VALUE IN HEALTH 16 (2013) 1074–1080

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

METHODOLOGICAL ARTICLES

Performance of the UKPDS Outcomes Model for Prediction of Myocardial Infarction and Stroke in the *ADDITION-Europe* Trial Cohort

Libo Tao, PhD[1], Edward C.F. Wilson, MSc[2], Simon J. Griffin, DM[1], Rebecca K. Simmons, PhD[1,], on behalf of the ADDITION-Europe study team*

[1]*MRC Epidemiology Unit, University of Cambridge, Cambridge, UK;* [2]*Health Economics Group, University of East Anglia, Norwich, UK*

ABSTRACT

Objectives: We assessed the performance of the UK Prospective Diabetes Study (UKPDS) outcomes model in predicting the risk of myocardial infarction (MI) and stroke in the Anglo-Danish-Dutch Study of Intensive Treatment in People with Screen Detected Diabetes in Primary Care (*ADDITION-Europe*) a trial cohort of patients with screen-detected type 2 diabetes from the United Kingdom, Denmark, and The Netherlands. **Methods:** We estimated the 5-year accumulated risk of MI and stroke for 2899 screen-detected people with type 2 diabetes by using the UKPDS outcomes model (version 1.3). We compared the predicted and actual risks by country and by intervention group (routine care; intensive multifactorial treatment). We assessed discrimination and goodness of fit by using area under receiver operating characteristic curves and the Hosmer-Lemeshow chi-square test. Multiple imputations were used to overcome missing data. **Results:** The UKPDS outcomes model overestimated the risk of MI and stroke. Mean predicted/actual ratios of 5-year accumulated risk were 2.31 for MI in the routine care group and 3.97 in the intensive multifactorial treatment group and 1.59 and 1.48 for stroke, respectively. The differences in absolute risk between the intervention groups were underestimated for MI (observed vs. predicted: 0.0127 vs. 0.0009) and slightly overestimated for stroke (–0.0013 vs. –0.0004). The area under the receiver operating characteristic curve was 0.72 (95% confidence interval 0.66–0.79) for MI and 0.70 (95% confidence interval 0.64–0.77) for stroke. The Hosmer-Lemeshow test statistic was nonsignificant in all groups. The model performed better in absolute risk prediction in Denmark and the United Kingdom than in The Netherlands. **Conclusions:** The UKPDS outcomes model has moderate discriminatory ability in the *ADDITION-Europe* trial cohort but overestimated absolute risk. The model may need updating for cardiovascular disease risk prediction in contemporary diabetes populations where patients may be diagnosed earlier in the disease trajectory and in whom cardiovascular risk is therefore lower.

Keywords: *ADDITION-Europe*, diabetes, UKPDS outcomes model, validation.

Background

Type 2 diabetes (T2D) is a chronic condition associated with significant disease burden and treatment costs, particularly for cardiovascular disease (CVD). Accurate prediction of CVD risk among people with T2D is important for targeting therapy to those at highest risk and for providing prognostic information. It is also a key component of economic evaluations of interventions aimed at improving the quality and length of life and reducing the disease burden of people with T2D.

The UK Prospective Diabetes Study (UKPDS) outcomes model (UKPDS-OM) is a cost-effectiveness analysis tool that was derived from the UKPDS, a multicenter randomized trial in 5102 newly diagnosed patients with T2D, recruited from 23 UK centers [1]. The UKPDS-OM has been validated both internally and externally among people with clinically diagnosed diabetes for prediction of CVD and other complications. In internal validation analysis, the observed and modeled cumulative incidence of CVD complications and all-cause mortality from diagnosis of diabetes to 12 years of follow-up was well matched [2]. In external validation studies, results varied from poor to moderate in terms of absolute risk, discrimination, and calibration [3–7]. In general, the UKPDS-OM tends to overestimate the risk of CVD [3,6,7].

The UKPDS-OM was developed by using data from patients who were recruited between 1977 and 1997. Since that time, the treatment of T2D and related conditions, as well as the cost of treatment, has changed. The variables assessed by the UKPDS-OM, including baseline characteristics, patient history, and clinical efficacy, are also likely to vary between countries. In addition, many health systems have introduced screening programs to detect diabetes earlier in the course of the disease. Consequently, there is a need to evaluate the performance of the UKPDS-OM in contemporary cohorts in different countries.

* *Address correspondence to:* Rebecca K. Simmons, MRC Epidemiology Unit, Institute of Metabolic Science, Box 285, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK.

E-mail: rebecca.simmons@mrc-epid.cam.ac.uk.

