BMJ Open

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005809
Article Type:	Research
Date Submitted by the Author:	28-May-2014
Complete List of Authors:	Hippisley-Cox, Julia; University of Nottingham, ; ClinRisk Ltd, Coupland, Carol; University of Nottingham, Division of Primary Care Brindle, Peter; Bristol Clinical Commissioning Group,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, General practice / Family practice, Health informatics
Keywords:	qresearch, cprd, Epidemiology < TROPICAL MEDICINE, qrisk2, prognosis, validation

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Research paper submitted to BMJOpen
May 2014

Authors

Julia Hippisley-Cox Professor of Clinical Epidemiology & General Practice[1]
Carol Coupland Associate Professor and Reader in Medical Statistics[1]
Peter Brindle Research and Development Programme Director[2]

Institutions

[1]Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham, NG2 7RD.

[2]Avon Primary Care Research Collaborative, Bristol Clinical Commissioning Group, South Plaza, Marlborough Street, Bristol BS1 3NX.

Guarantor & Author for correspondence - Julia Hippisley-Cox

Email: Julia.hippisley-cox@nottingham.ac.uk

Telephone: 0115 8466915

Fax: 0115 8466904

Abstract

Objectives To validate the performance of a set of risk prediction algorithms developed using the QResearch database, in an independent sample from general practices contributing to the Clinical Research Data Link (CPRD).

Setting Prospective open cohort study using practices contributing to the CPRD database and practices contributing to the QResearch database.

Participants The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices between 01 Jan 1998 and 31 July 2012. The validation statistics for QResearch were obtained from the original published papers which used a one third sample of practices separate to those used to derive the score. A cohort from QResearch was used to compare incidence rates and baseline characteristics and consisted of 6.8 million patients from 753 practices registered between 01 Jan 1998 and until 31 July 2013.

Outcome measures Incident events relating to seven different risk prediction scores: QRISK2 (cardiovascular disease); QStroke (ischaemic stroke); QDiabetes (type 2 diabetes); QFracture (osteoporotic fracture and hip fracture); QKidney (moderate and severe kidney failure); QThrombosis (venous thromboembolism); QBleed (intracranial bleed and upper gastrointestinal haemorrhage). Measures of discrimination and calibration were calculated.

Results Overall, the baseline characteristics of the CPRD and QResearch cohorts were similar though QResearch had higher recording levels for ethnicity and family history. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with the published results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch and the ROC value was 0.85 on both databases. The scores were well calibrated in CPRD.

Conclusion Each of the algorithms performed practically as well in the external independent CPRD validation cohorts as they had in the original published QResearch validation cohorts.

Strengths and limitations of this study

- This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system from that used to derive the algorithms.
- The discrimination and calibration statistics for each score were very similar in CPRD to those published from validation cohorts from QResearch. This supports their potential utility in the general population of patients in primary care.
- A strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice.
- The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible.
- Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using these algorithms in primary care.

1 Introduction

In the last 7 years, we have developed a series of risk prediction algorithms using the QResearch database. QResearch is a large research database containing pseudonymised individual level data from over 700 general practices using the EMIS clinical system. The algorithms predict outcomes such as cardiovascular disease(www.qrisk.org)[1], stroke (www.qstroke.org)[2], type 2 diabetes (www.qdiabetes.org)[3], osteoporotic fracture (www.qfracture.org)[4], moderate or severe kidney disease (www.qkidney.org)[5], venous thrombo-embolism (www.qthrombosis.org)[6], and emergency hospital admission (www.qadmissions.org)[7]. Generally, the "QPrediction" algorithms have been designed to systematically identify patients in primary care at high risk of a serious clinical outcome for whom further intervention to lower risk of that outcome might be possible. They are also designed to quantify absolute risk of serious outcomes in a way which patients can understand and which might help guide lifestyle and management decisions. A number of these algorithms are now integrated into GP clinical computer systems, included in national guidelines[1 4] and are in daily use across the NHS [1 3 8].

The algorithms were originally developed using a random two thirds sample of practices contributing to the QResearch database and validated on the remaining third. Whilst this represents a physically discrete population of patients and practices for validation, the practices all use the same clinical computer system (EMIS), which is in use in 53% of UK practices. A more stringent test of performance is to validate the algorithms on a fully external database derived from practices using a different but commonly used primary care computer system. This would help determine whether the predictions from the algorithms are likely to generalise to the whole population in England. Whilst some of the algorithms have been validated by an independent team using the THIN primary care database[9-12], there are currently no published validations of the algorithms using a primary care database which is routinely linked to mortality data in the same way as QResearch.

We therefore decided to validate the various QPrediction Scores using another database known as the Clinical Research Data Link (CPRD). The General Practice Research database (GPRD) was originally set up in 1988 and is of similar nature to QResearch although it is derived from practices using a different clinical computer system (Vision, which is used by 20% of GPs). It was extended to include linked data mortality data and data from secondary care and was renamed the Clinical Research Data Link (CPRD) in 2012. Our secondary objective was to compare the ascertainment of incident clinical events recorded in GP data alone with that recorded in either GP data or the linked mortality data in both the CPRD and QResearch.

2 Methods

2.1 CPRD Study population

For the validation using CPRD, we identified an open cohort of patients aged 25-99 years at baseline and followed this cohort up until 31st July 2012 (the latest date for which linked data were available at the time of analysis). We restricted the CPRD cohort to 357 practices in England which had linked ONS mortality and hospital admissions data. For each patient we determined an entry date to the cohort, which was the latest of the following dates: 25th birthday, date of registration with the practice plus one year, date on which the practice computer system was installed plus one year, and the beginning of the study period (01 January 1998). Patients were censored at the earliest date of the relevant outcome, de-registration with the practice, last upload of computerised data or the study end date (31 July 2012).

For the assessment of the two QBleed outcomes (intracranial bleed and upper gastrointestinal haemorrhage) we used a later cohort entry date of 01.01.2007 for comparability with the equivalent study period for the derivation of the algorithm on QResearch[13].

2.2 QResearch study population

For comparison of the validation statistics (ROC, D and R2 statistics), we extracted the original published values from the papers which had been calculated using a one third sample of practices from QResearch which were independent from the two thirds of practices used to derive the scores.

For comparison of the baseline characteristics, incidence rates and ascertainment rates we used the latest version of the QResearch database which is currently available (QResearch 38, 31st Dec 2013). We identified an open cohort in the same way as for CPRD, using all of the QResearch practices, and with follow-up until 31 July 2013.

2.3 Inclusion and exclusion criteria

For both databases, we excluded patients without a Townsend score (an area based measure of material deprivation derived from the post code) and temporary residents. For each score we then identified patients who were eligible to have the score calculated according to the relevant inclusion and exclusion criteria as summarised in Table 4

2.4 Risk scores included in validation

We validated the following risk prediction scores on CPRD:

1. QDiabetes - 10 year risk of type 2 diabetes[3]

2. QRISK2-2014 – 10 year risk of cardiovascular disease[1]
3. QStroke – 10 year risk of stroke or transient ischaemic attack (TIA)[5]
4. QFracture - 10 year risk of hip or osteoporotic fracture[4]
5. QThrombosis – 5 year risk of VTE[6]
6. QBleed – 5 year risk of upper gastrointestinal haemorrhage and intracranial haemorrhage[13]
7. QKidney – 5 year risk of moderate-severe kidney disease[5]

2.5 Clinical outcomes

We identified the relevant clinical outcome using the same definition as had been applied in the original derivation of the risk scores using QResearch. The data sources used to identify the clinical outcomes had varied over the six years during which the original studies had been undertaken due to the changing availability of linked hospital and mortality data over that time. In 2008, the QResearch database was linked to mortality records for 1997 onwards. In 2013, the QResearch database was linked to hospital admissions records with data for patients from 1998 onwards. For QRISK2-2014, the outcome was identified by the presence of the relevant Read code on the GP record or an ICD10 code recorded on the linked mortality record or the linked hospital admissions record. For QStroke, QDiabetes, QFracture and QThrombosis, the outcome was identified either by the presence of the relevant Read code recorded on the GP record or an ICD10 code recorded on the linked mortality record. For QKidney, the outcome was identified solely from information recorded in the GP record as in the original study as it required blood test values which were only present in the GP record. For QBleed, the outcome was identified in CPRD from events recorded either on the linked hospital admissions database or the linked mortality record in order to identify the events most likely to have serious clinical consequences for the patient.

We determined case ascertainment for each clinical outcome on both databases, by calculating the proportion of cases recorded on the GP record out of the total number of cases recorded on either the GP record or linked mortality record. We calculated the age standardised incidence rates of each outcome based on outcomes recorded on (1) the GP record alone and on (2) the GP record or linked mortality; (3) GP or linked mortality or hospital records. We standardised CPRD rates to the age distribution of the QResearch population in five year bands to ensure comparability.

2.6 Risk factors and missing values

We extracted data from CPRD for all the predictor variables included in one or more of the different algorithms using the same definitions as those used in the original QResearch studies to enable a direct comparison of the results. We developed a mapping between the Read and medication reference tables to identify the equivalent code in each database. This included the following variables recorded at entry to the cohort:

- **demographics** - age, sex, ethnicity, resident in care home, material deprivation (as measured by the Townsend score)
- **clinical values** - smoking status (non-smoker, ex-smoker, light smoker [1-9 cigarettes/day], moderate smoker [10-19 cigarettes/day], heavy smoker [20+ cigarettes/day];body mass index, systolic blood pressure, alcohol consumption
- **laboratory results** –cholesterol/HDL ratio, platelets
- **family history**- family history of osteoporosis or hip fracture in a first degree relative, coronary heart disease in first degree relative under the age of 60 years, diabetes in a first degree relative.
- **chronic diseases** – congestive cardiac failure, atrial fibrillation, coronary heart disease, cardiovascular disease, peripheral vascular disease, venous thromboembolism, diabetes, rheumatoid arthritis, SLE, hypertension, renal disease, renal stones, inflammatory bowel disease, dementia, Parkinson's disease, epilepsy, cancer, chronic liver disease or pancreatitis, oesophageal varices, prior haemorrhage, malabsorption endocrine diseases, asthma or COPD, history of falls, prior osteoporotic fracture, varicose vein surgery, emergency admissions or hip surgery in last 6 months.
- **prescribed medication**- antidepressants, antipsychotics, antiplatelets, oral NSAIDs, tamoxifen, oestrogen containing hormone replacement therapy (BNF chapter 6.4.1.1), systemic corticosteroids, combined oral contraceptive.

The combination of predictor variables required for each risk score varied with the score being validated as shown in Table 1. We used the clinical value recorded closest to the date on which the patient entered the study for body mass index, systolic blood pressure, smoking status, platelets, and total and HDL cholesterol.

2.7 Townsend scores

We used the Townsend score evaluated at output area as a proxy for material deprivation. The CPRD dataset differs from the QResearch dataset in that each patient in the CPRD dataset is allocated to a tenth of deprivation (as measured by the Townsend score) and only the category number is provided. In contrast, each patient in the QResearch dataset is allocated the individual Townsend score corresponding to their output area of residence (i.e. continuous data). In order to calculate risk scores in the CPRD cohort, we used the median value for each tenth as supplied by CPRD. Patients with missing Townsend scores were excluded from the cohorts.

2.8 Discrimination and calibration statistics

We used multiple imputation to replace missing values for body mass index, systolic blood pressure, smoking status, alcohol, and total and HDL cholesterol. We created five multiply imputed datasets and used Rubin's rules to combine effect estimates and standard errors to allow for the uncertainty due to imputing missing data[14 15].

We applied the algorithm for each score to eligible patients in the CPRD study cohort to obtain predicted risks for each of the relevant clinical outcomes. We calculated the estimated risk for eligible patients in the CPRD validation dataset over 5 years or 10 years depending on which score was used. We then tested the performance of each score in the CPRD cohort and compared it with the published results from the original QResearch validation cohorts.

In order to assess calibration (i.e. degree of similarity between predicted and observed risks), we calculated the mean predicted risk and the observed risk [16]obtained using the Kaplan-Meier estimate and compared the ratio of the mean predicted risk to the observed risk for patients in the validation cohort in each decile of predicted risk. We calculated the area under the Receiver Operator Curve (ROC) statistic to assess discrimination (i.e. ability of a risk prediction equation to distinguish between those who do and do not have an event during the follow-up period). We also calculated the D statistic[17] and an R squared statistic derived from the D statistic[18] which are measures of discrimination and explained variation appropriate for survival models. The D statistic has been developed as a new measure of discrimination specifically for censored survival data, higher values indicate improved discrimination, and an increase in the D statistic of at least 0.1 indicates an important difference in prognostic separation between different risk classification schemes. The R^2 statistic derived from the D statistic is a measure specific to censored survival data– it measures explained variation in time to the outcome event and higher values indicate more variation is explained[19].

We identified the proportion of patients in the CPRD validation cohort who were in the top decile of predicted risk and used this to calculate the sensitivity, specificity and observed risk at this threshold. We used the top decile for comparability across the scores and with previous studies though the choice of threshold for use in clinical practice will depend on the context and cost-effectiveness of relevant interventions. Analyses were conducted using Stata (version 13.1).

2.9 Sample size estimation

There is currently no clear guidance on sample size requirements for studies evaluating the performance (validation) of a multivariable risk score, but a commonly used rule-of-thumb is that it is desirable to seek a dataset with at least 100 patients with the outcome of interest. We used all the available data on the CPRD to maximize the power of the study.

3 Results

3.1 Study populations

The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices with linked data between 01 Jan 1998 and 31 July 2012. The QResearch cohort consisted of 6.8 million patients from 753 practices with linked data, registered between 01 Jan 1998 and until 31 July 2013. The numbers of patients in each geographical region are shown in Web Extra Table 1.

3.2 Baseline characteristics

Table 2 shows a comparison of the demographic characteristics for the CPRD and QResearch cohorts. The QResearch population was marginally younger with 34% of women and 33% of men aged 24-34 years compared with 28% and 27% for CPRD. QResearch had a higher proportion of patients with self-assigned ethnicity recorded compared with CPRD. For example, 60% of women had self-assigned ethnicity recorded on QResearch compared with 39% on CPRD. Recording of a positive family history of coronary heart disease and diabetes was more than twice as high in QResearch compared with CPRD. For example, 12% of women on QResearch had a family history of coronary heart disease recorded compared with 5% on CPRD.

Recording of cholesterol/HDL ratio was marginally higher on QResearch compared with CPRD. However the mean values for the various clinical values (BMI, systolic blood pressure, serum creatinine and cholesterol/HDL ratio) were extremely similar.

Table 3 shows prescribed medication and clinical diagnoses recorded in patients on or prior to entry to the study cohort. Overall, the prevalence of clinical diagnoses were similar on the two databases with CPRD having marginally higher prescribing rates.

The inclusion and exclusion criteria for each risk score are shown in Table 4 along with the numbers of patients eligible for each analysis on CPRD. For example, there were 3,177,192 patients aged 25-84 years. Of these, 99,189 had existing diabetes at baseline leaving 3,078,003 for the validation of QDiabetes.

3.3 Incidence rates of clinical outcomes

Table 5 shows the number of incident events for each clinical outcome in women recorded on GP data and those recorded on either GP data or cause specific mortality data for both the CPRD and QResearch cohorts. It also shows the age standardized incidence rates per 1000 person years. Table 6 shows the comparable information for men.

For example, there were 35,617 incident ischaemic stroke events for women on CPRD. Of these, 32,283 had been identified on the GP record with an additional 3,334 events identified on the linked ONS mortality record. The ascertainment of events on the GP record alone was therefore 32283/35617 i.e. 90.6%.

For QResearch, there were 70,477 incident stroke events recorded on either the GP or linked ONS mortality record of which 63,572 had been identified on the GP record. The ascertainment was therefore 90.2%. For thromboembolism, 91.1% of events on CPRD were identified on the GP record alone compared with 90.6% for QResearch. Similar results were obtained for men with levels of ascertainment between the two databases being extremely close suggesting similar recording patterns between the two groups of GP practices contributing to each database.

The age standardized incidence rates of events on CPRD tended to be marginally lower than those on QResearch as shown by the ratio of the CPRD rates to those in QResearch (Table 5). For example, the rate ratio for fractured neck of femur in women was 0.94 indicating that CPRD had a 6% lower incidence rate compared with QResearch. The effect was more marked for moderate or severe kidney failure where the incidence rates for CPRD were approximately 25% lower than those for QResearch in women and 16% lower in men.

The age standardized incidence rates of upper gastrointestinal haemorrhage and intracranial haemorrhage among patients prescribed anticoagulants and those not prescribed anticoagulants are shown in Web extra table 2. The rates are similar for CPRD and QResearch.

3.4 Validation statistics

Table 7 shows the discrimination statistics for each score in CPRD in men and women and also the published values from previous validations using QResearch. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch. The D statistic for women was 2.03 within CPRD compared with 2.08 for QResearch. The ROC value for women was 0.85 on both databases.

Of all the scores, QFracture (fractured neck of femur) had the best performance in men in CPRD with a ROC value of 0.89, R^2 value of 71% and D statistic of 3.17. The corresponding figures for QResearch in men were 0.89, 72% and 3.26.

QThrombosis had the lowest values for women in CPRD with an ROC value of 0.77, R^2 of 34.5 and D statistic of 1.49. The corresponding figures for women in QResearch were 0.75, 33.5 and 1.45.

Figure 1 compares the mean predicted risks and observed risks for each score across each tenth of predicted risk (1 representing the lowest risk and 10 the highest risk)

and demonstrates that the models are generally well calibrated for patients on CPRD.

The QKidney score (moderate or severe kidney failure) showed the observed risk was lower than the predicted risk. This might indicate a degree of over prediction of the score. Alternatively, it could be related to the lower incidence rate of kidney failure observed among women on the CPRD compared with QResearch.

3.5 Performance for the top decile of risk.

Table 8 shows the sensitivity, specificity and observed risk for patients in the top decile of each score on CPRD. The observed risk is higher than the threshold since this represents the observed risk within the top decile of predicted risk. For example, the cut off for the top tenth of risk for QFracture (fractured neck of femur) was a 10 year risk of 3.7%. At this threshold the sensitivity was 66.5%, specificity 90.4% and observed risk 9.4%. The results are similar to those obtained from QResearch (not shown).

4 Discussion

4.1 Summary of key findings

This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system (Vision system supplied by In Practice Systems) from the QResearch database (which is based on practices using EMIS clinical systems). Practices using the Vision system together with practices using EMIS make up approximately 75% of all the English general practices. The discrimination and calibration statistics for each score were remarkably similar in CPRD to those published from validation cohorts from QResearch.

Before a clinical risk score can be reliably used in clinical practice, evidence is needed that it can successfully predict the intended outcome in groups of patients other than ones used to develop the score but similar to ones in whom the score might be used. Not all risk scores perform well in external samples – this can be due to deficiencies in the design or modelling methods used to derive the algorithm, if the model is over fitted or if there is an important predictor which is absent[20]. Other reasons for poor performance include differences between the setting of patients in the new and derivation samples, differences in how information is recorded and differences in patient characteristics[20]. It is for these reasons, that we have meticulously assembled the CPRD cohort using the same inclusion/exclusion criteria, definitions of predictor and outcome variables as in the original derivation studies. Any differences observed are therefore more likely to be due to capture of information and underlying population characteristics. In this study, we have found marginal differences in incidence rates between QResearch and CPRD and higher rates of recording of family history and ethnicity in QResearch though these have not been large enough to materially affect our results.

4.2 Strengths and limitations

A strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice.

The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible.

The quality of information on CPRD is likely to be good since previous studies have validated similar outcomes and exposures and found levels of completeness and accuracy to be good[21 22].

4.3 Comparison with other studies

Our validation results confirm earlier studies undertaken on the THIN database (another general practice database which is derived from the Vision system but which isn't linked to mortality data). These earlier studies include external validations of QRISK2[10 11 23], QDiabetes[12], QFracture[9] and QKidney[24] by an independent team who were not involved in the development of the algorithms. These independent validations have demonstrated similar performance compared with the validations performed by study authors using the QResearch database. This study builds on previous validations by providing new information on the performance of scores not previously validated on an external database (QBleed and QThrombosis) and by utilising the linked data which was not available on the THIN database. Together with the present study (which includes a number of scores not previously tested in an external population), the results provide consistent evidence that these QPrediction scores are likely to provide appropriate estimates of disease risk in contemporary primary care populations in England and to discriminate between patients at different levels of risk with reasonable reliability.

4.4 Comparison of QResearch and CPRD baseline characteristics

Overall, our results show a striking similarity between CPRD and QResearch cohorts for nearly all baseline characteristics. There are two notable exceptions. First, recording of ethnicity was higher in QResearch than CPRD. Second, fewer patients in the CPRD cohort had a recorded family history of diabetes and coronary heart disease in a first degree relative under the age of 60 years. Given the similarity for the other risk factors and treatments, it is likely that the difference in ethnicity and family history recording reflects a difference in recording patterns between the two clinical computer systems rather than a true difference between the two cohorts. A similar pattern for recording of ethnicity and family history was also reported in the validation of QRISK on the Health Improvement Network (THIN database) [11 25]. This was thought to be due to different usage of clinical templates in the clinical system, with EMIS practices having ethnicity and family history included more often thereby prompting the user to enter this information in a more systematic fashion.

4.5 Comparison of QResearch and CPRD incidence rates

The age standardised incidence rates for each condition were generally marginally higher on QResearch than CPRD although the proportions of events identified on GP data alone (out of all possible events on either GP or linked mortality data) was very close. This suggests that patterns of recording of major clinical events are very similar between QResearch and CPRD although the absolute value varies by clinical condition. For example, 91% of ischaemic stroke events are identified on the GP record alone compared with 99% of hip fractures. We also note the lower levels of cardiovascular events in the GP clinical record alone which was between 13-15%. Some of this will reflect new sudden events where the first presentation was a

hospital admission or death whilst others may reflect some under-representation of existing cases not recorded in the GP record. Our study is unable to distinguish between these two scenarios, though the latter one potentially has clinical consequences if the patient is not identified as having cardiovascular disease as they may not be offered secondary prevention.

We think that the information on baseline characteristics and incidence rates will have a utility beyond the present study since it suggests that both databases are fundamentally similar in many aspects and likely to generate similar results for a range of epidemiological studies[26].

4.6 Summary

In summary, we have tested a set of QPrediction scores using an external independent cohort of practices contributing to the CPRD. The results demonstrate good performance, comparable to the results obtained from QResearch, meaning that the findings of studies performed in either database are likely to be applicable in England.

5 Supporting information

Approvals

The project was approved in accordance with the QResearch® agreement with Trent Research Ethics Committee (ref 03/04/021) and approved by the ISAC committee of the CPRD (ref 13_079).

Acknowledgements

We thank EMIS and EMIS practices for their contribution to the QResearch database. We thank CPRD and Vision Practices for allowing access to the CPRD for this study.

