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

Pagano, E.; Konings, S.R.A.; Cuonzo, D. di; Rosato, R.; Bruno, G.; Heijden, A.A. van der; ... ; Feenstra, T.L.

ORIGINAL ARTICLE

WILEY

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

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

[1]Unit of Clinical Epidemiology, "Città della Salute e della Scienza" Hospital and CPO Piemonte, Turin, Italy

[2]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), Groningen, The Netherlands

[3]Department of Psychology, University of Turin, Turin, Italy

[4]Laboratory of Diabetic Nephropathy, Department of Medical Sciences, University of Turin, 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, Oxford, UK

[9]University of Groningen, Faculty of Science and Engineering, Groningen Research Institute of Pharmacy, Groningen, The Netherlands

[10]RIVM, Bilthoven, The Netherlands

Correspondence
Eva Pagano, MSc, Unit of Clinical Epidemiology, "Città della Salute e della Scienza" Hospital, Via Santena 5, 10126, Torino, Italy.
Email: eva.pagano@cpo.it

Funding information
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

Abstract

Aim: To externally validate the UK Prospective Diabetes Study Outcomes Model version 2 (UKPDS-OM2) 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) cohort (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 the CMS and 0.12 in the 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. CHF incidence was predicted well for the DCS (5 years bias −0.001), but underestimated for the CMS cohort.

Conclusions: The 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.

KEYWORDS
cardiovascular disease, diabetes complications, health economics

1 | INTRODUCTION

Several health economic decision models have been developed and validated for type 2 diabetes (T2D) mellitus populations to support healthcare policy making. They are mainly used to extrapolate long-term clinical outcomes and costs from clinical trials, to assess cost-effectiveness and to estimate the expected total budget impact of introducing new types of treatment. These analyses aid decision makers as part of a Health Technology Assessment to support healthcare priority setting.[1]

The UK Prospective Diabetes Study Outcomes Model version 2 (UKPDS-OM2) is an individual-level model 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 UK Prospective Diabetes Study (UKPDS) data that account for a range of patient characteristics, such as age, sex, preexisting complications, laboratory values and lifestyle habits. The model, as described by Hayes et al[2] is transparent[3,4] and has been validated internally as well as externally.[3,5] The UKPDS-OM2 has significant advantages over version 1, as it is based on longer follow-up data (almost double follow-up time), simulates more outcomes, and captures more comprehensively the progression of diabetes.[2,6]

In view of a potential wide utilization of the UKPDS-OM2 in cost-effectiveness analysis and in the evaluation of strategies for the management of T2D at the European level in the future, external validation in data across European countries is of great interest. In particular, with the aim of using the UKPDS-OM2 to support cost-effectiveness studies of new biomarkers in Western and Southern European countries with relatively extensive diabetes care programmes in place, an external validation using real-world data at the European level was necessary.

As observed in a validation of the first version of the UKPDS-OM,[7] differences in health and healthcare among countries, reflected in differences in variables such as 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.

The previous release of the UKPDS-OM2, the UKPDS-OM1, has been validated in several settings.[7–10] Previous patient-level validation work of the UKPDS-OM2 equations has been done in the United States[11,12] and Germany.[13] They consistently report overprediction of all-cause mortality, but varied in terms of performance relative to other outcomes. Since these results are contradictory and the sample size was small (fewer than 500 participants), evidence on model performance in a European setting is currently quite limited.

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

2 | METHODS

The 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 ischaemic heart disease (IHD). The list of International Classification of Diseases, 9th revision (ICD-9) and 10th revision (ICD-10) codes used for defining fatal and non-fatal events was derived from the UKPDS and is provided in Table S1. 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; glycated haemoglobin (HbA1c); and estimated glomerular filtration rate (eGFR).

2.1 | Patient data

Two unselected observational cohorts, the CMS cohort (n = 1931)[14,15] and the DCS cohort (n = 5188),[16,17] were used to inform the UKPDS-OM2 with patient-level data. Details on data selection and handling of missing data are reported in Appendix S1 (see sections "Patient data", "Missing data" and Tables S2-S9).

