

Temporal Validation of the UKPDS Outcomes Model Using 10-Year Posttrial Monitoring Data

JOSE LEAL, DPHIL¹
ALISON J. HAYES, PHD²
ALASTAIR M. GRAY, PHD¹

RURY R. HOLMAN, MB, CHB, FRCP³
PHILIP M. CLARKE, PHD⁴

OBJECTIVE—To evaluate the accuracy of the UK Prospective Diabetes Study Outcomes Model (UKPDS-OM) in predicting clinical outcomes during the UKPDS posttrial monitoring (PTM) period.

RESEARCH DESIGN AND METHODS—At trial end in 1997, the 4,031 surviving UKPDS patients, of the 5,102 originally enrolled in the study, returned to their usual care providers, with no attempts made to maintain them in their randomized therapy groups. PTM risk factor data were collected for 5 years and clinical outcome data for 10 years. The UKPDS-OM was used firstly to forecast likely progression of HbA_{1c}, systolic blood pressure, total-to-HDL cholesterol ratio, and smoking status and secondly to estimate the likely first occurrence of seven major diabetes-related complications or death from any cause. Model predictions were compared against observed PTM data for risk factor time paths and survival probabilities for major diabetes complications.

RESULTS—UKPDS-OM-forecasted risk factor time paths were similar to those observed for HbA_{1c} (up to 3 years) and total-to-HDL cholesterol ratio but underestimated for systolic blood pressure and smoking status. Predicted 10-year event probabilities were similar to those observed for blindness, ischemic heart disease, myocardial infarction, and renal failure but were higher for heart failure and death from any cause and lower for stroke and amputation.

CONCLUSIONS—The UKPDS-OM has good predictive accuracy for two of four risk factor time paths and for 10-year clinical outcome probabilities with the exception of stroke, amputation, heart failure, and death from any cause. An updated version of the model incorporating PTM data is being developed.

Diabetes Care 36:1541–1546, 2013

Computer simulation models are a useful means of extrapolating data from clinical trials over longer periods of time or to other populations (1). Several such models have been developed for type 2 diabetes, and their usefulness in decision making is widely recognized (2). They allow estimation of longer-term clinical outcomes and costs and help decision makers make informed choices between competing interventions. The UK Prospective Diabetes Study Outcomes

Model (UKPDS-OM) is a simulation model based on patient data from the UKPDS that predicts the occurrence of diabetes-related complications over a life time and quantifies the respective (quality-adjusted) life expectancy and life-time health care costs (3). The UKPDS-OM was developed primarily to simulate, at a patient level, mortality and a profile of complications that can be used for health economic evaluations of diabetes-related interventions (4) but has also been

applied to health service planning (5) and used as a long-term prognostic tool (6).

The UKPDS-OM was shown to be internally valid and closely match the outcomes observed during the UKPDS trial itself (3). Additional information on risk factor levels and clinical outcomes collected during the UKPDS 10-year post-trial monitoring (PTM) study (7) provides the means to perform temporal validation (8) of predicted risk factor time paths and clinical outcomes against data not used in its construction. We undertook a prospective and blinded evaluation of the accuracy of the UKPDS-OM in predicting clinical outcomes during the 10-year PTM period.

RESEARCH DESIGN AND METHODS

The UKPDS involved 5,102 patients with newly diagnosed type 2 diabetes recruited between 1977 and 1991 and randomized to different blood glucose and blood pressure control regimens (9,10). Of these, 3,642 patients for whom annual risk factor data were available were used to construct the UKPDS-OM (3). Briefly, this is a probabilistic discrete-time model with annual cycles that is based on a system of parametric survival equations simulating patient-level outcomes. The model predicts an individual's absolute probability of first occurrence of seven complications (myocardial infarction [MI], other ischemic heart disease [IHD], heart failure, stroke, blindness, renal failure, and amputation) and death, conditional on the patient's characteristics of age, ethnicity, sex, and duration of diabetes and time-varying clinical risk factors (systolic blood pressure [SBP], HbA_{1c}, lipid levels, smoking status, and history of previous complications). Supplementary Table 1 provides the definitions of each complication. The model predicts both complications and time paths for clinical risk factors. If the patient survives a 1-year cycle, the risk factor time paths are updated and carried forward to the next model cycle together with any history of nonfatal complications that may have been