Source of funding:

The validation of the QPrediction scores is funded by the National Institute for Health Research's School for Primary Care Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funding body did not play a role in the study design, writing of the manuscript or in the decision to submit the manuscript for publication.

Statement

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in HEART editions and any other BMJPGL products to exploit all subsidiary rights.

Contributorship

JHC initiated the study, undertook the literature review, data extraction, data manipulation and primary data analysis and write the first draft of the paper. JHC is the guarantor. CC contributed to the design, analysis, interpretation and drafting of the paper. PB contributed to the development of core ideas, the analysis plan, interpretation of the results and the drafting of the paper.

Competing interests

JHC 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). JHC is also director of ClinRisk Ltd which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of Medical Statistics at the University of Nottingham and a consultant statistician for ClinRisk Ltd. PB has received financial support for undertaking the validation work from the National School for Primary Care Research. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations. There are no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

All the algorithms validated in this paper are published as open source software under the GNU Lesser Public License. No additional data are available.

6 References

1. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008:bmj.39609.449676.25.
2. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ 2013;**346**.
3. Hippisley-Cox J, Coupland C, Robson J, et al. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009;**338**:b880-.
4. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. BMJ 2009;**339**:b4229-.
5. Hippisley-Cox J, Coupland C. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney(R) Scores. BMC Family Practice 2010;**11**:49.
6. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. BMJ 2011;**343**:d4656.
7. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions BMJ 2013:(in press).
8. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.
9. Collins GS, Mallett S, Altman DG. Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. BMJ 2011;**342**:d3651.
10. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010;**340**:c2442-.
11. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;**339**:b2584-.
12. Collins GS, Altman DG. External validation of the QDScore for predicting the 10-year risk of developing Type 2 diabetes. Diabetic Medicine 2011;**28**:599-607.
13. Hippisley-C J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleeding with anticoagulants: cohort study to derive and validate the QBleed score. BMJ 2014;**(in press)**.
14. Gray A, Clarke P, Farmer A, et al. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). BMJ 2002;**325**(7369):860.
15. Royston P. Multiple imputation of missing values. Stata Journal 2004;**4**(3):227-41.
16. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007:bmj.39261.471806.55.

17. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;**23**:723-48.
18. Royston P. Explained variation for survival models. Stata J 2006;**6**:1-14.
19. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;**13**:33.
20. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;**338**(may28_1):b605-.
21. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**:4-14.
22. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;**60**:e128-36.
23. Collins GS, Altman A. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;**344**.
24. Collins GS, Altman D. Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney scores using a primary care database. British Journal of General Practice 2012;**62**(597):e243-e50.
25. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 2008;**94**:34-39.
26. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. BMJ Open 2014;**4**(4).
27. Hippisley-Cox J, Coupland C. QRISK2-2014 Annual Update Information, 2014:5.
28. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ 2012;**344**(may22 1):e3427-e27.
29. Hippisley-Cox J CC. QDiabetes 2013 Annual Update Information Secondary QDiabetes 2013 Annual Update Information 2013. http://qdiabetes.org/QDiabetes-2013-Annual-Update-Information.pdf.

Table 1 Summary of QPrediction scores including outcome and predictor variables

Score	Outcome	Predictors
QDiabetes[3]	10 year risk of type 2 diabetes[±]	**In men and women:** Age, sex, ethnicity, deprivation, smoking, family history of diabetes, cardiovascular disease, treated hypertension, steroid tables, body mass index
QRISK2[27]	10 year risk of CVD recorded[**]	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation.
QStroke[2]	10 year risk of ischaemic stroke or TIA[±]	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, congestive cardiac failure, valvular heart disease
QKidney[5]	5 year risk of moderate or severe kidney failure[µ]	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, body mass index, family history of kidney disease, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, SLE, peripheral vascular disease, kidney stones, NSAIDs
QThrombosis	5 year risk of venous thromboembolism[±]	**In men and women**: age, body mass index, smoking status, varicose veins, congestive cardiac failure, chronic renal disease, cancer, chronic obstructive pulmonary disease, inflammatory bowel disease, hospital admission in past six months, and current prescriptions for antipsychotic drugs. Additionally in women: combined oral contraceptives, tamoxifen, and hormone replacement therapy
QBleed[13]	5 year risk upper gastrointestinal bleed in patient starting anticoagulants vs others*	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; venous thromboembolism; congestive cardiac failure; treated hypertension; cancer; recent abnormal platelets (< 150µL or >480µL); current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QBleed[13]	5 year risk of intracranial bleed in patient starting anticoagulants vs others *	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; treated hypertension; recent abnormal platelets (< 150µL or >480µL); current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QFracture[28]	10 year risk of hip fracture[±] 10 year risk of osteoporotic fracture [µ]	**In women:** HRT usage, age, body mass index, smoking status, recorded alcohol use, parental history of osteoporosis, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls, menopausal symptoms, chronic liver disease, gastrointestinal malabsorption and other endocrine disorders. **In men:** age, body mass index, smoking status, recorded alcohol use, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls and liver disease.

[±] recorded either on GP record or linked ONS mortality record; µ recoded on the GP record.
*Recorded either on linked hospital admissions record or ONS mortality record
**Recorded either on linked hospital admissions record or ONS mortality or linked hospital admissions record

Table 2 Comparison of baseline characteristics of patients in CPRD validation cohort and QResearch comparison cohort

	CPRD	*CPRD*	*QResearch*	*QResearch*
	Men (%)	**Women (%)**	**Men (%)**	**Women (%)**
total	1,588,803	1,682,709	3,304,145	3,454,504
Ageband				
25-34 years	427975 (26.9)	467192 (27.8)	1083589 (32.8)	1179742 (34.2)
35-44 years	396680 (25.0)	364150 (21.6)	814988 (24.7)	731089 (21.2)
45-54 years	294274 (18.5)	277663 (16.5)	558553 (16.9)	516188 (14.9)
55-64 years	212817 (13.4)	211636 (12.6)	390229 (11.8)	389266 (11.3)
65-74 years	148180 (9.3)	164172 (9.8)	267997 (8.1)	298847 (8.7)
75+ years	108877 (6.9)	197896 (11.8)	188789 (5.7)	339372 (9.8)
mean Townsend score (SD)	-.5 (3.2)	-.5 (3.2)	.3 (3.6)	.2 (3.6)
Care home resident	1407 (0.1)	3466 (0.2)	2983 (0.1)	7411 (0.2)
Ethnicity recorded	587879 (37.0)	658835 (39.2)	1859462 (56.3)	2077181 (60.1)
White or not recorded	1515113 (95.4)	1602212 (95.2)	3010061 (91.1)	3149618 (91.2)
Indian	16442 (1.0)	16025 (1.0)	56156 (1.7)	50406 (1.5)
Pakistani	6606 (0.4)	6146 (0.4)	30632 (0.9)	23405 (0.7)
Bangladeshi	2419 (0.2)	1688 (0.1)	23017 (0.7)	17450 (0.5)
Other Asian	10795 (0.7)	11873 (0.7)	32513 (1.0)	36886 (1.1)
Caribbean	4989 (0.3)	6425 (0.4)	25782 (0.8)	32953 (1.0)
Black African	12883 (0.8)	14771 (0.9)	51980 (1.6)	56528 (1.6)
Chinese	2914 (0.2)	4176 (0.2)	16084 (0.5)	23043 (0.7)
Other ethnic group	16642 (1.0)	19393 (1.2)	57920 (1.8)	64215 (1.9)
Smoking status recorded	1442088 (90.8)	1595538 (94.8)	2943405 (89.1)	3219598 (93.2)
Non smoker	613833 (38.6)	834721 (49.6)	1449694 (43.9)	1973691 (57.1)
Ex-smoker	252873 (15.9)	222615 (13.2)	611837 (18.5)	545125 (15.8)
Light smoker (1-9/day)	104466 (6.6)	109864 (6.5)	472614 (14.3)	384482 (11.1)
Moderate smoker (10-19/day)	183000 (11.5)	179391 (10.7)	223631 (6.8)	202776 (5.9)
Heavy smoker (20+/day)	142438 (9.0)	87474 (5.2)	185629 (5.6)	113524 (3.3)
Smoker amount not recorded	145478 (9.2)	161473 (9.6)	0 (0.0)	0 (0.0)
Alcohol status recorded	1238110 (77.9)	1379002 (82.0)	2584335 (78.2)	2834426 (82.1)
Non drinker	163633 (10.3)	318880 (19.0)	583752 (17.7)	1035692 (30.0)
Trivial <1u/day	460091 (29.0)	726851 (43.2)	782985 (23.7)	1144469 (33.1)
Light 1-2u/day	411261 (25.9)	290547 (17.3)	481674 (14.6)	402750 (11.7)
Moderate 3-6/day	166328 (10.5)	36763 (2.2)	648549 (19.6)	237679 (6.9)
Heavy 7-9u/day	19612 (1.2)	2853 (0.2)	54083 (1.6)	7152 (0.2)
Very Heavy >/day	17185 (1.1)	3108 (0.2)	24468 (0.7)	5195 (0.2)
Family History				
family history CHD	68805 (4.3)	80985 (4.8)	326995 (9.9)	417537 (12.1)
family history of diabetes	96810 (6.1)	132390 (7.9)	357109 (10.8)	487397 (14.1)
family history osteoporosis	880 (0.1)	10062 (0.6)	1655 (0.1)	17529 (0.5)
family history kidney disease	1253 (0.1)	1586 (0.1)	2034 (0.1)	2769 (0.1)

Clinical Values recorded				
BMI recorded	1268235 (79.8)	1481918 (88.1)	2553514 (77.3)	2857742 (82.7)
mean BMI (SD)	29.6 (6.8)	28.2 (7.0)	29.4 (6.8)	28.2 (7.0)
SBP recorded	1359560 (85.6)	1590226 (94.5)	2755733 (83.4)	3190390 (92.4)
mean SBP(SD)	133.1 (23.6)	128.6 (22.7)	132.2 (18.3)	127.1 (20.8)
cholesterol/HDL ratio	587865 (37.0)	606035 (36.0)	1323503 (40.1)	1368180 (39.6)
mean cholesterol ratio (SD)	4.4 (1.4)	3.7 (1.2)	4.4 (1.4)	3.7 (1.2)
platelets recorded	223461 (14.1)	382799 (22.7)	478596 (14.5)	829702 (24.0)
platelets < 150 or > 480	11051 (0.7)	13282 (0.8)	23479 (0.7)	27009 (0.8)
Creatinine recorded	811779 (51.1)	997118 (59.3)	1714337 (51.9)	2053036 (59.4)
Mean creatinine (SD)	96.7 (149.1)	79.7 (911.6)	95.5 (30.7)	78 (23.7)

Table 3 Prescribed medication and clinical diagnoses recorded at baseline in CPRD validation cohort and QResearch comparison cohort

Prescribed medication	*CPRD men (%)*	*CPRD women (%)*	*QResearch men (%)*	*QResearch women (%)*
antidepressants	101553 (6.4)	235797 (14.0)	178532 (5.4)	398018 (11.5)
antipsychotics	33884 (2.1)	79514 (4.7)	47464 (1.4)	92307 (2.7)
antiplatelets	97475 (6.1)	92816 (5.5)	160910 (4.9)	153405 (4.4)
oral NSAIDs	246515 (15.5)	346416 (20.6)	396026(12.0)	556644 (16.1)
tamoxifen	n/a	9231 (0.5)	n/a	18343 (0.5)
oestrogen only Hormone replacement therapy	n/a	119373 (7.1)	n/a	208333 (6.0)
oral corticosteroids	45597 (2.9)	71352 (4.2)	54354 (1.6)	88205 (2.6)
oral contraceptive pill	n/a	174287 (10.4)	n/a	332696 (9.6)
Recorded Diagnoses				
congestive cardiac failure	15836 (1.0)	19707 (1.2)	24965 (0.8)	28852 (0.8)
atrial fibrillation	20125 (1.3)	20102 (1.2)	33499 (1.0)	32580 (0.9)
coronary heart disease	80377 (5.1)	57703 (3.4)	130220 (3.9)	88606 (2.6)
cardiovascular disease	101430 (6.4)	83167 (4.9)	165495 (5.0)	130214 (3.8)
peripheral vascular disease	17029 (1.1)	13101 (0.8)	25004 (0.8)	17078 (0.5)
venous thromboembolism	15072 (0.9)	23090 (1.4)	27086 (0.8)	40813 (1.2)
rheumatoid or SLE	7455 (0.5)	19010 (1.1)	21453 (0.6)	48447 (1.4)
rheumatoid arthritis	7243 (0.5)	17468 (1.0)	21142 (0.6)	45542 (1.3)
SLE	228 (0.0)	1756 (0.1)	351 (0.0)	3374 (0.1)
type 1 diabetes	6238 (0.4)	4924 (0.3)	12029 (0.4)	9612 (0.3)
type 2 diabetes	51634 (3.2)	43271 (2.6)	95401 (2.9)	79654 (2.3)
treated hypertension	123584 (7.8)	161709 (9.6)	210516 (6.4)	267076 (7.7)
chronic renal disease	3968 (0.2)	4082 (0.2)	8550 (0.3)	8995 (0.3)
moderate/severe kidney failure	14107 (0.9)	9500 (0.6)	30407 (0.9)	21509 (0.6)
severe kidney failure	1603 (0.1)	1125 (0.1)	3641 (0.1)	2672 (0.1)
renal stones	13415 (0.8)	6443 (0.4)	37422 (1.1)	29204 (0.8)
inflammatory bowel disease	8962 (0.6)	10208 (0.6)	17762 (0.5)	19502 (0.6)
dementia	6686 (0.4)	16634 (1.0)	12872 (0.4)	30497 (0.9)
parkinsons disease	4546 (0.3)	4676 (0.3)	6830 (0.2)	6611 (0.2)
epilepsy or anticonvulsants	47170 (3.0)	71171 (4.2)	56516 (1.7)	61561 (1.8)
cancer	26866 (1.7)	43908 (2.6)	51649 (1.6)	79326 (2.3)
liver disease	3959 (0.2)	2893 (0.2)	9947 (0.3)	6410 (0.2)
chronic liver disease or pancreatitis	5521 (0.3)	4051 (0.2)	13069 (0.4)	8729 (0.3)
oesophageal varices	469 (0.0)	340 (0.0)	1626 (0.0)	1388 (0.0)
prior haemorrhage	97562 (6.1)	79765 (4.7)	203278 (6.2)	147533 (4.3)
malabsorption	7343 (0.5)	9375 (0.6)	21042 (0.6)	26002 (0.8)
endocrine diseases	3082 (0.2)	14097 (0.8)	6026 (0.2)	27731 (0.8)
COPD	24029 (1.5)	20737 (1.2)	41281 (1.2)	34785 (1.0)
asthma or COPD	142974 (9.0)	169503 (10.1)	273768 (8.3)	310027 (9.0)
history of falls	28878 (1.8)	61905 (3.7)	34584 (1.0)	67465 (2.0)
prior fracture	24265 (1.5)	45752 (2.7)	62092 (1.9)	89000 (2.6)
varicose vein surgery	18979 (1.2)	47012 (2.8)	35651 (1.1)	85602 (2.5)
emergency admissions or hip op	3483 (0.2)	5266 (0.3)	3335 (0.1)	5508 (0.2)

Table 4 Numbers of patients eligible for each score in the CPRD validation cohort

Risk Score	*Clinical outcome*	*Eligible age range*	*exclusion criteria*	*total in age range*	*total with exclusions*	*total eligible for analysis*
QDiabetes	Type 2 diabetes	25-84	type 1 or 2 diabetes at study entry	3,177,192	99,189	3,078,003
QStroke	ischaemic stroke	25-84	existing stroke or anticoagulants at study entry	3,177,192	70,961	3,106,231
QRISK2	cardiovascular disease	25-84	existing CVD or statins at study entry	3,177,192	232,722	2,944,470
QThrombosis	thromboembolism	25-84	existing VTE or anticoagulants at study entry	3,177,192	53,904	3,123,288
QFracture	fractured neck of femur	30-99	none except age	2,852,381	0	2,852,381
QFracture	osteoporotic fracture	30-99	none except age	2,852,381	0	2,852,381
QKidney	moderate or severe kidney failure	35-74	existing moderate or severe kidney failure	2,069,572	10,518	2,059,054
QKidney	severe kidney failure	35-74	existing severe kidney failure	2,069,572	1,930	2,067,642
QBleed	upper gastro-intestinal bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413
QBleed	intracranial bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413

*entry date was 01.01.1998 except for upper GI bleed and intracranial bleed where entry date was 01.01.2007

Table 5 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in women

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	48,143	99.88	4.13 (4.10 to 4.17)	102,544	99.93	4.33 (4.31 to 4.36)	0.95
	GP or ONS	48,203	n/a	4.13 (4.10 to 4.17)	102,618	n/a	4.34 (4.31 to 4.36)	0.95
Ischaemic stroke	GP data	32,283	90.64	2.45 (2.42 to 2.48)	63,582	90.22	2.45 (2.44 to 2.47)	1.00
	GP or ONS	35,617	n/a	2.62 (2.59 to 2.64)	70,477	n/a	2.70 (2.68 to 2.72)	0.97
Cardiovascular disease	GP data	55,833	85.71	**5.41 (5.37 to 5.46)**	107,412	84.96	4.32 (4.30 to 4.35)	1.25
	GP or ONS	65,143	n/a	**6.32 (6.27 to 6.37)**	126,433	n/a	5.03 (5.01 to 5.06)	1.26
	GP or ONS or HES	**69,202**	n/a	**6.72 (6.67 to 6.77)**	140,510	n/a	5.63 (5.60 to 5.66)	1.19
Thromboembolism	GP data	18,199	91.1	1.52 (1.49 to 1.54)	35,971	90.55	1.46 (1.44 to 1.47)	1.04
	GP or ONS	19,978	n/a	1.64 (1.62 to 1.67)	39,727	n/a	1.60 (1.58 to 1.62)	1.03
Fractured neck of femur	GP data	17,529	99.98	1.32 (1.30 to 1.34)	34,821	99.99	1.40 (1.39 to 1.42)	0.94
	GP or ONS	17,533	n/a	1.32 (1.30 to 1.34)	34,825	n/a	1.40 (1.39 to 1.42)	0.94
Osteoporotic fracture	GP data	34,528	n/a	2.89 (2.58 to 3.20)	81,334	n/a	3.63 (3.61 to 3.66)	0.80
mod /severe kidney failure	GP data	19,902	n/a	2.06 (1.76 to 2.36)	48,665	n/a	2.81 (2.78 to 2.83)	0.73
severe kidney failure	GP data	1,737	n/a	0.18 (0.09 to 0.27)	4,150	n/a	0.24 (0.24 to 0.25)	0.74

Table 6 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in men

	CPRD				*QResearch*			
outcome1	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	60,731	99.92	5.84 (5.79 to 5.89)	128,234	99.94	5.97 (5.94 to 6.00)	0.98
	GP or ONS	60,782	n/a	5.84 (5.80 to 5.89)	128,317	n/a	5.98 (5.94 to 6.01)	0.98
ischaemic stroke	GP data	32,223	93.55	3.17 (3.14 to 3.20)	63,480	92.85	3.10 (3.08 to 3.13)	1.02
	GP or ONS	34,443	n/a	3.33 (3.30 to 3.37)	68,366	n/a	3.37 (3.34 to 3.40)	0.99
Cardiovascular disease	GP data	70,283	86.7	7.38 (7.33 to 7.44)	137,136	86.12	7.12 (7.08 to 7.16)	1.03
	GP or ONS	81,068	n/a	8.52 (8.46 to 8.58)	159,240	n/a	8.37 (8.33 to 8.41)	1.02
	GP or ONS or HES	84,620	n/a	8.90 (8.84 to 8.96)	174,405	n/a	9.17 (9.13 to 9.21)	0.97
thromboembolism	GP data	15,655	92.32	1.49 (1.46 to 1.51)	31,503	92.22	1.44 (1.43 to 1.46)	1.03
	GP or ONS	16,958	n/a	1.61 (1.59 to 1.63)	34,161	n/a	1.57 (1.56 to 1.59)	1.02
fractured neck of femur	GP data	5,706	99.98	0.65 (0.63 to 0.67)	12,435	99.98	0.71 (0.70 to 0.73)	0.91
	GP or ONS	5,707	n/a	0.65 (0.63 to 0.67)	12,438	v	0.71 (0.70 to 0.73)	0.91
osteoporotic fracture	GP data	11,169	n/a	1.29 (1.05 to 1.52)	28,555	n/a	1.54 (1.52 to 1.55)	0.84
Mod/severe kidney failure	GP data	37,597	n/a	4.88 (4.37 to 5.38)	86,649	n/a	5.82 (5.78 to 5.85)	0.84
severe kidney failure	GP data	3,472	n/a	0.54 (0.38 to 0.71)	7,372	n/a	0.47 (0.46 to 0.48)	1.15

Table 7 Performance of QPrediction scores on the CPRD validation cohort compared with published results for the QResearch validation cohort

		CPRD	*CPRD*	*QResearch*	*QResearch*
	statistic	**women mean (95%CI)**	**men mean (95%CI)**	**women mean (95%CI)**	**men mean (95%CI)**
QDiabetes 2013 (type 2 diabetes)[29]	ROC	0.846 (0.844 to 0.848)	0.818 (0.816 to 0.82)	0.853 (0.851 to 0.856)	0.837 (0.835 to 0.840)
	R^2 (%)	49.6 (49.2 to 50.1)	45.7 (45.3 to 46.2)	50.8 (50.3 to 51.4)	48.1 (47.6 to 48.6)
	D statistic	2.032 (2.015 to 2.049)	1.879 (1.863 to 1.895)	2.081 (2.058 to 2.104)	1.971 (1.951 to 1.991)
QKidney -2010[5] (moderate or severe kidney failure)	ROC	0.875 (0.87 to 0.879)	0.88 (0.878 to 0.883)	0.877 (0.873 to 0.880)	0.878 (0.874 to 0.882)
	R^2 (%)	58.3 (57.8 to 58.7)	57.5 (57.1 to 57.8)	56.45 (55.40 to 57.50)	58.29 (55.31 to 61.26)
	D statistic	2.418 (2.394 to 2.442)	2.379 (2.361 to 2.397)	2.33 (2.28 to 2.40)	2.42 (2.28 to 2.56)
QKidney -2010 (severe kidney failure)[5]	ROC	0.839 (0.822 to 0.855)	0.851 (0.84 to 0.862)	0.843 (0.825 to 0.860)	0.846 (0.829 to 0.862)
	R^2 (%)	51.4 (49.5 to 53.2)	53.8 (52.6 to 55.1)	55.39 (52.59 to 58.18)	56.65 (53.94 to 59.35)
	D statistic	2.103 (2.025 to 2.182)	2.21 (2.154 to 2.266)	2.28 (2.15 to 2.41)	2.34 (2.21 to 2.47)
QRISK2-2014[27] (cardiovascular disease)	ROC	0.883 (0.882 to 0.884)	0.859 (0.858 to 0.861)	0.892 (0.892 to 0.895)	0.871 (0.869 to 0.873)
	R^2 (%)	56.4 (56.1 to 56.7)	50.9 (50.6 to 51.2)	58.8 (58.4 to 59.1)	53.3 (52.9 to 53.7)
	D statistic	2.328 (2.313 to 2.343)	2.085 (2.071 to 2.098)	2.443 (2.423 to 2.463)	2.188 (2.171 to 2.205)
QStroke-2013[2] (ischaemic stroke or TIA)	ROC	0.882 (0.88 to 0.883)	0.869 (0.867 to 0.87)	0.877 (0.875 to 0.879)	0.866 (0.864 to 0.868)
	R^2 (%)	58.4 (58.1 to 58.8)	55.3 (54.9 to 55.7)	57.3 (56.8 to 57.8)	55.1 (54.6 to 55.7)
	D statistic	2.427 (2.408 to 2.446)	2.278 (2.259 to 2.297)	2.37 (2.35 to 2.40)	2.27 (2.24 to 2.30)
QThrombosis-2010[6] (venous thromboembolism)	ROC	0.756 (0.751 to 0.761)	0.765 (0.760 to 0.770)	0.75 (0.74 to 0.76)	0.75 (0.74 to 0.76)
	R^2 (%)	35.3 (34.5 to 36.1)	34.5 (33.7 to 35.4)	32.78 (31.08 to 34.48)	33.51 (31.71 to 35.30)

