPDS data.

In view of a potential wide utilisation of the UKPDS-OM2 in cost-effectiveness analysis and in the evaluation of strategies for the management of T2D at European level in the future, external validation in data across European countries is of great interest. As observed in a validation of the first version of the UKPDS-OM (3), differences in health and health care between countries, reflected in differences in parameters such as in life expectancy of the general population and mortality risk for T2D, are likely to determine biased estimates of the outcomes. An assessment of model behaviour in different contexts can provide information on the factors affecting validity in different settings according to population characteristics. This requires detailed insight into model performance for a variety of outcomes and performance measures.

The previous release of the UKPDS-OM2, the UKPDS-OM1, has been validated in several settings (3–6). Previous patient level validation work of the UKPDS-OM2 equations has been done in the US (7,8) and Germany (9). They consistently report over-prediction of all-cause mortality but varied in terms of performance relative to other outcomes. In the US studies, the risk of cardiovascular (CV) outcomes was found to be overestimated and generally poorly discriminated at 10 years of follow-up. In the German cohort, the UKPDS-OM2 over-predicted mortality at 10 years and showed poor discrimination for the CV outcomes despite no significant difference in observed and predicted CV outcomes. Since these results are contradictory and the sample size was small (less than 500 participants), evidence on model performance in an European setting is currently quite limited.

The aim of the present study therefore was to assess the performance of the UKPDS-OM2 in two large cohorts representative of different epidemiological scenarios within the European context, the Casale Monferrato Survey (CMS) from Italy and the Hoorn Diabetes Care System cohort (DCS) from the Netherlands.

## Methods

UKPDS-OM2 was used to simulate the DCS and CMS populations from baseline up to 10 and 15 years respectively, in order to compare its predicted cumulative incidences of T2D-related health outcomes with the observed cumulative incidences. The outcomes considered were all-cause mortality and the incidence of the following fatal and non-fatal events: myocardial infarction (MI), stroke, congestive heart failure (CHF), and other ischemic heart disease (IHD). The list of International Classification of Disease, 9$^{th}$ revision [ICD-9] and 10$^{th}$ revision [ICD-10] codes used for defining fatal and non-fatal events was derived from the UKPDS study and is provided in the Supplementary Table 1. Patients were included in the analysis if data were available at baseline on a predefined core set of risk factors, namely: sex; age; duration of diabetes (years); BMI; smoking

status (current smoker or not); total, HDL and LDL cholesterol; systolic blood pressure; HbA1c and eGFR.

*Casale Monferrato Survey*

The CMS started in 1988 (10) with the aim of assessing the prevalence of known diabetes in individuals living in the area of Casale Monferrato, Northern Italy. In 2000, a new survey included all members of the original cohort who were still alive and living in Casale Monferrato (N=860) plus all new diagnosis of T2D (N=2389) (11).

We identified 1931 patients meeting the inclusion criteria (out of a total of n=3249). Patients were excluded in case of presence of missing data in at least one of the core set of risk factors (N=995) and in case of lack of follow-up data due to record linkage problems (N=323). Table 1 reports the characteristics of the final cohort at baseline. Albuminuria, history of MI and history of stroke were partially missing. All patients were of European descent.

Outcomes during follow-up were identified from hospital discharge records (HDR), during the period 2000-2017, by means of a record linkage procedure based on an encrypted identification number. Vital status was last updated at the end of 2017. From 2000 to 2006 underlying causes of death were ascertained from death certificates and coded by two physicians. From 2007 to 2015 causes of death were derived directly from the national Death Registry. Fatal events were classified according ICD-9-CM [Clinical modification] codes. No risk factor data were available after baseline.

*Hoorn Diabetes Care System cohort*

The DCS cohort in the Netherlands is a primary care registry of individuals diagnosed with T2D (n=14756 in 2018). The DCS cohort started in 1998 and includes almost all people with T2D in the area of the DCS in West-Friesland. All participants visited the central diabetes center annually. Since 2008, a random subsample was invited for the DCS biobank (n=5946) (12). Participants in biobank subsample were asked for consent to link with medical records, to enable validation of self-reported

information on events with hospital records (13). Linkage took place in West-Friesland hospitals, during the period 2015-2018.We excluded 50 individuals for whom outcome data were not available, 500 individuals with year of entry after 2013, and another 208 with missing data on core variables at baseline, resulting 5188 individuals used for the analysis of all-cause mortality (Table 1 and Supplementary Table 2).

