



# The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56)

**Richard J. STEVENS, Viti KOTHARI, Amanda I. ADLER, Irene M. STRATTON and Rury R. HOLMAN on behalf of the United Kingdom Prospective Diabetes Study (UKPDS) Group**

Diabetes Trials Unit, Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, U.K.

## A B S T R A C T

A definitive model for predicting absolute risk of coronary heart disease (CHD) in male and female people with Type II diabetes is not yet available. This paper provides an equation for estimating the risk of new CHD events in people with Type II diabetes, based on data from 4540 U.K. Prospective Diabetes Study male and female patients. Unlike previously published risk equations, the model is diabetes-specific and incorporates glycaemia, systolic blood pressure and lipid levels as risk factors, in addition to age, sex, ethnic group, smoking status and time since diagnosis of diabetes. All variables included in the final model were statistically significant ( $P < 0.001$ , except smoking for which  $P = 0.0013$ ) in likelihood ratio testing. This model provides the estimates of CHD risk required by current guidelines for the primary prevention of CHD in Type II diabetes.

## INTRODUCTION

People with Type II diabetes have a risk of coronary heart disease (CHD) 2–4 times greater than the general population [1–3]. There is strong evidence that interventions can be beneficial in general populations [4–7], and increasingly in diabetic populations [8–10]. Patients at highest absolute risk have the most to gain from interventions, and it is desirable that all diabetic patients have their absolute CHD risk evaluated in order that optimal care can be determined. Treatment guidelines for primary prevention of CHD in primary care use absolute risk, alone or in conjunction with relative risk [11–14]. Obtaining estimates of absolute risk for developing CHD has been difficult in the absence of diabetes-specific risk equations. The model for CHD risk in Type II diabetes

presented here has wide applicability and will be of particular use to health care providers, insurers, planners, industry and government, in addition to clinicians and patients. Calculation of absolute risks also allows probabilities calculated from different models to be compared [15].

The U.K. Prospective Diabetes Study (UKPDS) is a landmark randomized controlled trial which showed that both intensive treatment of blood glucose and of blood pressure in diabetes can lower the risk of diabetes-related complications in individuals newly diagnosed with Type II diabetes [16,17]. In addition to answering therapy-related questions, the UKPDS cohort of 5102 patients, followed for a median of 10.7 years, provides an excellent opportunity to describe the natural history of treated disease. Presented here is a parametric model of the risk

**Key words:** coronary disease, diabetes mellitus, models, non-insulin-dependent, statistical.

**Abbreviations:** CHD, coronary heart disease; DCCT, Diabetes Control and Complications Trial; GDM, Global Diabetes Model; HDL, high-density lipoprotein; IMIB, Institute for Medical Informatics and Biostatistics; MI, myocardial infarction; UKPDS, U.K. Prospective Diabetes Study.

**Correspondence:** Dr Richard J. Stevens (e-mail richard.stevens@dtu.ox.ac.uk).

for CHD in patients with Type II diabetes. It provides formulae for incidence rates, estimates of probability for CHD complications, and the relative risks associated with potential risk factors. The model provides equations for absolute risk, incorporating the effect of multiple risk factors to give overall event rates, rather than relative risk as reported previously [18,19].

Previous models for CHD, such as the Framingham risk equations [20,21], have not been specifically designed for people with Type II diabetes. Although the accuracy of the Framingham equations for diabetic patients has been debated [22], they have been shown to apply well to U.K. populations [23], and many models for diabetic complications have been published that use them for cardiovascular risk [24–28]. Framingham and other models for CHD risk in the general population [29,30] use dichotomous variables for glycaemia, such as presence or absence of diabetes. In contrast, our diabetes-specific approach is advantageous as we include HbA<sub>1c</sub> as a continuous variable. We also replace age as a risk factor by two diabetes-specific variables: age at diagnosis of diabetes and time since diagnosis of diabetes, as previous UKPDS analyses have shown the importance of this distinction to diabetic complications [31]. There is also some evidence that diabetic dyslipidaemia is qualitatively different from dyslipidaemia in the general population [32].

## METHODS

### Study subjects

The UKPDS has been described previously [33]. Briefly, between 1977 and 1991, general practitioners in the catchment areas of 23 participating UKPDS hospitals

were asked to refer all patients aged 25–65 years presenting with newly diagnosed diabetes. Patients generally attended a UKPDS clinic within two weeks of referral. Inclusion criteria included, in addition to newly diagnosed diabetes, a fasting plasma glucose greater than 6 mmol/l (108 mg/dl) on two further occasions, and no recent history of myocardial infarction (MI), angina or heart failure. Exclusion criteria have been listed previously [33]. There were 5102 patients recruited to the UKPDS from the 7616 patients referred.

For this analysis, data from 4540 patients of White, Afro-Caribbean or Asian-Indian ethnic group were included. This excluded 39 patients of other ethnic groups, 248 patients with missing data for HbA<sub>1c</sub>, systolic blood pressure or lipids, and 275 patients with follow-up times too short for the model fitting process (see below). The characteristics of these patients are shown in Table 1.

