Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm

**Abstract**

**Background**
Pancreatic cancer has the worst survival for any cancer and is often diagnosed late when the cancer is advanced. Chances of survival are more likely if patients can be diagnosed earlier.

**Aim**
To derive and validate an algorithm to estimate absolute risk of having pancreatic cancer in patients with and without symptoms in primary care.

**Design and setting**
Cohort study using data from 375 UK QResearch® general practices for development and 189 for validation.

**Method**
Included patients were aged 30–84 years, free at baseline from a diagnosis of pancreatic cancer and had not had dysphagia, abdominal pain, abdominal distension, appetite loss, or weight loss recorded in the preceding 12 months. The primary outcome was incident diagnosis of pancreatic cancer recorded in the following 2 years. Risk factors examined included: age, body mass index, smoking status, alcohol, deprivation, diabetes, pancreatitis, previous diagnosis of cancer apart from pancreatic cancer, dysphagia, abdominal pain, abdominal distension, appetite loss, weight loss, diarrhoea, constipation, tiredness, itching, and anaemia. Cox proportional hazards models were used to develop separate risk equations in males and females. Measures of calibration and discrimination assessed performance in the validation cohort.

**Results**
There were a total of 1415 incident cases of pancreatic cancer from 4.1 million person-years in the derivation cohort. Independent predictors in both males and females were age, smoking, type 2 diabetes, chronic pancreatitis, abdominal pain, appetite loss, and weight loss. Abdominal distension was a predictor for females only; dysphagia and constipation were predictors for males only. On validation, the algorithms explained 59% of the variation in females and 62% in males. The receiver operating characteristic statistics were 0.84 (females) and 0.87 (males). The D statistic was 2.44 (females) and 2.61 (males). The 10% of patients with the highest predicted risks contained 62% of all pancreatic cancers diagnosed over the following 2 years.

**Conclusion**
The algorithm has good discrimination and calibration and could potentially be used to help identify those at highest risk of pancreatic cancer to facilitate early referral and investigation.

**Keywords**
diagnosis, pancreatic cancer, primary care, QResearch, risk prediction, symptoms.

**INTRODUCTION**
Pancreatic cancer is the 11th most common cancer in the UK.1 Diagnoses are often made late when the cancer is advanced.2 Less than 20% of patients are suitable for surgery and 84% of patients are likely to have died within a year of diagnosis,3 giving the worst survival rate for any cancer.4 However, the chance of survival is more likely if patients present at an early stage.1

There are only a few established risk factors for pancreatic cancer such as age,2,5 smoking,2,3 genetic factors,2 chronic pancreatitis,1 and alcohol.1,4 Diabetes may be a risk factor for pancreatic cancer or an early manifestation of a growing tumour.5,6 As there are few established risk factors and currently no reliable screening test, it is unlikely that there will be a national screening programme for pancreatic cancer; as such, it is likely that most pancreatic cancers will be diagnosed in patients who are symptomatic and presenting to primary care. The challenge is ensuring earlier diagnosis to help improve treatment options (for example, possibility of surgery) and prognosis. Earlier diagnosis could be helped by increased public awareness of symptoms that might indicate pancreatic cancer, such as weight loss, loss of appetite, and abdominal pain.2,4,5 Diagnosis could also be improved by more prompt investigation of patients who are symptomatic and present to their GP.6 In the UK, GPs will soon have better direct access to diagnostic investigations such as ultrasound, computerised tomography (CT) scanning, and magnetic resonance imaging (MRI) but they need better assessment tools to quantify a patient’s risk of different types of cancer and thereby ensure the right patients are sent for the right investigations. This would also make efficient use of scarce resources.6

Surprisingly, pancreatic cancer is not mentioned in current guidelines from the National Institute for Health and Clinical Excellence (NICE) on the referral of patients with suspected cancer.6 This might be because relatively little is known about the aetiology of pancreatic cancer or its attendant presenting features. Therefore, the pattern of presenting symptoms for patients with pancreatic cancer was investigated with a view to developing an algorithm to quantify the risk of a patient having pancreatic cancer, this algorithm would incorporate both symptoms and baseline risk factors such as age, chronic pancreatitis, and smoking.