The last year of enrollment was 2013 to enable at least 5 years of follow-up. Distribution by year of entry is reported in the Supplementary Figure 1. During follow-up, a maximum of 19.4% of risk factors was missing because of measurement errors, administrative mistakes, or missed annual checks. When more than one observation on a risk factor was available in a year, only the first measurement was considered. Due to censoring caused by entry after 2008, for 63% of individuals the 10 year follow-up time was not complete. A sensitivity analysis was performed with the subgroup of individuals followed for a full 10 years, i.e. using those individuals enrolled in 2008 (n=1931).

Causes of death were ascertained from national death records and last updated at the beginning of 2020. All outcomes were recorded using ICD-9 definitions and carefully aligned with the definitions of UKPDS-OM2 and CMS (see above).

*The UKPDS Outcomes Model version 2*

The UKPDS-OM2 is based on patient-level data from the United Kingdom Prospective Diabetes Study (2). We provide the characteristics of the UKPDS cohort used to inform the model, at 7 and 11 years of follow-up, in Table 1. The model simulates T2D populations modelling the occurrence of eighth diabetes-related complications (MI, IHD, stroke, CHF, amputation, renal failure, diabetic ulcer and blindness in one eye), second events (MI, stroke and amputation) and death to estimate (quality-adjusted) life expectancy, and costs. The UKPDS-OM2 predicts outcomes at patient level based on patient demographics (age, sex, ethnicity), duration of diabetes, risk factor levels over time, and history of T2D-related complications, using a probabilistic discrete-time state-transition model.

Further details on the model are reported in the Supplementary material.

*Missing data*

For individuals with missing data on variables other than core ones, we assumed these to be missing at random (MAR) and used multiple imputation to predict the missing values at patient level. In CMS, the percentage of individuals with missing values at baseline was 24% and in DCS 3.2%. For both cohorts, 25 datasets were imputed. Further details are reported in the Supplementary material (Supplementary Tables 3-8).

Following multiple imputation, values of the completely missing variables at baseline (Supplementary Table 4) were imputed by means of regression models developed in the UKPDS population.

In the simulation, we used the clinical risk factor levels as observed during the follow-up of the cohorts, if available. Missing risk factor data during follow-up were imputed using risk factor time-path equations developed by the UKPDS modelling team, based on the UKPDS cohort. As an additional analysis, risk factor trajectories were carried forward from last observed value (baseline in CMS).

*Model validation*

The model was run for each cohort using all patients with imputed data from time of entry into DCS and CMS cohorts up to 10 and 15 years of follow-up, respectively. Outcomes predicted by the UKPDS-OM2 were recorded. Further details are reported in the Supplementary material (Missing data section).

In predicting incidence of each T2D-related complication, only the first event after diagnosis was counted. We removed individuals with pre-existing events, resulting in specific sample sizes for each type of event (Supplementary Table 9).

Model validation was performed by comparing UKPDS-OM2 predictions with mean and 95% CI of the observed cumulative incidences in each cohort at 5, 10 and 15 (CMS only) years of follow-up, i.e. "calibration-in-the-large". UKPDS-OM2 was judged to be well calibrated for a particular outcome if the predicted probability fell within the 95% CI of the probability estimated from the observed data. A competing risk analysis was used to assess the cumulative incidence of outcome, with death for other causes as a competing event (SAS macro cuminc for CMS, R survival package for DCS).

We also calculated the difference between predicted and observed means in cumulative incidence: bias (difference between observed and predicted means), mean absolute error (arithmetic average of the absolute errors), root mean squared error (average of the squares of the errors) and mean absolute percentage error (average of the error in percentage terms) (14) (see online appendix for formulae). For all these measures values closer to zero describe a better accuracy.

Finally, predicted and observed cumulative incidence for all the outcomes at the different time points (5, 10 and, for CMS only, 15 years) were plotted together in one graph per cohort. We then estimated a linear regression for each cohort and report the resulting $R^2$.

Model discrimination, i.e. ability to distinguish individuals with different outcomes (15), was estimated with C-statistics using patient's observed survival time and predicted event-free survival at 5, 10 and 15 (CMS only) years, for each of the outcomes. In DCS data, C-statistic was calculated with the Uno's method in order to account for right censoring (16). In CMS data, there was no right censoring and the Harrell's C-statistic was used instead.

Analyses were performed with SAS 9.4 for CMS and R 4.0.0 for DCS.

*Sensitivity analysis*

All-cause mortality is a key outcome as it is a competing risk for the remaining outcomes. Hence, the following subgroup analyses were performed concerning the all-cause mortality outcome:

- Sex: male and female

- Age at baseline: up to 65 years old and above
- HbA1c at baseline: up to 7.4% (<59 mmol/mol) and 7.5% or higher (≥59mmol/mol)
- Duration of diabetes at baseline: 0-5, 6-10 and 11+ years
- BMI at baseline: up to 24.9, 25 to 29.9, 30+ kg/m2.

