UKPDS 60
Risk of Stroke in Type 2 Diabetes Estimated by the UK Prospective Diabetes Study Risk Engine

Viti Kothari, MSc; Richard J. Stevens, PhD; Amanda I. Adler, MD; Irene M. Stratton, MSc; Susan E. Manley, PhD; H. Andrew Neil, FRCP; Rudy R. Holman, FRCP; for the UK Prospective Diabetes Study Group

Background and Purpose—People with type 2 diabetes are at elevated risk of stroke compared with those without diabetes. Relative risks have been examined in earlier work, but there is no readily available method for predicting the absolute risk of stroke in a diabetic individual. We developed mathematical models to estimate the risk of a first stroke using data from 4549 newly diagnosed type 2 diabetic patients enrolled in the UK Prospective Diabetes Study.

Methods—During 30,700 person-years of follow-up, 188 first strokes (52 fatal) occurred. Model fitting was carried out by maximum likelihood estimation using the Newton-Raphson method. Diagnostic plots were used to compare survival probabilities calculated by the model with those calculated using nonparametric methods.

Results—Variables included in the final model were duration of diabetes, age, sex, smoking, systolic blood pressure, total cholesterol to high-density lipoprotein cholesterol ratio and presence of atrial fibrillation. Not included in the model were body mass index, hemoglobin A1c, ethnicity, and ex-smoking status. The use of the model is illustrated with a hypothetical study power calculation.

Conclusions—This model forecasts the absolute risk of a first stroke in people with type 2 diabetes using variables readily available in routine clinical practice. (Stroke. 2002;33:1776-1781.)

Key Words: blood pressure □ cholesterol □ diabetes mellitus □ statistics □ stroke
diagnosis of diabetes; 215 with missing data for blood pressure, lipids, or electrocardiography; and 262 with follow-up times too short (<4 years) for the model-fitting process. Characteristics of those included in the model are shown in Table 1.

Patients had blood pressure and biochemical measurements, including hemoglobin (Hb) A1c, lipid, and lipoprotein fractions recorded at diagnosis of diabetes, at the end of a 3- to 4-month dietary run-in period, and then at annual intervals. HbA1c was measured by high-performance liquid chromatography (Biorad Diamat Automated Glycosylated Hemoglobin Analyzer), nondiabetic range 4.5% to 6.2%.26,27 Cholesterol measurements were within the limits for accuracy of the lipid standardization program of the Centers for Disease Control and Prevention (CDC) in Atlanta, Ga, but when compared with the CDC reference laboratory in the United Kingdom were 2.5% lower and 1.2% lower (total cholesterol and HDL cholesterol, respectively).27-29 Systolic blood pressure (SBP) was recorded each year as the mean of three measurements, with a coefficient of variation <1.5%, taken at the same visit. AF was identified clinically or by electrocardiography.

End Points
Stroke in the UKPDS is defined as a neurological deficit with symptoms or signs lasting 1 month or more.25 No distinction was made between ischemic, embolic, and hemorrhagic strokes. In patients with multiple strokes, only the first stroke is considered here. The model treats death from causes other than stroke as censored.

Statistical Model and Methods
The mathematical form of the model and its derivation have been described elsewhere.30 Table 2 shows the variables considered for inclusion in the model. Model building then proceeded stepwise using likelihood ratio tests for significance. The following well-documented risk factors for stroke were tested for inclusion at a significance level of 0.05: sex, age at diagnosis of diabetes, SBP, atrial fibrillation, smoking, and ratio of total:HDL cholesterol.7,10,31

The following additional possible risk factors were tested at a significance level of 0.005, to reduce the danger of a type I error (erroneous inclusion) occurring: HbA1c, Afro-Caribbean ethnicity, Asian-Indian ethnicity, body mass index (BMI), and duration of diagnosed diabetes. Proportional-hazards assumptions were verified with log-cumulative hazard plots32 (data not shown). For AF, there were too few cases for a reliable log-cumulative hazard plot.