1098-3015/$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2013.06.001

The Anglo-Danish-Dutch Study of Intensive Treatment in People with Screen Detected Diabetes in Primary Care (*ADDITION-Europe*) is a primary care–based study of screening for T2D followed by a pragmatic open-label cluster randomized controlled trial comparing intensive multifactorial treatment (IT) with routine care (RC). The study was conducted over 5 years in three European countries: Denmark, the United Kingdom, and The Netherlands [8,9]. We aimed to assess the performance of the UKPDS-OM in predicting CVD risk in the *ADDITION-Europe* population, and thus to investigate its suitability in modeling longer term outcomes and costs for the *ADDITION-Europe* cohort.

Methods

Detailed information about the *ADDITION-Europe* study design and the main findings have been reported elsewhere [8,9]. In brief, 343 general practices were cluster randomized to screening plus RC of diabetes (176 general practitioners with 1379 participants) or screening followed by IT (167 general practitioners with 1678 participants). Allocation was concealed from patients throughout the trial. Following population-based stepwise screening programs among people aged 40 to 69 years (50–69 years in The Netherlands), 3057 of 3233 (95%) eligible participants with screen-detected diabetes agreed to take part (Denmark: 1533, United Kingdom: 1026, and The Netherlands: 498). Two participants withdrew during the trial period. The study was approved by local ethics committees in each center. All participants provided informed consent. The clinical trial registration number is NCT00237549.

In the IT group, the intensification of diabetes management was achieved through the addition of a number of features to existing diabetes care alongside lifestyle advice concerning diet, physical activity, and tobacco consumption and a stepwise target-led drug treatment regime to reduce hyperglycemia, blood pressure, hyperlipidemia, and microalbuminuria. RC practices followed national guidelines for diabetes management.

Measurement and Outcomes

Health assessments, at baseline and after 5 years, included biochemical and anthropometric measures, and were undertaken by centrally trained staff, who were blind to the study group allocation, following standard operating procedures. Standardized self-report questionnaires were used to collect information on sociodemographic characteristics (age, gender, and ethnicity), lifestyle habits (smoking status), and history of cardiovascular disease. Participants were followed for a median of 5.9 years. The primary outcome was time to cardiovascular event, including cardiovascular mortality, cardiovascular morbidity (nonfatal MI and nonfatal stroke), revascularization, and nontraumatic amputation. Participants' medical records and national registries were searched for potential end points by staff blind to group allocation. All events were independently adjudicated by two members of the local end point steering committee who were also blind to the group allocation according to an agreed protocol by using standardized case report forms. Because revascularization is not a component of the UKPDS-OM and only a single case of amputation was reported during follow-up, only MI and stroke events were examined in this analysis.

The 5-year accumulated absolute risks of MI and stroke were estimated for each participant in *ADDITION-Europe* by using the UKPDS-OM (version 1.3) [10]. This is a T2D-specific risk assessment tool that can be used to predict the annual risk of CVD events including MI and stroke. The model includes information on age at diagnosis, gender, ethnicity, duration of diabetes, weight, height, smoking status, presence or absence of atrial fibrillation (AF) and peripheral vascular disease (PVD), systolic blood pressure (SBP), glycosylated hemoglobin A_{1c} (HbA_{1c}), total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), and years since preexisting CVD events. Values of smoking status, SBP, HbA_{1c}, TC, and HDL-C were included at both baseline and 5-year follow-up.

Because information on AF and PVD was not collected at baseline in the *ADDITION-Europe* study, and given that all patients had newly diagnosed diabetes, these variables were set to 0. Number of years since preexisting ischemic heart disease, congestive heart failure, amputation, blindness, and renal failure were also set to 0 because this information was also not collected at baseline. Data on the number of years since any previous MI or stroke were collected in the *ADDITION-Europe* study and entered as appropriate into the model.

Statistical Analysis

We summarized baseline characteristics separately by center, and calculated actual 5-year rates of MI and stroke and compared them with the estimated risks over 5 years by using the UKPDS-OM. The baseline characteristics of UKPDS cohort were also included for comparison [11]. We used multiple imputation to deal with missing data [12,13] by using the Markov chain Monte Carlo method and assuming an arbitrary missing pattern. A multivariate normal distribution was used to impute missing values of age, gender, weight, height, smoking status, TC, HDL-C, SBP, and HbA_{1c}. It has been demonstrated in a previous study [14] that multivariate normal imputation is less biased than complete-case analysis, and produced similar results to other approaches despite the presence of binary and ordinal variables that did not follow a normal distribution. If the imputed HDL-C value was greater than the TC value, we assumed that HDL-C (mmol/l) = TC – 0.1 (five cases at baseline; one case at final follow-up). For each patient with missing data, we undertook five imputations and computed the average of the five risk estimates [15].