2.2 | UKPDS outcomes model version 2

The UKPDS-OM2 is based on patient-level data from the UKPDS.[2] It has been developed to substitute the UKPDS-OM1, since additional information has been collected during the UKPDS 10-year post-trial monitoring period, allowing data to be incorporated on new risk factors and outcomes. 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 eight 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 the 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 Appendix S1 (see section "The UKPDS Outcomes Model version 2").

2.3 | Model validation

The model was run for each cohort using all patients with imputed data from time of entry into the DCS and CMS cohorts up to 10 and 15 years of follow-up, respectively.

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 (Table S10 in Appendix S1).

Model validation was performed by comparing UKPDS-OM2 predictions with mean and 95% confidence interval (CI) of the observed cumulative incidences in each cohort at 5, 10 and 15 (CMS only) years of follow-up, that is, "calibration-in-the-large". The 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.

We also calculated the difference between predicted and observed means in cumulative incidence using measures from bias (difference between observed and predicted means) to mean absolute percentage error (MAPE; average of the error in percentage terms).[18] In general, MAPE is easier to compare across cohorts and outcomes as it is a relative measure. Values closer to zero described better accuracy.

Finally, predicted and observed cumulative incidence for all the outcomes at the different timepoints (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, that is, ability to distinguish individuals with different outcomes,[19] 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.

Further details are reported in Appendix S1 (See "Missing data", "Validation" and "Subgroup and sensitivity analyses" sections). Analyses were performed with SAS 9.4 for the CMS and R 4.0.0 for the DCS cohorts.

3 | RESULTS

Baseline characteristics of the cohorts are provided in Table 1. The CMS and DCS cohorts differed mainly in mean duration of diabetes, with the CMS cohort including more prevalent cases (mean duration of

TABLE 1 Baseline risk factors in the Casale Monferrato Survey, the Hoorn Diabetes Care System and UK Prospective Diabetes Study cohorts

	DCS cohort	CMS cohort	UKPDS[a] (duration 7 years)	UKPDS[a] (duration 11 years)
Number of subjects	5188	1931	4637	4145
Male, %	55.6	48.9	58.1	57.4
Mean (SD) age, years	64.8 (11.1)	67.8 (10.3)	58.5 (8.8)	62.2 (8.8)
Mean (SD) duration of diabetes, years	6.8 (5.9)	10.9 (8.0)	7.0 (0)	11.0 (0)
Mean (SD) BMI,[a] kg/m^2	30.3 (5.5)	28.5 (5.0)	29.2 (5.6) [4034]	29.3 (5.6) [3389]
Current smoker,[a] %	18.6	14.9	23.1 [4191]	18.7 [3526]
Mean (SD) HDL cholesterol,[a] mmol/L	1.2 (0.3)	1.4 (0.4)	1.1 (0.3) [3674]	1.1 (0.3) [1879]
Mean (SD) LDL cholesterol,[a] mmol/L	2.6 (0.9)	3.3 (0.9)	3.4 (1.0) [3672]	3.3 (0.9) [1879]
Mean (SD) systolic blood pressure,[a] mmHg	142.1 (20.1)	146.1 (16.4)	137.3 (18.7) [3997]	139.3 (19.2) [3387]
Mean (SD) HbA1c,[a] mmol/mol	50 (11)	53 (19)	63 (20) [3852]	67 (20) [3245]
Mean (SD) eGFR,[a] mL/min/1.73 m^2	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

[a]Numbers in square brackets refer to participants with available data for a particular risk factor.
Abbreviations: CMS, Casale Monferrato Survey; DCS, Hoorn Diabetes Care System; eGFR, estimated glomerular filtration rate; HbA1c, glycated haemoglobin; MI, myocardial infarction; SD, standard deviation; UKPDS, UK Prospective Diabetes Study.

10.9 years) and the DCS including more recently diagnosed cases (mean duration of 6.8 years). Age, systolic blood pressure and smoking prevalence, also differed slightly among the cohorts. All the other risk factors had comparable mean values for the DCS and CMS cohorts, while the 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 the CMS cohort, 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 CHF incidence well for the DCS cohort (bias −0.001 at 5 years), but underestimated CHF for the CMS cohort, with a bias ranging from −0.03 at 5 years to −0.07 at 15 years.