QResearch® primary care database was used to develop the risk prediction model as it contains robust data on many of the relevant exposures and outcomes. It is also representative of the population where such a model is likely to be used and has been used successfully to develop and validate a range of prediction models for use in primary care.9–12 Once validated, the models could be integrated into clinical computer systems to help systematically identify those at high risk and alert clinicians to those who might benefit most from further assessment or interventions.9–12 The algorithm could also be made available on
How this fits in
Pancreatic cancer is often diagnosed late when the cancer is advanced and the chance of survival is, therefore, poor. There is no reliable screening test so most diagnoses are likely to be made in patients who are symptomatic, as such, in order to make diagnoses earlier, there needs to be an increased awareness of symptoms among patients and earlier investigation of patients who are symptomatic by GPs. This study has developed a new algorithm which predicts the chances of having pancreatic cancer based on a combination of symptoms and baseline risk factors such as age, chronic pancreatitis, smoking, and diabetes. The algorithm performed well in an independent sample, both in terms of discrimination and calibration. The sensitivity was high — for example, if a threshold equivalent to the top 10% of patients with the highest risk is selected, this will account for 62% of all cases of pancreatic cancers occurring within the subsequent 2 years.

METHOD
Study design and data source
A prospective cohort study was carried out in a large population of primary care patients from an open cohort study using the QR esearch database (version 30). All practices in England and Wales who had been using their EMIS [Egton Medical Information Systems] computer system for at least a year were included. Two-thirds of practices were randomly allocated to the derivation dataset and the remaining one-third to a validation dataset. An open cohort of patients aged 30–84 years was identified, drawn from patients registered with practices between 1 January 2000 and 30 September 2010. The following were excluded: patients without a postcode-related Townsend score, those with a history of pancreatic cancer at baseline, and those with a recorded red flag symptom in the 12 months prior to the study entry date. For this study, a red-flag symptom was defined as one that might alarm the patient and indicate the presence of pancreatic cancer, that is, symptoms of dysphagia, loss of appetite, weight loss, abdominal pain, or abdominal distension. Jaundice was not included as this is relatively rare, usually considered a sign, and would have its own pathway for investigation. Patients entered the cohort on the latest of the study start date (1 January 2000) and 12 months after the patient registered with the practice; this ensured that all patients had a minimum of 12 months registration prior to study entry. For patients with incident dysphagia, appetite loss, weight loss, abdominal pain, or abdominal distension, the entry date was the date of first recorded onset within the study period.

Clinical outcome definition
The study outcome was pancreatic cancer, which was defined as incident diagnosis of pancreatic cancer during the 2 years following the entry date recorded on either the patient’s GP record using the relevant UK diagnostic Read Codes, or their linked cause of death record, according to the Office for National Statistics (ONS), using the relevant diagnostic codes from the International Classification of Diseases-9 (ICD-9) (157) or ICD-10 (C25). A 2-year period was used as this represents the period of time during which existing cancers are likely to become clinically manifest. Patients were censored at the earliest date of either diagnosis of pancreatic cancer, date of death, date of leaving the practice, or 2 years after their study entry date.

Predictor variables
Established predictor variables were examined, focusing on those that are likely to be recorded in the patient’s electronic record and that the patient is likely to know. Symptoms that might herald a diagnosis of pancreatic cancer were also included. Separate analyses were carried out in males and females. The following predictor variables were examined, using information recorded prior to study entry:

- currently consulting GP with first onset of dysphagia (yes/no);
- currently consulting GP with first onset of loss of appetite (yes/no);
- currently consulting GP with first onset of weight loss symptom (yes/no);
- currently consulting GP with first onset of abdominal pain (yes/no);
- currently consulting GP with first onset of abdominal distension (yes/no);
- recently consulted a GP with first onset of any of:
\[ e^{-0.3 (3.4)} -0.2 (3.6) \]

\[ 50.1(15.0) 50.1(14.9) \]

\[ 185 889 (50.2) 62 4 352 (50.2) \]

\[ 1178 682 (49.8) 629 932 (49.8) \]

\[ 26.4 (4.6) 26.4 (4.7) \]

\[ 877 243 (79.4) 1 009 931 (81.2) \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 495 561 (21.0) 258 963 (20.8) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 518 328 (21.9) 256 390 (20.6) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]

\[ 2 36 4 5 71 \]

\[ 518 328 (21.9) 256 390 (20.6) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]

\[ 2 36 4 5 71 \]