Calibration was also evaluated by plotting the observed (95% CI) and predicted mortality by decile of predicted cumulative mortality at year 10 of follow-up.

## Results

Baseline characteristics of the cohorts are provided in Table 1. The CMS and DCS cohorts differed mainly in the mean duration of diabetes, with CMS including more prevalent cases (mean duration of 10.9 years) and DCS including more recently diagnosed cases (mean duration of 6.8 years) (see Table 1). Age, SBP, smoking prevalence, also differed slightly among the cohorts. All the other risk factors had comparable mean values for DCS and CMS, while UKPDS showed lower rates of complications and higher HbA1c values.

Figure 1 shows the predicted and observed (with 95% CI) cumulative incidences in the two cohorts up to 10 years (DCS) and 15 years (CMS) of follow-up, for all the outcomes. Table 2 reports the comparison of observed (with 95%CI) and predicted cumulative incidence at 5, 10 and 15 (CMS only) years, and mean bias, by outcome and cohort. The model overestimated all-cause mortality in CMS, with a bias increasing from 0.04 at 5 years to 0.06 at 15 years of follow-up. Overestimation of mortality was larger for the DCS cohort, with a bias of 0.09 at 5 years and 0.14 at 10 years. The UKPDS-OM2 also overestimated MI incidence in both cohorts, with a bias of 0.05 at 5 years and 0.07-0.09 at 10 years. The model performed better for stroke and IHD in both cohorts. For stroke, bias was small (-0.01 and 0.01) at 5 years, slightly increasing with longer follow-up in both cohorts (-0.02 and 0.02). For IHD, similarly bias was small at 5 years (-0.01 and 0.001) and 10 years (-0,02

and 0.03).The model predicted heart failure incidence well for the DCS cohort (bias -0.001 at 5 years) but underestimated heart failure for CMS, with a bias ranging from -0.03 at 5 years to -0.07 at 15 years.

Supplementary Figure 2 plots the predicted vs observed cumulative incidence for all outcomes in one graph and all time point (at 5, 10 and, for CMS, 15 years), by cohort.

Results of the analysis of all-cause mortality by subgroup are reported graphically in Supplementary Figure 3. In CMS, the UKPDS-OM2 showed a reasonable performance (within the 95%CI of the observed rate) for the following subgroups: men, age below 65 years (up to 10 years of follow-up), a median duration of diabetes of 6-10 years, and BMI below 25. The model predicted particularly well in the subgroup with HbA1c above or equal to 7.5%. Similarly, in the DCS cohort, even if the overestimation was high for all the groups, the prediction was better for those under 65 years, with HBA1c above or equal to 7.5% and with BMI below 25.

In a sensitivity analysis carrying risk factors forward from last observed value (from baseline in CMS) during follow-up, the effect was negligible for the outcomes in the DCS. However, in the CMS, the overestimation of mortality and MI rates was reduced when risk factors were held constant (carried forward) from baseline (see Supplementary Figure 4).

Figure 2 shows the calibration plots of the observed all-cause mortality by decile of predictions at 10 years of follow-up. In CMS, model predictions were within the observed 95% CI for all but three subgroups but the predicted point estimate overestimated mortality in all subgroups. In DCS, model predictions were outside the observed 95% CI for all but one subgroup. As with CMS, the predicted point estimates overestimated observed mortality for all ten subgroups. However, when we used only individuals enrolled in 2008 (i.e. followed for the full 10 years), the model performance improved with overestimations being smaller for many subgroups.

Additional measures of calibration over time are reported in the Supplementary Table 10. In line with the "calibration in the large" results, calibration was worst for MI in CMS followed by all-cause mortality. In DCS, calibration was worst for all-cause mortality and MI, with small errors for CHF.

Supplementary Table 11 reports the C-statistics concerning the UKPDS-OM2's discriminatory capability in the CMS and DCS cohorts. For the CMS cohort, C-statistics values were above 70% for all-cause mortality across the three time points considered. For the remaining outcomes, the C-statistic values were around 60-65% at 5 and 10 years, and lower at 15 years for MI and IHD (59%). In DCS, the C-statistics indicated a reasonably good model performance (above 70%) for mortality, heart failure, AMI and stroke (at 10 years). The model predictions for IHD performed the worst in terms of discrimination in DCS (66% at 5 years and 10 years).