Figure S2 in Appendix S1 plots the predicted versus observed cumulative incidence for all outcomes in one graph and all timepoints (at 5, 10 and, for CMS, 15 years), by cohort. In both cohorts, predictions were strongly associated with observations, with R^2 above 0.9.

Results of the analysis of all-cause mortality by subgroup are reported graphically in Figure S3 in Appendix S1. In the CMS cohort, 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 to 10 years, and BMI below 25 kg/m^2. The model predicted particularly well in the subgroup with HbA1c above or equal to 58 mmol/mol. 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 58 mmol/mol and with BMI below 25 kg/m^2.

In a sensitivity analysis carrying risk factors during follow-up forward from the last observed values (from baseline in CMS), the effect was negligible for the outcomes in the DCS. However, in the CMS, where risk factors were held constant (carried forward) from baseline, the overestimation of mortality and MI rates was reduced (see Figure S4 in Appendix S1).

Figure 2 shows the calibration plots of the observed all-cause mortality by decile of predictions at 10 years of follow-up. In the CMS cohort, model predictions were within the observed 95% CI for all but three subgroups, but the predicted point estimate overestimated mortality in all subgroups. In the DCS cohort, model predictions were outside the observed 95% CI for all but one subgroup. As with the CMS, the predicted point estimates overestimated observed mortality for all 10 subgroups. However, when we used only individuals enrolled in 2008 (ie, 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 Table S11 in Appendix S1. In line with the "calibration in the large" results, calibration was worst for MI in the CMS cohort followed by all-cause mortality. In the DCS cohort, calibration was worst for all-cause mortality and MI, with small errors for CHF. MAPE was quite large for MI in both cohorts. Except for IHD and CHF, the UKPDS-OM2 did worse in the DCS than in the CMS cohort when assessed by MAPE.

Table S12 in Appendix S1 reports the C-statistics concerning the UKPDS-OM2's discriminatory capability in the CMS and DCS cohorts.
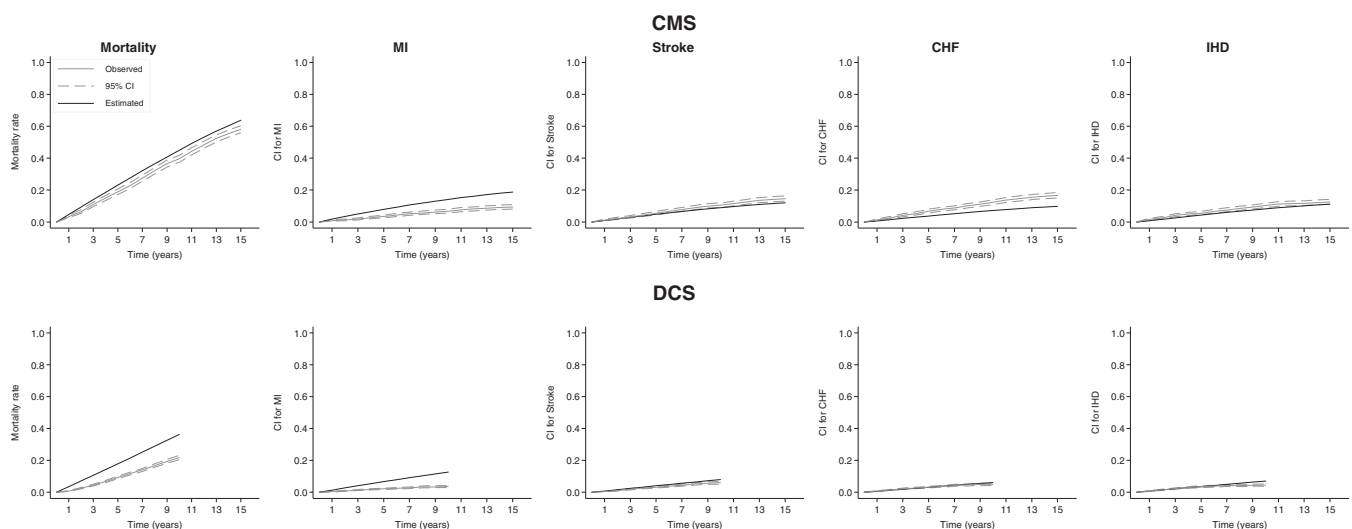
FIGURE 1 Observed (95% CI) and predicted cumulative incidence in the Casale Monferrato Survey (CMS) and Hoorn Diabetes Care System (DCS) cohorts during, respectively, 15 and 10 years of follow-up from the enrollment, by outcome. CHF, congestive heart failure; CI, confidence interval; IHD, ischaemic heart disease; MI, myocardial infarction