	D statistic	1.512 (1.485 to 1.538)	1.486 (1.458 to 1.513)	1.43 (1.37 to 1.49)	1.45 (1.39 to 1.51)
QBleed-2014[13] (upper gastrointestinal bleed)	ROC statistic	0.775 (0.770 to 0.781)	0.759 (0.753 to 0.764)	0.766 (0.758 to 0.775)	0.747 (0.738 to 0.756)
	R^2 (%)	44.7 (43.6 to 45.9)	41.6 (40.5 to 42.8)	40.7 (38.9 to 42.6)	36.9 (35.1 to 38.7)
	D statistic	1.842 (1.798 to 1.885)	1.729 (1.687 to 1.771)	1.70 (1.63 to 1.76)	1.57 (1.51 to 1.63)
QBleed-2014[13] (intracranial bleed)	ROC statistic	0.808 (0.801 to 0.816)	0.789 (0.780 to 0.797)	0.847 (0.838 to 0.856)	0.812 (0.80 to 0.824)
	R^2 (%)	51.7 (50.1 to 53.3)	50.0 (48.3 to 51.7)	58.0 (56.0 to 60.0)	53.3 (51.1 to 55.4)
	D statistic	2.118 (2.051 to 2.186)	2.046 (1.977 to 2.116)	2.40 (2.30 to 2.50)	2.19 (2.09 to 2.28)
QFracture-2012[28] (fractured neck of femur)	ROC	0.89 (0.888 to 0.892)	0.872 (0.867 to 0.877)	0.893 (0.890 to 0.896)	0.875 (0.868 to 0.883)
	R^2 (%)	70.6 (70.2 to 71)	69.2 (68.5 to 70)	71.73 (71.10 to 72.30)	70.37 (69.25 to 71.49)
	D statistic	3.171 (3.139 to 3.203)	3.07 (3.016 to 3.124)	3.26 (3.21 to 3.31)	3.15 (3.06 to 3.24)
QFracture -2012[28] (osteoporotic fracture: hip, spine, wrist, hip)	ROC	0.817 (0.814 to 0.819)	0.768 (0.763 to 0.773)	0.790 (0.787 to 0.793)	0.711 (0.703 to 0.719)
	R^2 (%)	56.3 (55.8 to 56.7)	49.8 (48.9 to 50.7)	51.9 (51.2 to 52.6)	38.20 (36.89 to 39.57)
	D statistic	2.322 (2.301 to 2.343)	2.038 (2.002 to 2.075)	2.13 (2.10 to 2.15)	1.61 (1.56 to 1.66)

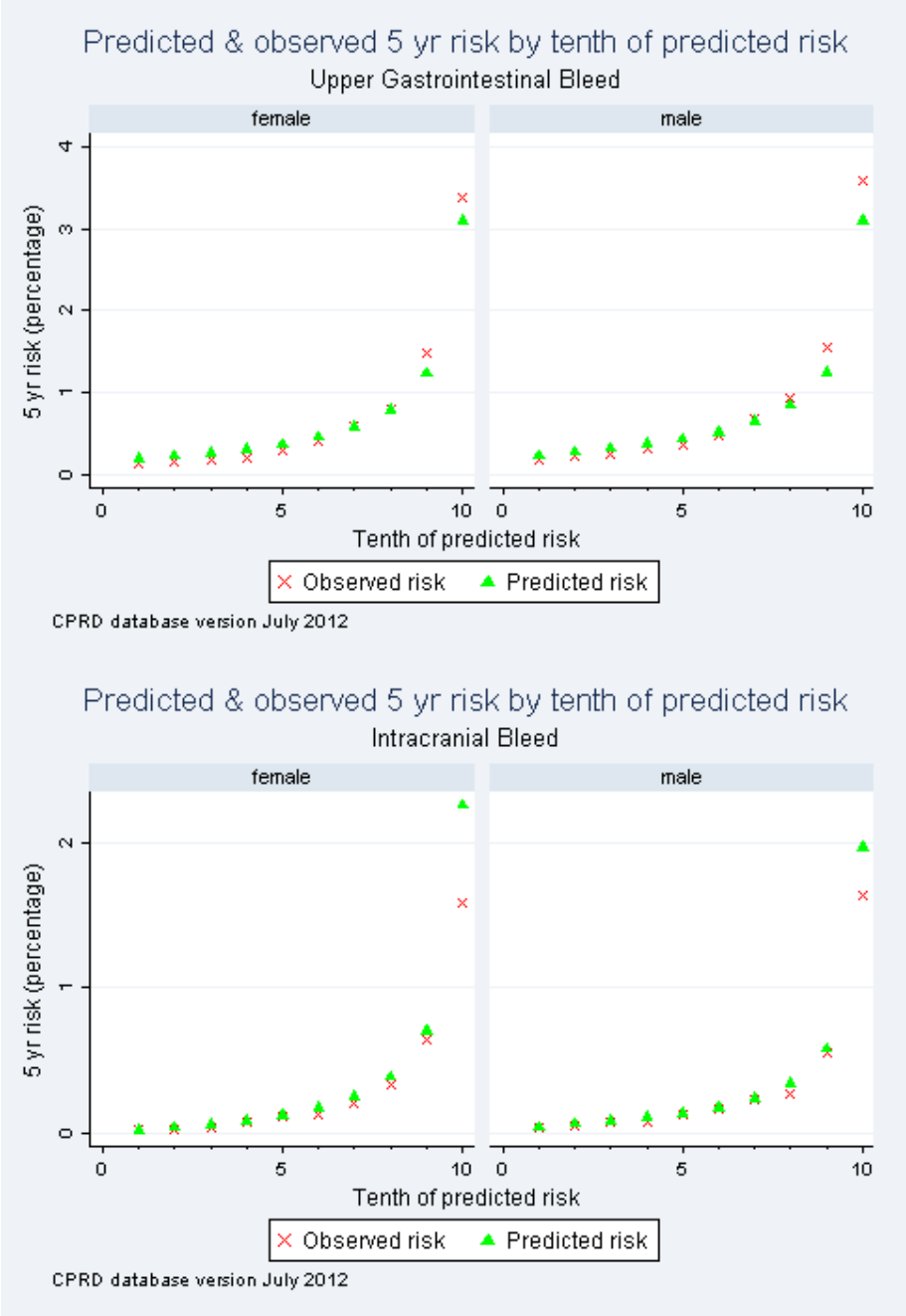
Notes on understanding validation statistics:

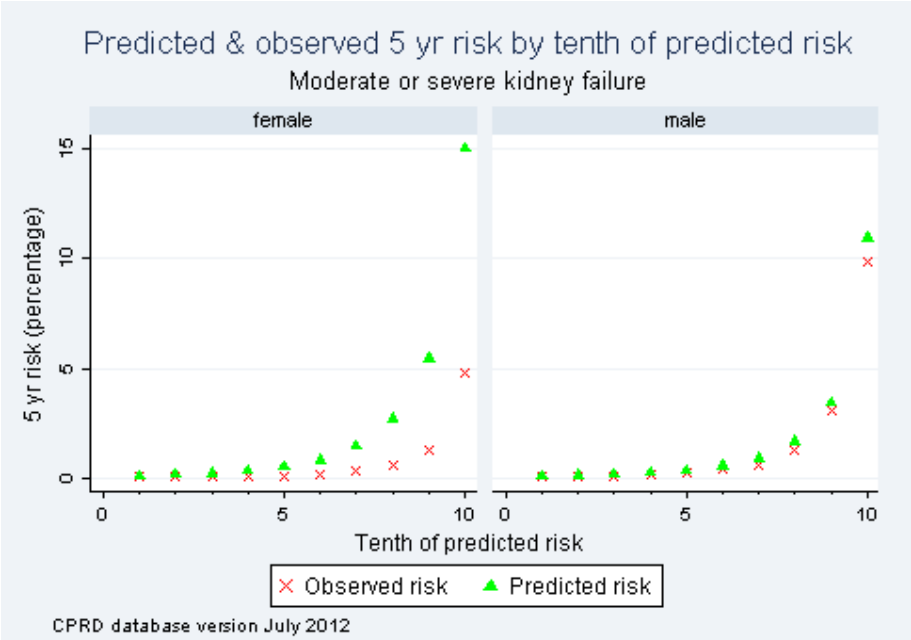
D statistic is a measure of discrimination - higher values indicate better discrimination
ROC statistic is a measure of discrimination - higher values indicate better discrimination
The R^2 statistic is a measure of explained variation - higher values indicate more variation is explained

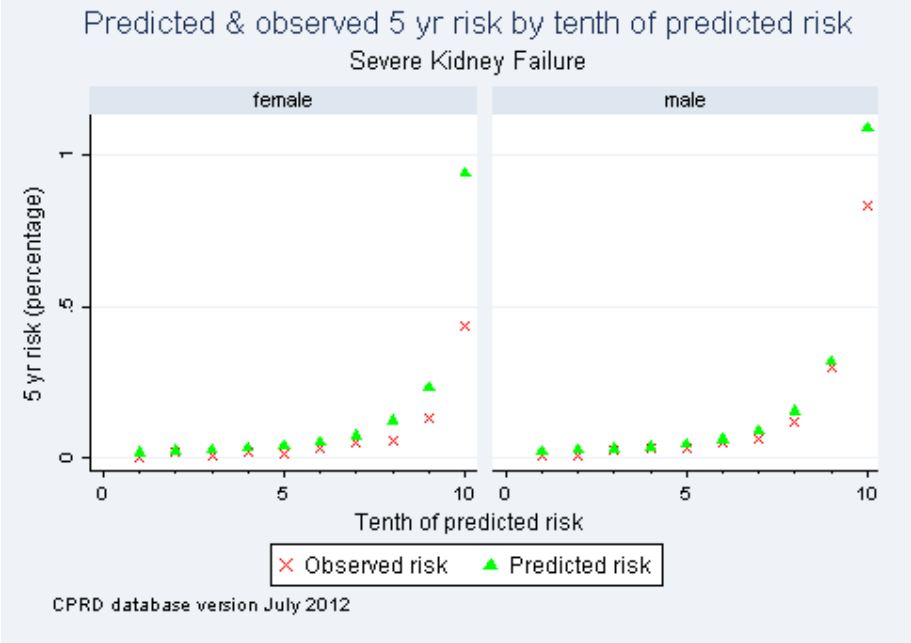
Table 8 Performance of each score for predicting the relevant outcome in the CPRD validation cohort. The cut off is the threshold of predicted risk for the top decile in the CPRD cohort.

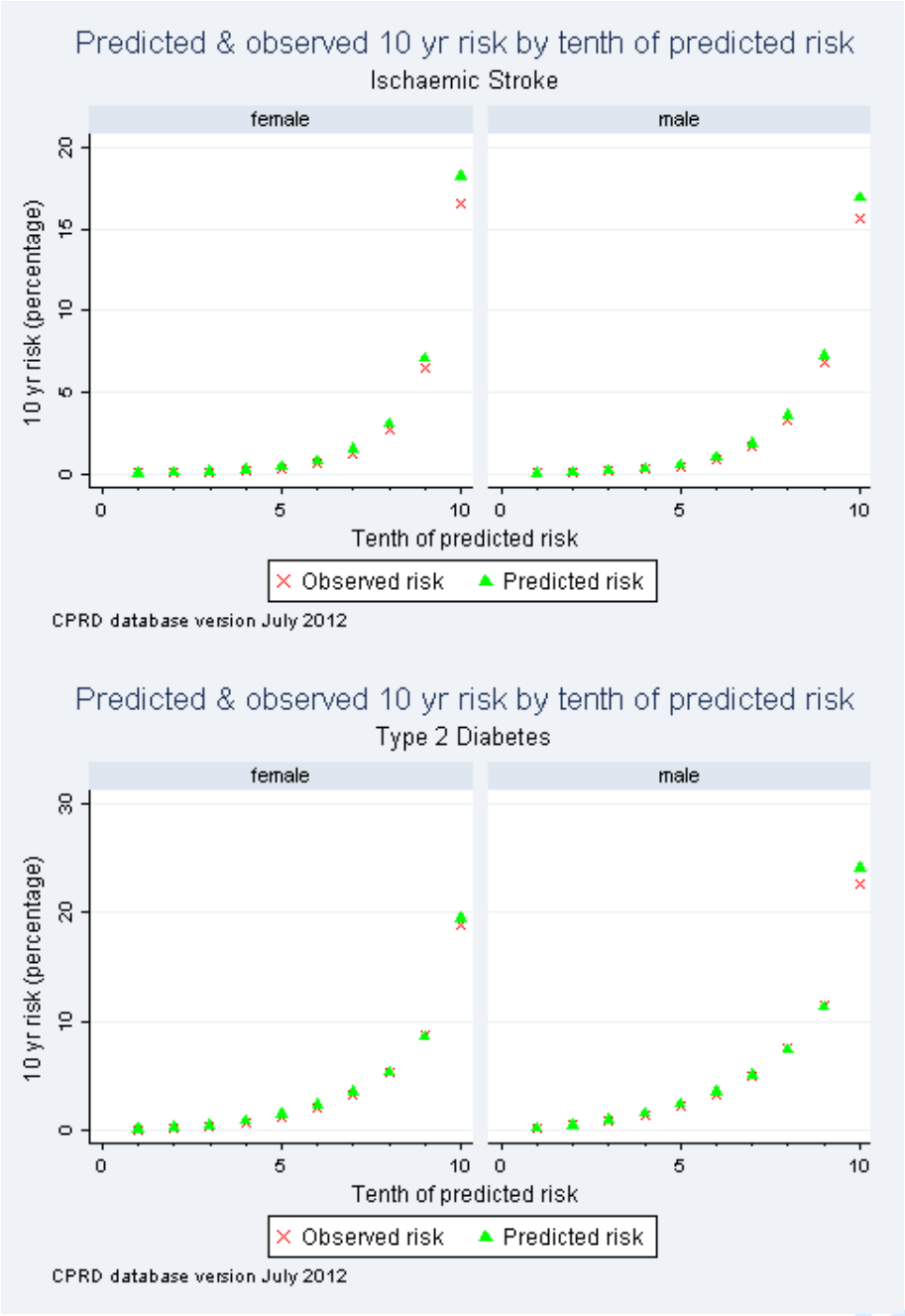
score	*outcome*	*duration*	*cut off (%) for top decile predicted risk*	*Sensitivity (%)*	*Specificity (%)*	*observed risk (%)*
QDiabetes	Type 2 diabetes	10 yr risk	13.0	44.8	91.0	20.8
QStroke	Ischaemic stroke	10 yr risk	10.5	54.7	90.8	16.1
QRISK2	cardiovascular disease	10 yr risk	20.7	49.9	91.9	31.8
QThrombosis	venous thromboembolism	5 yr risk	1.5	36.2	90.1	2.6
QKidney	moderate-severe kidney failure	5 yr risk	6.3	59.1	90.5	6.9
QKidney	severe kidney failure	5 yr risk	0.4	58.5	90.0	0.7
QBleed	upper GI bleed	5 yr risk	1.6	38.0	90.2	3.5
QBleed	intracranial bleed	5 yr risk	0.9	44.2	90.1	1.6
QFracture	fractured neck of femur	10 yr risk	3.7	66.5	90.4	9.4
QFracture	osteoporotic fracture	10 yr risk	7.8	49.6	90.5	13.1

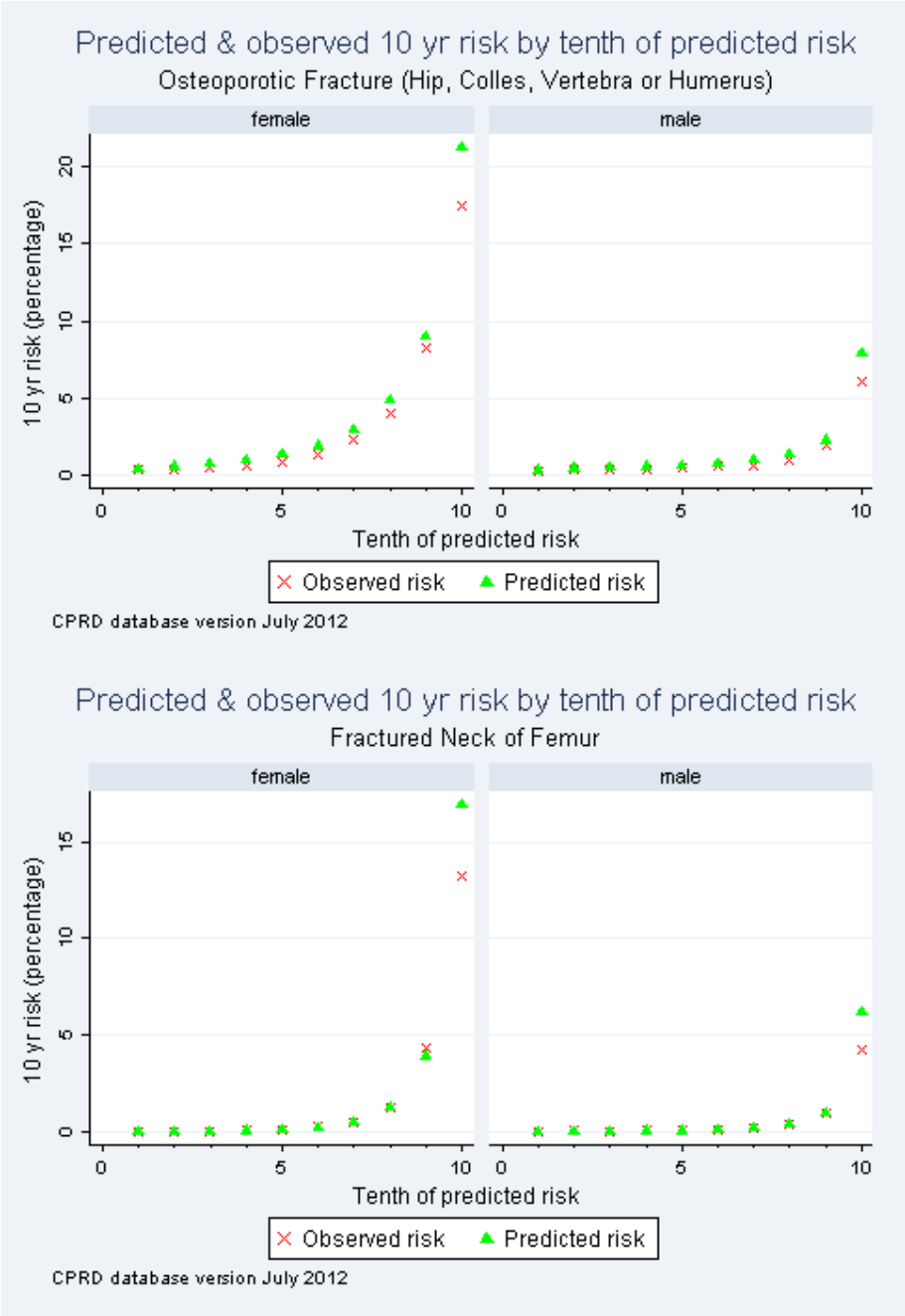
Figure 1 Calibration of each QPrediction score comparing the mean predicted risks with the observed risks in the CPRD cohort.

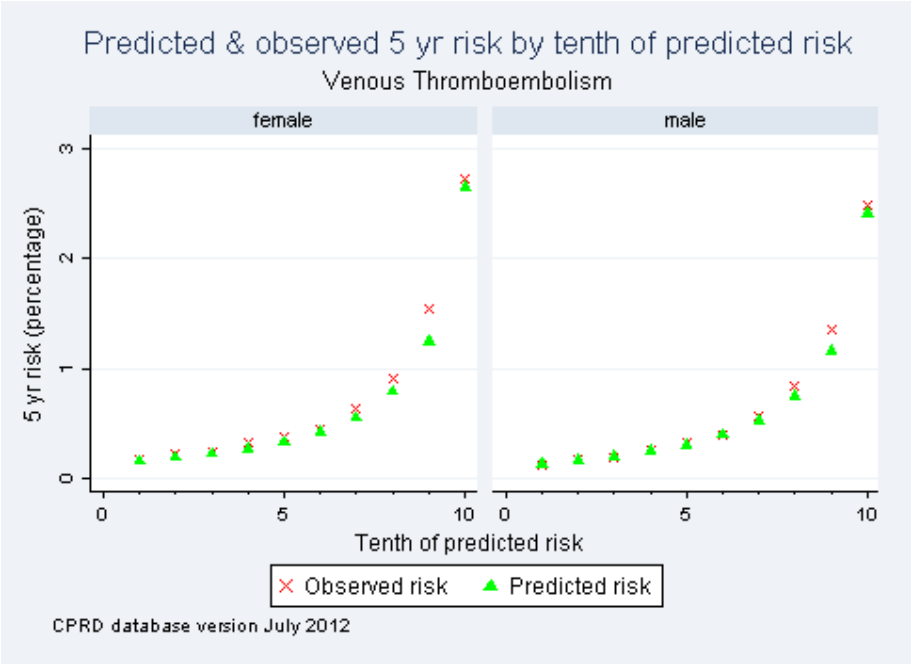












Web extra table 1. Numbers of patients in CPRD and QResearch by geographical area

	CPRD	*Col %*	*QResearch*	*Col %*
East Midlands	109,428	3.3	500,970	7.4
East of England	397,008	12.1	528,379	7.8
London	563,353	17.2	1,711,956	25.3
North East	59,558	1.8	269,695	4.0
North West	474,457	14.5	830,047	12.3
South Central	411,571	12.6	696,070	10.3
South East	362,319	11.1	545,811	8.1
South West	397,735	12.2	700,041	10.4
West Midlands	348,614	10.7	589,548	8.7
Yorks & Humber	147,469	4.5	386,132	5.7
Total	**3,271,512**	**100.00**	**6,758,649**	**100.00**

Web extra table 2 Number of cases of upper gastrointestinal bleed and intracranial bleed on CPRD and QResearch (one third sample database). Incidence rates per 1000 pyrs have been standardised to the QResearch population in 5 year bands.