In UKPDS-OM, the values of ethnicity were White-Caucasian, Afro-Caribbean, and Asian-Indian. In the *ADDITION-Europe* cohort, there were some unknown or unclassifiable values, for example, mixed white + African, mixed white + Asian, or others. These cases were unavoidably excluded from the analysis because ethnicity is unsuitable for multiple imputation (156 cases).

The ratio of mean predicted versus actual CVD rate was calculated and compared between intervention groups and countries. A *t* test was performed to compare the absolute difference between actual and mean predicted rates. We reported the predicted risks of MI and stroke for patients who did or did not experience the event in the trial. We examined the discrimination of the UKPDS-OM by computing the area under the receiver operating characteristic curve (aROC) and assessed goodness of fit by using the Hosmer and Lemeshow chi-square test. The same methods have been used in previous studies for the validation of the UKPDS and Framingham risk engines [4,6,16].

In UKPDS-OM, both MI and stroke are assigned a particular Weibull regression equation with age, gender, HbA_{1c}, and other variables as covariates [2]. To examine which distribution was the best fit for baseline risk of MI and stroke, we performed a number of survival regression analyses including exponential, log-normal, log-logistic, Weibull, and generalized gamma. We computed minimal Bayesian information criterion and Akaike's information criterion values to assess global model fit [17]. We then analyzed the covariates with the best fit distribution to determine whether they were significantly associated with MI and stroke. A *P* value of less than 0.10 was defined as statistically significant [18]. Because risk factor values were not available for every year of follow-up in the *ADDITION-Europe* trial cohort, the average of baseline and final follow-up values was used.

We also conducted a complete case sensitivity analysis to examine whether results were replicated, after excluding patients for whom there was missing data (n = 781/2899; 26.9%).

Statistical analyses were performed with the Statistics Analysis System (SAS, version 9.3).

Results

From the 3055 *ADDITION-Europe* participants, 156 patients were excluded because of unclassifiable or unknown ethnicity. Therefore, data on 2899 patients were used in this analysis. Age, gender, treatment group, and baseline HbA_{1c} values were not significantly different between the included and excluded patients.

The mean age of participants was 60.3 ± 6.9 years; 42.0% were female, and 96.3% were Caucasian (Table 1). The mean HbA_{1c} value at baseline was 7.0% ± 1.6%. Baseline characteristics were not significantly different between the three countries for age. There were larger numbers of women and individuals with Caucasian ethnicity in The Netherlands and Denmark than in the United Kingdom. Smoking rates were highest in Denmark and lowest in the United Kingdom. Conversely, the mean body mass index (BMI) was highest in the United Kingdom and lowest in Denmark. There was no clear trend for biochemical values across the centers, but systolic blood pressure was significantly higher in The Netherlands (165 mm Hg) than in the United Kingdom (149 mm Hg) and Denmark (143 mm Hg). Compared with the *ADDITION-Europe* cohort, the UKPDS cohort was younger, had slightly more men, had larger numbers of nonwhite participants, and lower levels of BMI, HDL-C cholesterol, and systolic blood pressure at baseline. HbA_{1c} levels were similar between the two cohorts (Table 1).

Actual and Predicted Event Rates

The MI rate in *ADDITION-Europe* was 0.0228. The overall predicted rate was 0.0681 ± 0.0486, and 0.0941 ± 0.0545 and 0.0658 ± 0.0474 for patients who did or did not experience an MI, respectively. For stroke, the actual rate was 0.0152 and the predicted rate was 0.0232 ± 0.0206 for patients who did experience stroke, and the actual rate was 0.0340 ± 0.0270 and the predicted rate was 0.0223 ± 0.0197 for patients who did not experience stroke.

Event rates predicted with the UKPDS-OM were higher than those observed in *ADDITION-Europe*. The ratio of predicted to actual 5-year accumulated risk was 2.31 in the RC group and 3.97 in the IT group for MI, and 1.59 and 1.48 for stroke, respectively. Because of the skewed distribution of predicted rates, the median was smaller than the mean; nevertheless, the UKPDS model also overestimated CVD event rates when the median was used. When data were disaggregated by country, the overestimation was higher in The Netherlands than in Denmark and the United Kingdom for both MI and stroke (Table 2).