Patients in the UKPDS had biochemical measurements, including HbA<sub>1c</sub>, blood pressures, and lipid and lipoprotein fractions, recorded at entry to the study, at randomization in the study after a three-month period of dietary therapy, and each year subsequently [33]. The systolic blood pressure recorded each year was the mean of three measurements taken at the same visit. Lipid ratio is defined here to be the ratio of total cholesterol to high-density lipoprotein (HDL) cholesterol. Biochemistry methodology has been reported previously [34]. HbA<sub>1c</sub> was measured by HPLC (Diamat Automated Glycosylated Haemoglobin Analyser, Bio-Rad), non-diabetic range 4.5–6.2% [34–35]. HbA<sub>1c</sub> measurements were certified comparable with the Diabetes Control and Complications Trial (DCCT) by the U. S. National Glycohemoglobin Standardization Program, with

$$\text{HbA}_{1c}(\text{UKPDS}) = 1.04 \times \text{HbA}_{1c}(\text{DCCT}) - 0.7336$$

**Table 1** Characteristics of patients, as mean (S.D.) or percentage (number), at diagnosis of diabetes (age, smoking status and body mass index) or mean of values 1–2 years after entry to study (HbA<sub>1c</sub>, systolic blood pressure and cholesterol measurements)

Variable	Values		
	Sex...	Men ( <i>n</i> = 2643)	Women ( <i>n</i> = 1897)
At diagnosis of diabetes			
Age (years)		51.5 (8.8)	52.7 (8.7)
White Caucasian (%)		81 (2151)	85 (1603)
Afro-Caribbean (%)		7.6 (201)	8.1 (153)
Asian-Indian (%)		11 (291)	7.4 (141)
Smoker (%)		34 (898)	25 (474)
Body-mass index		27.7 (4.6)	30.4 (6.3)
Mean of values one and two years after diagnosis of diabetes			
HbA <sub>1c</sub> (%)		6.6 (1.4)	6.9 (1.5)
Systolic blood pressure (mmHg)		133 (18)	139 (21)
Total cholesterol (mmol/l)		5.2 (1.0)	5.7 (1.1)
HDL cholesterol (mmol/l)		1.06 (0.23)	1.18 (0.27)

**Table 2** Risk factors included in the CHD model

Abbreviations, definitions and values used in equations in the text.

Abbreviations	Definitions/values
AGE	Age in years at diagnosis of diabetes
SEX	1 for female; 0 for male
AC	1 for Afro-Caribbean; 0 for Caucasian or Asian-Indian
SMOK	1 for a current smoker, of tobacco in any form, at diagnosis of diabetes; 0 otherwise
H	HbA <sub>1c</sub> (%), mean of values for years 1 and 2
SBP	Systolic blood pressure (mmHg), mean of values for years 1 and 2
LR	Total cholesterol/HDL cholesterol ratio, mean of values for years 1 and 2

where ( $r = 0.99$ ,  $n = 40$ ). Cholesterol and HDL cholesterol measurements were within the limits for accuracy of the lipid standardization program of the Centers for Disease Control and Prevention (Atlanta, GA, U.S.A.) [36], and were slightly lower than the Centers for Disease Control and Prevention reference laboratory in the U.K. [37];  $-2.5\%$  for total cholesterol and  $-1.2\%$  for HDL cholesterol [38].

### Model design

CHD is defined as the occurrence of fatal or non-fatal MI or sudden death, verified by two independent clinical assessors as described previously [33]. In patients with multiple CHD events, only the first event is considered in this study. During the early years of the study, all-cause mortality was not significantly different from that in the general population, possibly a consequence of exclusion criteria, such as previous vascular disease, malignant hypertension, ketonuria greater than 3 mmol/litre and severe retinopathy [39]. To avoid this possible selection bias, data from years 0–4 were not used in model fitting: that is, 275 patients with follow-up times less than four years or with MI prior to 4 years were not used in model fitting, and model fitting was carried out on data from the remaining 4540 patients using probabilities conditional on survival through the first 4 years. The baseline risk factors included in the model are listed in Table 2. Triacylglycerols are not included in the model as they were not significant ( $P > 0.5$  when added to the model above); excluded likewise were Asian-Indian ethnicity ( $P > 0.5$ ) and ex-smoking status ( $P > 0.2$ ).

To improve model stability, HbA<sub>1c</sub>, systolic blood pressure, total cholesterol and HDL cholesterol were taken to be the mean of values taken 1 year apart. Likelihood ratio comparisons, in an exploratory data analysis based on the Cox proportional hazards model, found these mean values to give a better fit than single values. Values at diagnosis have less predictive use, as many patients had treatment changed, by randomization, soon after diagnosis; we therefore use the mean of measurements taken at years 1 and 2 (Table 1).

Model fitting was carried out by maximum likelihood estimation, for which we used the Newton–Raphson method with numerical derivatives, as implemented by the Numerical Algorithms Group C library [40].

### Model equations

We give first the model equation for  $R(t)$ , the probability (risk) of a CHD event over  $t$  years, in a patient with newly diagnosed diabetes, in the absence of death from causes other than CHD:

$$R(t) = 1 - \exp \left\{ \frac{-q(1-d^t)}{1-d} \right\},$$

where  $d$  is the duration of diagnosed diabetes (see Table 3), and

$$q = q_0 \beta_1^{\text{AGE}-55} \beta_2^{\text{SEX}} \beta_3^{\text{AC}} \beta_4^{\text{SMOK}} \beta_5^{\text{H}-6.72} \beta_6^{(\text{SBP}-135.7)/10} \beta_7^{\ln(\text{LR})-1.59}$$

in which  $\ln$  denotes natural logarithm, and  $\beta_1$ – $\beta_7$  are risk factors defined in Tables 2 and 3. Notice that  $1 - R(t)$  is the survival probability for  $t$  years from diagnosis of diabetes. The survival probability for  $t$  years in a patient who has had diabetes for  $T$  years is also useful; it is  $R_T(t)$ , where

$$R_T(t) = 1 - \exp \left\{ -qd^T \left[ \frac{1-d^t}{1-d} \right] \right\}.$$

Equivalently, let  $P(t)$  be the probability of CHD in the year  $(t-1, t)$ , in a patient who has survived  $t-1$  years without CHD. Then the same model gives

$$P(t) = 1 - \exp(-qd^{t-1}).$$

The derivation of  $R(t)$  and of  $R_T(t)$  from this formula for  $P(t)$  is given in the Appendix.

