Quantifying the risk of type 2 diabetes in East London using the QDScore: a cross-sectional analysis

INTRODUCTION
The prevalence of diabetes is increasing rapidly, with current projections estimating a global increase in prevalence of 50.7% between 2011 and 2030.1 The growing burden of diabetes, alongside other non-communicable diseases such as cardiovascular disease, respiratory disease, and cancers, has been met with a call from the United Nations General Assembly to strengthen population-wide interventions, primary care services, and disease monitoring across the whole population.2 Currently in England, approximately 3.1 million people aged 16 years and above have diabetes. This figure is expected to increase to 4.6 million by 2030, as a result of ageing, changes in the ethnic composition of the population, and rising obesity levels.3 Social and ethnic differences are prominent. For example, South Asian and black groups have a rate of type 2 diabetes that is six- and threefold greater than the white population respectively,4,5 and those in the most affluent quintile of the population (<55 years) have half the prevalence of those in the most deprived quintile.6

There has been much interest in predicting the risk of developing type 2 diabetes and other non-communicable diseases, with the aim of preventing disease and reducing the burden on the health service.6,7 A recent systematic review identified 145 different models and scores for predicting the risk of developing type 2 diabetes and highlighted seven, suitable for routine use, that had a balanced combination of statistical accuracy and usability.8 Some were suitable for use with population-wide electronic primary care records. One of these, the QDScore5,9 also had high adaptability in socially complex populations, as its variables take account of ethnic diversity and socioeconomic deprivation. It is available online, as an integral part of EMIS Web (Egton Medical Information System),10 is open source (www.qdscore.org/), and has recently been renamed QDiabetes.

The aim of this cross-sectional study was to use the QDScore on an entire set of electronic primary care records in three deprived and ethnically-diverse inner-city boroughs, to generate population estimates of risk to inform possible interventions for targeting subgroups at high risk.

METHOD
The research was undertaken on 519 288 electronic primary care records across 135 out of 145 general practices in the boroughs of Tower Hamlets, Newham, and City and Hackney, in London, UK. The methods have been previously described.11,12 In summary, a cross-sectional analysis was undertaken in August 2011 of all adults aged 25–79 years without diabetes. Individuals with a Read Code for diabetes were excluded. Anonymised data were extracted from electronic health records in general practices using EMIS Web, via NHS N3 networks, and securely held by the Clinical Effectiveness Group at Queen Mary, University of London. The 10 clinical...
variables needed to calculate the QDScore were extracted from the records: age, sex, ethnicity, Townsend score of deprivation, family history of diabetes, personal history of cardiovascular disease, smoking status, hypertension treated with pharmacological therapy, current corticosteroid use, and body mass index (BMI). The additional clinical variables: QRisk score (www.qrisk.org — in those aged >30 years), diagnosed hypertension without pharmacological therapy, and estimated glomerular filtration rate (eGFR) were also extracted.

The QDScore, which undergoes regular updates, was supplied as an electronic batch processor by the authors (J Hippisley-Cox, personal communication, 2011) and used to calculate the 10-year risk of developing type 2 diabetes. Basic descriptive statistics were compiled using Stata (version 10). The risk of type 2 diabetes was categorised as low (0–9.99% risk at 10 years), medium (10–19.99%), or high (≥20%). Subgroup analyses using descriptive statistics, with variables used to calculate the QDScore (age, sex, ethnicity, deprivation, personal history of cardiovascular disease, hypertension treated with pharmacological therapy, and BMI), and the additional clinical variables not included in the score, were undertaken. These variables were selected after discussion among the research team, local GPs, and public health specialists, as to which would be most useful to inform commissioning of public health interventions.

Subgroups with higher risk were not unexpected, as certain predictor variables within the QDScore lead to a higher score (for example, certain ethnicities). Therefore, statistical collinearity between the outcome (% risk) and the predictor variables rendered tests of significance misleading. Additionally, geospatial maps were created to further assess the usefulness for commissioning.

How this fits in
Individual risk scores can be aggregated and used to map overall risk in a population. Identifying populations at high risk of chronic disease could improve targeting of risk-reduction interventions. Accurate description of subgroups at risk has a high potential to inform commissioning. The methodology described here could help support further research into interventions aimed at reducing the risk of chronic diseases.