From the ¹Health Economics Research Centre, Department of Public Health, University of Oxford, Oxford, U.K.; the ²Sydney School of Public Health, University of Sydney, Sydney, Australia; the ³Diabetes Trials Unit, University of Oxford, Oxford, U.K.; and the ⁴Melbourne School of Population Health, University of Melbourne, Melbourne, Australia.

Corresponding author: Jose Leal, jose.leal@dph.ox.ac.uk.

Received 12 June 2012 and accepted 14 November 2012.

DOI: 10.2337/dc12-1120

This article contains Supplementary Data online at <http://care.diabetesjournals.org/lookup/suppl/doi:10.2337/dc12-1120/-DC1>.

© 2013 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

predicted to occur. Holding all else constant, the absolute risk of a complication will generally increase with higher values of risk factors, with history of complications, and with duration of diagnosed diabetes. Estimates of increased risk, holding all else constant, for each complication and death associated with time-varying risk factors are provided in Supplementary Table 2. The methodology including a complete listing of equations is described in more detail in UKPDS 68 (3).

Running the UKPDS-OM requires setting the initial values concerning the patients' characteristics and clinical risk factors. The UKPDS-OM can predict the time paths of SBP, HbA_{1c}, total-to-HDL cholesterol, and smoking status for a given population or use time paths predefined by the model user. This allows, among other things, the simulation of different treatment modalities by directly modeling the respective changes in the risk factor time paths. The outcomes from the model include annual event probabilities, life expectancy, or quality-adjusted life expectancy.

When the UKPDS interventional trial closed on 30 September 1997, the surviving 4,031 patients entered the 10-year PTM observational study (7). These patients were returned to their community or hospital-based usual diabetes care providers, with no attempts made to maintain them in their randomized treatment groups. Annual risk factor and clinical outcome data were collected for 5 years in UKPDS clinics, and in years 6 to 10, owing to funding restrictions, patient and general practitioner questionnaires were used to collect clinical outcome data (7). Mortality information for all patients who were still living in the U.K. was obtained from the Office of National Statistics.

To ensure a blinded temporal validation of the UKPDS-OM, the modeling team (J.L., A.J.H., A.M.G., and P.M.C.) were given access only to the characteristics of the UKPDS patients at their point of entry into the PTM study (Table 1). Where baseline PTM risk factor data were unavailable, the average baseline values across the PTM cohort were used instead. Version 1.3 of the UKPDS-OM software (www.dtu.ox.ac.uk/outcomes-model/) was used to simulate risk factor time paths and clinical outcomes over 10 years from these patient characteristics and those at their time of diagnosis of diabetes. Monte-Carlo simulation error for

predicted outcomes was reduced by averaging 50,000 simulations per patient.

Model accuracy was assessed in a number of ways: 1) prediction of the risk factor trajectories during the PTM period from the PTM baseline data and comparison with the observed data (only for patients with observed data), 2) prediction of the clinical outcomes based on the simulated risk factor trajectories, and 3) prediction of clinical outcomes based on the observed risk factor trajectories, with last observation carried forward for data unavailable in later time periods.