### Robustness of the model

Model assumptions were checked with a series of diagnostic plots, comparing survival probabilities for the study population calculated by the model with survival probabilities for the study population calculated by non-parametric methods. The non-parametric method used

was the life-table method with one-year intervals, which also provides 95% confidence intervals [41]. The modelled survival probabilities were calculated as follows. Let  $S_{ij}$  denote the modelled survival probability at  $j$  years, equal to  $1 - R(j)$ , for patient  $i$ . For each year  $j = 1, \dots, 20$ , let  $S_j$  be the average of  $S_{ij}$  over all  $i = 1, \dots, 4540$ . Then  $S_j$  is the survival rate for the UKPDS population at  $j$  years predicted by the model.

Proportional hazards assumptions have been verified with log-cumulative hazard plots. Likelihood ratio tests were made for interactions between  $HbA_{1c}$ , systolic blood pressure and lipid ratio, and for interactions between each of these and age and sex.

## RESULTS

Median follow-up time from study entry to death was 10.7 years, and 10.3 years to death or MI; there were 29878 person-years of follow-up available for model fitting. The parameter estimates are shown in Table 3. All

variables included in the model were significant at  $P < 0.001$  in likelihood ratio tests, except for smoking ( $P = 0.0013$ ). No interactions were found between variables, with  $P > 0.25$  in all cases tested.

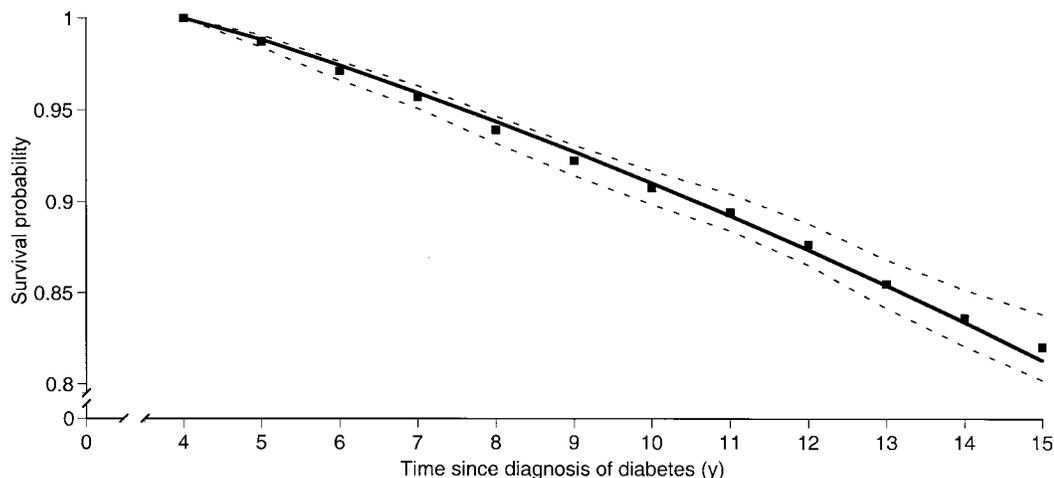
Although triacylglycerols were not significant when added to the final model ( $P < 0.5$ ), in a supplementary analysis they were added to a model that was not adjusted for lipid ratio, and found to be significant ( $P = 0.0014$ ; approximate risk ratio 1.10 per mmol/l). However, the  $P$ -value for the log-lipid ratio was stronger still ( $P < 10^{-10}$ ; parameter 3.845, as shown in Table 3).

Figure 1 shows that the survival rates predicted by the model lie close to the rates observed in the UKPDS, and well within the non-parametric confidence intervals. Figure 2 shows the observed and modelled survival rates separated by  $HbA_{1c}$ , systolic blood pressure and lipid ratio, confirming the ability of the model to adjust for these risk factors.

To illustrate the use of the equation, consider a White or Asian-Indian male non-smoker, with Type II diabetes newly-diagnosed at age 45 years, with  $HbA_{1c} = 7.5\%$ ,

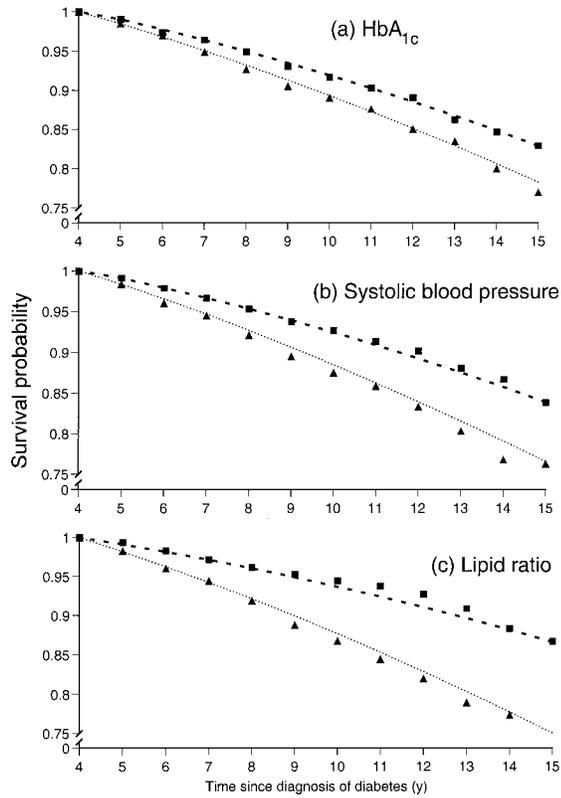
**Table 3** Parameter estimates by maximum likelihood

Parameter	Interpretation	Estimate	95% confidence interval
$q_0$	Intercept	0.0112	0.0082–0.014
$\beta_1$	Risk ratio for one year of age at diagnosis of diabetes	1.059	1.05–1.07
$\beta_2$	Risk ratio for female sex	0.525	0.42–0.63
$\beta_3$	Risk ratio for Afro-Caribbean ethnicity	0.390	0.19–0.59
$\beta_4$	Risk ratio for smoking	1.350	1.11–1.59
$\beta_5$	Risk ratio for 1% increase in $HbA_{1c}$	1.183	1.11–1.25
$\beta_6$	Risk ratio for 10 mmHg increase in systolic blood pressure	1.088	1.04–1.14
$\beta_7$	Risk ratio for unit increase in logarithm of lipid ratio	3.845	2.59–5.10
$d$	Risk ratio for each year increase in duration of diagnosed diabetes	1.078	1.05–1.11



**Figure 1** Observed survival rates from year 4, with 95% confidence intervals and modelled survival rates from year 4, for years 5–15

(■) Observed values, (---) 95% confidence intervals, (—) modelled.