<table>
<thead>
<tr>
<th>Table 1. Flow of data through the study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
</tr>
<tr>
<td>Registered with a general practice</td>
</tr>
<tr>
<td>Aged 25–79 years</td>
</tr>
<tr>
<td>Free from known diabetes</td>
</tr>
<tr>
<td>Data available for analysis</td>
</tr>
</tbody>
</table>
QRisk was closely associated with high risk of type 2 diabetes. For QRisk 0–9, 9.7% (n = 15 516) were at high risk for type 2 diabetes, compared to 31.1% (n = 12 487) for QRisk 10–19, and 47.7% (n = 9839) for QRisk ≥20 (Figure 2).

Similarly, vascular comorbidity, eGFR (<60 ml/min/1.73 m^2), chronic kidney disease stage >3), and increasing BMI all increased the chance of being high risk for developing type 2 diabetes (Table 3). QRisk, eGFR and diagnosed hypertension without pharmacological therapy were not included as predictor variables in the QDScore. All other variables presented were also predictor variables; therefore, the results are not unexpected.

Figure 3 shows the proportion of people at the level of an individual general practice at high risk of developing type 2 diabetes within 10 years. The prevalence of high diabetes risk varied from 4.1% to 16.7% across the region. The ‘East London diabetes belt’ stretches from Tower Hamlets in the west, with a large Bangladeshi population, to north-east Newham, where there is a high proportion of South Asian and black ethnic groups. Affluent riverside properties in the south, and parkside residences in the north, show low levels of risk for type 2 diabetes.

**DISCUSSION**

**Summary**

This research successfully used a risk-prediction model on half a million electronic records to describe the socially patterned risk of developing type 2 diabetes for an entire inner-city population. Overall risk was high, with 10% of all adults aged 25–79 years at high risk of developing type 2 diabetes; this is in addition to 7% of the population who are already known to have type 2 diabetes. Male sex, increasing age, South Asian and black ethnicity, deprivation, cardiovascular comorbidity, and obesity were all associated with increased risk. The overlap with QRisk was extensive. The extent of these findings locally was quantified. Although the results are not surprising, as QDScore assigns

<table>
<thead>
<tr>
<th>10-year risk of developing type 2 diabetes (%)</th>
<th>Number of people in category</th>
<th>% of sample</th>
<th>Sex, % male</th>
<th>Median score</th>
<th>Median age, years</th>
<th>White % (n = 214 542)</th>
<th>South Asian % (n = 135 000)</th>
<th>Black % (n = 82 036)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–9.9 (low)</td>
<td>410 801</td>
<td>79.1</td>
<td>53.0</td>
<td>1.8</td>
<td>34</td>
<td>83.2</td>
<td>69.4</td>
<td>72.4</td>
</tr>
<tr>
<td>10–19.9 (medium)</td>
<td>57 426</td>
<td>11.1</td>
<td>55.4</td>
<td>13.8</td>
<td>49</td>
<td>9.4</td>
<td>14.2</td>
<td>15.3</td>
</tr>
<tr>
<td>20–100 (high)</td>
<td>51 061</td>
<td>9.8</td>
<td>56.4</td>
<td>30.9</td>
<td>54</td>
<td>7.5</td>
<td>16.4</td>
<td>12.3</td>
</tr>
</tbody>
</table>

**Table 2. Proportion of individuals at low, high, and medium risk of developing type 2 diabetes over the next 10 years**
higher values to known risk factors, detailed population subgroup analyses have high potential to inform targeted interventions. Mapping risk revealed the ‘East London diabetes belt’ which, like the ‘diabetes belt’ in the southern states of the US, corresponds to high prevalence of known risk factors in the local population and has high potential to inform the work of commissioners locally and nationally, with a view to taking action to reduce the incidence of type 2 diabetes through locality-based interventions, particularly in regard to the ‘Olympic legacy, which holds the promotion of healthier lifestyles and the reduction of premature causes of death as key actions to be implemented by 2015. The predicted positive impacts on individual health will likely decrease diabetes risk across the east London population. In some areas, almost one in six adults fell into the ‘high-risk’ category.

Strengths and limitations
The study dataset had high completeness due to previous investment and long-standing supportive collaboration between the university, general practices, and the primary care trusts. This study demonstrates the type of population statistics that can be generated using a risk-prediction model on electronic records, and the rich level of detail that subgroup analysis can generate. For example, knowing the proportion of an ethnic group at high risk has the potential to inform targeted preventive measures and social marketing.