TABLE 2 Comparison of observed 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% CI)	Predicted	Bias	Observed	(95% CI)	Predicted	Bias	Observed	(95% CI)	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
MI	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	0.09, 0.10	0.18	0.09	0.22	0.21, 0.23	0.36	0.14	–	–	–	–
MI	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	–	–	–	–

Abbreviations: CHF, congestive heart failure; CI, confidence interval; CMS, Casale Monferrato Survey; DCS, Hoorn Diabetes Care System; IHD, ischaemic heart disease; MI, myocardial infarction.

For the CMS cohort, C-statistic values were above 70% for all-cause mortality across the three timepoints considered. For the remaining outcomes, the C-statistic values were approximately 60% to 65% at 5 and 10 years, and lower at 15 years for MI and IHD (59%). In the DCS cohort, 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 the DCS (66% at 5 and 10 years).

4 | DISCUSSION

To allow the wide utilization of the UKPDS-OM2 in T2D we need to assess its performance in cohorts of patients different from those used for model development. We externally validated the UKPDS-OM2 using individual patient-level data from two European cohorts, the Italian CMS cohort (South Europe) and the Dutch DCS cohort (Western Europe). We found the UKPDS-OM2 to overpredict the risk of all-cause mortality and MI in both cohorts, but to perform well for stroke and IHD outcomes. The predicted incidence of CHF 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 to or above 70%) compared to the CMS cohort (only mortality above 70%).

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. However, diabetes care changed significantly between 1977 and 2019, with 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.[20,21]

Differences in treatment strategies between the three cohorts—the UKPDS, CMS and DCS cohorts—could also explain the performance of the UKPDS-OM2. 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 10-year post-trial monitoring study. In contrast, both CMS and DCS populations received routine care, supplemented by a central diabetes centre.

The subgroup analyses on model performance for mortality indicated specifically room for improvement in elderly patients. This is reasonable, since the UKPDS cohort started with a population aged 58.5 years. In the CMS and DCS cohorts, the percentage of patients above 65 years of age was substantial (>50%), and 23% and 19% of patients, respectively, were aged above 75 years. New studies of the
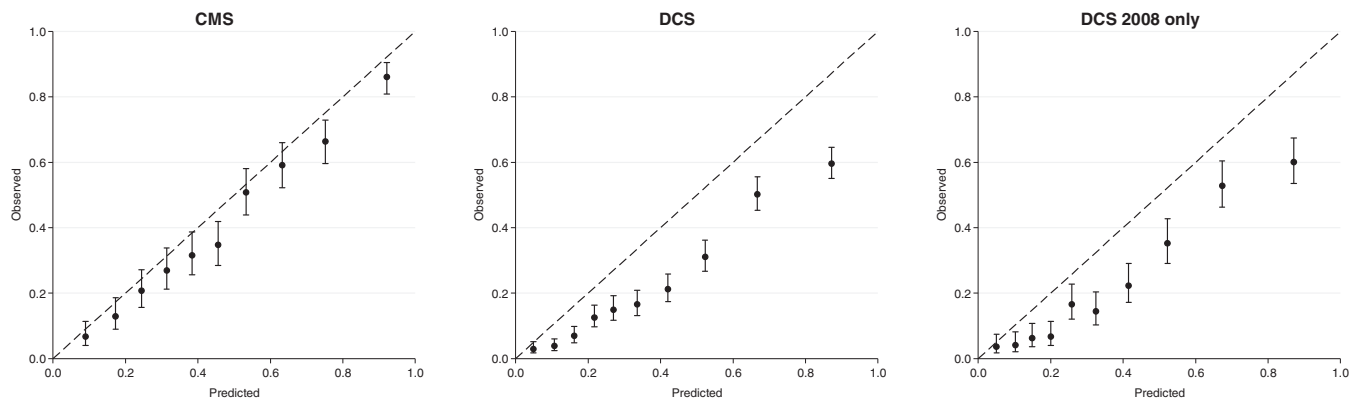
FIGURE 2 Calibration plots by decile of predicted risk: Observed mortality rates at year 10, with 95% confidence interval, against the predicted rates in the Casale Monferrato Survey (CMS) and Hoorn Diabetes Care System (DCS) cohorts and the DCS cohort enrolled in 2008