**Figure 2** Modelled and observed survival rates from 4 years after entry to the study

(a) HbA<sub>1c</sub> value. (■) Observed, HbA<sub>1c</sub> ≤ 7.0%; (---) modelled, HbA<sub>1c</sub> ≤ 7.0%; (▲) observed, HbA<sub>1c</sub> > 7.0%; (·····) modelled, HbA<sub>1c</sub> > 7.0%. (b) Systolic blood pressure. (■) Observed, SBP ≤ 140 mmHg; (---) modelled, SBP ≤ 140 mmHg; (▲) observed, SBP > 140 mmHg; (·····) modelled, SBP > 140 mmHg. (c) Lipid ratio. (■) Observed, lipid ratio ≤ 5.0; (---) modelled, lipid ratio ≤ 5.0; (▲) observed, lipid ratio > 5.0; (·····) modelled, lipid ratio > 5.0.

systolic blood pressure 160 mmHg, total cholesterol 4.9 mmol/l, HDL cholesterol 1.0 mmol/l, and no previous history of CHD. Firstly, lipid ratio = 4.9/1.0 = 4.9. Then, from the equations above:

$$\begin{aligned}
 q &= q_0 \beta_1^{45-55} \beta_2^0 \beta_3^0 \beta_4^0 \beta_5^{7.5-6.72} \beta_6^{(160-135.7)/10} \\
 &\quad \beta_7^{\ln(4.9)-1.59} = 0.112 \times 1.059^{-10} \times 0.525^0 \\
 &\quad \times 0.390^0 \times 1.350^0 \times 1.183^{0.78} \times 1.088^{2.43} \\
 &\quad \times 3.845^{00} = 0.112 \times 1.059^{-10} \times 1 \times 1 \times 1 \\
 &\quad \times 1.183^{0.78} \times 1.088^{2.43} \times 1 = 0.00883.
 \end{aligned}$$

Then the risk of CHD within 20 years is:

$$\begin{aligned}
 R(20) &= 1 - \exp\{-q[(1-d^{20})/(1-d)]\} \\
 &= 1 - \exp\{-0.0883[(1-1.078^{20})/ \\
 &\quad (1-1.078)]\} = 1 - \exp(-0.395) = 0.33
 \end{aligned}$$

so that there is a 33% chance of a CHD event within 20

years, assuming that death does not occur from non-CHD causes during those 20 years.

## DISCUSSION

Recent papers from the UKPDS group have demonstrated the continuous nature of glycaemia and blood pressure as risk factors for CHD [18,19]. New in this paper is a measure of dyslipidaemia, the ratio of total cholesterol to HDL cholesterol. The methodology of the present paper also differs from the previous papers in ways motivated by the accompanying 'Risk Engine' software project [42]. A fully parametric model combines hazard ratios and absolute event rates in a single equation, to allow estimation of event rates and survival probabilities in a variety of applications, such as resource use estimation for health planners, power calculations for clinical trials, and the estimation of effectiveness and cost-effectiveness in early stages of drug development cycles. The model could also be used to project risk profiles for a given patient, in the manner of the Sheffield and New Zealand Tables [43,44].

The importance of HbA<sub>1c</sub> as a continuous risk factor for CHD, in Type II diabetes, is underlined by the results of Table 4. A causal relationship between glycaemia, measured by HbA<sub>1c</sub>, and CHD has not been proved [16], but a predictive relationship has been established [18,45]. These equations model differences in risk factors, but do not include terms for any specific therapy. This is not necessary because the effects of the therapies used in the UKPDS were, for MI, found to be consistent with their effects on risk factors [18,19]. It is possible that the additional benefit with metformin observed for diabetes-related deaths in overweight patients may also apply to MI [9], but this affects only 308 of the 4541 patients in this analysis. The widely-used Framingham models are derived from a population with only 337 diabetic patients [20], and the Prospective Cardiovascular Munster ('PROCAM') model is derived only from the data from the men in that particular study [46]. Both these models, and the Lipid Research Clinics model [47], contain no measure of glycaemia beyond the presence or absence of diabetes itself, being designed for general rather than diabetes-specific application.

Previous silent MI is not considered as a risk factor as electrocardiography is not routinely performed on diabetic patients in clinical practice. Neither does our definition of CHD include silent MI, since silent MI was not an end point of the UKPDS. Some authors have suggested that the increased prevalence of silent MI in diabetes is in proportion to the increase in CHD [48], though evidence from the Framingham study suggests that this may vary with sex [49].

The example given in the Results section can be compared to risk estimates, for equivalent data, made by

**Table 4** Ten-year risk of CHD in patients with newly diagnosed diabetes, by the UKPDS Risk Engine, and by the Framingham risk profile as implemented by the Joint British Societies Cardiac Risk Assessor Computer Program [28]Values underlined indicate a risk greater than 30%.