Once the UKPDS-OM predictions had been produced, full access to the PTM data were provided. Predicted risk factor trajectories were compared visually with observed trajectories. Kaplan-Meier method event-free survival estimates were derived from the observed PTM data. Survival estimates were derived from the UKPDS-OM annual predictions using a life table approach and accounting for censoring due to death (11). Cumulative failure (complement of cumulative event-free survival: 1 minus Kaplan Meier) of each event was estimated from the observed and simulated PTM data. As the UKPDS-OM only predicts the first occurrence of a complication, patients were excluded from the analysis of each complication if they already had a previous history of that particular complication on entry to PTM. Mean PTM follow-up was estimated from the date of entry into the study (1 October 1997) until death or the closing date of the study (30 September 2007).

UKPDS-OM predictive accuracy for clinical outcomes was assessed as follows:

1. Model bias was estimated for each outcome by averaging the mean absolute error ($|\text{mean prediction} - \text{mean observed}|$ in years 1, 2, 3, etc.) across the 10 years of simulation. We also examined the proportion of years in which model predictions fell within observed 95% CIs and the respective mean absolute percentage error (MAPE) for all complications relative to the observed mean and 95% CI.
2. Calibration and discrimination (8). We compared the mean and 95% CI of the observed cumulative failure of each complication with the respective mean predicted cumulative failure in the full PTM cohort, i.e., "calibration-in-the-large." The model 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 (observed 95% CI). Calibration was further evaluated by dividing the PTM cohort into deciles of predicted risk for each complication at 10 years and plotting the observed cumulative failure of events at 10 years against the predicted probabilities. Model discrimination, i.e., ability to distinguish individuals with different outcomes, was assessed using Harrell *C* statistic estimated using each individual's survival time and the predicted event-free survival at 10 years (*somersd* package in STATA, version 11) (12). The *C* statistic is on the scale 0–1, where 1 represents perfect discrimination, and is expected to be at least 0.5 for the model prediction to be considered a positive predictor of survival time.

RESULTS—Mean PTM study patient follow-up time was 8.41 years (SD 2.8 [95% CI 8.31–8.49]) with an overall mortality of 33%. Model-predicted PTM study life expectancy was 8.34 years at 10 years of follow-up. The most frequent PTM outcome was death (33%), and the most frequent PTM complication was MI (15%), followed by stroke (6%) and other IHD (5%) (Supplementary Table 1). Predicted risk factor time paths were similar to those observed in the PTM cohort for HbA_{1c} (up to 3 years) and total-to-HDL cholesterol ratio (Supplementary Fig. 2).

Table 1—Characteristics of UKPDS patients at entry into the PTM study

<i>n</i>	4,031
Age (years)	62 (9.3)
Proportion of men	57
HbA _{1c}	8.2 (1.7)
Systolic blood pressure (mmHg)	139 (20)
LDL cholesterol (mmol/L)	3.3 (0.9)
HDL cholesterol (mmol/L)	1.10 (0.4)
Current smokers	20
BMI (kg/m ²)	29.3 (5.7)
Duration of diabetes (years)	11.0 (2.9)
History of MI	6.4
History of stroke	3.2
History of IHD	6.2
History of amputation	0.9
History of blindness	3.6
History of renal failure	0.5
History of heart failure	2.3

Data are means (SD) or percent unless otherwise indicated.

The model underpredicted the observed mean SBP by ~ 2.1 mmHg and considerably underpredicted the expected proportion of current smokers at the end of the simulation period (2% of the whole cohort) (Supplementary Fig. 5).

Overall, the UKPDS-OM showed reasonable predictive ability for the whole cohort compared with the PTM data (Fig. 1, Table 2, and Supplementary Table 4). The model predicted well the probabilities of blindness, renal failure (from 5 years onward), heart failure (up to 3 years), other IHD, MI (from 7 years onward), amputation (up to 5 years), stroke (up to 6 years) and mortality (up to 8 years). Model bias and MAPE estimates across all complications were low (Table 2). At 10 years of follow-up, the model overpredicted the probabilities of heart failure and mortality and underpredicted the probabilities of stroke and amputation. Nonetheless, the correlation between the observed and predicted probabilities across the eight complication categories was $>99\%$ (Supplementary Fig. 1).