Examination of the absolute event rates between intervention groups represents the impact of IT compared with RC. For MI, the difference between intervention groups was higher than that predicted by the UKPDS-OM: the actual difference was 0.0127 (RC − IT = 0.0297 − 0.0170), and the predicted difference was only 0.0009 (RC − IT = 0.0686 − 0.0677), which means that the UKPDS-OM underestimated the effect of intensive treatment. For stroke, the difference was much smaller: −0.0013 (RC − IT = 0.0145 − 0.0158) from the observed data and −0.0004 (RC − IT = 0.0230 − 0.0234) from the predicted data, indicating slight overestimation.

Discrimination Analysis

The aROC was 0.72 (95% confidence interval 0.66–0.79) for MI and 0.70 (95% confidence interval 0.64–0.77) for stroke, indicating that the UKPDS-OM had moderate discriminatory ability for MI and stroke (Table 3). The Netherlands had the lowest aROC for MI (0.69) and the highest for stroke (0.79). Results from the Hosmer-Lemeshow test were nonsignificant both in the overall trial population and in each country, suggesting that goodness of fit was acceptable.

Survival Regression Analysis

Survival analysis showed that for both MI and stroke, an exponential distribution provided the best fit for baseline risk based on Bayesian information criterion values. In addition, for Akaike's information criterion values, log-normal and exponential distributions provided the best fit for MI and stroke, respectively (Table 4). On this basis, the exponential distribution was used in survival regression analysis with a selection of covariates: age, gender, treatment group, smoking status, BMI, HbA_{1c}, SBP, and Ln (TC/HDL-C) for MI or TC/HDL-C for stroke. Results showed that age,

Table 1 – Baseline characteristics of the UKPDS and *ADDITION-Europe* trial cohorts.

Characteristic	UKPDS	ADDITION				P values (ANOVA or chi-square test)*
		All centers	Denmark	The Netherlands	United Kingdom	
N	3642	2899	1452	448	999	
Mean age ± SD (y)	53 ± 8.0	60.3 ± 6.9	60.0 ± 6.9	60.6 ± 5.3	60.6 ± 7.4	0.05
Female sex (%)	40	42.0	42.9	46.0	38.7	0.02
Caucasian ethnicity (%)	82	96.3	98.6	99.6	91.5	<0.001
Current smoker (%)	–	27.0	33.9	25.0	18.0	0.01
Mean BMI ± SD (kg/m^2)	27.7 ± 5.3	31.6 ± 5.6	30.8 ± 5.5	31.0 ± 5.3	33.0 ± 5.8	<0.001
Mean total cholesterol ± SD (mmol/L)	–	5.6 ± 1.2	5.7 ± 1.1	5.6 ± 1.1	5.4 ± 1.2	<0.001
Mean HDL-C ± SD (mmol/L)	1.1 ± 0.24	1.3 ± 0.4	1.4 ± 0.4	1.1 ± 0.4	1.2 ± 0.3	<0.001
Mean systolic blood pressure ± SD (mm Hg)	135.0 ± 19	149.1 ± 21.8	148.8 ± 20.2	164.7 ± 23.3	142.7 ± 19.8	<0.001
Mean HbA_{1c} ± SD (%)	7.1 ± 1.8	7.0 ± 1.6	6.8 ± 1.5	7.3 ± 1.4	7.3 ± 1.7	<0.001

ADDITION, Anglo-Danish-Dutch study of Intensive Treatment In PeOple with screeN detected diabetes in primary care; ANOVA, analysis of variance; BMI, body mass index; HbA_{1c}, hemoglobin A_{1c}; HDL-C, high-density lipoprotein cholesterol; UKPDS, UK Prospective Diabetes Study.
* *P* value from ANOVA or chi-square tests for the comparison of *ADDITION-Europe* centers.

Table 2 – Actual and predicted MI and stroke event rates, by country and intervention group.

Event	Center	Intervention	N	Actual rate	Predicted rate, mean ± SD	P/A*	P†
MI	All centers	RC	1314	0.0297	0.0686 ± 0.0477	2.31	<0.001
		IT	1585	0.0170	0.0677 ± 0.0494	3.97	<0.001
	Denmark	RC	601	0.0416	0.0646 ± 0.0443	1.55	<0.001
		IT	851	0.0165	0.0616 ± 0.0418	3.74	<0.001
	The Netherlands	RC	212	0.0094	0.0845 ± 0.0526	8.96	<0.001
		IT	236	0.0085	0.0874 ± 0.0626	10.32	<0.001
	United Kingdom	RC	501	0.0240	0.0667 ± 0.0481	2.79	<0.001
		IT	498	0.0221	0.0687 ± 0.0516	3.11	<0.001
Stroke	All centers	RC	1314	0.0145	0.0230 ± 0.0190	1.59	<0.001
		IT	1585	0.0158	0.0234 ± 0.0218	1.48	<0.001
	Denmark	RC	601	0.0133	0.0224 ± 0.0180	1.69	<0.001
		IT	851	0.0141	0.0205 ± 0.0178	1.46	<0.001
	The Netherlands	RC	212	0.0094	0.0300 ± 0.0224	3.18	<0.001
		IT	236	0.0085	0.0363 ± 0.0319	4.29	<0.001
	United Kingdom	RC	501	0.0180	0.0208 ± 0.0180	1.16	<0.001
		IT	498	0.0221	0.0221 ± 0.0198	1.00	0.490