\[ 518 328 (21.9) 256 390 (20.6) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]

\[ 2 36 4 5 71 \]

\[ 518 328 (21.9) 256 390 (20.6) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]

\[ 2 36 4 5 71 \]

\[ 518 328 (21.9) 256 390 (20.6) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]

\[ 2 36 4 5 71 \]

\[ 518 328 (21.9) 256 390 (20.6) \]

\[ 1 234 740 \]

\[ 2 36 4 5 71 \]

\[ 1 20 0 385 (50.8) 627 868 (50.5) \]

\[ 426 697 (18.0) 228 970 (18.4) \]

\[ 71 668 (3.0) 39 443 (3.2) \]

\[ 449 044 (6.3) 80 402 (6.5) \]

\[ 189 887 (7.4) 96 443 (7.8) \]

\[ 135 113 (5.7) 74 140 (6.0) \]

\[ 200 777 (8.5) 96 479 (7.8) \]
suitable for use with survival data where higher values indicate better discrimination, and the area under the receiver operating characteristic (ROC) curve (ROC curve statistic) at 2 years were calculated. Calibration was assessed by comparing the mean predicted risks at 2 years with the observed risk by tenth of predicted risk. The observed risk was obtained using the Kaplan–Meier estimate evaluated at 2 years. The validation cohort was used to define the thresholds for the 0.1%, 0.5%, 1%, 5%, and 10% of patients at highest estimated risk of pancreatic cancer at 2 years. Sensitivity, specificity, and positive and negative predictive values were calculated using these thresholds, restricting the analyses to patients who had the outcome within 2 years or had at least 2 years of follow-up. All available data on the database were used to maximise the power and generalisability of the results. Stata (version 11) was used for all analyses.

RESULTS

Overall study population

Overall, 564 QR research practices in England and Wales met the inclusion criteria, of which 375 were randomly assigned to the derivation dataset, with the remainder assigned to a validation cohort. A total of 2,538,615 patients aged 30–84 years were identified in the derivation cohort, and 124,458 (4.9%) patients without a recorded Townsend deprivation score were excluded; 161 (0.01%) patients with a history of pancreatic cancer were also excluded, as well as a further 49,425 (1.9%) patients with at least one red-flag symptom recorded in the 12 months prior to entry to the study, leaving 2,364,571 patients for analysis.

A total of 1,342,329 patients aged 30–84 years were identified in the validation cohort; of these, 70,847 (5.3%) patients without a recorded Townsend score were excluded, as well as 96 (0.01%) with a history of pancreatic cancer, and 27,646 (2.1%) with at least one red-flag symptom recorded in the 12 months prior to study entry, leaving 1,243,740 patients for analysis.

The baseline characteristics of each cohort were very similar, as shown in Table 1. As in another study, the patterns of missing data supported the use of multiple imputation to replace missing values for smoking status, alcohol intake, and body mass index (not shown, available from the authors).

Incidence rates for red-flag symptoms

Overall, in the derivation cohort, 15,648 patients with dysphagia were identified, 10,351 with appetite loss, 2,623 with weight loss, 2,325 with abdominal pain, and 7985 with abdominal distension. Figure 1 shows the age–sex incidence rates of each symptom. The incidence rates for abdominal distension, dysphagia, appetite loss, and weight loss were similar in males and females and increased steeply with age. Abdominal pain was more common in females; it tended to decrease with age in females but increase with age in males.

Incidence rates of pancreatic cancer

Overall in the derivation cohort, during the 2-year follow-up period, a total of 1415 incident cases of pancreatic cancer arising from...
4,169,461 person years of observation were identified, giving a crude rate of 34 cases per 100,000 person years. There were 1080 cases (76.3 % of 1415) identified using the GP record and an additional 335 (23.7%) identified from the linked death record. The incidence rates increased with age and tended to be higher in males than females [data not shown].

In the validation cohort, 781 incident cases of pancreatic cancer were identified arising from 2,184,336 person years of observation, giving a rate of 36 per 100,000 person years. There were 612 cases (78.4%) identified using the GP record and an additional 169 (21.6%) from the linked death record.

### Predictor variables

Table 2 shows the predictor variables selected for the final models for females and males. Independent predictors in both males and females were age, smoking status, type 2 diabetes, chronic pancreatitis, abdominal pain, appetite loss, weight loss. Abdominal distension was a predictor for females only; dysphagia and constipation were predictors for males only. The other variables examined were not independent risk factors so were not included in the final models.