Calibration plots per decile of predicted risk revealed that at 10 years of simulation, the UKPDS-OM overestimated the observed proportions of MI, stroke, other IHD, heart failure, and blindness in the subgroup with highest predicted risk (Fig. 2). These risk subgroups were typically characterized by individuals with history of complications at PTM baseline, such as history of IHD in the MI and stroke subgroups (comprising $\sim 50\%$ of the highest-risk subgroup) or history of blindness in the amputation subgroup (comprising $\sim 32\%$ of the highest-risk subgroup) (Supplementary Table 5). Excluding individuals with history of any complication at PTM entry reduced the UKPDS-OM bias for MI, heart failure, other IHD, amputation, and renal failure but increased the bias for stroke (Supplementary Figs. 3 and 4). In terms of discrimination ability, the UKPDS-OM was found to be a positive predictor of survival time for all outcomes at 10 years, ranging from 0.54 (95% CI 0.50–0.58) for other IHD to 0.71 (95% CI 0.67–0.76) for amputation.

Finally, model predictions using the observed risk factor data (carried forward if unavailable) were closer to the observed cumulative event probabilities for renal failure, stroke, and amputation (Supplementary Fig. 6) than predictions using the model-generated risk factor time paths.

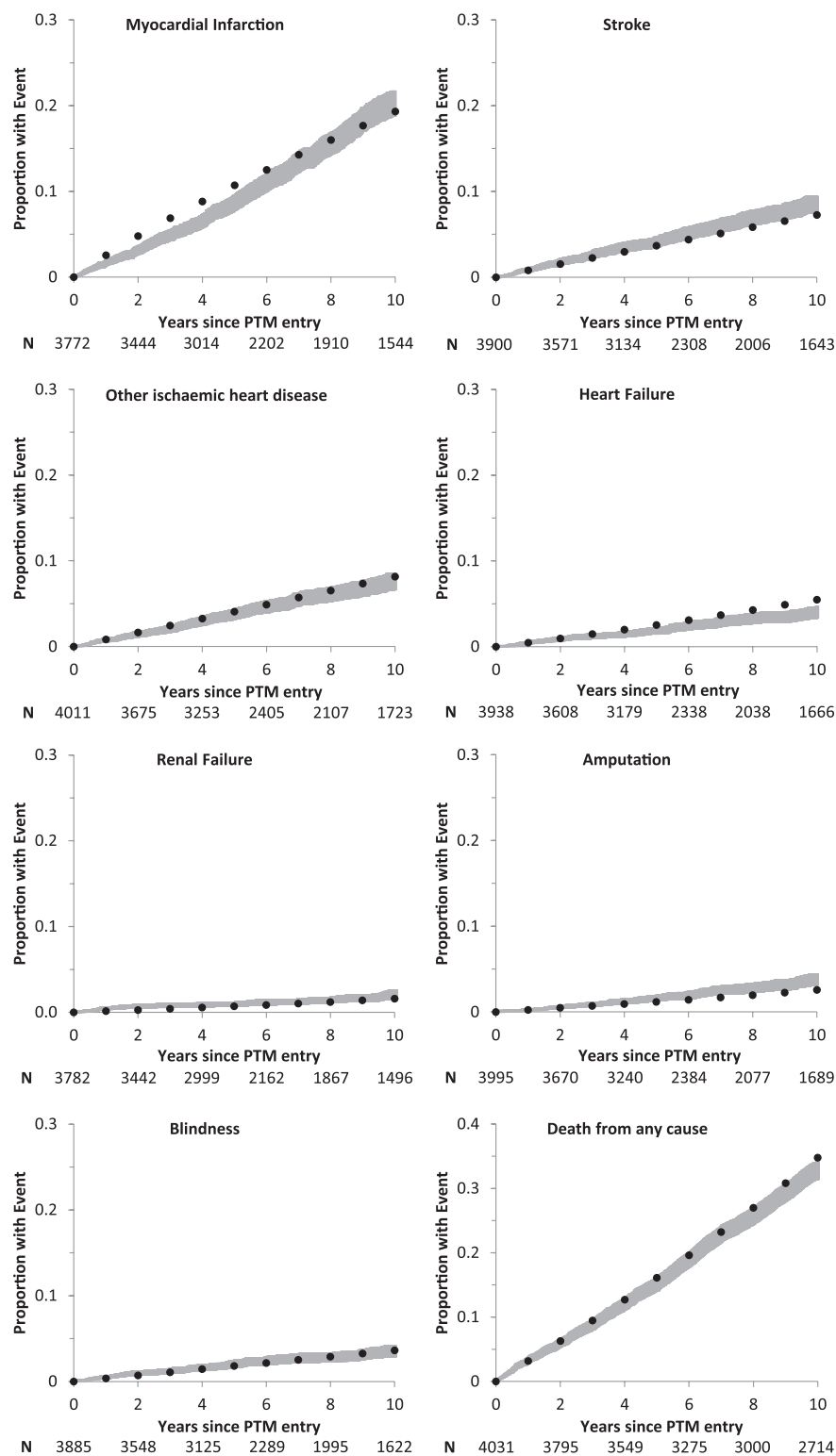


Figure 1—Observed and predicted cumulative probabilities of failure of first complications in PTM period. PTM curve (gray) comprises the observed 95% CI. Kaplan-Meier plots of cumulative probability of failure of each complication are shown with the respective 95% CI. Model curve (●) comprises the mean predicted estimate for each complication (using life table approach) based on the model-predicted risk factors. The number at risk at 2-year intervals is shown (N). “Years since PTM entry” comprises the time since the date of entry into the study (1 October 1997) until death or the closing date of the study.