IT, intensive multifactorial treatment group; MI, myocardial infarction; RC, routine care group.
* P/A, mean predicted rate/actual rate.
† From *t* tests comparing mean predicted and actual rates.

gender, and HbA_{1c} were significantly associated with MI, whereas only age and gender were significantly associated with stroke.

Complete Case Analysis

There were 2118 (from 2899) patients included in the complete case analysis. Compared with excluded patients because of missing data, the complete set had a lower mean age (60.1/60.8), a higher proportion of men (59.0%/55.3%), and a higher mean HbA_{1c} value (7.06/6.96). The difference in mean age was statistically significant. In general, results were similar compared with the main imputed analysis in terms of predicted/actual ratios and best fit distribution. The only exception was in survival analysis, where gender no longer remained significantly associated with stroke and HbA_{1c} became significant.

Discussion

The UKPDS-OM had moderate discriminatory ability among high-risk individuals in the *ADDITION-Europe* trial cohort. The model, however, tended to overestimate the absolute risk of both MI and stroke. Our results suggest that the model might need updating for predicting CVD risk in contemporary diabetes populations where patients may be diagnosed earlier in the disease trajectory and in whom cardiovascular risk is therefore lower.

The UKPDS-OM (version 1.3) overestimated 5-year absolute risk of MI and stroke in the *ADDITION-Europe* cohort, although the accuracy varied between countries, with particularly poor performance for The Netherlands compared with Denmark and the United Kingdom. One reason for the relatively poor prediction could be differences in baseline characteristics between the *ADDITION* and UKPDS cohorts. The UKPDS population had some baseline characteristics associated with a lower risk of CVD including lower mean age, mean BMI, and mean SBP, although some characteristics would lead to higher risk (more male participants, lower mean HDL-C, and slightly higher mean HbA_{1c}). The impact of baseline characteristics was therefore likely to be mixed. For The Netherlands, mean SBP was higher than in the UKPDS cohort (165 mm Hg vs. 135 mm Hg).

The purpose of the UKPDS model, however, is to supply the coefficients for the inputted risk factors. Therefore, the accuracy

Table 3 – Discrimination and calibration analysis results for the UKPDS outcomes model using *ADDITION-Europe* trial cohort data.

Event	Center	aROC (95% CI)	Chi-squared value from Hosmer-Lemeshow test	P value from Hosmer-Lemeshow test
MI	All centers	0.72 (0.66–0.79)	9.79	0.28
	Denmark	0.76 (0.66–0.85)	7.61	0.47
	The Netherlands	0.69 (0.45–0.94)	7.11	0.53
	United Kingdom	0.70 (0.59–0.81)	8.01	0.43
Stroke	All centers	0.70 (0.64–0.77)	9.10	0.33
	Denmark	0.76 (0.68–0.84)	7.59	0.47
	The Netherlands	0.79 (0.55–1.00)	7.17	0.52
	United Kingdom	0.65 (0.54–0.76)	6.07	0.64

ADDITION, Anglo-Danish-Dutch study of Intensive Treatment In PeOple with screeN detected diabetes in primary care; aROC, area under the receiver operating characteristic curve; CI, confidence interval; MI, myocardial infarction; UKPDS, UK Prospective Diabetes Study.

Table 4 – Goodness of fit of baseline risk and covariates of survival regression analysis using *ADDITION-Europe* data.