Variables						Risk by UKPDS Risk Engine (%)			Risk by Framingham (%)
Present age (years)	Sex	Smoking	Systolic BP (mmHg)	Lipid ratio	HbA <sub>1c</sub> (%)...	6	8	10	All values
55	Female	No	140	4.0		5.7	7.9	10.9	12.6
55	Female	No	140	8.0		13.9	18.9	25.4	25.6
55	Female	No	160	4.0		6.7	9.3	12.7	15.3
55	Female	No	160	8.0		16.2	21.9	29.3	29.4
55	Female	Yes	140	4.0		7.6	10.5	14.4	19.2
55	Female	Yes	140	8.0		18.2	24.6	<u>32.6</u>	<u>34.5</u>
55	Female	Yes	160	4.0		8.9	12.3	16.8	22.6
55	Female	Yes	160	8.0		21.2	28.4	<u>37.3</u>	<u>38.6</u>
55	Male	No	140	4.0		10.6	14.5	19.7	13.4
55	Male	No	140	8.0		24.8	<u>32.9</u>	<u>42.7</u>	26.7
55	Male	No	160	4.0		12.4	16.9	22.9	16.2
55	Male	No	160	8.0		28.6	<u>37.6</u>	<u>48.3</u>	<u>30.6</u>
55	Male	Yes	140	4.0		14.0	19.1	25.6	20.2
55	Male	Yes	140	8.0		<u>31.9</u>	<u>41.6</u>	<u>52.9</u>	<u>35.7</u>
55	Male	Yes	160	4.0		16.4	22.1	29.5	23.7
55	Male	Yes	160	8.0		<u>36.5</u>	<u>47.1</u>	<u>59.0</u>	<u>39.9</u>

the Institute for Medical Informatics and Biostatistics (IMIB) diabetes model and the Global Diabetes Model (GDM) [15]. First, the estimate from our model (33%) must be adjusted for the competing risk of death from non-CHD causes. Using a previously published simple life-expectancy model, this reduces the 20-year risk of CHD from 33% to 29% [50]. The IMIB model, derived from the Framingham equations, estimated a 23% chance for MI for equivalent data; since the IMIB endpoint does not include sudden cardiac death, this is similar to our 29% estimate. The GDM, derived from the Framingham equations and a Kaiser-Permanente data set, estimated a probability of only 10%. Explanations for the discrepancy between IMIB and the GDM results include a low rate of MI in the Kaiser-Permanente cohort, and differing assumptions about risk for CHD following another vascular event [15]. This marked difference, between two applications of the same (Framingham) model, illustrates the difficulties that can arise in the application, as well as in the development, of risk models.

Since HbA<sub>1c</sub>, blood pressure and lipid measurements are all subject to within-person variation, users of the model should consider the effect of regression dilution on estimates [51]. Where possible, we recommend using measurements of similar variability to our own (the mean of two measurements taken at different times) so that no correction for regression dilution is necessary. In the Appendix, we supply information on the adjustments to be made to the parameters when this is not possible.

The data from which our model derives is restricted to those patients recruited by the UKPDS for randomi-

zation in a clinical trial. Consequences include the restriction of the data to those aged 65 or under at diagnosis of diabetes, and the exclusion of those with recent major heart disease or stroke [33]. Ideally, the model would be derived from a large-scale epidemiological study of diabetic patients. Until such a model is published, our model has significant advantages for diabetic patients over other published models: particularly, the inclusion of HbA<sub>1c</sub>, as discussed above.

The exclusion from the model-fitting process of years 0–4 of the study data is one consequence of the use of a clinical trial, with selection criteria, rather than a truly population-based study. The first years of the study were known to have mortality rates lower than in the general population [39], and the exclusion of years 0–4 is both necessary (results not shown) and sufficient (Figures 1 and 2) for a good model fit. However, application for patients with less than 4 years of diagnosed diabetes involves backward extrapolation. The reliability of the model for forward extrapolation has been examined using temporal cross-validation methods, and is available from URL: <http://cs.portlandpress.com/cs/101/cs/1010671add.htm>. A more stringent test will be possible when data from UKPDS post-study monitoring are published, but this does not remove the need to test the generalizability of the model to other populations. We hope that the provision of the equations in software form [42] will encourage the comparison of the model to cohorts being studied elsewhere.

The potential applications of this model are many. A variation of the model was used to estimate increases in

life expectancy in UKPDS cost-effectiveness analyses [52–54]. The model can be used by health planners to estimate resource use, and, in conjunction with prevalence figures, the burden of complications. We intend to provide the model as a software package including the ‘Risk Engine’ that should prove useful to health care providers, health economists, clinicians and people with diabetes. The software will be distributed free of charge to non-profit organizations.

## ACKNOWLEDGMENTS

The co-operation of the patients and many NHS and non-NHS staff at the centres is much appreciated (participating centres: Radcliffe Infirmary, Oxford; Royal Infirmary, Aberdeen; University Hospital, Birmingham; St George’s Hospital and Hammersmith Hospital, London; City Hospital, Belfast; North Staffordshire Royal Infirmary, Stoke-on-Trent; Royal Victoria Hospital, Belfast; St Helier Hospital, Carshalton; Whittington Hospital, London; Norfolk and Norwich Hospital; Lister Hospital, Stevenage; Ipswich Hospital; Ninewells Hospital, Dundee; Northampton Hospital; Torbay Hospital; Peterborough General Hospital; Scarborough Hospital; Derbyshire Royal Infirmary; Manchester Royal Infirmary; Hope Hospital, Salford; Leicester General Hospital; Royal Devon and Exeter Hospital). The major grants for this study were from the U.K. Medical Research Council, British Diabetic Association, the U.K. Department of Health, The National Eye Institute, and The National Institute of Digestive, Diabetes and Kidney Disease in the National Institutes of Health, U.S.A., The British Heart Foundation, Novo-Nordisk, Bayer, Bristol-Myers Squibb, Hoechst, Eli Lilly, Lipha and Farmitalia Carlo Erba. Other funding companies and agencies, the supervising committees, and all participating staff are listed in a previous paper [16]. This work was supported by a grant from the Wellcome Trust (054470/Z/98/DG/NOS/fh). We are grateful to Andrew Neil, Carole Cull and Sue Manley for helpful comments.