Table 2—Comparison of model predictions with observed data from the PTM study

Complication	At 10 years			Over 10 years		
	Observed proportion with events	Predicted proportion with events	C statistic	Model bias (MAPE relative to mean)	Percent of years in which predictions fell within the observed 95% CI	MAPE relative to observed 95% CI (range)
Death from any cause	0.327 (0.313–0.342)	0.348	0.69 (0.68–0.71)	0.009 (5)	80	0.2 (0–2)
MI	0.202 (0.187–0.218)	0.193	0.67 (0.64–0.69)	0.014 (25)	40	10 (0–28)
Stroke	0.084 (0.074–0.095)	0.073	0.68 (0.65–0.71)	0.006 (12)	60	1 (0–5)
IHD	0.075 (0.065–0.086)	0.081	0.54 (0.50–0.58)	0.004 (14)	100	0
Amputation	0.037 (0.030–0.045)	0.026	0.71 (0.67–0.76)	0.005 (23)	50	6 (0–17)
Blindness in one eye	0.035 (0.028–0.042)	0.036	0.60 (0.55–0.65)	0.002 (10)	100	0
Renal failure	0.020 (0.015–0.026)	0.016	0.67 (0.60–0.74)	0.003 (30)	60	11 (0–41)
Heart failure	0.039 (0.032–0.047)	0.055	0.70 (0.65–0.75)	0.008 (35)	30	8 (0–18)

Data are means (95% CI) unless otherwise indicated. Boldface type indicates that the prediction is within the observed 95% CI.

CONCLUSIONS—This study shows that the UKPDS-OM predicted many of the diabetes-related complications that occurred during PTM, with 10-year estimates falling within the 95% CI of the observed survival probabilities for MI, blindness, IHD, and renal failure. Although the model slightly overpredicted all-cause mortality, predicted life expectancy over the PTM period was within the observed 95% CI and a high correlation was observed between observed and predicted probabilities of complications.