	CPRD validation		*QResearch validation*	
	cases on	**age standardised Incidence rate per 1000pyrs**	**cases**	**Age standardised Incidence rate per 1000pyrs**
no anticoagulants	13,314	1.41 (1.39 to 1.43)	6,447	1.33 (1.30 to 1.36)
anticoagulants	359	6.70 (4.06 to 9.34)	153	6.10 (3.20 to 8.98)
Intracranial bleed				
no anticoagulants	5,190	0.53 (0.51 to 0.54)	2,716	0.56 (0.54 to 0.58)
anticoagulants	233	2.45 (1.23 to 3.68)	104	2.87 (1.11 to 4.39)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page in manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2 abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 2, page 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 2, 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5 & 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*Pages 6*
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Page 8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Pages 6, 7, 8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Pages 7, 8
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6
		(*c*) Explain how missing data were addressed	Page 7
		(*d*) If applicable, explain how loss to follow-up was addressed	
		(*e*) Describe any sensitivity analyses	Page 7,8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 9
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Page 9 tables 2, 3
		(b) Indicate number of participants with missing data for each variable of interest	table 2,3
		(c) Summarise follow-up time (eg, average and total amount)	Table5
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9 table 5
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were	Pages 10, 11 Table 3, table 5

		included	
		(*b*) Report category boundaries when continuous variables were categorized	Page 6
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Page 11
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Page 11
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 12- 13,
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 13
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 15

BMJ Open

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005809.R1
Article Type:	Research
Date Submitted by the Author:	01-Jul-2014
Complete List of Authors:	Hippisley-Cox, Julia; University of Nottingham, ; ClinRisk Ltd, Coupland, Carol; University of Nottingham, Division of Primary Care Brindle, Peter; Bristol Clinical Commissioning Group,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, General practice / Family practice, Health informatics
Keywords:	qresearch, cprd, Epidemiology < TROPICAL MEDICINE, qrisk2, prognosis, validation

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Revised Research paper submitted to BMJ Open
Reference-2014-005809
June 2014

Authors

Julia Hippisley-Cox	Professor of Clinical Epidemiology & General Practice[1]
Carol Coupland	Associate Professor and Reader in Medical Statistics[1]
Peter Brindle	Research and Development Programme Director[2]

Institutions

[1]Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham, NG2 7RD.

[2]Avon Primary Care Research Collaborative, Bristol Clinical Commissioning Group, South Plaza, Marlborough Street, Bristol BS1 3NX.

Guarantor & Author for correspondence - Julia Hippisley-Cox

Email: Julia.hippisley-cox@nottingham.ac.uk

Telephone: 0115 8466915

Fax: 0115 8466904

Abstract

Objectives To validate the performance of a set of risk prediction algorithms developed using the QResearch database, in an independent sample from general practices contributing to the Clinical Research Data Link (CPRD).

Setting Prospective open cohort study using practices contributing to the CPRD database and practices contributing to the QResearch database.

Participants The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices between 01 Jan 1998 and 31 July 2012. The validation statistics for QResearch were obtained from the original published papers which used a one third sample of practices separate to those used to derive the score. A cohort from QResearch was used to compare incidence rates and baseline characteristics and consisted of 6.8 million patients from 753 practices registered between 01 Jan 1998 and until 31 July 2013.

Outcome measures Incident events relating to seven different risk prediction scores: QRISK2 (cardiovascular disease); QStroke (ischaemic stroke); QDiabetes (type 2 diabetes); QFracture (osteoporotic fracture and hip fracture); QKidney (moderate and severe kidney failure); QThrombosis (venous thromboembolism); QBleed (intracranial bleed and upper gastrointestinal haemorrhage). Measures of discrimination and calibration were calculated.

Results Overall, the baseline characteristics of the CPRD and QResearch cohorts were similar though QResearch had higher recording levels for ethnicity and family history. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with the published results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch and the ROC value was 0.85 on both databases. The scores were well calibrated in CPRD.

Conclusion Each of the algorithms performed practically as well in the external independent CPRD validation cohorts as they had in the original published QResearch validation cohorts.

Strengths and limitations of this study

- This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system from that used to derive the algorithms.
- The discrimination and calibration statistics for each score were very similar in CPRD to those published from validation cohorts from QResearch. This supports their potential utility in the general population of patients in primary care.
- A strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice.
- The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible.
- Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using these algorithms in primary care.

1 Introduction

In the last 7 years, we have developed a series of risk prediction algorithms using the QResearch database. QResearch is a large research database containing pseudonymised individual level data from over 700 general practices using the EMIS clinical system. The QResearch database consists of data collected from primary care (coded information on socio-demographic characteristics, diagnoses, symptoms, smoking/alcohol, clinical measurements, laboratory values, prescriptions and referrals) which has been linked to cause of death, hospital episodes and cancer registrations at individual patient level.

The algorithms predict outcomes such as cardiovascular disease(www.qrisk.org)[1], stroke (www.qstroke.org)[2], type 2 diabetes (www.qdiabetes.org)[3], osteoporotic fracture (www.qfracture.org)[4], moderate or severe kidney disease (www.qkidney.org)[5], venous thrombo-embolism (www.qthrombosis.org)[6], and emergency hospital admission (www.qadmissions.org)[7]. Generally, the "QPrediction" algorithms have been designed to systematically identify patients in primary care at high risk of a serious clinical outcome for whom further intervention to lower risk of that outcome might be possible. They are also designed to quantify absolute risk of serious outcomes in a way which patients can understand and which might help guide lifestyle and management decisions. A number of these algorithms are now integrated into GP clinical computer systems, included in national guidelines[1 4] and are in daily use across the NHS [1 3 8].

The algorithms were originally developed using a random two thirds sample of practices contributing to the QResearch database and validated on the remaining third. Whilst this represents a physically discrete population of patients and practices for validation, the practices all use the same clinical computer system (EMIS), which is in use in 53% of UK practices. A more stringent test of performance is to validate the algorithms on a fully external database derived from practices using a different but commonly used primary care computer system. This would help determine whether the predictions from the algorithms are likely to generalise to the whole population in England. Whilst some of the algorithms have been validated by an independent team using the THIN primary care database[9-12], there are currently no published validations of the algorithms using a primary care database which is routinely linked to mortality data in the same way as QResearch.

We therefore decided to validate the various QPrediction Scores using another database known as the Clinical Research Data Link (CPRD). The General Practice Research database (GPRD) was originally set up in 1988 and is of similar nature to QResearch although it is derived from practices using a different clinical computer system (Vision, which is used by 20% of GPs). It was extended to include linked mortality data and data from secondary care and was renamed the Clinical Research Data Link (CPRD) in 2012. Our secondary objective was to compare the ascertainment of incident clinical events recorded in GP data alone with that

recorded in either GP data or the linked mortality data in both the CPRD and QResearch.

2 Methods

2.1 CPRD Study population

For the validation using CPRD, we identified an open cohort of patients aged 25-99 years at entry to the cohort and followed this cohort up until 31st July 2012 (the latest date for which linked data were available at the time of analysis). We restricted the CPRD cohort to 357 practices in England which had linked ONS mortality and hospital admissions data. For each patient we determined an entry date to the cohort, which was the latest of the following dates: 25th birthday, date of registration with the practice plus one year, date on which the practice computer system was installed plus one year, and the beginning of the study period (01 January 1998). Patients were censored at the earliest date of the relevant outcome, de-registration with the practice, last upload of computerised data or the study end date (31 July 2012).

For the assessment of the two QBleed outcomes (intracranial bleed and upper gastrointestinal haemorrhage) we used a later cohort entry date of 01.01.2007 for comparability with the equivalent study period for the derivation of the algorithm on QResearch[13].

2.2 QResearch study population

For comparison of the validation statistics (ROC, D and R2 statistics), we extracted the original published values from the papers which had been calculated using a one third sample of practices from QResearch which were independent from the two thirds of practices used to derive the scores.

For comparison of the baseline characteristics, incidence rates and ascertainment rates we used the latest version of the QResearch database which is currently available (QResearch 38, 31st Dec 2013). We identified an open cohort in the same way as for CPRD, using all of the QResearch practices in England, and with follow-up until 31 July 2013.

2.3 Inclusion and exclusion criteria

For both databases, we excluded patients without a Townsend score (an area based measure of material deprivation derived from the post code) and temporary residents. For each score we then identified patients who were eligible to have the score calculated according to the relevant inclusion and exclusion criteria as summarised in Table 4

2.4 Risk scores included in validation

We validated the following risk prediction scores on CPRD:

1. QDiabetes - 10 year risk of type 2 diabetes[3]
2. QRISK2 – 10 year risk of cardiovascular disease[1]
3. QStroke – 10 year risk of stroke or transient ischaemic attack (TIA)[5]
4. QFracture - 10 year risk of hip or osteoporotic fracture[4]
5. QThrombosis – 5 year risk of venous thrombo-embolism (VTE)[6]
6. QBleed – 5 year risk of upper gastrointestinal haemorrhage and intracranial haemorrhage[13]
7. QKidney – 5 year risk of moderate-severe kidney disease[5]

2.5 Clinical outcomes

We identified the relevant clinical outcome using the same definition as had been applied in the original derivation of the risk scores using QResearch. The data sources used to identify the clinical outcomes had varied over the six years during which the original studies had been undertaken due to the changing availability of linked hospital and mortality data over that time. In 2008, the QResearch database was linked to mortality records for 1997 onwards. In 2013, the QResearch database was linked to hospital admissions records with data for patients from 1998 onwards. For the latest updated version of QRISK2 (QRISK2-2014), the outcome was identified by the presence of the relevant Read code on the GP record or an ICD10 code recorded on the linked mortality record or on the linked hospital admissions record. For QStroke, QDiabetes, QFracture and QThrombosis, the outcome was identified either by the presence of the relevant Read code recorded on the GP record or an ICD10 code recorded on the linked mortality record. For QKidney, the outcome was identified solely from information recorded in the GP record as in the original study as it required blood test values which were only present in the GP record. For QBleed, the outcome was identified in CPRD from events recorded either on the linked hospital admissions database or the linked mortality record in order to identify the events most likely to have serious clinical consequences for the patient.

We determined case ascertainment for each clinical outcome on both databases, by calculating the proportion of cases recorded on the GP record out of the total number of cases recorded on either the GP record or linked mortality record. We calculated the age standardised incidence rates of each outcome based on outcomes recorded on (1) the GP record alone and on (2) the GP record or linked mortality; (3) GP or linked mortality or hospital records. We standardised CPRD rates to the age distribution of the QResearch population in five year bands to ensure comparability.

2.6 Risk factors and missing values

We extracted data from CPRD for all the predictor variables included in one or more of the different algorithms using the same definitions as those used in the original

QResearch studies to enable a direct comparison of the results. We developed a mapping between the Read and medication reference tables to identify the equivalent code in each database. This included the following variables recorded at entry to the cohort:

- **demographics** – age (continuous), sex, ethnicity (9 categories – white, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean ,Black African, Chinese, Other ethnic group), resident in care home, material deprivation (as measured by the Townsend score)
- **clinical values** - smoking status (non-smoker, ex-smoker, light smoker [1-9 cigarettes/day], moderate smoker [10-19 cigarettes/day], heavy smoker [20+ cigarettes/day];body mass index, systolic blood pressure, alcohol consumption-none drinker, trivial (<1u/day), light (1-2u/day), moderate (3-6u/day), heavy (7-9u/day), very heavy (>9 day).
- **laboratory results** –cholesterol/HDL ratio, platelets
- **family history**- family history of osteoporosis or hip fracture in a first degree relative, coronary heart disease in first degree relative under the age of 60 years, diabetes in a first degree relative.
- **chronic diseases** – congestive cardiac failure, atrial fibrillation, coronary heart disease, cardiovascular disease, periperal vascular disease, venous thromboembolism, diabetes, rheumatoid arthritis, SLE, hypertension, renal disease, renal stones, inflammatory bowel disease, dementia, Parkinson's disease, epilepsy, cancer, chronic liver disease or pancreatitis, oesophageal varices, prior haemorrhage, malabsorption endocrine diseases, asthma or COPD, history of falls, prior osteoporotic fracture, varicose vein surgery, emergency admissions or hip surgery in last 6 months.
- **prescribed medication**- anticoagulants, antidepressants, antipsychotics, antiplatelets, oral NSAIDs, tamoxifen, oestrogen containing hormone replacement therapy (BNF chapter 6.4.1.1), systemic corticosteroids, combined oral contraceptive.

The combination of predictor variables required for each risk score varied with the score being validated as shown in Table 1. We used the clinical value recorded closest to the date on which the patient entered the study for body mass index, systolic blood pressure, smoking status, platelets, and total and HDL cholesterol. Patients were considered to be exposed to medication at entry to the cohort if they had at least 2 prescriptions for the relevant medication prescribed prior to the study entry date with the most recent one occurring within 28 days of the study entry date.

2.7 Townsend scores

We used the Townsend score evaluated at output area as a proxy for material deprivation. The CPRD dataset differs from the QResearch dataset in that each patient in the CPRD dataset is allocated to a tenth of deprivation (as measured by the Townsend score) and only the category number is provided. In contrast, each

patient in the QResearch dataset is allocated the individual Townsend score corresponding to their output area of residence (i.e. continuous data). In order to calculate risk scores in the CPRD cohort, we used the median value for each tenth as supplied by CPRD. Patients with missing Townsend scores were excluded from the cohorts.

2.8 Discrimination and calibration statistics

We used chained equations with the ICE procedure in STATA[14] to perform multiple imputation to replace missing values for body mass index, systolic blood pressure, smoking status, alcohol, and total and HDL cholesterol. We created five multiply imputed datasets and used Rubin's rules to combine effect estimates and standard errors to allow for the uncertainty due to imputing missing data[15 16].

We applied the algorithm for each score to eligible patients in the CPRD study cohort to obtain predicted risks for each of the relevant clinical outcomes. We calculated the estimated risk for eligible patients in the CPRD validation dataset over 5 years or 10 years depending on which score was used. We then tested the performance of each score in the CPRD cohort and compared it with the published results from the original QResearch validation cohorts.

In order to assess calibration (i.e. degree of similarity between predicted and observed risks), we calculated the mean predicted risk and the observed risk [17]obtained using the Kaplan-Meier estimate and compared the ratio of the mean predicted risk to the observed risk for patients in the validation cohort in each decile of predicted risk. We calculated the area under the Receiver Operator Curve (ROC) statistic to assess discrimination (i.e. ability of a risk prediction equation to distinguish between those who do and do not have an event during the follow-up period). We also calculated the D statistic[18] and an R squared statistic derived from the D statistic[19] which are measures of discrimination and explained variation appropriate for survival models. The D statistic has been developed as a new measure of discrimination specifically for censored survival data, higher values indicate improved discrimination, and an increase in the D statistic of at least 0.1 indicates an important difference in prognostic separation between different risk classification schemes. The R^2 statistic derived from the D statistic is a measure specific to censored survival data– it measures explained variation in time to the outcome event and higher values indicate more variation is explained[20]. We also repeated the assessment of discrimination by restricting the analysis for each score to patients without missing data for relevant clinical or laboratory measures used in the risk score (ie those with complete data for all predictor variables in the risk score).

We identified the proportion of patients in the CPRD validation cohort who were in the top decile of predicted risk and used this to calculate the sensitivity, specificity and observed risk at this threshold. We used the top decile for comparability across the scores and with previous studies though the choice of threshold for use in clinical practice will depend on the context and cost-effectiveness of relevant interventions. Analyses were conducted using Stata (version 13.1).

2.9 Sample size estimation

There is currently no clear guidance on sample size requirements for studies evaluating the performance (validation) of a multivariable risk score, but a commonly used rule-of-thumb is that it is desirable to seek a dataset with at least 100 patients with the outcome of interest. We used all the available data on the CPRD to maximize the power of the study.

3 Results

3.1 Study populations

The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices with linked data between 01 Jan 1998 and 31 July 2012. The QResearch cohort consisted of 6.8 million patients from 753 practices with linked data, registered between 01 Jan 1998 and until 31 July 2013. The numbers of patients in each geographical region are shown in Web Extra Table 1.

3.2 Baseline characteristics

Table 2 shows a comparison of the demographic characteristics for the CPRD and QResearch cohorts.

The QResearch population was marginally younger with 34.2% of women and 32.8% of men aged 24-34 years compared with 27.8% and 26.9% for CPRD.

3.2.1 Recording of ethnicity

QResearch had a higher proportion of patients with self-assigned ethnicity recorded compared with CPRD both overall (58.2% vs 38.1%) and in each of the 10 geographical areas within England (web extra table 2). We repeated the analysis restricting information on QResearch to that recorded prior to 31 July 2012 (for comparability with the calendar time available on CPRD). Of the 6,758,649 patients in the QResearch cohort, 3,856,244 (57.1%) had ethnicity recorded prior to this date.

3.2.2 Recording of family history

Recording of a positive family history of coronary heart disease and diabetes was more than twice as high in QResearch compared with CPRD. For example, for family history of coronary heart disease, 11.0% of patients had a value recorded for QResearch compared with 4.6% for CPRD (web extra table 2). Restricting information to that recorded prior to July 2012 for QResearch, then 6,758,649 (10.7%) had a positive family history of coronary heart disease recorded.

3.2.3 Recording of alcohol and smoking levels

Recording of alcohol levels was very similar in both QResearch and CPRD. For example, 82.1% of women had alcohol level recorded in both databases. Recording of smoking status was marginally higher in women compared with men in QResearch (93.2% vs 89.1%) and also CPRD (94.8% vs 90.8%).

3.2.4 Recording of clinical values

Recording of cholesterol/HDL ratio was marginally higher on QResearch compared with CPRD (40.1% vs 36.0%). Recording of body mass index and systolic blood pressure tended to be marginally higher on CPRD than QResearch. However the mean values for the various clinical values (BMI, systolic blood pressure, serum creatinine and cholesterol/HDL ratio) were extremely similar.

Table 3 shows prescribed medication and clinical diagnoses recorded in patients on or prior to entry to the study cohort. Overall, the prevalence of clinical diagnoses were similar on the two databases with CPRD having marginally higher prescribing rates.

The inclusion and exclusion criteria for each risk score are shown in Table 4 along with the numbers of patients eligible for each analysis on CPRD. For example, there were 3,177,192 patients aged 25-84 years. Of these, 99,189 had existing diabetes at baseline leaving 3,078,003 for the validation of QDiabetes. Table 4 also shows the numbers and percentage out of those eligible for inclusion with complete data for risk factors necessary for calculation of the score which would otherwise need to be imputed (i.e. laboratory or clinical values). The amount of missing data varies substantially between the scores with scores requiring multiple laboratory or clinical values (such as QRISK2) having the lowest levels of completeness.

3.2.5 Comparison between CPRD linked and unlinked data

Web extra Table 3 shows characteristics for CPRD cohort with linked data with CPRD cohort without linked data. The CPRD cohort with linked data tended to have higher recorded of ethnicity compared with the CPRD cohort without linked data (38.1% vs 28.4%). Recording of smoking, alcohol, body mass index, systolic blood pressure, cholesterol and platelets were all higher on the CPRD cohort with linked data than those without linked data.

3.3 Incidence rates of clinical outcomes

Table 5 shows the number of incident events for each clinical outcome in women recorded on GP data and those recorded on either GP data or cause specific mortality data for both the CPRD and QResearch cohorts. It also shows the age standardized incidence rates per 1000 person years. Table 6 shows the comparable information for men.

For example, there were 35,617 incident ischaemic stroke or TIA events for women on CPRD. Of these, 32,283 had been identified on the GP record with an additional 3,334 events identified on the linked ONS mortality record. The ascertainment of events on the GP record was therefore 32283/35617 i.e. 90.6%. For QResearch, there were 70,477 incident stroke events recorded on either the GP or linked ONS mortality record of which 63,572 had been identified on the GP record. The ascertainment was therefore 90.2%.

For thromboembolism in women, 91.1% of events recorded on either the GP or linked ONS mortality record on CPRD were identified on the GP record compared with 90.6% for QResearch. Similar results were obtained for men with levels of ascertainment between the two databases being extremely close suggesting similar recording patterns between the two groups of GP practices contributing to each database.

The age standardized incidence rates of events on CPRD tended to be marginally lower than those on QResearch as shown by the ratio of the CPRD rates to those in QResearch (Table 5). For example, the rate ratio for fractured neck of femur in women was 0.94 indicating that CPRD had a 6% lower incidence rate compared with QResearch. The effect was more marked for moderate or severe kidney failure where the incidence rates for CPRD were approximately 25% lower than those for QResearch in women and 16% lower in men.

The age standardized incidence rates of upper gastrointestinal haemorrhage and intracranial haemorrhage among patients prescribed anticoagulants and those not prescribed anticoagulants are shown in Web extra table 4. The rates are similar for CPRD and QResearch.

3.4 Validation statistics

Table 7 shows the discrimination statistics for each score in CPRD in men and women and also the published values from previous validations using QResearch. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch. The D statistic for women was 2.03 within CPRD compared with 2.08 for QResearch. The ROC value for women was 0.85 on both databases.

Of all the scores, QFracture (fractured neck of femur) had the best performance in men in CPRD with a ROC value of 0.89, R^2 value of 71% and D statistic of 3.17. The corresponding figures for QResearch in men were 0.89, 72% and 3.26.

QThrombosis had the lowest values for women in CPRD with an ROC value of 0.77, R^2 of 34.5 and D statistic of 1.49. The corresponding figures for women in QResearch were 0.75, 33.5 and 1.45.

Figure 1(a-j) compares the mean predicted risks and observed risks for each score across each tenth of predicted risk (1 representing the lowest risk and 10 the highest risk) and demonstrates that the models are generally well calibrated for patients on CPRD.

The QKidney score (moderate or severe kidney failure) showed the observed risk was lower than the predicted risk. This might indicate a degree of over prediction of the score. Alternatively, it could be related to the lower incidence rate of kidney failure observed among women on the CPRD compared with QResearch.

Web extra table 5 presents the ROC, D and R^2 statistic for each score restricted to patients from CPRD with complete recording of laboratory and risk factor data for each score. The results were very similar to the results obtained using multiply imputed dataset for the majority of scores except for QRISK2 and QStroke where values tended to be lower.