## APPENDIX

### Choice of model

This model form was chosen for its interpretability in diabetic patients, and for its close fit to the UKPDS data. It was motivated by the approximate multiplicative effect of age on risk observed in the UKPDS. For example, categorizing age at entry to study into ten-year groups, univariate Cox proportional hazards modelling finds risk ratios relative to the 35–45 age group are 1.73 for age group 45–55 and 2.98 for age group 55–65. Since  $1.73^2 = 2.99$ , this strongly suggests a model of the form

$P(t) \propto \beta^{AGE}$  for some parameter  $\beta$ . The model we have used has the approximate property  $P(t) \propto \beta_1^{AGE} \cdot d^t$ , so that risk increases by a factor of  $\beta_1$  with each year of age at diagnosis of diabetes and by a factor  $d$  with each year subsequent to diagnosis of diabetes. The model may therefore be interpreted as one in which the risk of CHD increases with age, but the risk increases more per year after diagnosis of diabetes, as is suggested by previous analyses of data on Q-wave infarction and hypertension [31].

### Regression dilution

The term ‘regression dilution’ describes the behaviour of parameter estimates in regression and in survival analysis when a predictor variable cannot be measured precisely, for example, due to within-subject bio-variability. As uncertainty increases in the measurement of the predictor variable, the apparent size of the effect parameter is decreased or ‘diluted’ [55]. We used a maximum likelihood method to estimate the extent of this effect for HbA<sub>1c</sub>, systolic blood pressure and log-lipid ratio in our model. The method estimates a correction factor for the log of each effect parameter [51]. The method is not exact for a proportional hazards model, but the approximation is good for a study with high levels of censoring [56]. The estimated correction factors are 1.33 for HbA<sub>1c</sub>, 1.25 for systolic blood pressure and 1.22 for the lipid ratio (total cholesterol to HDL cholesterol). These imply that, were the model to be applied to perfect measures of each variable,  $\beta_5$  should be increased to  $1.183^{1.33} = 1.250$ ,  $\beta_6$  to  $1.088^{1.25} = 1.111$ , and  $\beta_7$  to  $3.845^{1.22} = 5.171$ . Since variables are centred around their means in the model equation (for example, HbA<sub>1c</sub> is used as  $H - 6.72$ ), no adjustment is required to the intercept parameter  $q_0$ .

These corrections have not been incorporated into the parameter estimates in Table 3, as users of the model will in general also have imprecise measurements of HbA<sub>1c</sub>, blood pressure and lipids for any given individual. Assuming the values entered into the model were recorded with similar precision to those on which the model was built, no correction is necessary. The corrected parameters should only be used where all bio-variation has been removed; for example, by taking the average of a very large number of measurements.

Conversely, users may have estimates less precise than our own, such as a single measurement of HbA<sub>1c</sub> where we have used a mean of two values. In this case the parameter values in Table 3 would be over-estimates, and  $\beta_5$ ,  $\beta_6$ , and  $\beta_7$  should be decreased to 1.144, 1.073 and 3.11 respectively. To derive these, note that it follows from the derivation of the maximum likelihood method that if the correction factor for the mean of two readings of a variable is  $1 + (\lambda/2)$ , then the correction factor for the same variable measured with a single reading is  $(1 + \lambda)$ .

For  $HbA_{1c}$ , for example, the correction factor for the two-point mean was  $1.33 = 1 + (0.66/2)$ ; now observe that  $1.144^{(1+0.66)} = 1.250$ , and so 1.144 is the estimated value that  $\beta_5$  would have taken had it been derived from single measures of  $HbA_{1c}$ , rather than two-point means.

### Mathematical properties of the model equation

Recall the formula

$$P(t) = 1 - \exp(-qd^{t-1}),$$

where

$$q = q_0 \beta_1^{AGE-55} \beta_2^{SEX} \beta_3^{AC} \beta_4^{SMOK} \beta_5^{H-6.72} \beta_6^{(BP-135.7)/10} \beta_7^{\ln(LR)-1.59}.$$

From the Taylor expansion

$$\exp(-x) = 1 - x + (x^2/2) + O(x^3),$$

where  $O(x^3)$  denotes terms in  $x^3$  and higher powers of  $x$ , it follows that

$$1 - \exp(-x) = x + O(x^2),$$

and so for typically small values  $P(t)$  is approximately equal to  $qd^{t-1}$ . For example,  $P(t) = 0.049$  when  $qd^{t-1}$ . Then

$$P(t) \approx qd^{t-1} = q_0 \beta_1^{AGE-55} \beta_2^{SEX} \beta_3^{AC} \beta_4^{SMOK} \beta_5^{H-6.72} \beta_6^{(BP-135.7)/10} \beta_7^{\ln(LR)-1.59} d^{t-1},$$

and the model is approximately a proportional hazards model on discrete time with baseline hazard function  $q_0 d^{t-1}$  and risk ratios  $\beta_1, \dots, \beta_7$ . It also follows that  $q_0$  has an interpretation as the approximate probability of CHD in the first year of diagnosis for a 55-year old, non-smoking White or Asian-Indian male with  $H = 6.72$ ,  $SBP = 135.7$  and  $LR = \exp(1.59) = 4.9$ . To see that the formula for  $P(t)$  is equivalent to the formula for  $R_T(t)$ , define  $S_T(t) = 1 - R_T(t)$ . Then

$$S_T(t) = [1 - P(T+1)][1 - P(T+2)] \dots \\ [1 - P(i+t)] = \exp[-q(d^T + d^{T+1} + \dots + d^{T+t-1})] = \exp[-qd^T(1-d^t)/(1-d)]$$

by application of the standard formula for a geometric progression, and the formula for  $R_T(t)$  follows. Then  $R(t)$  is the special case  $T = 0$ .