The QDScore has only been validated to estimate the risk of diabetes for individuals aged 25–79 years. A large proportion of those registered with a GP (\(n=313,053; 35.5\%\)) was outside this age range, reflecting the young population. Type 2 diabetes is increasingly common in younger age groups, making this a weakness of the

![Figure 3. Proportion of people at high risk of type 2 diabetes (≥20%) by general practice: The ‘East London diabetes belt’.](image)

Table 3. Proportion of people at high risk in various comorbidity groups

<table>
<thead>
<tr>
<th>Comorbidity Group</th>
<th>Number of people at high risk</th>
<th>Number of people in category</th>
<th>Proportion at high risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVD (IHD/stroke/TIA)</td>
<td>5637</td>
<td>9864</td>
<td>57.1</td>
</tr>
<tr>
<td>BMI ≥30 kg/m²</td>
<td>32,564</td>
<td>76,162</td>
<td>42.8</td>
</tr>
<tr>
<td>Hypertension (diagnosed ± pharmacological therapy)</td>
<td>23,102</td>
<td>48,169</td>
<td>48.0</td>
</tr>
<tr>
<td>eGFR &lt;60 ml/min/1.73 m²</td>
<td>2905</td>
<td>7026</td>
<td>41.3</td>
</tr>
<tr>
<td>QRisk ≥20</td>
<td>9839</td>
<td>20,629</td>
<td>47.7</td>
</tr>
</tbody>
</table>

CVD = cardiovascular disease. eGFR = estimated glomerular filtration rate. IHD = ischaemic heart disease. TIA = transient ischaemic attack. *RRS of 9864 had the diagnosis included in the QDScore calculator, as nine codes were added later for subgroup analysis which should have been included in the original score calculation. **37.3% had treated hypertension with pharmacological therapy and this diagnosis was included in the QDScore calculation.
study.

collinearity between outcome and predictor variables prevented tests of statistical significance, but as the principle purpose of the descriptive statistics is for service planning, and trends were very clear (for example, association with ethnicity), this is unlikely to impact on the routine use of such data.

The QDScore is also available for incorporation into the routine electronic primary care record system, which is informative for individual patient consultations. The use of the score for individual patient management was outwith the scope of this population-level research. It was also not the intention to validate the score or assess the impact of components on risk, as this has already been undertaken in prospective independent samples.2–6

Comparison with existing literature

While many studies describing risk models and scores exist,7–9,16 relatively few validated scores have described 'real-world' applications as in this study. The Finnish Type 2 Diabetes Risk Score (FINDRISC) was used in a survey of 400 adults aged 20–73 years in Libya; approximately 12% were at high or very high risk of developing diabetes over 10 years.10 Both FINDRISC and the Indian Diabetes Risk Score were used on 198 migrants in Norway identified through mosques, Norwegian classes, and directly in shops and on the street; with FINDRISC, 29% were at great/extreme risk (the two highest categories) over 10 years.20 However, these studies were small and exploratory, and were focused mainly on identifying individuals rather than managing risk in entire populations.

Despite the emerging ability to quantify diabetes risk, reliable information on evaluated and effective interventions for those at high risk is lacking. The evidence base for public health interventions that reliably reduce the risk (and therefore incidence) of type 2 diabetes is currently mainly limited to those with abnormal glycaemic indices. People with impaired glucose tolerance can reduce their risk of developing type 2 diabetes via lifestyle interventions that promote dietary improvement and more physical activity, and by pharmacological therapy.21 At the population level for individuals without diabetes (who may or may not have abnormal glycaemic indices), research indicates that the more behavioural goals that can be attained over time (controlling weight, diet, and physical activity), the lower the incidence of type 2 diabetes in the long term.22 The effectiveness of population-level strategies for reducing the risk of type 2 diabetes could benefit from further study and evaluation.

Implications for practice and research

In May 2011, the UK National Institute for Health and Clinical Excellence [NICE] produced guidance on population and community interventions aimed at preventing diabetes.3

Several generic principles relevant to many public health areas underpin the NICE guidance for populations and communities:4

- behaviour change through education;
- supportive personal health plans;
- weight management through healthier eating (for example 'five a day') and interventions aimed at weight reduction that are measured, specific, and individual;
- physical activity including 30 minutes 5 days a week; and
- cultural sensitivity to ensure that interventions take account of language, educational barriers, religion, and cultural norms.

Eleven specific recommendations are made. Many are expressed in generic terms, covering strategy, health promotion, education, physical activity, healthy eating, and targeted prevention.