Event	Distribution	BIC	AIC	Significant covariates ($P > 0.10$, from the exponential distribution)
MI	Gamma	656.12	585.50	
	Exponential	641.45*	583.43	
	Log-logistic	648.95	585.13	Age, gender, HbA_{1c}
	Log-normal	646.98	583.16*	
	Weibull	649.24	585.43	
Stroke	Gamma	434.02	364.41	
	Exponential	418.54*	360.50*	
	Log-logistic	426.21	362.39	Age, gender
	Log-normal	426.27	362.45	
	Weibull	426.21	362.40	

ADDITION, Anglo-Danish-Dutch study of Intensive Treatment In PeOple with screeN detected diabetes in primary care; AIC, Akaike's information criterion; BIC, Bayesian information criterion; HbA1c, hemoglobin A1c; MI, myocardial infarction.
* The exponential distribution provided the overall best fit model.

of predicted outcomes should be broadly insensitive to the baseline characteristics: it would be reasonable to reject a model on the basis of baseline characteristics only if the two populations were very different (e.g., predicting the risk of heart disease on the basis of a cohort of 80-year-olds and applying it to a cohort of 20-year-olds). The most likely explanation for the overestimation of events is due to improvements in treatment and care between the time the UKPDS data were collected (enrolment period 1977–1997) and the *ADDITION-Europe* data collection (2002–2006). Furthermore, the *ADDITION-Europe* intervention comprised a much more comprehensive set of preventative therapies administered from an earlier stage in the disease than would be routinely offered. Finally, our findings are consistent with other studies exploring the applicability of the UKPDS and Framingham models in populations with low disease rates [6,7].

When inputting data into the UKPDS risk engine, we set AF and PVD to 0 because these data were not collected at baseline in the *ADDITION-Europe* trial. PVD was a risk factor only for amputation (odds ratio = 11.4) in the UKPDS-OM. Because we focus on MI and stroke in our analysis, the omission of this data will not affect our results. AF is a risk factor for stroke (odds ratio = 4.2) in the UKPDS-OM [2], suggesting that missing data for this variable would likely underestimate the risk of stroke. Previous studies, however, show that the prevalence of AF ranges from 1.2% to 2.8% in people aged 60 years [19,20], and so missing data for this variable are unlikely to have a large impact on our findings.

Although this analysis showed that the UKPDS-OM tended to overestimate the risk of MI and stroke in each arm of the *ADDITION-Europe* study, its performance varied in predicting the difference between intervention groups. The assessment of the cost-effectiveness of health care interventions requires knowledge of the increment between the comparators, rather than the absolute levels of each. In our study, the UKPDS-OM underestimated the incremental incidence in MI between arms (RC-IT) and slightly overestimated the difference for stroke. Whether these differences could be sufficient to render the UKPDS unsuitable to extrapolate the *ADDITION* data remains ultimately a subjective decision. The insight that this analysis provides is that if the UKPDS-OM is unadjusted, it will underestimate any reductions in MI attributable to intensive treatment, therefore potentially underestimating the cost-effectiveness of the intensive intervention. The small disparity in the predicted rate versus the actual rate of stroke may be of limited consequence due to the effects of discounting.

The UKPDS-OM had moderate discriminatory ability for predicting MI and stroke. The aROC ranged from 0.65 to 0.79 [21]. None of the Hosmer-Lemeshow tests was statistically significant, showing that the goodness of fit was not unacceptable. We found a discrepancy between absolute risk prediction and model predicted risk. One possible explanation for this discrepancy is the insensitivity of the aROC and Hosmer-Lemeshow tests; previous studies have shown that different pairs of predicted and actual risk can have the same contribution to the aROC [22,23]. Despite their shortcomings, these tests are routinely adopted in analyses of this sort. We therefore present the results but caution against overinterpretation.

The UKPDS-OM was based on a set of survival regression equations. It is essential to choose a suitable baseline distribution and covariates for the model to perform successfully. In the UKPDS-OM, the Weibull distribution was applied for all CVD events [2]. The best-fit model for the *ADDITION-Europe* data, however, was an exponential distribution. This showed that age, gender, and HbA_{1c} were significant predictors of MI, while age and gender were significant for stroke. In previous studies, many risk factors such as age, gender, blood pressure, AF, renal failure, dyslipidemia, and heart failure have been linked to CVD risk [24,25]. Our analyses indicated that age, gender, and HbA_{1c} were important covariates and should be included in the UKPDS-OM. In the UKPDS cohort, the researchers collected biochemistry data every year and applied the values taken at the year of the event in their regression analysis [2]. Because we did not collect data annually in *ADDITION-Europe*, we averaged baseline and follow-up values of SBP, HDL-C, TC, and HbA_{1c} and applied these values at the year of the event in the regression analysis. This pragmatic decision may have contributed to the observed differences between the UKPDS-OM and our results.