3.5 Performance for the top decile of risk.

Table 8 shows the sensitivity, specificity and observed risk for patients in the top decile of each score on CPRD. The observed risk is higher than the risk threshold value since this represents the observed risk within the top decile of predicted risk. For example, the cut off for the top tenth of risk for QFracture (fractured neck of femur) was a 10 year risk of 3.7%. At this threshold the sensitivity was 66.5%, specificity 90.4% and observed risk 9.4%. The results are similar to those obtained from QResearch (not shown).

4 Discussion

4.1 Summary of key findings

This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system (Vision system supplied by In Practice Systems) from the QResearch database (which is based on practices using EMIS clinical systems). Practices using the Vision system together with practices using EMIS make up approximately 75% of all the English general practices. The discrimination and calibration statistics for each score were remarkably similar in CPRD to those published from validation cohorts from QResearch. Our paper also provides updated information on a direct comparison between two of the world's largest general practice databases which have both been linked to mortality and second care data.

Before a clinical risk score can be reliably used in clinical practice, evidence is needed that it can successfully predict the intended outcome in groups of patients other than ones used to develop the score but similar to ones in whom the score might be used. Not all risk scores perform well in external samples – this can be due to deficiencies in the design or modelling methods used to derive the algorithm, if the model is over fitted or if there is an important predictor which is absent[21]. Other reasons for poor performance include differences between the setting of patients in the new and derivation samples, differences in how information is recorded and differences in patient characteristics[21]. It is for these reasons, that we have meticulously assembled the CPRD cohort using the same inclusion/exclusion criteria, definitions of predictor and outcome variables as in the original derivation studies. Any differences observed are therefore more likely to be due to capture of information and underlying population characteristics. In this study, we have found marginal differences in incidence rates between QResearch and CPRD and higher rates of recording of family history and ethnicity in QResearch though these have not been large enough to materially affect our results.

4.2 Strengths and limitations

One strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice. CPRD only had Townsend score for patients recorded for approximately half their practices (unlike QResearch where Townsend score is included for all practices) so we had to limit the validation cohort in CPRD for this analysis to those practices with linked Townsend scores. We undertook a comparison between patients registered with CPRD practices with and without linked data. We found marginally higher recording for ethnicity, smoking, alcohol, clinical values for the CPRD cohort with linked data compared with the unlinked data but similar characteristics for demographics, comorbidities,

medication and clinical values (results not shown) so we have no reason to believe this would have biased our results.

Another strength of general practice databases is the large volume of patients who tend to be representative of the general population. A limitation of routinely collected data is that not all patients will have all clinical and laboratory data recorded leading to missing data values in some of the parameters needed to calculate the risk scores. We have reported performance in all patients using multiple imputation to replace missing values and restricted to patients without missing values and found very similar results for the majority of algorithms tested. There was some degradation of performance associated with large amounts of missing data although not sufficient to affect our conclusion. The software used to implement QPrediction scores in clinical practice includes algorithms to estimate body mass index, systolic blood pressure and cholesterol/HDL ratio which can be used where relevant data is not recorded to generate an estimate risk score. The clinician can then enter the relevant data fields once the patient is assessed to calculate an actual risk score using recorded values.

The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible. The CPRD database uses the same clinical coding system as QResearch for clinical values (it uses Read version 2). However, there is a third clinical system in use in England (SystmOne) which uses a different coding system known as Clinical terms version 3(CTV3). Whilst there is a mapping between Read codes and CTV3, we have not tested the algorithms on a database using CTV3 in this study so are unable to draw conclusions regarding the generalisability of the results of the validation to practices using this system.

The quality of information on CPRD is likely to be good since previous studies have validated similar outcomes and exposures and found levels of completeness and accuracy to be good[22 23].

4.3 Comparison with other studies

The aim of this study was to validate a collection of QPrediction tools. The details of the derivation and first validation of each prediction tool have been separately published in the peer reviewed literature including information on definitions of predictor variables with supplementary information available on the relevant websites. We haven't duplicated information in the present paper but have provided the relevant links and references.

Our validation results confirm earlier studies undertaken on the THIN database (another general practice database which is derived from the Vision system but

which isn't linked to mortality data). These earlier studies include external validations of QRISK2[10 11 24], QDiabetes[12], QFracture[9] and QKidney[25] by an independent team who were not involved in the development of the algorithms. These independent validations have demonstrated similar performance compared with the validations performed by study authors using the QResearch database. This study builds on previous validations by providing new information on the performance of scores not previously validated on an external database (QBleed and QThrombosis) and by utilising the linked data which was not available on the THIN database. Together with the present study (which includes a number of scores not previously tested in an external population), the results provide consistent evidence that these QPrediction scores are likely to provide appropriate estimates of disease risk in contemporary primary care populations in England and to discriminate between patients at different levels of risk with reasonable reliability.

4.4 Comparison of QResearch and CPRD baseline characteristics

Overall, our results show a striking similarity between CPRD and QResearch cohorts for nearly all baseline characteristics. There are two notable exceptions. First, recording of ethnicity was higher in QResearch than CPRD. Second, fewer patients in the CPRD cohort had a recorded family history of diabetes and coronary heart disease in a first degree relative under the age of 60 years. Recording differences in ethnicity and family history were not explained by geographic differences or difference in data capture period between the two databases. Given the similarity for the other risk factors and treatments, it is likely that the difference in ethnicity and family history recording reflects a difference in recording patterns between the two clinical computer systems rather than a true difference between the two cohorts. A similar pattern for recording of ethnicity and family history was also reported in the validation of QRISK on the Health Improvement Network (THIN database) [11 26]. This was thought to be due to different usage of clinical templates in the clinical system, with EMIS practices having ethnicity and family history included more often thereby prompting the user to enter this information in a more systematic fashion.

4.5 Comparison of QResearch and CPRD incidence rates

The age standardised incidence rates for each condition were generally marginally higher on QResearch than CPRD although the proportions of events identified on GP data (out of all events recorded on either GP or linked mortality data) was very close. This suggests that patterns of recording of major clinical events are very similar between QResearch and CPRD although the absolute value varies by clinical condition. For example, 91% of ischaemic stroke events recorded on either GP or linked mortality data are identified on the GP record compared with 99% of hip fractures. We also note the lower levels of total cardiovascular events in the GP clinical record which was between 13-15% lower than the total recorded on either the GP record, the linked mortality record or the linked hospital admissions record.

Some of this will reflect new sudden events where the first presentation was a hospital admission or death whilst others may reflect some under-representation of existing cases not recorded in the GP record. Our study is unable to distinguish between these two scenarios, though the latter one potentially has clinical consequences if the patient is not identified as having cardiovascular disease as they may not be offered secondary prevention.

We think that the information on baseline characteristics and incidence rates will have a utility beyond the present study since it suggests that both databases are fundamentally similar in many aspects and likely to generate similar results for a range of epidemiological studies[27].

4.6 Summary

In summary, we have tested a set of QPrediction scores using an external independent cohort of practices contributing to the CPRD. The results demonstrate good performance, comparable to the results obtained from QResearch, meaning that the findings of studies performed in either database are likely to be applicable in England.

5 Supporting information

Approvals
The project was approved in accordance with the QResearch® agreement with Trent Research Ethics Committee (ref 03/04/021) and approved by the ISAC committee of the CPRD (ref 13_079).

Acknowledgements
We thank EMIS and EMIS practices for their contribution to the QResearch database. We thank CPRD and Vision Practices for allowing access to the CPRD for this study.

Source of funding:
The validation of the QPrediction scores is funded by the National Institute for Health Research's School for Primary Care Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funding body did not play a role in the study design, writing of the manuscript or in the decision to submit the manuscript for publication.

Contributorship
JHC initiated the study, undertook the literature review, data extraction, data manipulation and primary data analysis and write the first draft of the paper. JHC is the guarantor. CC contributed to the design, analysis, interpretation and drafting of the paper. PB contributed to the development of core ideas, the analysis plan, interpretation of the results and the drafting of the paper.

Competing interests
JHC 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). JHC is also director of ClinRisk Ltd which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of Medical Statistics at the University of Nottingham and a consultant statistician for ClinRisk Ltd. PB has received financial support for undertaking the validation work from the National School for Primary Care Research. This work and any views expressed within it are solely those of the

co-authors and not of any affiliated bodies or organisations. There are no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

No additional data are available.

References

1. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008:bmj.39609.449676.25.
2. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ 2013;**346**.
3. Hippisley-Cox J, Coupland C, Robson J, et al. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009;**338**:b880-.
4. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. BMJ 2009;**339**:b4229-.
5. Hippisley-Cox J, Coupland C. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney(R) Scores. BMC Family Practice 2010;**11**:49.
6. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. BMJ 2011;**343**:d4656.
7. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions BMJ 2013:(in press).
8. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.
9. Collins GS, Mallett S, Altman DG. Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. BMJ 2011;**342**:d3651.
10. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010;**340**:c2442-.
11. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;**339**:b2584-.
12. Collins GS, Altman DG. External validation of the QDScore for predicting the 10-year risk of developing Type 2 diabetes. Diabetic Medicine 2011;**28**:599-607.
13. Hippisley-C J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleeding with anticoagulants: cohort study to derive and validate the QBleed score. BMJ 2014;**(in press)**.
14. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. Journal of statistical software 2011;**45**:1-20.
15. Gray A, Clarke P, Farmer A, et al. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). BMJ 2002;**325**(7369):860.
16. Royston P. Multiple imputation of missing values. Stata Journal 2004;**4**(3):227-41.

17. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007:bmj.39261.471806.55.
18. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;**23**:723-48.
19. Royston P. Explained variation for survival models. Stata J 2006;**6**:1-14.
20. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;**13**:33.
21. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;**338**(may28_1):b605-.
22. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**:4-14.
23. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;**60**:e128-36.
24. Collins GS, Altman A. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;**344**.
25. Collins GS, Altman D. Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney scores using a primary care database. British Journal of General Practice 2012;**62**(597):e243-e50.
26. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 2008;**94**:34-39.
27. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. BMJ Open 2014;**4**(4).
28. Hippisley-Cox J, Coupland C. QRISK2-2014 Annual Update Information, 2014:5.
29. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ 2012;**344**(may22 1):e3427-e27.
30. Hippisley-Cox J CC. QDiabetes 2013 Annual Update Information Secondary QDiabetes 2013 Annual Update Information 2013. http://qdiabetes.org/QDiabetes-2013-Annual-Update-Information.pdf.

Table 1 Summary of QPrediction scores including outcome and predictor variables

Score	Weblink∞	Outcome	Predictors
QDiabetes[3]	www.qdiabetes.org	10 year risk of type 2 diabetes±	**In men and women:** Age, sex, ethnicity, deprivation, smoking, family history of diabetes, cardiovascular disease, treated hypertension, steroid tables, body mass index
QRISK2[28]	www.qrisk.org	10 year risk of CVD **	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation.
QStroke[2]	www.qstroke.org	10 year risk of ischaemic stroke or TIA±	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, congestive cardiac failure, valvular heart disease
QKidney[5]	www.qkidney.org	5 year risk of moderate or severe kidney failureμ	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, body mass index, family history of kidney disease, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, SLE, peripheral vascular disease, kidney stones, NSAIDs
QThrombosis	www.qthrombosis.org	5 year risk of venous thromboembolism±	**In men and women**: age, body mass index, smoking status, varicose veins, congestive cardiac failure, chronic renal disease, cancer, chronic obstructive pulmonary disease, inflammatory bowel disease, hospital admission in past six months, and current prescriptions for antipsychotic drugs. Additionally in women: combined oral contraceptives, tamoxifen, and hormone replacement therapy
QBleed[13]	www.qbleed.org	5 year risk of upper gastrointestinal bleed in patient starting anticoagulants vs others*	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; venous thromboembolism; congestive cardiac failure; treated hypertension; cancer; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QBleed[13]	www.qbleed.org	5 year risk of intracranial bleed in patient starting anticoagulants vs others *	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; treated hypertension; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QFracture[29]	www.qfracture.org	10 year risk of hip fracture± 10 year risk of	**In women:** HRT usage, age, body mass index, smoking status, recorded alcohol use, parental history of osteoporosis, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls, menopausal symptoms, chronic liver disease, gastrointestinal malabsorption and other endocrine disorders.

		osteoporotic fracture [µ]	**In men:** age, body mass index, smoking status, recorded alcohol use, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls and liver disease.

∞the web link has the relevant calculator, links to academic papers, additional information including links to the open source software

[±] recorded either on GP record or linked ONS mortality record;

µ recorded on the GP record.

*Recorded either on linked hospital admissions record or ONS mortality record

**Recorded either on linked hospital admissions record or ONS mortality or linked hospital admissions record

Table 2 Comparison of baseline characteristics of patients in CPRD validation cohort and QResearch comparison cohort

	CPRD	*CPRD*	*QResearch*	*QResearch*
	Men (%)	**Women (%)**	**Men (%)**	**Women (%)**
total	1,588,803	1,682,709	3,304,145	3,454,504
Ageband				
25-34 years	427975 (26.9)	467192 (27.8)	1083589 (32.8)	1179742 (34.2)
35-44 years	396680 (25.0)	364150 (21.6)	814988 (24.7)	731089 (21.2)
45-54 years	294274 (18.5)	277663 (16.5)	558553 (16.9)	516188 (14.9)
55-64 years	212817 (13.4)	211636 (12.6)	390229 (11.8)	389266 (11.3)
65-74 years	148180 (9.3)	164172 (9.8)	267997 (8.1)	298847 (8.7)
75+ years	108877 (6.9)	197896 (11.8)	188789 (5.7)	339372 (9.8)
mean Townsend score (SD)	-.5 (3.2)	-.5 (3.2)	.3 (3.6)	.2 (3.6)
Care home resident	1407 (0.1)	3466 (0.2)	2983 (0.1)	7411 (0.2)
Ethnicity recorded	587879 (37.0)	658835 (39.2)	1859462 (56.3)	2077181 (60.1)
White or not recorded	1515113 (95.4)	1602212 (95.2)	3010061 (91.1)	3149618 (91.2)
Indian	16442 (1.0)	16025 (1.0)	56156 (1.7)	50406 (1.5)
Pakistani	6606 (0.4)	6146 (0.4)	30632 (0.9)	23405 (0.7)
Bangladeshi	2419 (0.2)	1688 (0.1)	23017 (0.7)	17450 (0.5)
Other Asian	10795 (0.7)	11873 (0.7)	32513 (1.0)	36886 (1.1)
Caribbean	4989 (0.3)	6425 (0.4)	25782 (0.8)	32953 (1.0)
Black African	12883 (0.8)	14771 (0.9)	51980 (1.6)	56528 (1.6)
Chinese	2914 (0.2)	4176 (0.2)	16084 (0.5)	23043 (0.7)
Other ethnic group	16642 (1.0)	19393 (1.2)	57920 (1.8)	64215 (1.9)
Smoking status recorded	1442088 (90.8)	1595538 (94.8)	2943405 (89.1)	3219598 (93.2)
Non smoker	613833 (38.6)	834721 (49.6)	1449694 (43.9)	1973691 (57.1)
Ex-smoker	252873 (15.9)	222615 (13.2)	611837 (18.5)	545125 (15.8)
Light smoker (1-9/day)	104466 (6.6)	109864 (6.5)	472614 (14.3)	384482 (11.1)
Moderate smoker (10-19/day)	183000 (11.5)	179391 (10.7)	223631 (6.8)	202776 (5.9)
Heavy smoker (20+/day)	142438 (9.0)	87474 (5.2)	185629 (5.6)	113524 (3.3)
Smoker amount not recorded	145478 (9.2)	161473 (9.6)	0 (0.0)	0 (0.0)
Alcohol status recorded	1238110 (77.9)	1379002 (82.0)	2584335 (78.2)	2834426 (82.1)
Non drinker	163633 (10.3)	318880 (19.0)	583752 (17.7)	1035692 (30.0)
Trivial <1u/day	460091 (29.0)	726851 (43.2)	782985 (23.7)	1144469 (33.1)
Light 1-2u/day	411261 (25.9)	290547 (17.3)	481674 (14.6)	402750 (11.7)
Moderate 3-6/day	166328 (10.5)	36763 (2.2)	648549 (19.6)	237679 (6.9)
Heavy 7-9u/day	19612 (1.2)	2853 (0.2)	54083 (1.6)	7152 (0.2)
Very Heavy >/day	17185 (1.1)	3108 (0.2)	24468 (0.7)	5195 (0.2)
Family History				
family history of CHD	68805 (4.3)	80985 (4.8)	326995 (9.9)	417537 (12.1)
family history of diabetes	96810 (6.1)	132390 (7.9)	357109 (10.8)	487397 (14.1)
family history osteoporosis	880 (0.1)	10062 (0.6)	1655 (0.1)	17529 (0.5)
family history kidney disease	1253 (0.1)	1586 (0.1)	2034 (0.1)	2769 (0.1)

Clinical Values recorded				
BMI recorded	1268235 (79.8)	1481918 (88.1)	2553514 (77.3)	2857742 (82.7)
mean BMI (SD)	29.6 (6.8)	28.2 (7.0)	29.4 (6.8)	28.2 (7.0)
SBP recorded	1359560 (85.6)	1590226 (94.5)	2755733 (83.4)	3190390 (92.4)
mean SBP(SD)	133.1 (23.6)	128.6 (22.7)	132.2 (18.3)	127.1 (20.8)
cholesterol/HDL ratio recorded	587865 (37.0)	606035 (36.0)	1323503 (40.1)	1368180 (39.6)
mean cholesterol ratio (SD)	4.4 (1.4)	3.7 (1.2)	4.4 (1.4)	3.7 (1.2)
platelets recorded	223461 (14.1)	382799 (22.7)	478596 (14.5)	829702 (24.0)
platelets < 150 or > 480	11051 (0.7)	13282 (0.8)	23479 (0.7)	27009 (0.8)
Creatinine recorded	811779 (51.1)	997118 (59.3)	1714337 (51.9)	2053036 (59.4)
Mean creatinine (SD)	96.7 (149.1)	79.7 (911.6)	95.5 (30.7)	78 (23.7)

Table 3 Prescribed medication and clinical diagnoses recorded at baseline in CPRD validation cohort and QResearch comparison cohort

	CPRD men (%)	*CPRD women (%)*	*QResearch men (%)*	*QResearch women (%)*
Prescribed medication				
anticoagulants	15955 (1.0)	13077 (0.8)	27024 (0.8)	22178 (0.6)
antidepressants	101553 (6.4)	235797 (14.0)	178532 (5.4)	398018 (11.5)
antipsychotics	33884 (2.1)	79514 (4.7)	47464 (1.4)	92307 (2.7)
antiplatelets	97475 (6.1)	92816 (5.5)	160910 (4.9)	153405 (4.4)
oral NSAIDs	246515 (15.5)	346416 (20.6)	396026(12.0)	556644 (16.1)
tamoxifen	n/a	9231 (0.5)	n/a	18343 (0.5)
oestrogen only Hormone replacement therapy	n/a	119373 (7.1)	n/a	208333 (6.0)
oral corticosteroids	45597 (2.9)	71352 (4.2)	54354 (1.6)	88205 (2.6)
oral contraceptive pill	n/a	174287 (10.4)	n/a	332696 (9.6)
Recorded Diagnoses				
congestive cardiac failure	15836 (1.0)	19707 (1.2)	24965 (0.8)	28852 (0.8)
atrial fibrillation	20125 (1.3)	20102 (1.2)	33499 (1.0)	32580 (0.9)
coronary heart disease	80377 (5.1)	57703 (3.4)	130220 (3.9)	88606 (2.6)
cardiovascular disease	101430 (6.4)	83167 (4.9)	165495 (5.0)	130214 (3.8)
peripheral vascular disease	17029 (1.1)	13101 (0.8)	25004 (0.8)	17078 (0.5)
venous thromboembolism	15072 (0.9)	23090 (1.4)	27086 (0.8)	40813 (1.2)
rheumatoid or SLE	7455 (0.5)	19010 (1.1)	21453 (0.6)	48447 (1.4)
rheumatoid arthritis	7243 (0.5)	17468 (1.0)	21142 (0.6)	45542 (1.3)
SLE	228 (0.0)	1756 (0.1)	351 (0.0)	3374 (0.1)
type 1 diabetes	6238 (0.4)	4924 (0.3)	12029 (0.4)	9612 (0.3)
type 2 diabetes	51634 (3.2)	43271 (2.6)	95401 (2.9)	79654 (2.3)
treated hypertension	123584 (7.8)	161709 (9.6)	210516 (6.4)	267076 (7.7)
chronic renal disease	3968 (0.2)	4082 (0.2)	8550 (0.3)	8995 (0.3)
moderate/severe kidney failure	14107 (0.9)	9500 (0.6)	30407 (0.9)	21509 (0.6)
severe kidney failure	1603 (0.1)	1125 (0.1)	3641 (0.1)	2672 (0.1)
renal stones	13415 (0.8)	6443 (0.4)	37422 (1.1)	29204 (0.8)
inflammatory bowel disease	8962 (0.6)	10208 (0.6)	17762 (0.5)	19502 (0.6)
dementia	6686 (0.4)	16634 (1.0)	12872 (0.4)	30497 (0.9)
parkinsons disease	4546 (0.3)	4676 (0.3)	6830 (0.2)	6611 (0.2)
epilepsy or anticonvulsants	47170 (3.0)	71171 (4.2)	56516 (1.7)	61561 (1.8)
cancer	26866 (1.7)	43908 (2.6)	51649 (1.6)	79326 (2.3)
liver disease	3959 (0.2)	2893 (0.2)	9947 (0.3)	6410 (0.2)
chronic liver disease or pancreatitis	5521 (0.3)	4051 (0.2)	13069 (0.4)	8729 (0.3)
oesophageal varices	469 (0.0)	340 (0.0)	1626 (0.0)	1388 (0.0)
prior haemorrhage	97562 (6.1)	79765 (4.7)	203278 (6.2)	147533 (4.3)
malabsorption	7343 (0.5)	9375 (0.6)	21042 (0.6)	26002 (0.8)
endocrine diseases	3082 (0.2)	14097 (0.8)	6026 (0.2)	27731 (0.8)
COPD	24029 (1.5)	20737 (1.2)	41281 (1.2)	34785 (1.0)
asthma or COPD	142974 (9.0)	169503 (10.1)	273768 (8.3)	310027 (9.0)
history of falls	28878 (1.8)	61905 (3.7)	34584 (1.0)	67465 (2.0)
prior fracture	24265 (1.5)	45752 (2.7)	62092 (1.9)	89000 (2.6)

varicose vein surgery	18979	(1.2)	47012	(2.8)	35651	(1.1)	85602	(2.5)
emergency admissions or hip op	3483	(0.2)	5266	(0.3)	3335	(0.1)	5508	(0.2)

Table 4 **Numbers of patients eligible for each score in the CPRD validation cohort and number of patients with complete risk factor recording not requiring multiple imputation.**