### REFERENCES

- Barrett-Connor, E. L., Cohn, B. A., Wingard, D. L. and Edelstein, S. L. (1991) Why is diabetes mellitus a stronger risk factor for fatal ischemic heart disease in women than in men? *JAMA J. Am. Med. Assoc.* **265**, 627–631
- Koskinen, P., Manttari, M., Manninen, V., Huttunen, J. K., Heinonen, O. P. and Frick, M. H. (1992) Coronary heart disease incidence in NIDDM patients in the Helsinki Heart Study. *Diabetes Care* **15**, 820–825
- Manson, J. A. E., Colditz, G. A., Stampfer, M. J. et al. (1991) A prospective study of maturity-onset diabetes mellitus and risk of coronary heart disease and stroke in women. *Arch. Intern. Med.* **151**, 1141–1147
- Medical Research Council Working Party. (1985) MRC trial of treatment of mild hypertension: principal results. *Br. Med. J.* **291**, 97–104
- Shepherd, J., Cobbe, S. M., Ford, I. et al. (1995) Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. West of Scotland Coronary Prevention Study Group. *N. Engl. J. Med.* **333**, 1301–1307
- Dahlof, B., Lindhol, L., Hansson, L., Scherstenn, B., Ekblom, T. and Western, P. O. (1991) Morbidity and mortality in the Swedish trial in old patients with hypertension (STOP-Hypertension). *Lancet* **338**, 1281–1285
- Downs, J. R., Clearfield, M., Weis, S. et al. (1998) Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. Air Force/Texas Coronary Atherosclerosis Prevention Study. *JAMA J. Am. Med. Assoc.* **279**, 1615–1622
- Pyörälä, K., Pedersen, T. R., Kjekshus, J., Fågerman, O., Olsson, A. G. and Thorgeirsson, G. (1997) Cholesterol lowering with simvastatin improves prognosis of diabetic patients with coronary heart disease. A subgroup analysis of the Scandinavian Simvastatin (4S). *Diabetes Care* **20**, 614–620
- UKPDS Group. (1998) Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). *Lancet* **352**, 854–865
- Heart Outcomes Prevention Evaluation Study Investigators. (2000) Effects of ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. *Lancet* **355**, 253–259
- British Cardiac Society, British Hyperlipidaemia Association, British Hypertension Society and British Diabetic Association. (1998) Joint British recommendations on prevention of coronary heart disease in clinical practice. *Heart* **80**, S1–S29
- Wood, D., Backer, G. D., Faergeman, O., Graham, I., Mancia, G. and Pyörälä, K. (1998) Prevention of coronary heart disease in clinical practice: Recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. *Atherosclerosis* **140**, 199–270
- Grundy, S. M., Balady, G. J., Criqui, M. H. et al. (1998) Primary prevention of coronary heart disease: guidance from Framingham. A statement for healthcare professionals from the AHA Task Force on Risk Reduction. *Circulation* **97**, 1876–1887
- Expert Panel on Detection Evaluation and Treatment of High Blood Cholesterol in Adults. (2001) Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA J. Am. Med. Assoc.* **285**, 2486–2497
- Brown, J. B., Palmer, A. J., Bisgaard, P., Chan, W., Pedula, K. and Russell, A. (2000) The Mt Hood challenge: cross-testing two diabetes simulation models. *Diabetes Res. Clin. Prac.* **50**, S57–S64
- UKPDS Group. (1998) 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* **352**, 837–853
- UKPDS Group. (1998) Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes (UKPDS 38). *Br. Med. J.* **317**, 703–713
- UKPDS Group. (2000) Association of glycaemia with macrovascular and microvascular complications of Type 2 diabetes (UKPDS 35). *Br. Med. J.* **321**, 405–412
- UKPDS Group. (2000) Association of systolic blood pressure with macrovascular and microvascular complications of type 2 diabetes (UKPDS 36). *Br. Med. J.* **321**, 412–419