## Discussion

In an external validation of the UKPDS-OM2 with individual patient level data from two European cohorts, the Italian CMS cohort (South Europe) and the Dutch DCS cohort (Western Europe), the UKPDS-OM2 over-predicted the risk of all-cause mortality and MI in both cohorts but performed well for stroke and IHD outcomes. The predicted incidence of heart failure was accurate in the Dutch cohort but was considerably underestimated in the Italian cohort. Furthermore, model performance deteriorated the longer the period of analysis. In terms of model discrimination, the UKPDS-OM2 performed better in the DCS cohort (all but one outcome with C-statistic equal or above 70%) compared to the CMS cohort (only mortality above 70%). The model was best at discriminating all-cause mortality in both cohorts and heart failure, MI and stroke in the DCS cohort. However, from a policy making perspective, given the applications of health economic decision models, the main interest is on performance in terms of cumulative risk prediction rather than discrimination.

The performance of the UKPDS-OM2 in both cohorts may be partly explained by the effects of differences in the time-periods of observation. The DCS cohort consisted of individuals with T2D followed between 2008 and 2018 whereas the CMS cohort consisted of individuals followed between 2000 and 2015. In contrast, the UKPDS-OM2 was built using data from a cohort of newly diagnosed individuals followed between 1977 and 2007. This meant that diabetes care changed significantly during the period covered by the three studies. During this time, we had the arrival of new antidiabetic agents, increased levels of statin use and improvements in the prevention and care of diabetes-related complications such as stroke and MI. This has been accompanied by a reduction in the absolute and relative mortality over time in people with T2D (17,18).

Another difference between the three cohorts UKPDS, CMS and DCS, could be related to the study effects. The UKPDS population was randomized into intensive glucose/blood pressure control compared to conventional care between 1977 and 1997. When the UKPDS trial finished in 1997, all surviving UKPDS patients entered into a ten-year post-trial monitoring study. In contrast, both CMS and DCS populations received routine care, supplemented by a central diabetes center.

The subgroup analyses on model performance for mortality indicated specifically room for improvement in elderly patients. This is reasonable, since the UKPDS study cohort started with a population aged 58.5 years. In CMS and DCS, the percentage of patients above 65 years of age was substantial (>50%), and 23% and 19% of patients, respectively, were aged above 75. New studies of the risk in elderly patients would be quite relevant and could potentially inform targeted treatment for elderly individuals with T2D.

Another focus for efforts to improve the UKPDS-OM2 should be the risk of myocardial infarction which showed poor fit in both cohorts. This confirms findings in previous studies in the US and Germany (7–9). Changes in cardiovascular risk management as well as in medical treatment after a cardiovascular event may explain this, and underline the need for calibrating the UKPDS-OM2 or estimating new risk equations.

The main strength of our analysis is the use of individual patient level data from two observational cohorts of people with T2D. Both cohorts were sampled in routine care, including almost all individuals with T2D in the region of interest, and had large sample sizes of nearly 2000 and 5000 individuals respectively, and long follow-up. With more than 150 and 135 events respectively for MI at 10 years – the rarest cardiovascular event - sample sizes were large enough to allow for assessment of performance for all the outcomes.

Our results add to previous findings suggesting that the UKPDS-OM2 overpredicts mortality and certain cardiovascular outcomes such as MI (11-13).

To perform the present validation a very significant effort was undertaken to harmonize data to inform the UKPDS-OM2 and to extract the incidence of events in both cohorts. Particular commitment was necessary to identify the core set of risk factors to be used as inclusion criteria, in order to maximize the number of patients to be included in both cohorts. Key variables in several cases needed to be recoded to harmonize cohorts data with UKPDS-OM2 requirements.

The potential misalignment of model outcomes with the outcomes recorded in the two cohorts is worth noting. We sought to align UKPDS-OM2 outcomes to diagnostic codes (ICD-9 and ICD-10) in the administrative records of both cohorts but could not be certain of their exact match. For example, in the CMS cohort, fatal events were obtained from two different data sources: 1) causes of death validated by clinicians up to 2006 and 2) the Death Registry from 2007. This resulted in two different approaches in coding the main cause of death and possible misclassification between the cardiovascular diseases. This could explain the observed higher rates of heart failure up to 2006 compared to after 2006. In the DCS cohort, while most cardiovascular events could be verified in medical records after linkage, this was not possible for about 5% of events, since they occurred outside of the linkage setting (after 2018, or in an external hospital). Comparing self-reported to validated events (for a subgroup of 453 participants) showed that the sensitivity of self-report was 86%, while specificity was 90%[5]. Nonetheless, this way of using self-reported information during

the annual exam might have led to an underestimation of cardiovascular events in the DCS cohort, in particular due to underreporting of fatal events.