Risk Score	*Clinical outcome*	*Eligible age range*	*exclusion criteria at study entry*	*total in age range*	*total with exclusions*	*total eligible for analysis*	*Total complete data*	*% complete data*
QDiabetes	Type 2 diabetes	25-84	type 1 or 2 diabetes at study entry	3,177,192	99,189	3,078,003	2,467,642	80.2
QStroke	ischaemic stroke	25-84	existing stroke or anticoagulants at study entry	3,177,192	70,961	3,106,231	1,032,184	33.2
QRISK2	cardiovascular disease	25-84	existing CVD or statins at study entry	3,177,192	232,722	2,944,470	906,781	30.8
QThrombosis	thromboembolism	25-84	existing VTE or anticoagulants at study entry	3,177,192	53,904	3,123,288	2,513,347	80.5
QFracture	fractured neck of femur	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QFracture	osteoporotic fracture	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QKidney	moderate or severe kidney failure	35-74	existing moderate or severe kidney failure	2,069,572	10,518	2,059,054	1,146,619	55.7
QKidney	severe kidney failure	35-74	existing severe kidney failure	2,069,572	1,930	2,067,642	1,153,979	55.8
QBleed	upper gastro-intestinal bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0
QBleed	intracranial bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0

*entry date was 01.01.1998 except for upper GI bleed and intracranial bleed where entry date was 01.01.2007

Table 5 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in women

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	48,143	99.88	4.13 (4.10 to 4.17)	102,544	99.93	4.33 (4.31 to 4.36)	0.95
	GP or ONS	48,203	n/a	4.13 (4.10 to 4.17)	102,618	n/a	4.34 (4.31 to 4.36)	0.95
Ischaemic stroke	GP data	32,283	90.64	2.45 (2.42 to 2.48)	63,582	90.22	2.45 (2.44 to 2.47)	1.00
	GP or ONS	35,617	n/a	2.62 (2.59 to 2.64)	70,477	n/a	2.70 (2.68 to 2.72)	0.97
Cardiovascular disease	GP data	55,833	85.71	5.41 (5.37 to 5.46)	107,412	84.96	4.32 (4.30 to 4.35)	1.25
	GP or ONS	65,143	n/a	6.32 (6.27 to 6.37)	126,433	n/a	5.03 (5.01 to 5.06)	1.26
	GP or ONS or HES	69,202	n/a	6.72 (6.67 to 6.77)	140,510	n/a	5.63 (5.60 to 5.66)	1.19
Thromboembolism	GP data	18,199	91.1	1.52 (1.49 to 1.54)	35,971	90.55	1.46 (1.44 to 1.47)	1.04
	GP or ONS	19,978	n/a	1.64 (1.62 to 1.67)	39,727	n/a	1.60 (1.58 to 1.62)	1.03
Fractured neck of femur	GP data	17,529	99.98	1.32 (1.30 to 1.34)	34,821	99.99	1.40 (1.39 to 1.42)	0.94
	GP or ONS	17,533	n/a	1.32 (1.30 to 1.34)	34,825	n/a	1.40 (1.39 to 1.42)	0.94
Osteoporotic fracture	GP data	34,528	n/a	2.89 (2.58 to 3.20)	81,334	n/a	3.63 (3.61 to 3.66)	0.80
mod /severe kidney failure	GP data	19,902	n/a	2.06 (1.76 to 2.36)	48,665	n/a	2.81 (2.78 to 2.83)	0.73
severe kidney failure	GP data	1,737	n/a	0.18 (0.09 to 0.27)	4,150	n/a	0.24 (0.24 to 0.25)	0.74

Table 6 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in men

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	60,731	99.92	5.84 (5.79 to 5.89)	128,234	99.94	5.97 (5.94 to 6.00)	0.98
	GP or ONS	60,782	n/a	5.84 (5.80 to 5.89)	128,317	n/a	5.98 (5.94 to 6.01)	0.98
ischaemic stroke	GP data	32,223	93.55	3.17 (3.14 to 3.20)	63,480	92.85	3.10 (3.08 to 3.13)	1.02
	GP or ONS	34,443	n/a	3.33 (3.30 to 3.37)	68,366	n/a	3.37 (3.34 to 3.40)	0.99
Cardiovascular disease	GP data	70,283	86.7	7.38 (7.33 to 7.44)	137,136	86.12	7.12 (7.08 to 7.16)	1.03
	GP or ONS	81,068	n/a	8.52 (8.46 to 8.58)	159,240	n/a	8.37 (8.33 to 8.41)	1.02
	GP or ONS or HES	84,620	n/a	8.90 (8.84 to 8.96)	174,405	n/a	9.17 (9.13 to 9.21)	0.97
thromboembolism	GP data	15,655	92.32	1.49 (1.46 to 1.51)	31,503	92.22	1.44 (1.43 to 1.46)	1.03
	GP or ONS	16,958	n/a	1.61 (1.59 to 1.63)	34,161	n/a	1.57 (1.56 to 1.59)	1.02
fractured neck of femur	GP data	5,706	99.98	0.65 (0.63 to 0.67)	12,435	99.98	0.71 (0.70 to 0.73)	0.91
	GP or ONS	5,707	n/a	0.65 (0.63 to 0.67)	12,438	n/a	0.71 (0.70 to 0.73)	0.91
osteoporotic fracture	GP data	11,169	n/a	1.29 (1.05 to 1.52)	28,555	n/a	1.54 (1.52 to 1.55)	0.84
Mod/severe kidney failure	GP data	37,597	n/a	4.88 (4.37 to 5.38)	86,649	n/a	5.82 (5.78 to 5.85)	0.84
severe kidney failure	GP data	3,472	n/a	0.54 (0.38 to 0.71)	7,372	n/a	0.47 (0.46 to 0.48)	1.15

Table 7 Performance of QPrediction scores on the CPRD validation cohort compared with published results for the QResearch validation cohort

		CPRD	*CPRD*	*QResearch*	*QResearch*
		women	**men**	**women**	**men**
	statistic	**mean (95%CI)**	**mean (95%CI)**	**mean (95%CI)**	**mean (95%CI)**
QDiabetes 2013 (type 2 diabetes)[30]	ROC	0.846 (0.844 to 0.848)	0.818 (0.816 to 0.82)	0.853 (0.851 to 0.856)	0.837 (0.835 to 0.840)
	R^2 (%)	49.6 (49.2 to 50.1)	45.7 (45.3 to 46.2)	50.8 (50.3 to 51.4)	48.1 (47.6 to 48.6)
	D statistic	2.032 (2.015 to 2.049)	1.879 (1.863 to 1.895)	2.081 (2.058 to 2.104)	1.971 (1.951 to 1.991)
QKidney -2010[5] (moderate or severe kidney failure)	ROC	0.875 (0.87 to 0.879)	0.88 (0.878 to 0.883)	0.877 (0.873 to 0.880)	0.878 (0.874 to 0.882)
	R^2 (%)	58.3 (57.8 to 58.7)	57.5 (57.1 to 57.8)	56.45 (55.40 to 57.50)	58.29 (55.31 to 61.26)
	D statistic	2.418 (2.394 to 2.442)	2.379 (2.361 to 2.397)	2.33 (2.28 to 2.40)	2.42 (2.28 to 2.56)
QKidney -2010 (severe kidney failure)[5]	ROC	0.839 (0.822 to 0.855)	0.851 (0.84 to 0.862)	0.843 (0.825 to 0.860)	0.846 (0.829 to 0.862)
	R^2 (%)	51.4 (49.5 to 53.2)	53.8 (52.6 to 55.1)	55.39 (52.59 to 58.18)	56.65 (53.94 to 59.35)
	D statistic	2.103 (2.025 to 2.182)	2.21 (2.154 to 2.266)	2.28 (2.15 to 2.41)	2.34 (2.21 to 2.47)
QRISK2-2014[28] (cardiovascular disease)	ROC	0.883 (0.882 to 0.884)	0.859 (0.858 to 0.861)	0.892 (0.892 to 0.895)	0.871 (0.869 to 0.873)
	R^2 (%)	56.4 (56.1 to 56.7)	50.9 (50.6 to 51.2)	58.8 (58.4 to 59.1)	53.3 (52.9 to 53.7)
	D statistic	2.328 (2.313 to 2.343)	2.085 (2.071 to 2.098)	2.443 (2.423 to 2.463)	2.188 (2.171 to 2.205)
QStroke-2013[2] (ischaemic stroke or TIA)	ROC	0.882 (0.88 to 0.883)	0.869 (0.867 to 0.87)	0.877 (0.875 to 0.879)	0.866 (0.864 to 0.868)
	R^2 (%)	58.4 (58.1 to 58.8)	55.3 (54.9 to 55.7)	57.3 (56.8 to 57.8)	55.1 (54.6 to 55.7)
	D statistic	2.427 (2.408 to 2.446)	2.278 (2.259 to 2.297)	2.37 (2.35 to 2.40)	2.27 (2.24 to 2.30)
QThrombosis-2010[6] (venous thromboembolism)	ROC	0.756 (0.751 to 0.761)	0.765 (0.760 to 0.770)	0.75 (0.74 to 0.76)	0.75 (0.74 to 0.76)
	R^2 (%)	35.3 (34.5 to 36.1)	34.5 (33.7 to 35.4)	32.78 (31.08 to 34.48)	33.51 (31.71 to 35.30)

	D statistic	1.512 (1.485 to 1.538)	1.486 (1.458 to 1.513)	1.43 (1.37 to 1.49)	1.45 (1.39 to 1.51)
QBleed-2014[13] (upper gastrointestinal bleed)	ROC statistic	0.775 (0.770 to 0.781)	0.759 (0.753 to 0.764)	0.766 (0.758 to 0.775)	0.747 (0.738 to 0.756)
	R^2 (%)	44.7 (43.6 to 45.9)	41.6 (40.5 to 42.8)	40.7 (38.9 to 42.6)	36.9 (35.1 to 38.7)
	D statistic	1.842 (1.798 to 1.885)	1.729 (1.687 to 1.771)	1.70 (1.63 to 1.76)	1.57 (1.51 to 1.63)
QBleed-2014[13] (intracranial bleed)	ROC statistic	0.808 (0.801 to 0.816)	0.789 (0.780 to 0.797)	0.847 (0.838 to 0.856)	0.812 (0.80 to 0.824)
	R^2 (%)	51.7 (50.1 to 53.3)	50.0 (48.3 to 51.7)	58.0 (56.0 to 60.0)	53.3 (51.1 to 55.4)
	D statistic	2.118 (2.051 to 2.186)	2.046 (1.977 to 2.116)	2.40 (2.30 to 2.50)	2.19 (2.09 to 2.28)
QFracture-2012[29] (fractured neck of femur)	ROC	0.89 (0.888 to 0.892)	0.872 (0.867 to 0.877)	0.893 (0.890 to 0.896)	0.875 (0.868 to 0.883)
	R^2 (%)	70.6 (70.2 to 71)	69.2 (68.5 to 70)	71.73 (71.10 to 72.30)	70.37 (69.25 to 71.49)
	D statistic	3.171 (3.139 to 3.203)	3.07 (3.016 to 3.124)	3.26 (3.21 to 3.31)	3.15 (3.06 to 3.24)
QFracture -2012[29] (osteoporotic fracture: hip, spine, wrist,humerus)	ROC	0.817 (0.814 to 0.819)	0.768 (0.763 to 0.773)	0.790 (0.787 to 0.793)	0.711 (0.703 to 0.719)
	R^2 (%)	56.3 (55.8 to 56.7)	49.8 (48.9 to 50.7)	51.9 (51.2 to 52.6)	38.20 (36.89 to 39.57)
	D statistic	2.322 (2.301 to 2.343)	2.038 (2.002 to 2.075)	2.13 (2.10 to 2.15)	1.61 (1.56 to 1.66)

Notes on understanding validation statistics:

D statistic is a measure of discrimination - higher values indicate better discrimination
ROC statistic is a measure of discrimination - higher values indicate better discrimination
The R^2 statistic is a measure of explained variation - higher values indicate more variation is explained

Table 8 Performance of each score for predicting the relevant outcome in the CPRD validation cohort. The cut off is the threshold of predicted risk for the top decile in the CPRD cohort.

score	*outcome*	*duration*	*cut off (%) for top decile predicted risk*	*Sensitivity (%)*	*Specificity (%)*	*observed risk (%)*
QDiabetes	Type 2 diabetes	10 yr risk	13.0	44.8	91.0	20.8
QStroke	Ischaemic stroke	10 yr risk	10.5	54.7	90.8	16.1
QRISK2	cardiovascular disease	10 yr risk	20.7	49.9	91.9	31.8
QThrombosis	venous thromboembolism	5 yr risk	1.5	36.2	90.1	2.6
QKidney	moderate-severe kidney failure	5 yr risk	6.3	59.1	90.5	6.9
QKidney	severe kidney failure	5 yr risk	0.4	58.5	90.0	0.7
QBleed	upper GI bleed	5 yr risk	1.6	38.0	90.2	3.5
QBleed	intracranial bleed	5 yr risk	0.9	44.2	90.1	1.6
QFracture	fractured neck of femur	10 yr risk	3.7	66.5	90.4	9.4
QFracture	osteoporotic fracture	10 yr risk	7.8	49.6	90.5	13.1

FIGURE LEGENDS

Figure 1 Calibration of each QPrediction score comparing the mean predicted risks with the observed risks in the CPRD cohort.

1a QThrombosis (venous thromboembolism)

1b QFracture (hip)

1c QFracture (hip, colles, spine, shoulder)

1d QStroke (ischaemic stroke)

1e QDiabetes (type 2 diabetes)

1f QBleed (upper gastrointestinal haemorrhage)

1g QBleed (intracranial haemorrhage)

1h QKidney (moderate or severe kidney failure)

1i QKidney(severe kidney failure)

1j QRisk2 (cardiovascular disease)

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Revised Research paper submitted to BMJ Open
Reference-2014-005809
June 2014

Authors

Julia Hippisley-Cox	Professor of Clinical Epidemiology & General Practice[1]
Carol Coupland	Associate Professor and Reader in Medical Statistics[1]
Peter Brindle	Research and Development Programme Director[2]

Institutions

[1]Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham, NG2 7RD.

[2]Avon Primary Care Research Collaborative, Bristol Clinical Commissioning Group, South Plaza, Marlborough Street, Bristol BS1 3NX.

Guarantor & Author for correspondence - Julia Hippisley-Cox

Email: Julia.hippisley-cox@nottingham.ac.uk

Telephone: 0115 8466915

Fax: 0115 8466904

Abstract

Objectives To validate the performance of a set of risk prediction algorithms developed using the QResearch database, in an independent sample from general practices contributing to the Clinical Research Data Link (CPRD).

Setting Prospective open cohort study using practices contributing to the CPRD database and practices contributing to the QResearch database.

Participants The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices between 01 Jan 1998 and 31 July 2012. The validation statistics for QResearch were obtained from the original published papers which used a one third sample of practices separate to those used to derive the score. A cohort from QResearch was used to compare incidence rates and baseline characteristics and consisted of 6.8 million patients from 753 practices registered between 01 Jan 1998 and until 31 July 2013.

Outcome measures Incident events relating to seven different risk prediction scores: QRISK2 (cardiovascular disease); QStroke (ischaemic stroke); QDiabetes (type 2 diabetes); QFracture (osteoporotic fracture and hip fracture); QKidney (moderate and severe kidney failure); QThrombosis (venous thromboembolism); QBleed (intracranial bleed and upper gastrointestinal haemorrhage). Measures of discrimination and calibration were calculated.

Results Overall, the baseline characteristics of the CPRD and QResearch cohorts were similar though QResearch had higher recording levels for ethnicity and family history. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with the published results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch and the ROC value was 0.85 on both databases. The scores were well calibrated in CPRD.

Conclusion Each of the algorithms performed practically as well in the external independent CPRD validation cohorts as they had in the original published QResearch validation cohorts.

Strengths and limitations of this study

- This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system from that used to derive the algorithms.
- The discrimination and calibration statistics for each score were very similar in CPRD to those published from validation cohorts from QResearch. This supports their potential utility in the general population of patients in primary care.
- A strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice.
- The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible.
- Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using these algorithms in primary care.

1 Introduction

In the last 7 years, we have developed a series of risk prediction algorithms using the QResearch database. QResearch is a large research database containing pseudonymised individual level data from over 700 general practices using the EMIS clinical system. The QResearch database consists of data collected from primary care (coded information on socio-demographic characteristics, diagnoses, symptoms, smoking/alcohol, clinical measurements, laboratory values, prescriptions and referrals) which has been linked to cause of death, hospital episodes and cancer registrations at individual patient level.

The algorithms predict outcomes such as cardiovascular disease(www.qrisk.org)[1], stroke (www.qstroke.org)[2], type 2 diabetes (www.qdiabetes.org)[3], osteoporotic fracture (www.qfracture.org)[4], moderate or severe kidney disease (www.qkidney.org)[5], venous thrombo-embolism (www.qthrombosis.org)[6], and emergency hospital admission (www.qadmissions.org)[7]. Generally, the "QPrediction" algorithms have been designed to systematically identify patients in primary care at high risk of a serious clinical outcome for whom further intervention to lower risk of that outcome might be possible. They are also designed to quantify absolute risk of serious outcomes in a way which patients can understand and which might help guide lifestyle and management decisions. A number of these algorithms are now integrated into GP clinical computer systems, included in national guidelines[1 4] and are in daily use across the NHS [1 3 8].

The algorithms were originally developed using a random two thirds sample of practices contributing to the QResearch database and validated on the remaining third. Whilst this represents a physically discrete population of patients and practices for validation, the practices all use the same clinical computer system (EMIS), which is in use in 53% of UK practices. A more stringent test of performance is to validate the algorithms on a fully external database derived from practices using a different but commonly used primary care computer system. This would help determine whether the predictions from the algorithms are likely to generalise to the whole population in England. Whilst some of the algorithms have been validated by an independent team using the THIN primary care database[9-12], there are currently no published validations of the algorithms using a primary care database which is routinely linked to mortality data in the same way as QResearch.

We therefore decided to validate the various QPrediction Scores using another database known as the Clinical Research Data Link (CPRD). The General Practice Research database (GPRD) was originally set up in 1988 and is of similar nature to QResearch although it is derived from practices using a different clinical computer system (Vision, which is used by 20% of GPs). It was extended to include linked mortality data and data from secondary care and was renamed the Clinical Research Data Link (CPRD) in 2012. Our secondary objective was to compare the ascertainment of incident clinical events recorded in GP data alone with that

recorded in either GP data or the linked mortality data in both the CPRD and QResearch.

2 Methods

2.1 CPRD Study population

For the validation using CPRD, we identified an open cohort of patients aged 25-99 years at entry to the cohort and followed this cohort up until 31st July 2012 (the latest date for which linked data were available at the time of analysis). We restricted the CPRD cohort to 357 practices in England which had linked ONS mortality and hospital admissions data. For each patient we determined an entry date to the cohort, which was the latest of the following dates: 25th birthday, date of registration with the practice plus one year, date on which the practice computer system was installed plus one year, and the beginning of the study period (01 January 1998). Patients were censored at the earliest date of the relevant outcome, de-registration with the practice, last upload of computerised data or the study end date (31 July 2012).

For the assessment of the two QBleed outcomes (intracranial bleed and upper gastrointestinal haemorrhage) we used a later cohort entry date of 01.01.2007 for comparability with the equivalent study period for the derivation of the algorithm on QResearch[13].

2.2 QResearch study population

For comparison of the validation statistics (ROC, D and R2 statistics), we extracted the original published values from the papers which had been calculated using a one third sample of practices from QResearch which were independent from the two thirds of practices used to derive the scores.

For comparison of the baseline characteristics, incidence rates and ascertainment rates we used the latest version of the QResearch database which is currently available (QResearch 38, 31st Dec 2013). We identified an open cohort in the same way as for CPRD, using all of the QResearch practices in England, and with follow-up until 31 July 2013.

2.3 Inclusion and exclusion criteria

For both databases, we excluded patients without a Townsend score (an area based measure of material deprivation derived from the post code) and temporary residents. For each score we then identified patients who were eligible to have the score calculated according to the relevant inclusion and exclusion criteria as summarised in Table 4

2.4 Risk scores included in validation

We validated the following risk prediction scores on CPRD:

1. QDiabetes - 10 year risk of type 2 diabetes[3]
2. QRISK2 – 10 year risk of cardiovascular disease[1]
3. QStroke – 10 year risk of stroke or transient ischaemic attack (TIA)[5]
4. QFracture - 10 year risk of hip or osteoporotic fracture[4]
5. QThrombosis – 5 year risk of venous thrombo-embolism (VTE)[6]
6. QBleed – 5 year risk of upper gastrointestinal haemorrhage and intracranial haemorrhage[13]
7. QKidney – 5 year risk of moderate-severe kidney disease[5]

2.5 Clinical outcomes

We identified the relevant clinical outcome using the same definition as had been applied in the original derivation of the risk scores using QResearch. The data sources used to identify the clinical outcomes had varied over the six years during which the original studies had been undertaken due to the changing availability of linked hospital and mortality data over that time. In 2008, the QResearch database was linked to mortality records for 1997 onwards. In 2013, the QResearch database was linked to hospital admissions records with data for patients from 1998 onwards. For the latest updated version of QRISK2 (QRISK2-2014), the outcome was identified by the presence of the relevant Read code on the GP record or an ICD10 code recorded on the linked mortality record or on the linked hospital admissions record. For QStroke, QDiabetes, QFracture and QThrombosis, the outcome was identified either by the presence of the relevant Read code recorded on the GP record or an ICD10 code recorded on the linked mortality record. For QKidney, the outcome was identified solely from information recorded in the GP record as in the original study as it required blood test values which were only present in the GP record. For QBleed, the outcome was identified in CPRD from events recorded either on the linked hospital admissions database or the linked mortality record in order to identify the events most likely to have serious clinical consequences for the patient.

We determined case ascertainment for each clinical outcome on both databases, by calculating the proportion of cases recorded on the GP record out of the total number of cases recorded on either the GP record or linked mortality record. We calculated the age standardised incidence rates of each outcome based on outcomes recorded on (1) the GP record alone and on (2) the GP record or linked mortality; (3) GP or linked mortality or hospital records. We standardised CPRD rates to the age distribution of the QResearch population in five year bands to ensure comparability.