To improve model stability, SBP, total cholesterol, HDL cholesterol, and HbA1c were each taken to be the mean of values taken 1 and 2 years after diagnosis of diabetes, as in our previous article.30 Model fitting was carried out by maximum likelihood estimation, using the Newton-Raphson method as implemented in the Numerical Algorithms Group C Library.33

In a supplementary analysis, we fitted alternative models in which the ratio T:H was replaced with each of the following: total cholesterol, HDL cholesterol, low-density lipoprotein (LDL) cholesterol, triglycerides, and the ratio LDL:HDL cholesterol. During the early years of the study, all-cause mortality was not significantly different from that in the general population, possibly because people with life-threatening illnesses were excluded from the study.34 To avoid selection bias, years 0 to 4 of data were not used in model fitting; 262 patients with follow-up times <4 years were excluded as described above, and the fitting process for the remaining data used likelihoods conditional on surviving to 4 years without a stroke.

Model adequacy was checked with diagnostic plots, comparing modeled survival probabilities, calculated as the average across the 4549 patients, to survival probabilities for the study population calculated by nonparametric (life-table) methods.32

External validity of the model was examined by comparison with the rates of fatal stroke reported in 1370 patients of the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) older-onset cohort, who were followed for an average of 8.3 years.35 The risk equation of this article was used to calculate 8.3-year risk of stroke in smoking and nonsmoking men and women with type 2 diabetes and characteristics as reported by WESDR.36-38 Because the blood pressure and cholesterol values taken are the mean of a large number, correction for regression dilution was incorporated. These risk estimates were then combined in a weighted average, reflecting the male-to-female mix and the prevalence of smoking in the WESDR cohort, to give an estimate of the overall risk of stroke, for comparison with the rates of fatal stroke reported by WESDR. Rates of nonfatal stroke were not reported by WESDR.

Results
Median follow-up time was 10.5 years. There were 30 700 person-years of follow-up available for model fitting, during which 52 fatal and 136 nonfatal first strokes occurred.

Table 3 shows the parameters included in the final model. Not significant were HbA1c (P=0.086), Afro-Caribbean

### TABLE 1. Characteristics of Subjects: Mean±sd, or n (Percentage)

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>n (%)</td>
<td>2671 (59%)</td>
<td>1878 (41%)</td>
</tr>
<tr>
<td>At diagnosis of diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years</td>
<td>51.5±.8</td>
<td>52.6±.8</td>
</tr>
<tr>
<td>Ethnic group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2171 (81%)</td>
<td>1583 (84%)</td>
</tr>
<tr>
<td>Afro-Caribbean</td>
<td>198 (7%)</td>
<td>153 (8%)</td>
</tr>
<tr>
<td>Asian Indian</td>
<td>302 (11%)</td>
<td>142 (8%)</td>
</tr>
<tr>
<td>Current smoker</td>
<td>908 (34%)</td>
<td>471 (25%)</td>
</tr>
<tr>
<td>AF</td>
<td>18 (0.7%)</td>
<td>10 (0.5%)</td>
</tr>
<tr>
<td>Body mass index, kg/m²</td>
<td>26.5±5.0</td>
<td>28.6±6.0</td>
</tr>
</tbody>
</table>

Mean of values taken 1 and 2 years after diagnosis of diabetes

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>HbA1c (%)</td>
<td>6.6±1.4</td>
<td>6.9±1.5</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>133±18</td>
<td>139±21</td>
</tr>
<tr>
<td>Total cholesterol, mmol/L</td>
<td>5.2±1.0</td>
<td>5.7±1.1</td>
</tr>
<tr>
<td>HDL cholesterol, mmol/L</td>
<td>1.06±0.23</td>
<td>1.18±0.26</td>
</tr>
<tr>
<td>T:H</td>
<td>5.2±1.4</td>
<td>5.1±1.5</td>
</tr>
</tbody>
</table>

AF indicates atrial fibrillation; T:H, ratio of total:HDL cholesterol.