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 MI, which showed poor fit in both cohorts. This confirms findings in previous studies in the United States and Germany.[11–13] 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 the present analysis is the use of individual patient-level data, with a long follow-up (10 and 15 years), from two observational unselected 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 up to 15 years of data. With more than 150 and 135 events, respectively, for MI at 10 years—the rarest cardiovascular event—sample sizes were large enough to allow assessment of performance for all the included outcomes.

Our results add to previous findings suggesting that the UKPDS-OM2 overpredicts mortality and certain cardiovascular outcomes such as MI.[11–13] Other models incorporating equations from the UKPDS-OM2 have also reported validation exercises.[18,22,23] However, these validation studies were based on aggregate data from published studies and the choice of using individual data allowed us to better match outcomes, to capture variation in patient characteristics and outcomes, to address missing risk factor values, and to compare several subgroups of interest.

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 cohort 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: (a) causes of death validated by clinicians up to 2006, and (b) 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 approximately 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%.[16] Nonetheless, this way of using self-reported information during the annual examination 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 (eg, atrial fibrillation, peripheral vascular disease, heart rate) and were imputed based on their association with available risk factors derived using UKPDS data.

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 obtained using the imputed risk factor time paths. The imputation process allowed us to maximize the number of patients available for analysis, but could result in a possible overestimation of the differences between predicted and observed events if there is a significant mismatch between imputed and actual (but unobserved) risk factor time paths. However, using risk equations from the UKPDS-OM to impute missing values allowed us to test the model as it is likely to be used by

other researchers. The equations were provided by the authors of the UKPDS-OM and have yet to be published.

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 the CMS cohort (data not shown), while for the DCS cohort only the percentage of smokers and the level of eGFR was higher in the exclusion group (data not shown).

Another limitation was the exclusion of microvascular endpoints (eg, amputation, renal failure) and cardiovascular death from the analysis. However, we focused on the main cost drivers and drivers of loss in quality of life for T2D and most microvascular endpoints are rare in contemporary cohorts such as ours. Furthermore, cause of death is less robustly assessed in real-world cohorts and was judged not to be sufficiently robust for validation purposes in both cohorts.

Finally, contrary to validation studies performed for single risk prediction equations, in the present study no recalibration of the UKPDS-OM2 as a starting point for validation was performed. Because the model is informed by 15 risk prediction equations in combination and the recalibration of each individual risk equation will affect other risk equations due to the outcomes being interrelated, standard recalibration 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.

Beside all the mentioned limitations, this study represents an attempt to validate a decision model using unselected per-patient data, as they are usually available in the real world, that is with missing values, self-reported data and few observable outcomes.