2.6 Risk factors and missing values

We extracted data from CPRD for all the predictor variables included in one or more of the different algorithms using the same definitions as those used in the original

QResearch studies to enable a direct comparison of the results. We developed a mapping between the Read and medication reference tables to identify the equivalent code in each database. This included the following variables recorded at entry to the cohort:

- **demographics** – age (continuous), sex, ethnicity (9 categories – white, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean ,Black African, Chinese, Other ethnic group), resident in care home, material deprivation (as measured by the Townsend score)
- **clinical values** - smoking status (non-smoker, ex-smoker, light smoker [1-9 cigarettes/day], moderate smoker [10-19 cigarettes/day], heavy smoker [20+ cigarettes/day];body mass index, systolic blood pressure, alcohol consumption- none drinker, trivial (<1u/day), light (1-2u/day), moderate (3-6u/day), heavy (7-9u/day), very heavy (>9 day).
- **laboratory results** –cholesterol/HDL ratio, platelets
- **family history**- family history of osteoporosis or hip fracture in a first degree relative, coronary heart disease in first degree relative under the age of 60 years, diabetes in a first degree relative.
- **chronic diseases** – congestive cardiac failure, atrial fibrillation, coronary heart disease, cardiovascular disease, periperal vascular disease, venous thromboembolism, diabetes, rheumatoid arthritis, SLE, hypertension, renal disease, renal stones, inflammatory bowel disease, dementia, Parkinson's disease, epilepsy, cancer, chronic liver disease or pancreatitis, oesophageal varices, prior haemorrhage, malabsorption endocrine diseases, asthma or COPD, history of falls, prior osteoporotic fracture, varicose vein surgery, emergency admissions or hip surgery in last 6 months.
- **prescribed medication**- anticoagulants, antidepressants, antipsychotics, antiplatelets, oral NSAIDs, tamoxifen, oestrogen containing hormone replacement therapy (BNF chapter 6.4.1.1), systemic corticosteroids, combined oral contraceptive.

The combination of predictor variables required for each risk score varied with the score being validated as shown in Table 1. We used the clinical value recorded closest to the date on which the patient entered the study for body mass index, systolic blood pressure, smoking status, platelets, and total and HDL cholesterol. Patients were considered to be exposed to medication at entry to the cohort if they had at least 2 prescriptions for the relevant medication prescribed prior to the study entry date with the most recent one occurring within 28 days of the study entry date.

2.7 Townsend scores

We used the Townsend score evaluated at output area as a proxy for material deprivation. The CPRD dataset differs from the QResearch dataset in that each patient in the CPRD dataset is allocated to a tenth of deprivation (as measured by the Townsend score) and only the category number is provided. In contrast, each

patient in the QResearch dataset is allocated the individual Townsend score corresponding to their output area of residence (i.e. continuous data). In order to calculate risk scores in the CPRD cohort, we used the median value for each tenth as supplied by CPRD. Patients with missing Townsend scores were excluded from the cohorts.

2.8 Discrimination and calibration statistics

We used chained equations with the ICE procedure in STATA[14] to perform multiple imputation to replace missing values for body mass index, systolic blood pressure, smoking status, alcohol, and total and HDL cholesterol. We created five multiply imputed datasets and used Rubin's rules to combine effect estimates and standard errors to allow for the uncertainty due to imputing missing data[15 16].

We applied the algorithm for each score to eligible patients in the CPRD study cohort to obtain predicted risks for each of the relevant clinical outcomes. We calculated the estimated risk for eligible patients in the CPRD validation dataset over 5 years or 10 years depending on which score was used. We then tested the performance of each score in the CPRD cohort and compared it with the published results from the original QResearch validation cohorts.

In order to assess calibration (i.e. degree of similarity between predicted and observed risks), we calculated the mean predicted risk and the observed risk [17]obtained using the Kaplan-Meier estimate and compared the ratio of the mean predicted risk to the observed risk for patients in the validation cohort in each decile of predicted risk. We calculated the area under the Receiver Operator Curve (ROC) statistic to assess discrimination (i.e. ability of a risk prediction equation to distinguish between those who do and do not have an event during the follow-up period). We also calculated the D statistic[18] and an R squared statistic derived from the D statistic[19] which are measures of discrimination and explained variation appropriate for survival models. The D statistic has been developed as a new measure of discrimination specifically for censored survival data, higher values indicate improved discrimination, and an increase in the D statistic of at least 0.1 indicates an important difference in prognostic separation between different risk classification schemes. The R^2 statistic derived from the D statistic is a measure specific to censored survival data– it measures explained variation in time to the outcome event and higher values indicate more variation is explained[20]. We also repeated the assessment of discrimination by restricting the analysis for each score to patients without missing data for relevant clinical or laboratory measures used in the risk score (ie those with complete data for all predictor variables in the risk score).

We identified the proportion of patients in the CPRD validation cohort who were in the top decile of predicted risk and used this to calculate the sensitivity, specificity and observed risk at this threshold. We used the top decile for comparability across the scores and with previous studies though the choice of threshold for use in clinical practice will depend on the context and cost-effectiveness of relevant interventions. Analyses were conducted using Stata (version 13.1).

2.9 Sample size estimation

There is currently no clear guidance on sample size requirements for studies evaluating the performance (validation) of a multivariable risk score, but a commonly used rule-of-thumb is that it is desirable to seek a dataset with at least 100 patients with the outcome of interest. We used all the available data on the CPRD to maximize the power of the study.

3 Results

3.1 Study populations

The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices with linked data between 01 Jan 1998 and 31 July 2012. The QResearch cohort consisted of 6.8 million patients from 753 practices with linked data, registered between 01 Jan 1998 and until 31 July 2013. The numbers of patients in each geographical region are shown in Web Extra Table 1.

3.2 Baseline characteristics

Table 2 shows a comparison of the demographic characteristics for the CPRD and QResearch cohorts.

The QResearch population was marginally younger with 34.2% of women and 32.8% of men aged 24-34 years compared with 27.8% and 26.9% for CPRD.

3.2.1 Recording of ethnicity

QResearch had a higher proportion of patients with self-assigned ethnicity recorded compared with CPRD both overall (58.2% vs 38.1%) and in each of the 10 geographical areas within England (web extra table 2). We repeated the analysis restricting information on QResearch to that recorded prior to 31 July 2012 (for comparability with the calendar time available on CPRD). Of the 6,758,649 patients in the QResearch cohort, 3,856,244 (57.1%) had ethnicity recorded prior to this date.

3.2.2 Recording of family history

Recording of a positive family history of coronary heart disease and diabetes was more than twice as high in QResearch compared with CPRD. For example, for family history of coronary heart disease, 11.0% of patients had a value recorded for QResearch compared with 4.6% for CPRD (web extra table 2). Restricting information to that recorded prior to July 2012 for QResearch, then 6,758,649 (10.7%) had a positive family history of coronary heart disease recorded.

3.2.3 *Recording of alcohol and smoking levels*

Recording of alcohol levels was very similar in both QResearch and CPRD. For example, 82.1% of women had alcohol level recorded in both databases. Recording of smoking status was marginally higher in women compared with men in QResearch (93.2% vs 89.1%) and also CPRD (94.8% vs 90.8%).

3.2.4 *Recording of clinical values*

Recording of cholesterol/HDL ratio was marginally higher on QResearch compared with CPRD (40.1% vs 36.0%). Recording of body mass index and systolic blood pressure tended to be marginally higher on CPRD than QResearch. However the mean values for the various clinical values (BMI, systolic blood pressure, serum creatinine and cholesterol/HDL ratio) were extremely similar.

Table 3 shows prescribed medication and clinical diagnoses recorded in patients on or prior to entry to the study cohort. Overall, the prevalence of clinical diagnoses were similar on the two databases with CPRD having marginally higher prescribing rates.

The inclusion and exclusion criteria for each risk score are shown in Table 4 along with the numbers of patients eligible for each analysis on CPRD. For example, there were 3,177,192 patients aged 25-84 years. Of these, 99,189 had existing diabetes at baseline leaving 3,078,003 for the validation of QDiabetes. Table 4 also shows the numbers and percentage out of those eligible for inclusion with complete data for risk factors necessary for calculation of the score which would otherwise need to be imputed (i.e. laboratory or clinical values). The amount of missing data varies substantially between the scores with scores requiring multiple laboratory or clinical values (such as QRISK2) having the lowest levels of completeness.

3.2.5 *Comparison between CPRD linked and unlinked data*

Web extra Table 3 shows characteristics for CPRD cohort with linked data with CPRD cohort without linked data. The CPRD cohort with linked data tended to have higher recorded of ethnicity compared with the CPRD cohort without linked data (38.1% vs 28.4%). Recording of smoking, alcohol, body mass index, systolic blood pressure, cholesterol and platelets were all higher on the CPRD cohort with linked data than those without linked data.

3.3 Incidence rates of clinical outcomes

Table 5 shows the number of incident events for each clinical outcome in women recorded on GP data and those recorded on either GP data or cause specific mortality data for both the CPRD and QResearch cohorts. It also shows the age standardized incidence rates per 1000 person years. Table 6 shows the comparable information for men.

For example, there were 35,617 incident ischaemic stroke or TIA events for women on CPRD. Of these, 32,283 had been identified on the GP record with an additional 3,334 events identified on the linked ONS mortality record. The ascertainment of events on the GP record was therefore 32283/35617 i.e. 90.6%. For QResearch, there were 70,477 incident stroke events recorded on either the GP or linked ONS mortality record of which 63,572 had been identified on the GP record. The ascertainment was therefore 90.2%.

For thromboembolism in women, 91.1% of events recorded on either the GP or linked ONS mortality record on CPRD were identified on the GP record compared with 90.6% for QResearch. Similar results were obtained for men with levels of ascertainment between the two databases being extremely close suggesting similar recording patterns between the two groups of GP practices contributing to each database.

The age standardized incidence rates of events on CPRD tended to be marginally lower than those on QResearch as shown by the ratio of the CPRD rates to those in QResearch (Table 5). For example, the rate ratio for fractured neck of femur in women was 0.94 indicating that CPRD had a 6% lower incidence rate compared with QResearch. The effect was more marked for moderate or severe kidney failure where the incidence rates for CPRD were approximately 25% lower than those for QResearch in women and 16% lower in men.

The age standardized incidence rates of upper gastrointestinal haemorrhage and intracranial haemorrhage among patients prescribed anticoagulants and those not prescribed anticoagulants are shown in Web extra table 4. The rates are similar for CPRD and QResearch.

3.4 Validation statistics

Table 7 shows the discrimination statistics for each score in CPRD in men and women and also the published values from previous validations using QResearch. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch. The D statistic for women was 2.03 within CPRD compared with 2.08 for QResearch. The ROC value for women was 0.85 on both databases.

Of all the scores, QFracture (fractured neck of femur) had the best performance in men in CPRD with a ROC value of 0.89, R^2 value of 71% and D statistic of 3.17. The corresponding figures for QResearch in men were 0.89, 72% and 3.26.

QThrombosis had the lowest values for women in CPRD with an ROC value of 0.77, R^2 of 34.5 and D statistic of 1.49. The corresponding figures for women in QResearch were 0.75, 33.5 and 1.45.

Figure 1 compares the mean predicted risks and observed risks for each score across each tenth of predicted risk (1 representing the lowest risk and 10 the highest risk) and demonstrates that the models are generally well calibrated for patients on CPRD.

The QKidney score (moderate or severe kidney failure) showed the observed risk was lower than the predicted risk. This might indicate a degree of over prediction of the score. Alternatively, it could be related to the lower incidence rate of kidney failure observed among women on the CPRD compared with QResearch.

Web extra table 5 presents the ROC, D and R^2 statistic for each score restricted to patients from CPRD with complete recording of laboratory and risk factor data for each score. The results were very similar to the results obtained using multiply imputed dataset for the majority of scores except for QRISK2 and QStroke where values tended to be lower.

3.5 Performance for the top decile of risk.

Table 8 shows the sensitivity, specificity and observed risk for patients in the top decile of each score on CPRD. The observed risk is higher than the risk threshold value since this represents the observed risk within the top decile of predicted risk. For example, the cut off for the top tenth of risk for QFracture (fractured neck of femur) was a 10 year risk of 3.7%. At this threshold the sensitivity was 66.5%, specificity 90.4% and observed risk 9.4%. The results are similar to those obtained from QResearch (not shown).

4 Discussion

4.1 Summary of key findings

This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system (Vision system supplied by In Practice Systems) from the QResearch database (which is based on practices using EMIS clinical systems). Practices using the Vision system together with practices using EMIS make up approximately 75% of all the English general practices. The discrimination and calibration statistics for each score were remarkably similar in CPRD to those published from validation cohorts from QResearch. Our paper also provides updated information on a direct comparison between two of the world's largest general practice databases which have both been linked to mortality and second care data.

Before a clinical risk score can be reliably used in clinical practice, evidence is needed that it can successfully predict the intended outcome in groups of patients other than ones used to develop the score but similar to ones in whom the score might be used. Not all risk scores perform well in external samples – this can be due to deficiencies in the design or modelling methods used to derive the algorithm, if the model is over fitted or if there is an important predictor which is absent[21]. Other reasons for poor performance include differences between the setting of patients in the new and derivation samples, differences in how information is recorded and differences in patient characteristics[21]. It is for these reasons, that we have meticulously assembled the CPRD cohort using the same inclusion/exclusion criteria, definitions of predictor and outcome variables as in the original derivation studies. Any differences observed are therefore more likely to be due to capture of information and underlying population characteristics. In this study, we have found marginal differences in incidence rates between QResearch and CPRD and higher rates of recording of family history and ethnicity in QResearch though these have not been large enough to materially affect our results.

4.2 Strengths and limitations

One strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice. CPRD only had Townsend score for patients recorded for approximately half their practices (unlike QResearch where Townsend score is included for all practices) so we had to limit the validation cohort in CPRD for this analysis to those practices with linked Townsend scores. We undertook a comparison between patients registered with CPRD practices with and without linked data. We found marginally higher recording for ethnicity, smoking, alcohol, clinical values for the CPRD cohort with linked data compared with the unlinked data but similar characteristics for demographics, comorbidities,

medication and clinical values (results not shown) so we have no reason to believe this would have biased our results.

Another strength of general practice databases is the large volume of patients who tend to be representative of the general population. A limitation of routinely collected data is that not all patients will have all clinical and laboratory data recorded leading to missing data values in some of the parameters needed to calculate the risk scores. We have reported performance in all patients using multiple imputation to replace missing values and restricted to patients without missing values and found very similar results for the majority of algorithms tested. There was some degradation of performance associated with large amounts of missing data although not sufficient to affect our conclusion. The software used to implement QPrediction scores in clinical practice includes algorithms to estimate body mass index, systolic blood pressure and cholesterol/HDL ratio which can be used where relevant data is not recorded to generate an estimate risk score. The clinician can then enter the relevant data fields once the patient is assessed to calculate an actual risk score using recorded values.

The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible. The CPRD database uses the same clinical coding system as QResearch for clinical values (it uses Read version 2). However, there is a third clinical system in use in England (SystmOne) which uses a different coding system known as Clinical terms version 3(CTV3). Whilst there is a mapping between Read codes and CTV3, we have not tested the algorithms on a database using CTV3 in this study so are unable to draw conclusions regarding the generalisability of the results of the validation to practices using this system.

The quality of information on CPRD is likely to be good since previous studies have validated similar outcomes and exposures and found levels of completeness and accuracy to be good[22 23].

4.3 Comparison with other studies

The aim of this study was to validate a collection of QPrediction tools. The details of the derivation and first validation of each prediction tool have been separately published in the peer reviewed literature including information on definitions of predictor variables with supplementary information available on the relevant websites. We haven't duplicated information in the present paper but have provided the relevant links and references.

Our validation results confirm earlier studies undertaken on the THIN database (another general practice database which is derived from the Vision system but

which isn't linked to mortality data). These earlier studies include external validations of QRISK2[10 11 24], QDiabetes[12], QFracture[9] and QKidney[25] by an independent team who were not involved in the development of the algorithms. These independent validations have demonstrated similar performance compared with the validations performed by study authors using the QResearch database. This study builds on previous validations by providing new information on the performance of scores not previously validated on an external database (QBleed and QThrombosis) and by utilising the linked data which was not available on the THIN database. Together with the present study (which includes a number of scores not previously tested in an external population), the results provide consistent evidence that these QPrediction scores are likely to provide appropriate estimates of disease risk in contemporary primary care populations in England and to discriminate between patients at different levels of risk with reasonable reliability.

4.4 Comparison of QResearch and CPRD baseline characteristics

Overall, our results show a striking similarity between CPRD and QResearch cohorts for nearly all baseline characteristics. There are two notable exceptions. First, recording of ethnicity was higher in QResearch than CPRD. Second, fewer patients in the CPRD cohort had a recorded family history of diabetes and coronary heart disease in a first degree relative under the age of 60 years. Recording differences in ethnicity and family history were not explained by geographic differences or difference in data capture period between the two databases. Given the similarity for the other risk factors and treatments, it is likely that the difference in ethnicity and family history recording reflects a difference in recording patterns between the two clinical computer systems rather than a true difference between the two cohorts. A similar pattern for recording of ethnicity and family history was also reported in the validation of QRISK on the Health Improvement Network (THIN database) [11 26]. This was thought to be due to different usage of clinical templates in the clinical system, with EMIS practices having ethnicity and family history included more often thereby prompting the user to enter this information in a more systematic fashion.

4.5 Comparison of QResearch and CPRD incidence rates

The age standardised incidence rates for each condition were generally marginally higher on QResearch than CPRD although the proportions of events identified on GP data (out of all events recorded on either GP or linked mortality data) was very close. This suggests that patterns of recording of major clinical events are very similar between QResearch and CPRD although the absolute value varies by clinical condition. For example, 91% of ischaemic stroke events recorded on either GP or linked mortality data are identified on the GP record compared with 99% of hip fractures. We also note the lower levels of total cardiovascular events in the GP clinical record which was between 13-15% lower than the total recorded on either the GP record, the linked mortality record or the linked hospital admissions record.

Some of this will reflect new sudden events where the first presentation was a hospital admission or death whilst others may reflect some under-representation of existing cases not recorded in the GP record. Our study is unable to distinguish between these two scenarios, though the latter one potentially has clinical consequences if the patient is not identified as having cardiovascular disease as they may not be offered secondary prevention.

We think that the information on baseline characteristics and incidence rates will have a utility beyond the present study since it suggests that both databases are fundamentally similar in many aspects and likely to generate similar results for a range of epidemiological studies[27].

4.6 Summary

In summary, we have tested a set of QPrediction scores using an external independent cohort of practices contributing to the CPRD. The results demonstrate good performance, comparable to the results obtained from QResearch, meaning that the findings of studies performed in either database are likely to be applicable in England.

5 Supporting information

Approvals
The project was approved in accordance with the QResearch® agreement with Trent Research Ethics Committee (ref 03/04/021) and approved by the ISAC committee of the CPRD (ref 13_079).

Acknowledgements
We thank EMIS and EMIS practices for their contribution to the QResearch database. We thank CPRD and Vision Practices for allowing access to the CPRD for this study.

Source of funding:
The validation of the QPrediction scores is funded by the National Institute for Health Research's School for Primary Care Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funding body did not play a role in the study design, writing of the manuscript or in the decision to submit the manuscript for publication.

Statement
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees

to permit this article (if accepted) to be published in HEART editions and any other BMJPGL products to exploit all subsidiary rights.

Contributorship

JHC initiated the study, undertook the literature review, data extraction, data manipulation and primary data analysis and write the first draft of the paper. JHC is the guarantor. CC contributed to the design, analysis, interpretation and drafting of the paper. PB contributed to the development of core ideas, the analysis plan, interpretation of the results and the drafting of the paper.

Competing interests

JHC 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). JHC is also director of ClinRisk Ltd which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of Medical Statistics at the University of Nottingham and a consultant statistician for ClinRisk Ltd. PB has received financial support for undertaking the validation work from the National School for Primary Care Research. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations. There are no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

All the algorithms validated in this paper are published as open source software under the GNU Lesser Public License. No additional data are available.

6 References

1. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008:bmj.39609.449676.25.
2. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ 2013;**346**.
3. Hippisley-Cox J, Coupland C, Robson J, et al. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009;**338**:b880-.
4. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. BMJ 2009;**339**:b4229-.
5. Hippisley-Cox J, Coupland C. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney(R) Scores. BMC Family Practice 2010;**11**:49.
6. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. BMJ 2011;**343**:d4656.
7. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions BMJ 2013:(in press).
8. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.
9. Collins GS, Mallett S, Altman DG. Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. BMJ 2011;**342**:d3651.
10. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010;**340**:c2442-.
11. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;**339**:b2584-.
12. Collins GS, Altman DG. External validation of the QDScore for predicting the 10-year risk of developing Type 2 diabetes. Diabetic Medicine 2011;**28**:599-607.
13. Hippisley-C J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleeding with anticoagulants: cohort study to derive and validate the QBleed score. BMJ 2014;**(in press)**.
14. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. Journal of statistical software 2011;**45**:1-20.
15. Gray A, Clarke P, Farmer A, et al. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). BMJ 2002;**325**(7369):860.
16. Royston P. Multiple imputation of missing values. Stata Journal 2004;**4**(3):227-41.

17. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007:bmj.39261.471806.55.
18. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;**23**:723-48.
19. Royston P. Explained variation for survival models. Stata J 2006;**6**:1-14.
20. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;**13**:33.
21. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;**338**(may28_1):b605-.
22. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**:4-14.
23. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;**60**:e128-36.
24. Collins GS, Altman A. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;**344**.
25. Collins GS, Altman D. Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney scores using a primary care database. British Journal of General Practice 2012;**62**(597):e243-e50.
26. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 2008;**94**:34-39.
27. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. BMJ Open 2014;**4**(4).
28. Hippisley-Cox J, Coupland C. QRISK2-2014 Annual Update Information, 2014:5.
29. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ 2012;**344**(may22 1):e3427-e27.
30. Hippisley-Cox J CC. QDiabetes 2013 Annual Update Information Secondary QDiabetes 2013 Annual Update Information 2013. http://qdiabetes.org/QDiabetes-2013-Annual-Update-Information.pdf.

Table 1 Summary of QPrediction scores including outcome and predictor variables

Score	Weblink∞	Outcome	Predictors
QDiabetes[3]	www.qdiabetes.org	10 year risk of type 2 diabetes±	**In men and women:** Age, sex, ethnicity, deprivation, smoking, family history of diabetes, cardiovascular disease, treated hypertension, steroid tables, body mass index
QRISK2[28]	www.qrisk.org	10 year risk of CVD**	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation.
QStroke[2]	www.qstroke.org	10 year risk of ischaemic stroke or TIA±	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, congestive cardiac failure, valvular heart disease
QKidney[5]	www.qkidney.org	5 year risk of moderate or severe kidney failureμ	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, body mass index, family history of kidney disease, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, SLE, peripheral vascular disease, kidney stones, NSAIDs
QThrombosis	www.qthrombosis.org	5 year risk of venous thromboembolism±	**In men and women**: age, body mass index, smoking status, varicose veins, congestive cardiac failure, chronic renal disease, cancer, chronic obstructive pulmonary disease, inflammatory bowel disease, hospital admission in past six months, and current prescriptions for antipsychotic drugs. Additionally in women: combined oral contraceptives, tamoxifen, and hormone replacement therapy
QBleed[13]	www.qbleed.org	5 year risk of upper gastrointestinal bleed in patient starting anticoagulants vs others*	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; venous thromboembolism; congestive cardiac failure; treated hypertension; cancer; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QBleed[13]	www.qbleed.org	5 year risk of intracranial bleed in patient starting anticoagulants vs others *	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; treated hypertension; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QFracture[29]	www.qfracture.org	10 year risk of hip fracture± 10 year risk of	**In women:** HRT usage, age, body mass index, smoking status, recorded alcohol use, parental history of osteoporosis, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls, menopausal symptoms, chronic liver disease, gastrointestinal malabsorption and other endocrine disorders.

		osteoporotic fracture [µ]	**In men:** age, body mass index, smoking status, recorded alcohol use, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls and liver disease.