Risk of pancreatic cancer in females was significantly associated with increasing age as shown in Figure 2. Risk also increased with the amount smoked. For example, compared with non-smokers, the risks were increased by 2.4-fold for heavy smokers and 1.5-fold for light smokers (Table 2). The risks were also elevated in females with type 2 diabetes (2.1-fold higher), chronic pancreatitis (3.2-fold higher), abdominal pain (4.1-fold higher), abdominal distension (3.0-fold higher), appetite loss (3.9-fold higher), and weight loss (3.3-fold higher) (Table 2).

The magnitudes of the hazard ratios in males were generally similar to those for females, as shown in Table 2), but the relative increase in risk with increasing age was less steep (Figure 2). The risks were also elevated with appetite loss (2.5-fold higher), abdominal pain (5.2-fold higher), dysphagia (2.6-fold higher), and constipation (1.9-fold higher). There was also a significant interaction between weight loss and age in males, as shown in Figure 3. Unlike in females, abdominal distension was not a significant predictor in males.

### Validation

The validation statistics (Table 3) showed that the risk-prediction equations explained 59% of the variation in time to diagnosis in females and 62% of the variation in males. The D statistic was 2.44 for females and 2.61 for males. The area under the ROC statistics were 0.84 for females and 0.87 for males.

Figure 4 shows the mean predicted scores and the observed risks at 2 years within each tenth of predicted risk, in order to assess the calibration of the model in the validation cohort. Overall, the model was well calibrated. There was close correspondence between predicted and observed 2-year risks within each model tenth for males and females, with a small degree of over-prediction in the highest tenth in males.

### Individual risk assessment and thresholds

One potential use for this algorithm is within consultations with individual patients particularly if they present with new onset of an alarm symptom such as dysphagia, abdominal pain, weight loss, or appetite loss. Some clinical examples are shown in Box 1. The results could help inform the decision to undertake further investigations such as abdominal ultrasound, MRI, CT scan, or endoscopic retrograde cholangiopancreatography.

The algorithm could also be used for systematic risk stratification for a population of patients aged 30–84 years. Software implementing the algorithm could be integrated into GP computer systems to calculate the risk of a patient having an existing but as yet undiagnosed pancreatic cancer based on information already recorded in the patient’s electronic health record. Patients at highest risk could be identified and recalled for a clinical assessment.

The 90th centile defined a high-risk group with a 2-year risk score of >0.2% (Table 4). There were 487 new cases of pancreatic cancer within this group out of 781 new cases identified in the validation cohort, which accounted for 62.4% of all new cases of pancreatic cancer (sensitivity). The positive predictive value (PPV) with this threshold was 0.6%. Alternatively, using a threshold based on the top 5% of risk (that is, a 2-year risk score >0.3%) had a sensitivity of 45.3% and a PPV of 0.9%. The sensitivity of an approach based on single symptoms ranged from 1.2% for abdominal distension to 39.8% for abdominal pain; 18.6% of cases of pancreatic cancer had a diagnosis of type 2 diabetes at baseline.

### DISCUSSION

**Summary**

This research has developed and validated a new algorithm designed to quantify the...
The absolute risk of having pancreatic cancer, which is either currently present or likely to become manifest within 2 years. To the authors’ knowledge, this is the first algorithm of its kind and is based on age, smoking status, alcohol status, chronic pancreatitis, type 2 diabetes, loss of appetite, weight loss, abdominal pain, abdominal distension, dysphagia, and constipation. The algorithm is based on simple clinical variables which can be ascertained in clinical practice. The algorithm performed well in a separate validation sample with good discrimination and calibration.

Strengths and limitations

Key strengths of the study include size, duration of follow-up, representativeness, and lack of selection, recall, and responder bias. UK general practices have good levels of accuracy and completeness in recording clinical diagnoses.\(^26\) The authors consider that the study has good face validity since it has been conducted in the setting where the majority of patients in the UK are assessed, treated, and followed-up. The algorithm has good face validity as it confirms the significance of established risk factors such as age, smoking, chronic pancreatitis, type 2 diabetes, and associated symptoms such as weight loss, appetite loss, and abdominal pain.