Strengths, Limitations, and Further Study

In this study, we examined the performance of the UKPDS-OM for both MI and stroke in different countries and between intervention groups. The strengths of the *ADDITION-Europe* trial cohort include the large sample size with participants from three European countries and robust CVD end point ascertainment. We used multiple imputations to handle missing data, which allowed the best use of the available data while taking into account the uncertainty in missing values. Alternatively, the exclusion of incomplete observations would lead to considerable loss of data. Previous research has suggested that for this reason multiple imputation is the preferable method for handling missing data. It is, however, at the expense of a greater number of assumptions, which might influence the analysis [26,27]. In our study, both the imputed data set and the complete case data set produced similar results, which confirmed that our analysis was robust. Limitations include the lack of detailed annual risk factor measurement and the lack of information on AF, PVD,

and years since preexisting CVD conditions, blindness, and renal failure at baseline. This might have impacted the efficiency of the model and introduced some bias in the model prediction, mostly likely contributing to an underestimation of CVD risk.

Furthermore, the follow-up time was 5 years in the *ADDITION-Europe* trial cohort. We validated the UKPDS-OM by using short-term data with a view of extrapolating our data to a longer time horizon to estimate the lifetime incremental cost-effectiveness of intensive multifactorial therapy. As follow-up continues, more events will occur. Once longer term data are available, we will be able to revise our estimate of the longer term cost-effectiveness of the intervention.

Conclusions

The UKPDS-OM overestimated the 5-year cumulative absolute risk of MI and stroke in the *ADDITION-Europe* cohort but had moderate discriminatory ability for predicting MI and stroke. The difference in absolute risk between the intervention groups was underestimated for MI and slightly overestimated for stroke. Use of the model to extrapolate *ADDITION-Europe* data may be possible, but any reductions in MI may be underestimated. Interpretation of the results of any cost-effectiveness analysis should take account of this caveat, and appropriate sensitivity analyses should be conducted adjusting for the difference in risk.

Acknowledgments

We gratefully acknowledge the contribution of all participants, practice nurses, and general practitioners in the *ADDITION-Europe* study. We are grateful to the independent end point committee in Denmark (Professors Kristian Thygesen, Hans Ibsen, Ole Færgeman, and Birger Thorsteinsson), the United Kingdom (Professor Jane Armitage and Dr. Louise Bowman), and The Netherlands (Professors Cees Tack and Jaap Kappelle). We thank Dr. Clare Boothby (MRC Epidemiology Unit, UK) for her assistance with data cleaning, the Diabetes Trials Unit at Oxford University for the UKPDS-OM license, and the Health Economics Group at the University of East Anglia for methodological advice.

Source of financial support: *ADDITION-Cambridge* was supported by the Wellcome Trust (grant reference no. G061895); the Medical Research Council (grant reference no. G0001164); the National Institute for Health Research (NIHR) Health Technology Assessment Program (grant reference no. 08/116/300); National Health Service R&D support funding (including the Primary Care Research and Diabetes Research Networks); and the National Institute for Health Research. S.J.G. receives support from the Department of Health NIHR Program Grant funding scheme (RP-PG-0606-1259). The views expressed in this publication are those of the authors and not necessarily those of the UK Department of Health. Bio-Rad provided equipment for HbA_{1c} testing during the screening phase. *ADDITION-Denmark* was supported by the National Health Services in the counties of Copenhagen, Aarhus, Ringkøbing, Ribe, and South Jutland in Denmark; the Danish Council for Strategic Research; the Danish Research Foundation for General Practice; Novo Nordisk Foundation; the Danish Centre for Evaluation and Health Technology Assessment; the diabetes fund of the National Board of Health; the Danish Medical Research Council; and the Aarhus University Research Foundation. The trial has been given unrestricted grants from Novo Nordisk AS, Novo Nordisk Scandinavia AB, Novo Nordisk UK, ASTRA Denmark, Pfizer Denmark, GlaxoSmithKline Pharma Denmark, Servier Denmark A/S, and HemoCue Denmark A/S. Parts of the grants from the Novo Nordisk Foundation, the Danish Council for Strategic Research, and Novo Nordisk were transferred to the other centers. *ADDITION-Leicester* was supported by the Department of Health and ad hoc Support Sciences; the NIHR Health Technology Assessment Program (grant reference no. 08/116/300); National Health Service R&D support funding (including the Primary Care Research and Diabetes Research Network, and LNR CLAHRC); and the National Institute for Health Research. M.J.D. and K.K. receive support from the Department of Health NIHR Program Grant funding scheme (RP-PG-0606- 1272). *ADDITION-Netherlands* was supported by unrestricted grants from Novo Nordisk, Glaxo Smith Kline, and Merck and by the Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht.