### TABLE 2. Variables Considered for Inclusion in the Model

<table>
<thead>
<tr>
<th></th>
<th>Age in years at diagnosis of diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>1 for female; 0 for male</td>
</tr>
<tr>
<td>Afro-Caribbean</td>
<td>1 for Afro-Caribbean; 0 for white or Asian-Indian</td>
</tr>
<tr>
<td>Asian-Indian</td>
<td>1 for Asian-Indian; 0 for white or Afro-Caribbean</td>
</tr>
<tr>
<td>Smoking</td>
<td>1 for a current smoker at diagnosis of diabetes; 0 otherwise</td>
</tr>
<tr>
<td>Ex-smoking</td>
<td>1 for a former smoker at diagnosis of diabetes; 0 if a current or nonsmoker</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index (kg/m²) at diagnosis of diabetes</td>
</tr>
<tr>
<td>AF</td>
<td>Atrial fibrillation, detected by electrocardiography, at diagnosis of diabetes (1 for yes, 0 for no)</td>
</tr>
<tr>
<td>HbA1c</td>
<td>HbA1c (%), mean of values measured 1 and 2 years after diagnosis</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic blood pressure (mm Hg), mean of values measured 1 and 2 years after diagnosis</td>
</tr>
<tr>
<td>T:H</td>
<td>Mean of values measured 1 and 2 years after diagnosis</td>
</tr>
<tr>
<td>T</td>
<td>Duration of diabetes (years)</td>
</tr>
</tbody>
</table>

T:H indicates ratio of total:HDL cholesterol.
TABLE 3. Parameters Included in Model Equation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.00186</td>
<td>0.000999–0.00271</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Duration of diagnosed diabetes, per year</td>
<td>1.145</td>
<td>1.094–1.196</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age at diagnosis of diabetes, per year</td>
<td>1.092</td>
<td>1.067–1.117</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Female sex</td>
<td>0.700</td>
<td>0.486–0.914</td>
<td>0.021</td>
</tr>
<tr>
<td>Smoking at diagnosis of diabetes</td>
<td>1.547</td>
<td>1.082–2.011</td>
<td>0.0052</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>8.554</td>
<td>2.744–14.36</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Systolic blood pressure, per 10 mm Hg</td>
<td>1.122</td>
<td>1.040–1.204</td>
<td>0.0025</td>
</tr>
<tr>
<td>Lipid ratio, T:H</td>
<td>1.138</td>
<td>1.034–1.242</td>
<td>0.0085</td>
</tr>
</tbody>
</table>

T:H indicates ratio of total/HDL cholesterol.

(P > 0.5) and Asian-Indian (P > 0.5) ethnicity, and ex-smoking status (P > 0.25). BMI achieved conventional significance (P = 0.0085) but did not meet our criteria for inclusion in the model.

In a patient who has had no stroke in the first T years of diagnosed diabetes, the probability of a stroke in the next t years is

\[ R(t \mid T) = \exp[-q \times 1.145^T \times (1 - 1.145^T) / (1 - 1.145)] \]

where

\[ q = 0.00186 \times 1.092^{sex} \times 0.700^{smoking} \times 1.547^{smoking} \times 8.554^{AF} \times 1.122^{SBP-135/10} \times 1.138^{T:H-5.1} \]

When the input values are of greater or lesser stability than those used to develop the model—for example, blood pressure on a single occasion, rather than averaged over 2 occasions—the effect of regression dilution must be considered, as indicated in the Appendix.

This model uses the ratio T:H for compatibility with our existing model software. However, it was found that models of similarly good fit to the data could be built using either total cholesterol (P = 0.0044), LDL cholesterol (P = 0.0037), or the ratio LDL:HDLC cholesterol (P = 0.0013) as the only measure of dyslipidemia. Nonnested models cannot be compared by likelihood ratio tests, but graphical comparisons found these models to be very similar to the T:H ratio model (data not shown). Triglycerides were significant (P = 0.037) as the sole lipid measure, but not in models already containing LDL cholesterol, LDL:HDLC cholesterol ratio, or T:H ratio. HDL cholesterol was not significant in any model (P > 0.1).