Our work has some limitations. First, some risk factor data needed to inform the UKPDS-OM2 were not available. Of these, some risk factors were completely missing in the cohorts (e.g. atrial fibrillation, PVD, heart rate) and were imputed based on their association with available risk factors derived using UKPDS data. Partially missing data at baseline were handled with multiple imputation following the missing at random assumption.

Second, risk factor data were completely missing at follow-up in the CMS cohort and censored in the DCS cohort. These missing risk factor time paths were imputed using risk factor time-path equations developed by the UKPDS modelling team, and based on the UKPDS cohort However, in sensitivity analysis, when we carried forward the last observed values the findings were similar to those using the imputed risk factor time-paths.

Third, our study populations were those with complete observations at baseline on core risk factors. Thus we were not able to assess model performance for patients with missing data at baseline. This could have affected the generalizability of our cohorts, however, we expect this effect to be relatively limited. Differences between included and excluded subjects in available variables were low for CMS (data not shown), while for DCS only the percentage of smokers and the level of eGFR was higher in the exclusion group (data not shown).

Finally, contrary to validation studies performed for single risk prediction equations, in the present study no re-calibration of the UKPDS-OM2 as a starting point for validation was performed. Since the model is informed by 15 risk prediction equations in combination and the re-calibration of each individual risk equation will affect other risk equations due to the outcomes being inter-related, standard re-calibration was infeasible. Rather than calibration, transferability of the decision model is at stake and we assessed this by considering calibration, calibration in the large and discrimination.

Our results indicate that this transferability is not satisfactory for all outcomes and new or adjusted risk equations may be needed before applying the model to the Italian or Dutch setting. This requires careful and stepwise adjustments and is a topic for further research.

To allow the wide utilisation of the UKPDS-OM2 in T2D we need to assess its performance in cohorts of patients different from those used for model development. In this study, we showed the UKPDS-OM2 to overpredict the risk of MI and all-cause mortality. Model calibration of the UKPDS-OM2 may be justifiable to ensure its relevance to policy makers outside the UK. The current results provide support and direction for such calibrations.

Present results highlight the importance of external validity assessment and need for possible adjustments before applying outcome simulation models to context that are different from the ones used for their development.

## Acknowledgments

*Funding*

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115881 (RHAPSODY). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This work is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0097. The opinions expressed and arguments employed herein do not necessarily reflect the official views of these funding bodies."

*Competing interests*

The authors declare that they have no competing interests.

*Authors' contributions*

Design: EP SK TF JL AvH GB

Conduct/data collection: EP SK TF JL AvH GB

Analysis: EP SK TF JL DDC, RR, RS, JB

Writing manuscript: EP SK TF JL, AvH, GB, DDC, RR, RS, JB

## References


1. Kuntz KM, Russell LB, Owens DK, Sanders GD, Trikalinos TA, Salomon JA. Decision Models in Cost-Effectiveness Analysis. In: Neumann PJ, Ganiats TG, Russell LB, Sanders GD, Siegel JE, editors. Cost-Effectiveness in Health and Medicine [Internet]. Oxford University Press; 2016 [cited 2019 Jun 5]. p. 105–36. Available from: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780190492939.001.0001/acprof-9780190492939-chapter-5

2. 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 Sep;56(9):1925–33.

3. Pagano E, Gray A, Rosato R, Gruden G, Perin PC, Merletti F, et al. Prediction of mortality and macrovascular complications in type 2 diabetes: validation of the UKPDS Outcomes Model in the Casale Monferrato Survey, Italy. Diabetologia. 2013 Aug;56(8):1726–34.

4. Reynoso-Noverón N, Mehta R, Almeda-Valdes P, Rojas-Martinez R, Villalpando S, Hernández-Ávila M, et al. Estimated incidence of cardiovascular complications related to type 2 diabetes in Mexico using the UKPDS outcome model and a population-based survey. Cardiovasc Diabetol. 2011 Jan 7;10:1.

5. Song M, Alexander CM, Mavros P, Lopez VA, Malik S, Phatak HM, et al. Use of the UKPDS Outcomes Model to predict all-cause mortality in U.S. adults with type 2 diabetes mellitus: comparison of predicted versus observed mortality. Diabetes Res Clin Pract. 2011 Jan;91(1):121–6.

6. Palmer AJ, Mount Hood 5 Modeling Group, Clarke P, Gray A, Leal J, Lloyd A, et al. Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2013 Jun;16(4):670–85.