The UKPDS-OM performed less well when predicting the observed risk factor time paths. This is most likely due to changes in practice that have resulted in an intensification of treatment in recent years. The model successfully predicted the levels of HbA_{1c} during the first 3 years of follow-up but failed to predict the fall in hemoglobin observed in the PTM population during the subsequent 2 years. The reduction in HbA_{1c} was observed, on average, across all PTM patients regardless of their regimen of blood glucose control at the start of the UKPDS trial (7) and is thought to be due to the implementation of clinical guidelines for stricter glucose control after the publication of the UKPDS results (10,13,14). As this sharp decline was only observed during the PTM period, the UKPDS-OM did not have information sufficient for stimulating it. The predicted levels of SBP were consistently below the observed values by a small margin, but the predicted trend was similar to the observed one. Finally, the proportion of current smokers was significantly underpredicted by the UKPDS-OM. This was due to the specification of the smoking predictive equation (3). When the observed smoking status

data (carried forward if unavailable) were used instead of the predicted values, the UKPDS-OM produced estimates closer to the observed survival probabilities for MI and stroke (data not shown).

Previous history of complications was also shown to affect the UKPDS-OM performance in predicting MI, heart failure, amputation, and renal failure. For example, the model significantly overpredicted the probability of MI at 10 years in the subgroup of the PTM cohort composed mostly by individuals with a previous history of other IHD (>50%). This could be explained by having insufficient statistical power to accurately estimate the interactions between the different complications when estimating the risk equations.

Validation of computer simulation models of chronic diseases is important, as they are used increasingly by decision makers and clinicians to assess outcomes and costs of interventions in cost-effectiveness analysis, to estimate the future burden of disease, and to produce prognostic life tables with information on potential benefits from better management of risk factors. Temporal validation builds confidence that the model is able to predict risk using data that were not used in the construction of the model. The results reported add further evidence to previous external validations of UKPDS-OM equations in predicting risk of major complications. For example, model predictions have been shown to closely match the 4-year total event probability of acute coronary events reported by the Collaborative Atorvastatin Diabetes Study (CARDS) (2). Similarly, when simulating the PROactive study results, the UKPDS-OM estimated a 3-year relative risk of 0.87

for a composite of all-cause mortality, nonfatal MI, and stroke—well within the observed 95% CI (0.72–0.98) (15). The UKPDS-OM was shown to give all-cause mortality predictions among U.S. National Health and Nutritional Examination Survey participants with characteristics similar to those of UKPDS patients that were comparable with observed probabilities (16).

A limitation to this study is that the current UKPDS-OM version does not explicitly predict second and subsequent events of the diabetes-related complications; as a result, patients with a previous history of a particular complication at PTM baseline had to be excluded from the analysis of that complication. Also, although the PTM study provided >26,000 person-years of observational data for nonfatal events, it is possible that some events may not have been captured by the study questionnaires (7). Fatal events were directly reported by the Office of National Statistics and were, therefore, less likely to be missed except for patients no longer living in the U.K. Also, since risk factor data were not available for the full follow-up period, we cannot reliably determine whether event differences were the result of the event prediction equations or differences between predicted and actual risk factor time paths. Our results also suggest that additional risk factor information specific to the population being simulated is likely to improve prediction. Finally, none of the validation measures available for evaluating the UKPDS-OM against censored data such as the PTM study were completely satisfactory. Measures of calibration (e.g., modified Hosmer-Lemeshow for censored data [17]) and discrimination