Example

Consider a power calculation for a hypothetical trial of lipid-lowering therapy in primary prevention of stroke, among white men with type 2 diabetes. Based on data from WESDR, assume a cohort with mean SBP 147 mm Hg, total cholesterol 5.65 mmol/L, HDL cholesterol 1.11 mmol/L, age 55 years at diagnosis of diabetes, and with 12 years of diagnosed diabetes.35,36 For a nonsmoking man, with characteristics as above and assuming no AF, calculations are as follows:

\[ q = 0.00186 \times 1.092^{sex} \times 0.700^{smoking} \times 1.547^{smoking} \times 8.554^{AF} \times 1.122^{SBP-135/10} \times 1.138^{T:H-5.1} \times 0.00212 \times (1 - 1.145^{T:H}) = 0.009, \]

so that such a patient has a 6.9% probability of a stroke within the next 5 years, conditional on not dying from causes other than stroke. For a patient who smokes, the probability is 10.5%. Assuming a 30% smoking prevalence, the average 5-year risk in the population is (0.3 × 10.5%) + (0.7 × 6.9%) = 8.0%. To have 80% power to detect an effect of a therapy that lowers risk by 15%, a 5-year trial would need approximately 15,000 patients. These event rates are likely to be overestimates for the whole population, because not all patients will remain in the study for 5 years, as a result of deaths from cardiac and other causes as well as loss to follow-up. A more thorough analysis would correct also for regression dilution and account for the variation in risk factors across the population rather than use mean values.

Validation

Figure 1 shows that survival rates predicted by the model lie close to the rates observed in the UKPDS. Observed and

---

Figure 1. Observed survival rates (●), with 95% CIs (dashed line) and modeled survival rates (solid line), from year 4.

Figure 2. Modeled and observed survival rates from year 4 by SBP (a) and lipid ratio (b). a, ● indicates observed (SBP ≥ 135 mm Hg); dashed line, modeled (SBP ≥ 135 mm Hg); ○, observed (SBP > 135 mm Hg); and dotted line, modeled (SBP > 135 mm Hg). b, ■ indicates observed (lipid ratio ≥ 5.0); dashed line, modeled (lipid ratio ≥ 5.0); ●, observed (lipid ratio > 5.0); and dotted line, modeled (lipid ratio > 5.0).
modeled survival rates are shown in Figure 2a and 2b, split by SBP and by the ratio T:H, confirming the ability of the model to adjust for these risk factors.

For the WESDR cohort, in which 59 deaths from stroke were observed, the equation derived here predicts a 13.2% risk of fatal and nonfatal stroke in 8.3 years, corresponding to 197 strokes in 1370 patients. Previous studies have reported stroke fatality rates between 23% and 30% in diabetes.3,7,39 Applying a 26.5% fatality rate to our prediction of 197 strokes gives 52 fatal strokes. For comparison, the Framingham equations for mortality predict 33 fatal strokes in the same cohort.41

Discussion
To our knowledge, this is the first time a model has been developed to predict the risk of stroke specifically in a population with type 2 diabetes. Because our equations are based on a cohort identified at diagnosis of diabetes and without previous stroke, the model predicts risk of stroke as a primary event. The inputs required by the model are all readily measured in clinical practice, and routinely reported in research, for maximum applicability. The model has been incorporated in the UKPDS Risk Engine software, which is available to noncommercial users without charge from our website.42

Several limitations arise from the use of clinical trial rather than epidemiologic data, although we have attempted to correct for the impact of trial selection criteria, as described above. For time periods <4 years or >20 years after diagnosis of diabetes, or for ages >65 at diagnosis of diabetes, predictions from this model are extrapolations. Our definition of nonfatal stroke is more stringent than is sometimes used, requiring signs or symptoms to persist for more than a month. Our model does not distinguish between fatal and nonfatal strokes or between ischemic and hemorrhagic strokes. It would be ideal to test for any effect of anti-hypertensive and other therapies, but this is not practical because many patients took many different therapies over the years.