- 20 Anderson, K. M., Wilson, P. W., Odell, P. M. and Kannel, W. B. (1991) An updated coronary risk profile. A statement for health professionals. *Circulation* **83**, 356–362
- 21 Anderson, K. M., Odell, P. M., Wilson, P. W. and Kannel, W. B. (1991) Cardiovascular disease risk profiles. *Am. Heart J.* **121**, 293–298
- 22 Zambanini, A., Smith, M. R. and Feher, M. D. (1999) Prediction of cardiovascular risk. Program is not suitable for diabetic patients. *Br. Med. J.* **318**, 1418
- 23 Ramachandram, S., French, J. M., Vanderpump, M. P. J., Croft, P. and Neary, R. H. (2000) Using the Framingham model to predict heart disease in the United Kingdom: retrospective study. *Br. Med. J.* **320**, 676–677
- 24 Eastman, R. C., Javitt, J. C., Herman, W. H. et al. (1997) Model of complications of NIDDM. I. Model construction and assumptions. *Diabetes Care* **20**, 725–734
- 25 Eastman, R. C., Javitt, J. C., Herman, W. H. et al. (1997) Model of complications of NIDDM. II. Analysis of the health benefits and cost-effectiveness of treating NIDDM with the goal of normoglycaemia. *Diabetes Care* **20**, 735–744
- 26 Palmer, A. J., Brandt, A., Gozzoli, V., Weiss, C., Stock, H. and Wenzel, H. (2000) Outline of a diabetes management model: principles and applications. *Diabetes Res. Clin. Prac.* **50**, S47–S56
- 27 Brown, J. B., Russell, A., Chan, W., Pedula, K. and Aickin, M. (2000) The global diabetes model: user friendly version 3.0. *Diabetes Res. Clin. Prac.* **50**, S15–S46
- 28 British Hypertension Society. (1998) Cardiac Risk Assessor for use with the Joint Recommendations of the British Cardiac Society, British Hyperlipidemia Association, British Hypertension Society and British Diabetic Association on Prevention of Coronary Heart Disease in Clinical Practice, University of Manchester, Manchester. <http://hyp.ac.uk/bhs/managmt.html>
- 29 International Task Force for Prevention of Coronary Heart Disease. (1998) Coronary heart disease: reducing the risk. *Nutr. Metab. Cardiovas. Dis.* **8**, 205–271
- 30 Grover, S. A., Coupal, L., Zowall, H. and Dorais M. (2000) Cost-effectiveness of treating hyperlipidemia in the presence of diabetes: who should be treated? *Circulation* **102**, 722–727
- 31 UKPDS Group. (1997) UK Prospective Diabetes Study 22: Effect of age at diagnosis on diabetic tissue damage during the first 6 years of NIDDM. *Diabetes Care* **20**, 1435–1441
- 32 Feher, M. D. and Elkeles, R. S. (1999) Lipid modification and coronary heart disease in type 2 diabetes: different from the general population? *Heart* **81**, 10–11
- 33 UKPDS Group. (1991) UK Prospective Diabetes Study VIII: Study design, progress and performance. *Diabetologia* **34**, 877–890
- 34 UKPDS Group. (1994) UK Prospective Diabetes Study XI: Biochemical risk factors in type 2 diabetic patients at diagnosis compared with age-matched normal subjects. *Diabetic Med.* **11**, 534–544
- 35 Cull, C. A., Manley, S. E., Stratton, I. M. et al. (1997) Approach to maintaining comparability of biochemical data during long-term clinical trials. *Clin. Chem.* **43**, 1913–1918
- 36 Laboratory Standardization Panel. (1988) Current status of blood cholesterol measurement in clinical chemistry laboratories in the United States. *Clin. Chem.* **34**, 193–201
- 37 Packard, C. J., Bell, M. A., Eaton, R. H., Dagen, M. M., Cassidy, M. and Shepherd, J. A. (1993) A pilot scheme for improving the accuracy of serum cholesterol measurement in Scotland and Northern Ireland. *Ann. Clin. Biochem.* **30**, 387–393
- 38 UKPDS Group. (1997) UK Prospective Diabetes Study 27: Plasma lipids and lipoproteins at diagnosis of NIDDM: effects of age and gender. *Diabetes Care* **20**, 1683–1687
- 39 UKPDS Group. (1998) Risk factors for coronary artery disease in non-insulin dependent diabetes (UKPDS 23). *Br. Med. J.* **316**, 823–828
- 40 Numerical Algorithms Group. (1998) Numerical Algorithms Group C Library, Oxford
- 41 Collett, D. (1994) Modelling Survival Data in Medical Research, pp 27–28, Chapman & Hall, London
- 42 UKPDS Group. (2000) UKPDS Risk Engine, <http://www.dtu.ox.ac.uk/ukpds/riskengine.html>
- 43 Ramsay, L. E., Haq, I. U., Jackson, P. R., Yeo, W. W., Pickin, D. M. and Payne, J. N. (1996) Target lipid-lowering drug therapy for primary prevention of coronary heart disease: an updated Sheffield table. *Lancet* **348**, 387–388
- 44 Jackson, R. (2000) Updated New Zealand cardiovascular risk-benefit prediction guide. *Br. Med. J.* **320**, 709–710
- 45 Khaw, K.-T., Wareham, N., Luben, R. et al. (2001) Glycated haemoglobin, diabetes and mortality in men in Norfolk cohort of European Prospective Investigation of Cancer and Nutrition (EPIC-Norfolk). *Br. Med. J.* **322**, 15–18
- 46 Assmann, G., Cullen, P., Jossa, F., Lewis, B. and Mancini, M. (1999) Coronary heart disease: reducing the risk. The scientific background to primary and secondary prevention of coronary heart disease. A worldwide view. *Arterioscler. Thromb. Vasc. Biol.* **19**, 1819–1824
- 47 Grover, S. A., Paquet, S., Levinton, C., Coupal, L. and Zowall, H. (1998) Estimating the benefits of modifying risk factors of cardiovascular disease. *Arch. Int. Med.* **158**, 655–662
- 48 Airaksinen, K. E. J. (2001) Silent coronary artery disease in diabetes – a feature of autonomic neuropathy or accelerated atherosclerosis? *Diabetologia* **44**, 259–266
- 49 Kannel, W. B. (1986) Silent myocardial ischemia and infarction: insights from the Framingham study. *Cardiol. Clin.* **4**, 583–591
- 50 Stevens, R., Adler, A., Gray, A., Briggs, A. and Holman, R. (2000) Life-expectancy projection by modelling and computer simulation (UKPDS 46). *Diabetes Res. Clin. Prac.* **50**, S5–S13
- 51 Frost, C. and Thompson, S. (2000) Correcting for regression dilution bias: comparison of methods for a single predictor variable. *J. R. Stat. Soc. (series A)* **163**, 173–190
- 52 UKPDS Group. (1998) Cost-effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes (UKPDS 40). *Br. Med. J.* **317**, 720–726
- 53 UKPDS Group. (2000) Cost-effectiveness of an intensive blood glucose control policy in patients with type 2 diabetes: economic analysis alongside randomised controlled trial (UKPDS 41). *Br. Med. J.* **320**, 1373–1378
- 54 UKPDS Group. (2001) Cost-effective analysis of intensive blood glucose control with metformin in overweight patients with type 2 diabetes (UKPDS 51). *Diabetologia* **44**, 298–304
- 55 MacMahon, S., Peto, R., Cutler, J. et al. (1990) Blood pressure, stroke, and coronary heart disease. Part 1: prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet* **335**, 765–774
- 56 Hughes, M. D. (1993) Regression dilution in the proportional hazards model. *Biometrics* **49**, 1056–1066

Received 18 December 2000/23 May 2001; accepted 21 August 2001