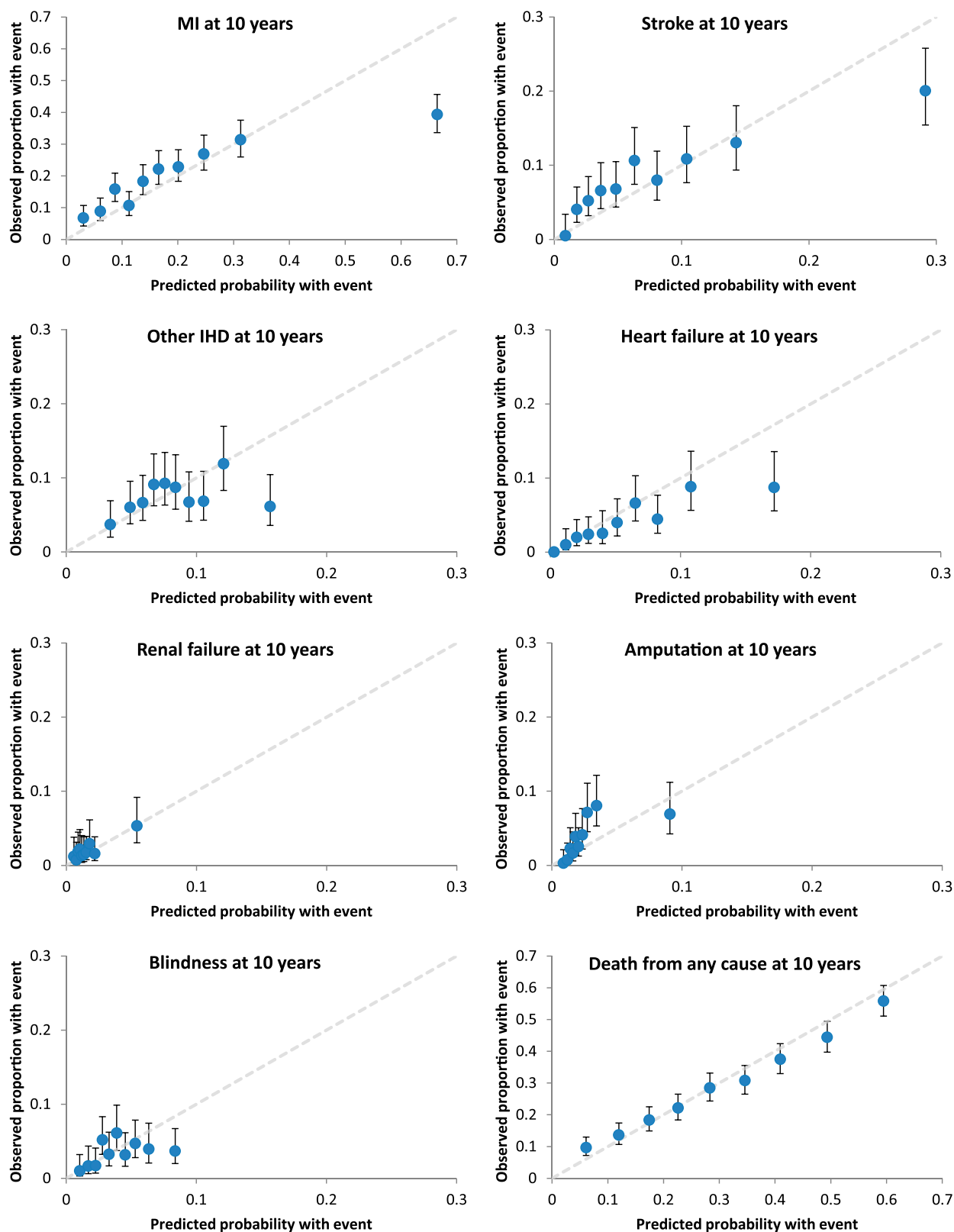


Figure 2—Calibration plots per deciles of risk at 10 years. Circles indicate the observed cumulative failure per deciles of predicted risk for seven complications and death, with vertical lines representing observed 95% CI. Cumulative failure for each complication at year 10 of PTM was estimated using Kaplan-Meier analysis. Simulation results are based on the modeled risk factor time paths. The 45° line (gray) represents perfect correlation over all deciles of predicted risk. (A high-quality color representation of this figure is available in the online issue.)