Although our derived figure, 52 fatal strokes, is lower than the 59 fatal strokes, this is to be expected because we model first stroke only. There are few other published models for the risk of stroke, and none specific to diabetes. The best-known previously published risk equations for stroke are derived from the Framingham cohort,20,41 and have been used in guidelines for clinical practice49 and in health economic modeling in the general population49 and in diabetes.21–23 The 5734 individuals in Framingham were older than in the UKPDS, (mean age 66 years) and experienced a higher event rate, 472 strokes in 10 years, but fewer than 10% had diabetes. The Framingham profile adjusts for left ventricular hypertrophy. Exclusion of the latter from the present paper gives our equations greater applicability, at the cost of some discriminatory power. The British Regional Heart Study model for risk of stroke does not adjust for presence of diabetes,50 and a recent cardiovascular risk score provides only relative risk in the case of stroke.51

Although guidelines for primary prevention often use risk of coronary heart disease as a surrogate for risk of cardiovascular disease,48,52 accurate estimation of stroke as well as coronary risk is preferable, for example, in hypertension management13 and for patients of African origin.15 The model here is suitable for use in a clinical setting.

Risk estimators also have a role to play in research and in policy formulation. The example in the Results section has demonstrated the use of the model in a power calculation, indicating that in primary prevention the large numbers needed to establish a lipid-lowering effect on stroke may require large-scale meta-analyses.53,54 The literature applying the Framingham equation to cost-effectiveness analyses in diabetes demonstrates a role for a diabetes-specific risk estimator.22,23,55 We hope that the present article will be of use to researchers, planners, and clinicians in the management of cerebrovascular risk in diabetes.

Appendix: Regression Dilution
The term regression dilution describes the behavior of parameter estimates in regression and in survival analysis when a predictor variable is subject to biovariability or assay variability.56 The risk equations reported in this article is recommended for use with estimates of blood pressure, and of total and HDL cholesterol, of similar precision to those with which the model was built, that is, the mean of 6 measurements of blood pressure, the mean of 2 measurements of total cholesterol, and the mean of 2 measurements of HDL cholesterol.

Adjustment for regression dilution is appropriate if the available data are not subject to biovariability, for example, if the blood pressure and cholesterol levels are the means over very large populations. Based on a maximum likelihood method,57 the following corrections are appropriate. For blood pressure, replace the parameter estimate 1.122 with 1.122−1.25. For lipid ratio, replace the parameter estimate 1.138 with 1.138−1.181. Because variables are centered around their means in the model equation (for example, blood pressure is used as SBP−135.5), no adjustment is necessary to the intercept parameter.

We use a method reported previously58 to estimate the adjustment necessary if the input data are of greater biovariability than ours. If the available blood pressure is the mean of just 3 measurements, replace 1.122 with 1.101. If blood pressure is a single measurement, replace 1.122 with 1.060. If the lipid ratio is based on single measurements of HDL and total cholesterol, replace the parameter 1.138 with 1.111.
Acknowledgments

The major grants for this study were from the United Kingdom Medical Research Council; British Diabetic Association; UK Department of Health; National Eye Institute and National Institute of Digestive, Diabetes, and Kidney Disease of the National Institutes of Health; British Heart Foundation; Novo Nordisk; Bayer; Bristol-Myers Squibb; Hoechst; Lilly; Lipta; and Farmalita Carlo Erba. R.J.S. was supported by Wellcome Trust Fellowship 054470/Z/98/Z/DG/NO/SH.

The cooperation of the patients and staff of the UKPDS is much appreciated. The participating centers in the UKPDS were Radcliffe Infirmary, Oxford; Royal Infirmary, Aberdeen; University Hospital, Birmingham; St. George’s Hospital and Hammersmith Hospital, London; Manchester City Hospital, Belfast; North Staffordshire Royal Infirmary, Stoke-on-Trent; Royal Victoria Hospital, Belfast; St. Helier Hospital, Carshalton; Whittington Hospital, London; Norfolk & 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 & Exeter Hospital. We are grateful to Prof Peter Bennet and to Dr Andrew Farmer for helpful comments.

References


UKPDS 60: Risk of Stroke in Type 2 Diabetes Estimated by the UK Prospective Diabetes Study Risk Engine

Viti Kothari, Richard J. Stevens, Amanda I. Adler, Irene M. Stratton, Susan E. Manley, H. Andrew Neil and Rudy R. Holman

for the UK Prospective Diabetes Study Group

Stroke. 2002;33:1776-1781
doi: 10.1161/01.STR.0000020091.07144.C7

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2002 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/33/7/1776