Limitations of the study include lack of formally-adjudicated outcomes, information bias, and missing data. Not all patients with symptoms will attend their GP, and in those who do, not all symptoms will be reported or recorded.

The database has linked cause of death from the UK ONS and the study is therefore likely to have picked up the majority of cases of pancreatic cancer, thereby minimising ascertainment bias. The incidence rate in this study’s population is close to European estimates suggesting good ascertainment of cases.\(^27\) It is higher than rates reported in UK cancer registry reports which probably reflects the improved ascertainment resulting from including of diagnoses recorded in either the primary care or the cause of death record. Patients who die of pancreatic cancer will be included on the linked cause of death data. Patients diagnosed with pancreatic cancer in hospital will have the information recorded in hospital discharge letters which are sent to the GP and then entered into the patient’s electronic record.

Comparison with existing literature

While the study was reliant on accuracy of information recorded by primary care

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Risk threshold</th>
<th>True negative</th>
<th>False negative</th>
<th>False positive</th>
<th>True positive</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>Positive predictive value, %</th>
<th>Negative predictive value, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 2 diabetes</td>
<td>n/a</td>
<td>940 800</td>
<td>638</td>
<td>30 125</td>
<td>145</td>
<td>18.6</td>
<td>96.9</td>
<td>0.5</td>
<td>99.9</td>
</tr>
<tr>
<td>Chronic pancreatitis</td>
<td>n/a</td>
<td>970 094</td>
<td>777</td>
<td>831</td>
<td>4</td>
<td>0.5</td>
<td>99.9</td>
<td>0.5</td>
<td>99.9</td>
</tr>
<tr>
<td>Abdominal pain</td>
<td>n/a</td>
<td>877 133</td>
<td>470</td>
<td>93 792</td>
<td>311</td>
<td>39.8</td>
<td>90.3</td>
<td>0.3</td>
<td>99.9</td>
</tr>
<tr>
<td>Abdominal distension</td>
<td>n/a</td>
<td>967 478</td>
<td>772</td>
<td>3447</td>
<td>9</td>
<td>1.2</td>
<td>99.6</td>
<td>0.3</td>
<td>99.9</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>n/a</td>
<td>965 494</td>
<td>770</td>
<td>5431</td>
<td>11</td>
<td>1.4</td>
<td>99.4</td>
<td>0.2</td>
<td>99.9</td>
</tr>
<tr>
<td>Appetite loss</td>
<td>n/a</td>
<td>967 570</td>
<td>754</td>
<td>3335</td>
<td>27</td>
<td>3.5</td>
<td>99.7</td>
<td>0.8</td>
<td>99.9</td>
</tr>
<tr>
<td>Weight loss</td>
<td>n/a</td>
<td>961 571</td>
<td>720</td>
<td>9354</td>
<td>61</td>
<td>7.8</td>
<td>99.0</td>
<td>0.6</td>
<td>99.9</td>
</tr>
</tbody>
</table>

The positive predictive values (PPVs) are an average for patients in each category. PPVs for individuals can be calculated using the web calculator to take their characteristics into account. n/a = not applicable. \(^*\)Criterion not met does not have disease. \(^*\)Criterion not met does have disease. \(^*\)Criterion met does not have disease. \(^*\)Criterion met does have disease.
Box 1. Clinical examples

- A 70-year-old male with type 2 diabetes, who is a heavy smoker and presents with abdominal pain, loss of appetite, and dysphagia has a 10.2% estimated risk of having pancreatic cancer. If he also has a history of chronic pancreatitis, the estimated risk would be 54.4%.
- A 75-year-old female, who is a moderate smoker and has abdominal pain, abdominal distension, and loss of appetite, has an 11.3% estimated risk of having pancreatic cancer. If she also has weight loss, the estimated risk would be 44.1%.
- A 50-year-old male, who is a heavy smoker with type 2 diabetes and no symptoms has a 0.1% estimated risk of developing pancreatic cancer in the next 2 years. If the same male had abdominal pain, appetite loss, dysphagia, and constipation the estimated risk would be 3.9%.

Funding
This study was undertaken by ClinRisk Ltd. There was no external funding.

Ethics committee
All QResearch® studies are independently reviewed in accordance with the QResearch® agreement with Trent Multi-Centre Ethics Committee (UK).

Provenance
Freely submitted; externally peer reviewed.