(e.g., Harrell *C* statistic [12,17]) are widely accepted but sensitive to the characteristics of the population (e.g., sample size, spread of risk in the population, censoring patterns). They are mostly used for comparative purposes such as choosing between different models with different predictors rather than to evaluate, as in our case, the absolute accuracy of a single model over a period of time (8,17,18). Hence, calibration plots per deciles of predicted risk and discrimination statistics at a single point in time as reported in this paper provide interesting but limited insights into the UKPDS-OM performance. As the aim was to evaluate the performance of the UKPDS-OM for the whole PTM cohort across 10 years of simulation and account for the uncertainty around the observed data, preference was given to measures such as model bias (mean absolute error) and MAPE, which are commonly used to assess forecasting accuracy.

The UKPDS-OM has been shown to have good predictive accuracy beyond the data used to estimate it. This provides more evidence of the potential usefulness of the model as a tool for decision makers and clinicians. The PTM data will also provide an excellent opportunity to update the UKPDS-OM and improve its predictions over the life time of patients with type 2 diabetes.

Acknowledgments—R.R.H. and A.M.G. are National Institute for Health Research (NIHR) senior investigators. J.L. was funded by the John Fell Oxford University Press Research Fund and by Medical Research Council Grant, An Outcomes Model for Type 2 Diabetes (S0801042). The Health Economics Research Centre is grateful to the NIHR for some of its funding. A.J.H. was funded by National Health and Medical Research Council (NHMRC) of Australia Grants 512463 and 571372. P.M.C. was funded by an NHMRC Career Development Award (571122).

No potential conflicts of interest relevant to this article were reported.

J.L. researched data; wrote the manuscript; contributed to the study conception, design, analysis, and interpretation of data; and reviewed and edited the manuscript. A.J.H., A.M.G., R.R.H., and P.M.C. contributed to the study conception, design, analysis, and interpretation of data and reviewed and edited the manuscript. J.L. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. American Diabetes Association Consensus Panel. Guidelines for computer modeling of diabetes and its complications. *Diabetes Care* 2004;27:2262–2265
2. Mount Hood 4 Modelling Group. Computer modeling of diabetes and its complications: a report of the Fourth Mount Hood Challenge Meeting. *Diabetes Care* 2007;30:1638–1646
3. Clarke PM, Gray AM, Briggs A, et al.; UK Prospective Diabetes Study (UKPDS) Group. A model to estimate the life time health outcomes of patients with type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68). *Diabetologia* 2004;47:1747–1759
4. O'Reilly D, Hopkins R, Blackhouse G, et al. Long-term cost-utility analysis of a multidisciplinary primary care diabetes management program in Ontario. *Can J Diabetes* 2007;31:205–214
5. 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
6. Leal J, Gray AM, Clarke PM. Development of life-expectancy tables for people with type 2 diabetes. *Eur Heart J* 2009;30:834–839
7. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med* 2008;359:1577–1589
8. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:b605
9. UKPDS Group. UK Prospective Diabetes Study (UKPDS). VIII. Study design, progress and performance. *Diabetologia* 1991;34:877–890
10. UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998;352: 837–853
11. Cutler SJ, Ederer F. Maximum utilization of the life table method in analyzing survival. *J Chronic Dis* 1958;8:699–712
12. Newson RB. Comparing the predictive powers of survival models using Harrell's *C* or Somers' *D*. *Stata J* 2010;10:339–358
13. UK Prospective Diabetes Study (UKPDS) Group. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* 1998;352: 854–865
14. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ* 1998;317:703–713
15. Holman RR, Retnakaran R, Farmer A, Stevens R. PROactive study. *Lancet* 2006; 367:25–26; author reply 26–27
16. 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;91:121–126
17. D'Agostino RB, Nam BH. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In *Handbook of Statistics*. Vol. 23. Balakrishnan N, Rao CR, Eds. London, Elsevier Science B.V., 2004, p. 1–25
18. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009;338:b604