In the present study, we showed that the UKPDS-OM2 over-predicted the risk of MI and all-cause mortality. Model calibration of the UKPDS-OM2 may be needed to ensure its relevance to policy makers outside the United Kingdom. Our findings provide support and direction for such calibrations. The present results highlight the importance of external validity assessments and need for possible adjustments before applying decision models to settings different from the ones used for their development, especially when their application aims at predicting absolute outcomes and effects.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Author contributions to the paper were as follows. E. Pagano: design; conduct/data collection; analysis; writing manuscript. S. R. A. Konings: design; conduct/data collection; analysis; writing manuscript. D. Di Cuonzo: analysis; writing manuscript. Rosalba rosato: analysis; writing manuscript. G. Bruno: design; conduct/data collection; writing manuscript. A. A. van der Heijden: design; conduct/data collection; writing manuscript. J. Beulens: analysis; writing manuscript. R. Slieker: analysis; writing manuscript. J. Leal: design; conduct/data collection; analysis; writing manuscript. T. L. Feenstra: design; conduct/data collection; analysis; writing manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/dom.14311.

DATA AVAILABILITY STATEMENT

Restrictions apply to the availability of data generated or analyzed during this study to preserve patient confidentiality or because they were used under license. The corresponding author will on request detail the restrictions and any conditions under which access to some data may be provided.

ORCID

Eva Pagano https://orcid.org/0000-0001-7552-2901
Jose Leal https://orcid.org/0000-0001-7870-6730
Talitha L. Feenstra https://orcid.org/0000-0002-5788-0454

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, eds. *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press; 2016:105-136. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780190492939.001.0001/acprof-9780190492939-chapter-5. Accessed June 5, 2019.
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;56(9):1925-1933.
3. Palmer AJ, Si L, Tew M, et al. Computer modeling of diabetes and its transparency: a report on the eighth Mount Hood challenge. *Value Health*. 2018;21(6):724-731.
4. Kent S, Becker F, Feenstra T, et al. The challenge of transparency and validation in health economic decision Modelling: a view from Mount Hood. *Pharmacoeconomics*. 2019;37(11):1305-1312.
5. Si L, Willis MS, Asseburg C, et al. Evaluating the ability of economic models of diabetes to simulate new cardiovascular outcomes trials: a report on the ninth Mount Hood diabetes challenge. *Value Health*. 2020;23(9):1163-1170.

6. Leal J, Hayes AJ, Gray AM, Holman RR, Clarke PM. Temporal validation of the UKPDS outcomes model using 10-year posttrial monitoring data. *Diabetes Care*. 2013;36(6):1541-1546.

7. Pagano E, Gray A, Rosato R, 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;56(8):1726-1734.

8. Reynoso-Noverón N, Mehta R, Almeda-Valdes P, 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;10:1.

9. Song M, Alexander CM, Mavros P, 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;91(1):121-126.

10. Palmer AJ, Mount Hood 5 Modeling Group, Clarke P, et al. Computer modeling of diabetes and its complications: a report on the fifth Mount Hood challenge meeting. *Value Health*. 2013;16(4): 670-685.

11. 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):788-798.

12. Basu S, Sussman JB, Berkowitz SA, 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.

13. Laxy M, Schöning VM, Kurz 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;37(12):1485-1494.

14. 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;35(9):851-856.

15. Bruno G, Merletti F, Bargero G, 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*. 2008;18(1):39-45.

16. van't Riet E, Schram MT, Abbink EJ, et al. The diabetes pearl: diabetes biobanking in The Netherlands. *BMC Public Health*. 2012;12:949.

17. van der Heijden AA, Rauh SP, Dekker JM, 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;7(5):e015599.

18. 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):375-396.

19. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138.

20. Gulliford MC, Charlton J. Is relative mortality of type 2 diabetes mellitus decreasing? *Am J Epidemiol*. 2009;169(4):455-461.

21. Tancredi M, Rosengren A, Svensson A-M, et al. Excess mortality among persons with type 2 diabetes. *N Engl J Med*. 2015;373(18): 1720-1732.

22. Lundqvist A, Steen Carlsson K, Johansen P, Andersson E, Willis M. Validation of the IHE cohort model of type 2 diabetes and the impact of choice of macrovascular risk equations. *PLoS One*. 2014;9(10):e110235.

23. McEwan P, Ward T, Bennett H, Bergenheim K. Validation of the UKPDS 82 risk equations within the Cardiff diabetes model. *Cost Eff Resour Alloc*. 2015 Aug 4;13(1):12.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.