7. Basu S, Sussman JB, Berkowitz SA, Hayward RA, Yudkin JS. Development and validation of Risk Equations for Complications Of type 2 Diabetes (RECODe) using individual participant data from randomised trials. Lancet Diabetes Endocrinol. 2017;5(10):788–98.

8. 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):586–95.

9. Laxy M, Schöning VM, Kurz C, Holle R, Peters A, Meisinger C, et al. Performance of the UKPDS Outcomes Model 2 for Predicting Death and Cardiovascular Events in Patients with Type 2 Diabetes Mellitus from a German Population-Based Cohort. PharmacoEconomics. 2019 Dec;37(12):1485–94.

10. Bruno G, Bargero G, Vuolo A, Pisu E, Pagano G. A population-based prevalence survey of known diabetes mellitus in northern Italy based upon multiple independent sources of ascertainment. Diabetologia. 1992 Sep;35(9):851–6.

11. Bruno G, Merletti F, Bargero G, Melis D, Masi I, Ianni A, et al. Changes over time in the prevalence and quality of care of type 2 diabetes in Italy: the Casale Monferrato surveys, 1988 and 2000. Nutr Metab Cardiovasc Dis NMCD. 2008 Jan;18(1):39–45.

12. van't Riet E, Schram MT, Abbink EJ, Admiraal WM, Dijk-Schaap MW, Holleman F, et al. The Diabetes Pearl: Diabetes biobanking in The Netherlands. BMC Public Health. 2012 Nov 6;12:949.

13. van der Heijden AA, Rauh SP, Dekker JM, Beulens JW, Elders P, 't Hart LM, et al. The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the Netherlands. BMJ Open. 2017 06;7(5):e015599.

14. Willis M, Johansen P, Nilsson A, Asseburg C. Validation of the Economic and Health Outcomes Model of Type 2 Diabetes Mellitus (ECHO-T2DM). PharmacoEconomics. 2017;35(3):375–96.

15. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiol Camb Mass. 2010 Jan;21(1):128–38.

16. 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 May 10;30(10):1105–17.

17. Gulliford MC, Charlton J. Is relative mortality of type 2 diabetes mellitus decreasing? Am J Epidemiol. 2009 Feb 15;169(4):455–61.

18. Tancredi M, Rosengren A, Svensson A-M, Kosiborod M, Pivodic A, Gudbjörnsdottir S, et al. Excess Mortality among Persons with Type 2 Diabetes. N Engl J Med. 2015 Oct 29;373(18):1720–32.

Figure 1. Observed (95%CI) and predicted cumulative incidence in CMS and DCS cohorts during respectively 15 and 10 years of follow up from the enrollment, by outcome.