Web calculator
Here is a simple web calculator to implement the QCancer® (pancreatic) algorithm, which is publically available alongside the paper and open source software http://www.qcancer.org/pancreas.

Competing interests
Julia Hippisley-Cox is professor of clinical epidemiology at the University of Nottingham and co-director of QResearch® — a not-for-profit organisation which is a joint partnership between the University of Nottingham and EMIS (leading commercial supplier of IT for 60% of general practices in the UK). Julia Hippisley-Cox is also a paid director of ClinRisk Ltd, which produces software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. Carol Coupland is associate professor of medical statistics at the University of Nottingham and a paid consultant statistician for ClinRisk Ltd. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations.

Acknowledgements
We acknowledge the contribution of EMIS practices who contribute to QResearch® and EMIS for expertise in establishing, developing and supporting the database. The algorithms presented in this paper will be released as Open Source Software under the GNU Lesser GPL v3.

Discuss this article
Contribute and read comments about this article on the Discussion Forum: http://www.rcgp.org.uk/bjgp-discuss

physicians, the quality of information is likely to be good since previous studies have validated similar outcomes and exposures using questionnaire data and found levels of completeness and accuracy in similar GP databases to be good.29,30 For example, one systematic review reported that on average 89% of diagnoses recorded on the GP electronic record are confirmed from other data sources.29 Currently however, there is limited information on the QResearch database regarding the precise type of cancer which means it was not possible to include the precise type of pancreatic cancer in the outcome (that is, distinguish between endocrine and exocrine tumours), the stage, or the grade. The QResearch database will be linked with information from cancer registries in the near future which is likely to increase the accuracy and completeness of this information.

Implications for research and practice
This study could help raise awareness of symptom complexes predictive of pancreatic cancer especially since current NICE guidelines on suspected cancer fail to mention pancreatic cancer despite its substantial morbidity and mortality. Further research is needed to assess whether use of this symptom based tool can lead to earlier identification of pancreatic cancer at a stage where curative treatment is more likely to be possible.

Although primarily designed to be used by GPs at the point of care to assess risk in symptomatic patients, and inform the decision to investigate or refer, the algorithm could be also be used by members of the public via a simple web calculator which could then prompt symptomatic patients to consult their GP. The algorithm could be integrated into GP clinical computer systems and used to generate a list of high risk patients who could then be recalled and systematically assessed. For example, the algorithm can identify 10% of the population in which approximately 62% of all new pancreatic cancer cases are likely to be diagnosed over the next 2 years. Table 4 shows the possible thresholds along with the sensitivity, specificity, positive and negative predictive values. This is intended to inform subsequent cost-effectiveness modelling and the choice of thresholds which is outside the scope of this paper.

The algorithms have been developed in one cohort and validated in a separate cohort representative of the patients likely to be considered for preventative measures. The algorithm performed well with good discrimination and calibration. Following independent external validation and cost-effectiveness modelling (which is outside the scope of the present study), the algorithm could potentially be used in clinical practice to identify those at highest risk of having pancreatic cancer to facilitate early referral and investigation and so help earlier identification of patients with pancreatic cancer.

While the study was not designed to determine whether type 2 diabetes causes pancreatic cancer it was found that pancreatic cancer is twice as likely to occur in patients with type 2 diabetes than those without it, although the absolute risk across all patients with diabetes is low (0.3%). No increased risk of pancreatic cancer among patients with type 1 diabetes was found, although there were only one-tenth of the number of patients with type 1 diabetes compared with type 2 diabetes so a type 2 error cannot be excluded. Further research into the potential mechanisms underlying this association for type 2 diabetes may be warranted.

Several additional symptoms not traditionally thought to be associated with pancreatic cancer (such as, abdominal distension in females, dysphagia, and constipation in males) were also identified but which remained independently predictive on multivariate analysis. Given the location of the pancreas, it is possible that a tumour could result in pressure on the fundus of the stomach or lower end of the oesophagus resulting in dysphagia. If confirmed by other studies, this study's results suggest that investigation of the pancreas should be considered alongside other upper gastrointestinal investigations for patients presenting with dysphagia. There may be other symptoms which could predict pancreatic cancer that were not included in this study, such as vomiting, nausea, fever, dyspepsia, backache, and depression. These could be explored in future versions of the algorithm to determine whether they improve its performance or results in reclassification around a risk threshold.
REFERENCES