NICE concluded that less expensive population-wide interventions have to be combined with effective individual interventions targeted at those at high risk.4 This approach is consistent with the Marmot review's recommendation of proportionate universalism: that is, in order to tackle health inequalities across all of society, public health action should be appropriate for everyone, but proportionally more for those whose need is highest.23

This study offers an approach that could be used to help achieve this dual objective. Black and minority ethnic groups and people from deprived areas are known to be at much greater risk of type 2 diabetes, and the presence of pre-existing cardiovascular disease puts such people in ‘triple jeopardy’. Analysis of electronic records allows health planners to identify ‘locality hotspots’, specific high-risk individuals, and the level of risk across the entire population, with a view to achieving proportionate universalism in a chosen disease area.

Although beyond the findings in this paper, it is of interest that a second set
The recommendations in this report consider two broad areas: identification of high-risk individuals; and individual interventions to reduce risk. The former, using a validated algorithm such as the QDScore, could be incorporated into the English NHS Health Check programme (http://www.healthcheck.nhs.uk/). Using a validated tool such as the QDScore could result in more accurate identification of high-risk individuals for further follow-up and interventions. The use of electronic ‘batch processors’ to calculate population estimates of risk at practice or small area level, for use by providers and commissioners, thus complements individual risk ascertainment by informing neighbourhood and practice risk profiles.

This study has demonstrated the significant overlap between being at high risk of diabetes and being at high risk of cardiovascular disease. For example, of those with a QRisk score ≥20 (meaning ≥20% risk of developing cardiovascular disease in 10 years), 48% were also at high risk (≥20%) for developing type 2 diabetes. This underlines a potential need to combine preventive interventions for these common conditions with overlapping risk factors, pending further study.

At local level, risk-scoring systems can illustrate detailed population-level risk distributions, including high-risk groups, and have the potential to guide targeted interventions for high-risk neighbourhoods and ethnic groups. These descriptive findings have high potential for informing local populations, commissioners, and also future studies of population-level interventions.
REFERENCES


Appendix 1. Data supplement

The QDScore electronic batch processor (risk engine v. 1.0.0.1; 2009.05.06) was used. Missing values for certain variables are handled by the QDScore by assigning the risk associated with an imputed or default value: ethnicity defaults to white; smoking status defaults to non-smoker; Townsend score of deprivation defaults to a score of 0 (the national average); BMI is substituted with an imputed value based on a prediction algorithm using age, sex, ethnicity, smoking status, treated hypertension, and cardiovascular disease. BMI values <15 or >54 were substituted with these values respectively. Other values only contributed to the risk score if positive.

Completeness of variables that were routinely collected on the general practice electronic records (n = 519,288) was as follows: age (100%), sex (100%), ethnicity (91.6%), Townsend deprivation score (99.8%), BMI (76.5%), and smoking status (96.4%). Other predictor variables were only recorded if positive: family history of diabetes (22.9%), personal history of cardiovascular disease (1.9%), treated hypertension with pharmacological therapy (7.2%), and current corticosteroid usage (1.0%). Additional variables were also only recorded if positive: diagnosed hypertension without pharmacological therapy (2.1%), and estimated glomerular filtration rate <60 (1.4%). QRisk was recorded for 53.8% of individuals >30 years (n = 410,874), and 24.2% of the sample were smokers. Of the predictor variables with missing values where a substituted value was used: ethnicity defaulted to white (10.7%); smoking was classified as smoker (3.6%); Townsend score defaulted to 0 (0.2%); BMI used an imputed value (23.5%; ranging from BMI 21.8–31.4); BMI of 15 or 54 was substituted for values above (0.14%) or below (0.08%) these thresholds.

In this dataset missing smoking status (n = 18,550; 3.6%) was classified as smoker rather than non-smoker. Replacing the missing status with non-smoker had no substantial effect on our results. Category of risk remained the same to one decimal place (low 79.1%; medium 11.1%; high 9.8%). Median risk for the entire dataset remained the same at 2.84. Effect on imputed BMIs was also very small (mean BMI increased from 26.21 to 26.25, and proportion at high risk with a BMI ≥30 decreased from 42.76% to 42.74%).

\( ^{a} \)In this dataset missing smoking status (n = 18,550; 3.6%) was classified as smoker rather than non-smoker. Replacing the missing status with non-smoker had no substantial effect on our results. Category of risk remained the same to one decimal place (low 79.1%; medium 11.1%; high 9.8%). Median risk for the entire dataset remained the same at 2.84. Effect on imputed BMIs was also very small (mean BMI increased from 26.21 to 26.25, and proportion at high risk with a BMI ≥30 decreased from 42.76% to 42.74%). \(^{b} \)Includes not stated, refused or unusable code.