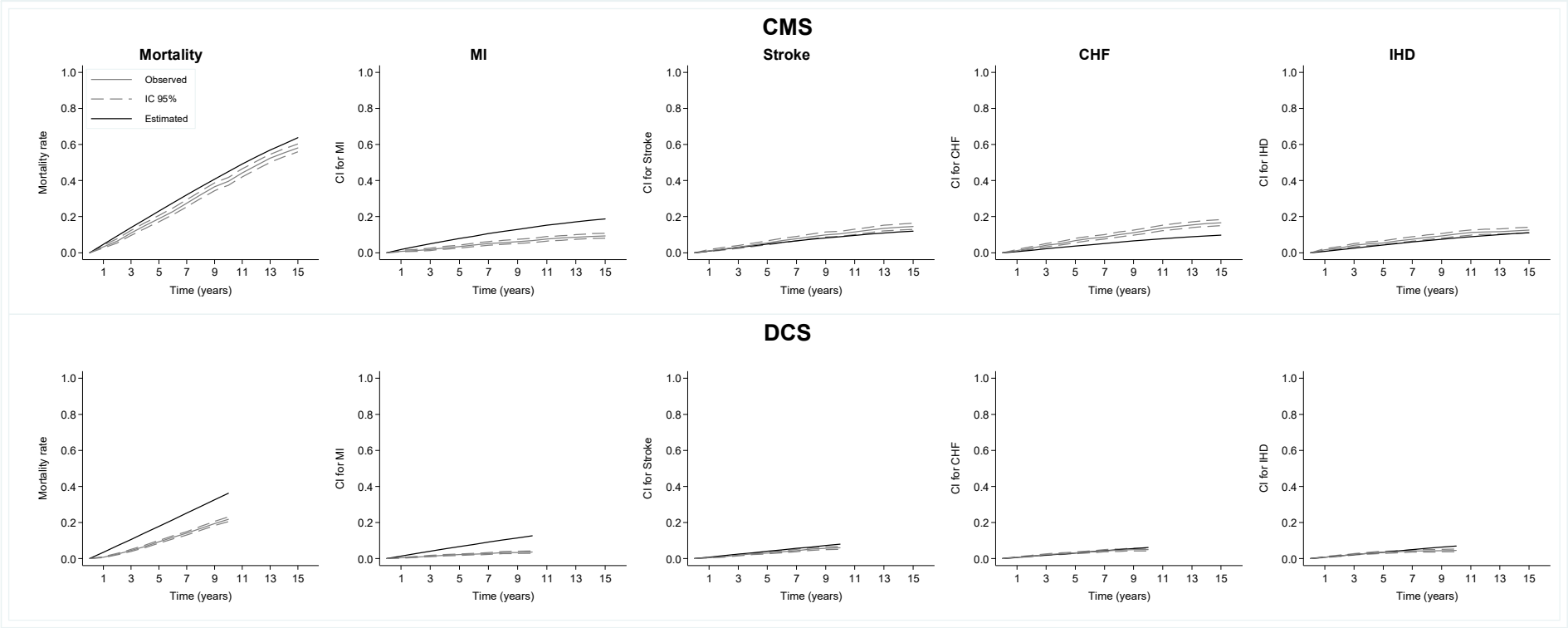

Figure 2. Calibration plots by decile of predicted risk: observed mortality rates at year 10 with 95% confidence interval against the predicted, in CMS, DCS and DCS enrolled in 2008 cohorts.

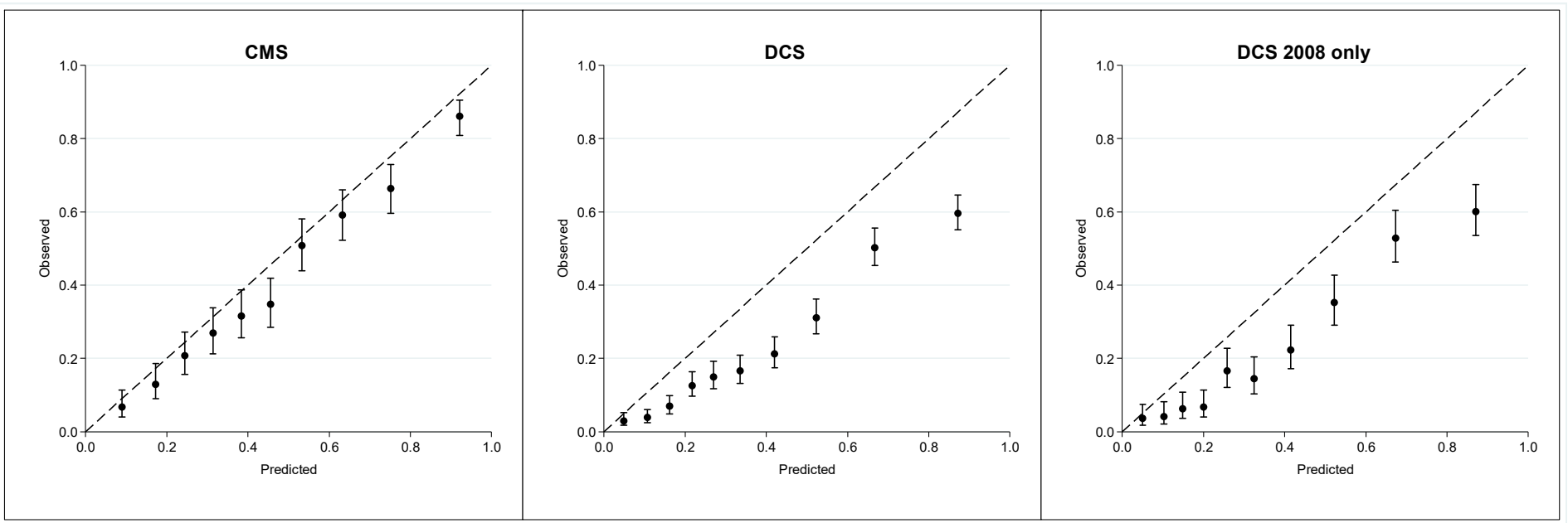

Table 1. Baseline risk factors in CMS, DCS and UKPDS cohorts.