REFERENCES

[1] UK Prospective Diabetes Study (UKPDS). Study design, progress and performance. Diabetologia 1991;34:877–90.
[2] Clarke PM, Gray AM, Briggs A, et al. A model to estimate the lifetime health outcomes of patients with type 2 diabetes: The United Kingdom Prospective Diabetes Study (UKPDS) outcomes model. Diabetologia 2004;47:1747–59.
[3] McEwan P, Bergenheim K, Currie C. External validation of the UKPDS outcomes model equations and the UKPDS risk engine equations in forecasting cardiovascular outcomes in people with type 2 diabetes. Value Health 2006;9:2.
[4] Simmons RK, Coleman RL, Price HC, et al. Performance of the UK Prospective Diabetes Study risk engine and the Framingham risk equations in estimating cardiovascular disease in the EPIC-Norfolk cohort. Diabetes Care 2009;32:708–13.
[5] Guzder RN, Gatling W, Mullee MA, et al. Prognostic value of the Framingham cardiovascular risk equation and the UKPDS risk engine for coronary heart disease in newly diagnosed type 2 diabetes: results from a United Kingdom study. Diabet Med 2005;22:554–62.
[6] Coleman RL, Stevens RJ, Retnakaran R, et al. Framingham score and DECODE risk equations do not provide reliable cardiovascular risk estimates in type 2 diabetes. Diabetes Care 2007;30:1292–3.
[7] Kengne AP, Patel A, Colagiuri S, et al. The Framingham and UK Prospective Diabetes Study (UKPDS) risk equations do not reliably estimate the probability of cardiovascular events in a large ethnically diverse sample of patients with diabetes. Diabetologia 2010;53:821–31.
[8] Griffin SJ, Borch-Johnsen K, Davies MJ, et al. Effect of early intensive multifactorial therapy on 5-year cardiovascular outcomes in individuals with type 2 diabetes detected by screening (ADDITION-Europe): a cluster-randomised trial. Lancet 2011;11:156–67.
[9] Lauritzen T, Griffin S, Borch-Johnsen K, et al. The ADDITION study: proposed trial of the cost-effectiveness of an intensive multifactorial intervention on morbidity and mortality among people with type 2 diabetes detected by screening. Int J Obes Relat Metab Disord 2000;24 (Suppl. 3):S6–11.
[10] UKPDS outcome model. Version 1.3 © Isis innovation ltd 2010. Available from: http://www.dtu.ox.ac.uk/outcomesmodel. [Accessed December 7, 2011].
[11] Stratton IM, Adler AI, Neil HA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. BMJ 2000;321:405–12.
[12] Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10:585–98.
[13] Yuan Y. Multiple imputation using SAS software. J Stat Softw 2011;45:1–25.
[14] Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Am J Epidemiol 2010;171:624–32.
[15] Rubin DB. Mutliple Imputation for Nonresponse in Surveys. Hoboken, NJ: Wiley, 2004.
[16] Van Dieren S, Peelen LM, Nothlings U, et al. External validation of the UK Prospective Diabetes Study (UKPDS) risk engine in patients with type 2 diabetes. Diabetologia 2011;54:264–70.
[17] Jones RH. Bayesian information criterion for longitudinal and clustered data. Stat Med 2011;30:3050–6.
[18] Velentgas P, Dreyer NA, Nourjah P, et al., eds. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Rockville, MD: Agency for Healthcare Research and Quality (AHRQ), 2013. http://effectivehealthcare.ahrq.gov/ehc/products/440/1166/User-Guide-Observational-CER-130113.pdf. [Accessed July 10, 2013].

[19] Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) study. JAMA 2001;285:2370–5.
[20] Davis RC, Hobbs FD, Kenkre JE, et al. Prevalence of atrial fibrillation in the general population and in high-risk groups: the ECHOES study. Europace 2012;14:1553–9.
[21] Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285–93.
[22] Meistrell ML. Evaluation of neural network performance by receiver operating characteristic (ROC) analysis: examples from the biotechnology domain. Comput Methods Programs Biomed 1990;32:73–80.
[23] Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. Clin Chem 2008;54:17–23.
[24] Merida-Rodrigo L, Poveda-Gomez F, Camafort-Babkowski M, et al. Long-term survival of ischemic stroke. Rev Clin Esp 2012;212:223–8.
[25] Ukena C, Dobre D, Mahfoud F, et al. Hypo- and hyperglycemia predict outcome in patients with left ventricular dysfunction after acute myocardial infarction: data from ephesus. J Card Fail 2012;18:439–45.
[26] Sinharay S, Stern HS, Russell D. The use of multiple imputation for the analysis of missing data. Psychol Methods 2001;6:317–29.
[27] Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;339:b2398.