| | DCS cohort | CMS cohort | UKPDS* (duration 7 years) | UKPDS * (duration 11 years) |
|---|---|---|---|---|
| **N** | 5188 | 1931 | 4637 | 4145 |
| **%Male** | 55.6% | 48.9% | 58.1% | 57.4% |
| **Age (sd)** | 64.8 (11.1) | 67.8 (10.3) | 58.5 (8.8) | 62.2 (8.8) |
| **Mean years of duration of diabetes (sd)** | 6.8 (5.9) | 10.9 (8.0) | 7.0 (0) | 11.0 (0) |
| **BMI** | 30.3 (5.5) | 28.5 (5.0) | 29.2 (5.6) [4034] | 29.3 (5.6) [3389] |
| **Current smoker %** | 18.6% | 14.9% | 23.1% [4191] | 18.7% [3526] |
| **HDL, mean mmol/l (sd)** | 1.2 (0.3) | 1.4 (0.4) | 1.1 (0.3) [3674] | 1.1 (0.3) [1879] |
| **LDL, mean mmol/l (sd)** | 2.6 (0.9) | 3.3 (0.9) | 3.4 (1.0) [3672] | 3.3 (0.9) [1879] |
| **Systolic blood pressure, mean mmHg (sd)** | 142.1 (20.1) | 146.1 (16.4) | 137.3 (18.7) [3997] | 139.3 (19.2) [3387] |
| **HbA1c, mean % (sd)** | 6.7% (1.0%) | 7.0% (1.7%) | 7.9% (1.8%) [3852] | 8.3% (1.8%) [3245] |
| **eGFR, mean ml/min/1.73m^2 (sd)** | 79.7 (20.9) | 80.9 (23.9) | 74.7 (17.5) [3423] | 72.2 (17.7) [3284] |
| **History of MI** | 7.7% | 7.9% | 4.1% | 6.3% |
| **History of stroke** | 6.0% | 6.7% | 1.9% | 3.0% |

*numbers in square brackets refer to participants with available data for a particular risk factor

Table 2. Comparison of observed (95%CI) and UKPDS-OM2 predicted cumulative incidence at 5, 10 and 15 years, and relative bias, by outcome and cohort.

| | 5 years | | | | 10 years | | | | 15 years | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Observed** | **(95%IC)** | | **Predicted** | **Bias** | **Observed** | **(95%IC)** | | **Predicted** | **Bias** | **Observed** | **(95%IC)** | | **Predicted** | **Bias** |
| **CMS** | | | | | | | | | | | | | | | |
| *Overall mortality* | 0.19 | 0.17 | 0.21 | 0.23 | 0.04 | 0.40 | 0.37 | 0.42 | 0.45 | 0.05 | 0.58 | 0.60 | 0.56 | 0.64 | 0.06 |
| *Myocardial Infarction* | 0.03 | 0.03 | 0.04 | 0.08 | 0.05 | 0.07 | 0.06 | 0.08 | 0.14 | 0.07 | 0.03 | 0.03 | 0.04 | 0.08 | 0.05 |
| *Stroke* | 0.05 | 0.04 | 0.07 | 0.05 | -0.01 | 0.10 | 0.09 | 0.12 | 0.09 | -0.02 | 0.15 | 0.13 | 0.16 | 0.12 | -0.03 |
| *CHF* | 0.07 | 0.06 | 0.08 | 0.04 | -0.03 | 0.12 | 0.11 | 0.14 | 0.07 | -0.05 | 0.17 | 0.15 | 0.18 | 0.10 | -0.07 |
| *IHD* | 0.05 | 0.05 | 0.07 | 0.04 | -0.01 | 0.10 | 0.09 | 0.12 | 0.08 | -0.02 | 0.13 | 0.11 | 0.14 | 0.11 | -0.01 |
| **DCS** | | | | | | | | | | | | | | | |
| *Overall mortality* | 0.09 | 0.10 | 0.09 | 0.18 | 0.09 | 0.22 | 0.23 | 0.21 | 0.36 | 0.14 | - | - | - | - | - |
| *Myocardial Infarction* | 0.02 | 0.02 | 0.03 | 0.07 | 0.05 | 0.04 | 0.03 | 0.04 | 0.13 | 0.09 | - | - | - | - | - |
| *Stroke* | 0.03 | 0.03 | 0.04 | 0.04 | 0.01 | 0.06 | 0.05 | 0.07 | 0.08 | 0.02 | - | - | - | - | - |
| *CHF* | 0.03 | 0.03 | 0.04 | 0.03 | -0.001 | 0.05 | 0.04 | 0.06 | 0.06 | 0.01 | - | - | - | - | - |
| *IHD* | 0.03 | 0.03 | 0.04 | 0.04 | 0.001 | 0.05 | 0.04 | 0.05 | 0.07 | 0.03 | - | - | - | - | - |