∞the web link has the relevant calculator, links to academic papers, additional information including links to the open source software

[±] recorded either on GP record or linked ONS mortality record;

µ recorded on the GP record.

*Recorded either on linked hospital admissions record or ONS mortality record

**Recorded either on linked hospital admissions record or ONS mortality or linked hospital admissions record

Table 2 Comparison of baseline characteristics of patients in CPRD validation cohort and QResearch comparison cohort

	CPRD	*CPRD*	*QResearch*	*QResearch*
	Men (%)	**Women (%)**	**Men (%)**	**Women (%)**
total	1,588,803	1,682,709	3,304,145	3,454,504
Ageband				
25-34 years	427975 (26.9)	467192 (27.8)	1083589 (32.8)	1179742 (34.2)
35-44 years	396680 (25.0)	364150 (21.6)	814988 (24.7)	731089 (21.2)
45-54 years	294274 (18.5)	277663 (16.5)	558553 (16.9)	516188 (14.9)
55-64 years	212817 (13.4)	211636 (12.6)	390229 (11.8)	389266 (11.3)
65-74 years	148180 (9.3)	164172 (9.8)	267997 (8.1)	298847 (8.7)
75+ years	108877 (6.9)	197896 (11.8)	188789 (5.7)	339372 (9.8)
mean Townsend score (SD)	-.5 (3.2)	-.5 (3.2)	.3 (3.6)	.2 (3.6)
Care home resident	1407 (0.1)	3466 (0.2)	2983 (0.1)	7411 (0.2)
Ethnicity recorded	587879 (37.0)	658835 (39.2)	1859462 (56.3)	2077181 (60.1)
White or not recorded	1515113 (95.4)	1602212 (95.2)	3010061 (91.1)	3149618 (91.2)
Indian	16442 (1.0)	16025 (1.0)	56156 (1.7)	50406 (1.5)
Pakistani	6606 (0.4)	6146 (0.4)	30632 (0.9)	23405 (0.7)
Bangladeshi	2419 (0.2)	1688 (0.1)	23017 (0.7)	17450 (0.5)
Other Asian	10795 (0.7)	11873 (0.7)	32513 (1.0)	36886 (1.1)
Caribbean	4989 (0.3)	6425 (0.4)	25782 (0.8)	32953 (1.0)
Black African	12883 (0.8)	14771 (0.9)	51980 (1.6)	56528 (1.6)
Chinese	2914 (0.2)	4176 (0.2)	16084 (0.5)	23043 (0.7)
Other ethnic group	16642 (1.0)	19393 (1.2)	57920 (1.8)	64215 (1.9)
Smoking status recorded	1442088 (90.8)	1595538 (94.8)	2943405 (89.1)	3219598 (93.2)
Non smoker	613833 (38.6)	834721 (49.6)	1449694 (43.9)	1973691 (57.1)
Ex-smoker	252873 (15.9)	222615 (13.2)	611837 (18.5)	545125 (15.8)
Light smoker (1-9/day)	104466 (6.6)	109864 (6.5)	472614 (14.3)	384482 (11.1)
Moderate smoker (10-19/day)	183000 (11.5)	179391 (10.7)	223631 (6.8)	202776 (5.9)
Heavy smoker (20+/day)	142438 (9.0)	87474 (5.2)	185629 (5.6)	113524 (3.3)
Smoker amount not recorded	145478 (9.2)	161473 (9.6)	0 (0.0)	0 (0.0)
Alcohol status recorded	1238110 (77.9)	1379002 (82.0)	2584335 (78.2)	2834426 (82.1)
Non drinker	163633 (10.3)	318880 (19.0)	583752 (17.7)	1035692 (30.0)
Trivial <1u/day	460091 (29.0)	726851 (43.2)	782985 (23.7)	1144469 (33.1)
Light 1-2u/day	411261 (25.9)	290547 (17.3)	481674 (14.6)	402750 (11.7)
Moderate 3-6/day	166328 (10.5)	36763 (2.2)	648549 (19.6)	237679 (6.9)
Heavy 7-9u/day	19612 (1.2)	2853 (0.2)	54083 (1.6)	7152 (0.2)
Very Heavy >/day	17185 (1.1)	3108 (0.2)	24468 (0.7)	5195 (0.2)
Family History				
family history of CHD	68805 (4.3)	80985 (4.8)	326995 (9.9)	417537 (12.1)
family history of diabetes	96810 (6.1)	132390 (7.9)	357109 (10.8)	487397 (14.1)
family history osteoporosis	880 (0.1)	10062 (0.6)	1655 (0.1)	17529 (0.5)
family history kidney disease	1253 (0.1)	1586 (0.1)	2034 (0.1)	2769 (0.1)

Clinical Values recorded				
BMI recorded	1268235 (79.8)	1481918 (88.1)	2553514 (77.3)	2857742 (82.7)
mean BMI (SD)	29.6 (6.8)	28.2 (7.0)	29.4 (6.8)	28.2 (7.0)
SBP recorded	1359560 (85.6)	1590226 (94.5)	2755733 (83.4)	3190390 (92.4)
mean SBP(SD)	133.1 (23.6)	128.6 (22.7)	132.2 (18.3)	127.1 (20.8)
cholesterol/HDL ratio recorded	587865 (37.0)	606035 (36.0)	1323503 (40.1)	1368180 (39.6)
mean cholesterol ratio (SD)	4.4 (1.4)	3.7 (1.2)	4.4 (1.4)	3.7 (1.2)
platelets recorded	223461 (14.1)	382799 (22.7)	478596 (14.5)	829702 (24.0)
platelets < 150 or > 480	11051 (0.7)	13282 (0.8)	23479 (0.7)	27009 (0.8)
Creatinine recorded	811779 (51.1)	997118 (59.3)	1714337 (51.9)	2053036 (59.4)
Mean creatinine (SD)	96.7 (149.1)	79.7 (911.6)	95.5 (30.7)	78 (23.7)

Table 3 Prescribed medication and clinical diagnoses recorded at baseline in CPRD validation cohort and QResearch comparison cohort

	CPRD men (%)	*CPRD women (%)*	*QResearch men (%)*	*QResearch women (%)*
Prescribed medication				
anticoagulants	15955 (1.0)	13077 (0.8)	27024 (0.8)	22178 (0.6)
antidepressants	101553 (6.4)	235797 (14.0)	178532 (5.4)	398018 (11.5)
antipsychotics	33884 (2.1)	79514 (4.7)	47464 (1.4)	92307 (2.7)
antiplatelets	97475 (6.1)	92816 (5.5)	160910 (4.9)	153405 (4.4)
oral NSAIDs	246515 (15.5)	346416 (20.6)	396026(12.0)	556644 (16.1)
tamoxifen	n/a	9231 (0.5)	n/a	18343 (0.5)
oestrogen only Hormone replacement therapy	n/a	119373 (7.1)	n/a	208333 (6.0)
oral corticosteroids	45597 (2.9)	71352 (4.2)	54354 (1.6)	88205 (2.6)
oral contraceptive pill	n/a	174287 (10.4)	n/a	332696 (9.6)
Recorded Diagnoses				
congestive cardiac failure	15836 (1.0)	19707 (1.2)	24965 (0.8)	28852 (0.8)
atrial fibrillation	20125 (1.3)	20102 (1.2)	33499 (1.0)	32580 (0.9)
coronary heart disease	80377 (5.1)	57703 (3.4)	130220 (3.9)	88606 (2.6)
cardiovascular disease	101430 (6.4)	83167 (4.9)	165495 (5.0)	130214 (3.8)
peripheral vascular disease	17029 (1.1)	13101 (0.8)	25004 (0.8)	17078 (0.5)
venous thromboembolism	15072 (0.9)	23090 (1.4)	27086 (0.8)	40813 (1.2)
rheumatoid or SLE	7455 (0.5)	19010 (1.1)	21453 (0.6)	48447 (1.4)
rheumatoid arthritis	7243 (0.5)	17468 (1.0)	21142 (0.6)	45542 (1.3)
SLE	228 (0.0)	1756 (0.1)	351 (0.0)	3374 (0.1)
type 1 diabetes	6238 (0.4)	4924 (0.3)	12029 (0.4)	9612 (0.3)
type 2 diabetes	51634 (3.2)	43271 (2.6)	95401 (2.9)	79654 (2.3)
treated hypertension	123584 (7.8)	161709 (9.6)	210516 (6.4)	267076 (7.7)
chronic renal disease	3968 (0.2)	4082 (0.2)	8550 (0.3)	8995 (0.3)
moderate/severe kidney failure	14107 (0.9)	9500 (0.6)	30407 (0.9)	21509 (0.6)
severe kidney failure	1603 (0.1)	1125 (0.1)	3641 (0.1)	2672 (0.1)
renal stones	13415 (0.8)	6443 (0.4)	37422 (1.1)	29204 (0.8)
inflammatory bowel disease	8962 (0.6)	10208 (0.6)	17762 (0.5)	19502 (0.6)
dementia	6686 (0.4)	16634 (1.0)	12872 (0.4)	30497 (0.9)
parkinsons disease	4546 (0.3)	4676 (0.3)	6830 (0.2)	6611 (0.2)
epilepsy or anticonvulsants	47170 (3.0)	71171 (4.2)	56516 (1.7)	61561 (1.8)
cancer	26866 (1.7)	43908 (2.6)	51649 (1.6)	79326 (2.3)
liver disease	3959 (0.2)	2893 (0.2)	9947 (0.3)	6410 (0.2)
chronic liver disease or pancreatitis	5521 (0.3)	4051 (0.2)	13069 (0.4)	8729 (0.3)
oesophageal varices	469 (0.0)	340 (0.0)	1626 (0.0)	1388 (0.0)
prior haemorrhage	97562 (6.1)	79765 (4.7)	203278 (6.2)	147533 (4.3)
malabsorption	7343 (0.5)	9375 (0.6)	21042 (0.6)	26002 (0.8)
endocrine diseases	3082 (0.2)	14097 (0.8)	6026 (0.2)	27731 (0.8)
COPD	24029 (1.5)	20737 (1.2)	41281 (1.2)	34785 (1.0)
asthma or COPD	142974 (9.0)	169503 (10.1)	273768 (8.3)	310027 (9.0)
history of falls	28878 (1.8)	61905 (3.7)	34584 (1.0)	67465 (2.0)
prior fracture	24265 (1.5)	45752 (2.7)	62092 (1.9)	89000 (2.6)

varicose vein surgery	18979	(1.2)	47012	(2.8)	35651	(1.1)	85602	(2.5)
emergency admissions or hip op	3483	(0.2)	5266	(0.3)	3335	(0.1)	5508	(0.2)

Table 4 Numbers of patients eligible for each score in the CPRD validation cohort and number of patients with complete risk factor recording not requiring multiple imputation.

Risk Score	*Clinical outcome*	*Eligible age range*	*exclusion criteria at study entry*	*total in age range*	*total with exclusions*	*total eligible for analysis*	*Total complete data*	*% complete data*
QDiabetes	Type 2 diabetes	25-84	type 1 or 2 diabetes at study entry	3,177,192	99,189	3,078,003	2,467,642	80.2
QStroke	ischaemic stroke	25-84	existing stroke or anticoagulants at study entry	3,177,192	70,961	3,106,231	1,032,184	33.2
QRISK2	cardiovascular disease	25-84	existing CVD or statins at study entry	3,177,192	232,722	2,944,470	906,781	30.8
QThrombosis	thromboembolism	25-84	existing VTE or anticoagulants at study entry	3,177,192	53,904	3,123,288	2,513,347	80.5
QFracture	fractured neck of femur	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QFracture	osteoporotic fracture	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QKidney	moderate or severe kidney failure	35-74	existing moderate or severe kidney failure	2,069,572	10,518	2,059,054	1,146,619	55.7
QKidney	severe kidney failure	35-74	existing severe kidney failure	2,069,572	1,930	2,067,642	1,153,979	55.8
QBleed	upper gastro-intestinal bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0
QBleed	intracranial bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0

*entry date was 01.01.1998 except for upper GI bleed and intracranial bleed where entry date was 01.01.2007

Table 5 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in women

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	48,143	99.88	4.13 (4.10 to 4.17)	102,544	99.93	4.33 (4.31 to 4.36)	0.95
	GP or ONS	48,203	n/a	4.13 (4.10 to 4.17)	102,618	n/a	4.34 (4.31 to 4.36)	0.95
Ischaemic stroke	GP data	32,283	90.64	2.45 (2.42 to 2.48)	63,582	90.22	2.45 (2.44 to 2.47)	1.00
	GP or ONS	35,617	n/a	2.62 (2.59 to 2.64)	70,477	n/a	2.70 (2.68 to 2.72)	0.97
Cardiovascular disease	GP data	55,833	85.71	5.41 (5.37 to 5.46)	107,412	84.96	4.32 (4.30 to 4.35)	1.25
	GP or ONS	65,143	n/a	6.32 (6.27 to 6.37)	126,433	n/a	5.03 (5.01 to 5.06)	1.26
	GP or ONS or HES	69,202	n/a	6.72 (6.67 to 6.77)	140,510	n/a	5.63 (5.60 to 5.66)	1.19
Thromboembolism	GP data	18,199	91.1	1.52 (1.49 to 1.54)	35,971	90.55	1.46 (1.44 to 1.47)	1.04
	GP or ONS	19,978	n/a	1.64 (1.62 to 1.67)	39,727	n/a	1.60 (1.58 to 1.62)	1.03
Fractured neck of femur	GP data	17,529	99.98	1.32 (1.30 to 1.34)	34,821	99.99	1.40 (1.39 to 1.42)	0.94
	GP or ONS	17,533	n/a	1.32 (1.30 to 1.34)	34,825	n/a	1.40 (1.39 to 1.42)	0.94
Osteoporotic fracture	GP data	34,528	n/a	2.89 (2.58 to 3.20)	81,334	n/a	3.63 (3.61 to 3.66)	0.80
mod /severe kidney failure	GP data	19,902	n/a	2.06 (1.76 to 2.36)	48,665	n/a	2.81 (2.78 to 2.83)	0.73
severe kidney failure	GP data	1,737	n/a	0.18 (0.09 to 0.27)	4,150	n/a	0.24 (0.24 to 0.25)	0.74

Table 6 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in men

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	60,731	99.92	5.84 (5.79 to 5.89)	128,234	99.94	5.97 (5.94 to 6.00)	0.98
	GP or ONS	60,782	n/a	5.84 (5.80 to 5.89)	128,317	n/a	5.98 (5.94 to 6.01)	0.98
ischaemic stroke	GP data	32,223	93.55	3.17 (3.14 to 3.20)	63,480	92.85	3.10 (3.08 to 3.13)	1.02
	GP or ONS	34,443	n/a	3.33 (3.30 to 3.37)	68,366	n/a	3.37 (3.34 to 3.40)	0.99
Cardiovascular disease	GP data	70,283	86.7	7.38 (7.33 to 7.44)	137,136	86.12	7.12 (7.08 to 7.16)	1.03
	GP or ONS	81,068	n/a	8.52 (8.46 to 8.58)	159,240	n/a	8.37 (8.33 to 8.41)	1.02
	GP or ONS or HES	84,620	n/a	8.90 (8.84 to 8.96)	174,405	n/a	9.17 (9.13 to 9.21)	0.97
thromboembolism	GP data	15,655	92.32	1.49 (1.46 to 1.51)	31,503	92.22	1.44 (1.43 to 1.46)	1.03
	GP or ONS	16,958	n/a	1.61 (1.59 to 1.63)	34,161	n/a	1.57 (1.56 to 1.59)	1.02
fractured neck of femur	GP data	5,706	99.98	0.65 (0.63 to 0.67)	12,435	99.98	0.71 (0.70 to 0.73)	0.91
	GP or ONS	5,707	n/a	0.65 (0.63 to 0.67)	12,438	n/a	0.71 (0.70 to 0.73)	0.91
osteoporotic fracture	GP data	11,169	n/a	1.29 (1.05 to 1.52)	28,555	n/a	1.54 (1.52 to 1.55)	0.84
Mod/severe kidney failure	GP data	37,597	n/a	4.88 (4.37 to 5.38)	86,649	n/a	5.82 (5.78 to 5.85)	0.84
severe kidney failure	GP data	3,472	n/a	0.54 (0.38 to 0.71)	7,372	n/a	0.47 (0.46 to 0.48)	1.15

Table 7 Performance of QPrediction scores on the CPRD validation cohort compared with published results for the QResearch validation cohort

		CPRD	*CPRD*	*QResearch*	*QResearch*
		women	**men**	**women**	**men**
	statistic	**mean (95% CI)**	**mean (95% CI)**	**mean (95% CI)**	**mean (95% CI)**
QDiabetes 2013 (type 2 diabetes)[30]	ROC	0.846 (0.844 to 0.848)	0.818 (0.816 to 0.82)	0.853 (0.851 to 0.856)	0.837 (0.835 to 0.840)
	R^2 (%)	49.6 (49.2 to 50.1)	45.7 (45.3 to 46.2)	50.8 (50.3 to 51.4)	48.1 (47.6 to 48.6)
	D statistic	2.032 (2.015 to 2.049)	1.879 (1.863 to 1.895)	2.081 (2.058 to 2.104)	1.971 (1.951 to 1.991)
QKidney -2010[5] (moderate or severe kidney failure)	ROC	0.875 (0.87 to 0.879)	0.88 (0.878 to 0.883)	0.877 (0.873 to 0.880)	0.878 (0.874 to 0.882)
	R^2 (%)	58.3 (57.8 to 58.7)	57.5 (57.1 to 57.8)	56.45 (55.40 to 57.50)	58.29 (55.31 to 61.26)
	D statistic	2.418 (2.394 to 2.442)	2.379 (2.361 to 2.397)	2.33 (2.28 to 2.40)	2.42 (2.28 to 2.56)
QKidney -2010 (severe kidney failure)[5]	ROC	0.839 (0.822 to 0.855)	0.851 (0.84 to 0.862)	0.843 (0.825 to 0.860)	0.846 (0.829 to 0.862)
	R^2 (%)	51.4 (49.5 to 53.2)	53.8 (52.6 to 55.1)	55.39 (52.59 to 58.18)	56.65 (53.94 to 59.35)
	D statistic	2.103 (2.025 to 2.182)	2.21 (2.154 to 2.266)	2.28 (2.15 to 2.41)	2.34 (2.21 to 2.47)
QRISK2-2014[28] (cardiovascular disease)	ROC	0.883 (0.882 to 0.884)	0.859 (0.858 to 0.861)	0.892 (0.892 to 0.895)	0.871 (0.869 to 0.873)
	R^2 (%)	56.4 (56.1 to 56.7)	50.9 (50.6 to 51.2)	58.8 (58.4 to 59.1)	53.3 (52.9 to 53.7)
	D statistic	2.328 (2.313 to 2.343)	2.085 (2.071 to 2.098)	2.443 (2.423 to 2.463)	2.188 (2.171 to 2.205)
QStroke-2013[2] (ischaemic stroke or TIA)	ROC	0.882 (0.88 to 0.883)	0.869 (0.867 to 0.87)	0.877 (0.875 to 0.879)	0.866 (0.864 to 0.868)
	R^2 (%)	58.4 (58.1 to 58.8)	55.3 (54.9 to 55.7)	57.3 (56.8 to 57.8)	55.1 (54.6 to 55.7)
	D statistic	2.427 (2.408 to 2.446)	2.278 (2.259 to 2.297)	2.37 (2.35 to 2.40)	2.27 (2.24 to 2.30)
QThrombosis-2010[6] (venous thromboembolism)	ROC	0.756 (0.751 to 0.761)	0.765 (0.760 to 0.770)	0.75 (0.74 to 0.76)	0.75 (0.74 to 0.76)
	R^2 (%)	35.3 (34.5 to 36.1)	34.5 (33.7 to 35.4)	32.78 (31.08 to 34.48)	33.51 (31.71 to 35.30)

	D statistic	1.512 (1.485 to 1.538)	1.486 (1.458 to 1.513)	1.43 (1.37 to 1.49)	1.45 (1.39 to 1.51)
QBleed-2014[13] (upper gastrointestinal bleed)	ROC statistic	0.775 (0.770 to 0.781)	0.759 (0.753 to 0.764)	0.766 (0.758 to 0.775)	0.747 (0.738 to 0.756)
	R^2 (%)	44.7 (43.6 to 45.9)	41.6 (40.5 to 42.8)	40.7 (38.9 to 42.6)	36.9 (35.1 to 38.7)
	D statistic	1.842 (1.798 to 1.885)	1.729 (1.687 to 1.771)	1.70 (1.63 to 1.76)	1.57 (1.51 to 1.63)
QBleed-2014[13] (intracranial bleed)	ROC statistic	0.808 (0.801 to 0.816)	0.789 (0.780 to 0.797)	0.847 (0.838 to 0.856)	0.812 (0.80 to 0.824)
	R^2 (%)	51.7 (50.1 to 53.3)	50.0 (48.3 to 51.7)	58.0 (56.0 to 60.0)	53.3 (51.1 to 55.4)
	D statistic	2.118 (2.051 to 2.186)	2.046 (1.977 to 2.116)	2.40 (2.30 to 2.50)	2.19 (2.09 to 2.28)
QFracture-2012[29] (fractured neck of femur)	ROC	0.89 (0.888 to 0.892)	0.872 (0.867 to 0.877)	0.893 (0.890 to 0.896)	0.875 (0.868 to 0.883)
	R^2 (%)	70.6 (70.2 to 71)	69.2 (68.5 to 70)	71.73 (71.10 to 72.30)	70.37 (69.25 to 71.49)
	D statistic	3.171 (3.139 to 3.203)	3.07 (3.016 to 3.124)	3.26 (3.21 to 3.31)	3.15 (3.06 to 3.24)
QFracture -2012[29] (osteoporotic fracture: hip, spine, wrist,humerus)	ROC	0.817 (0.814 to 0.819)	0.768 (0.763 to 0.773)	0.790 (0.787 to 0.793)	0.711 (0.703 to 0.719)
	R^2 (%)	56.3 (55.8 to 56.7)	49.8 (48.9 to 50.7)	51.9 (51.2 to 52.6)	38.20 (36.89 to 39.57)
	D statistic	2.322 (2.301 to 2.343)	2.038 (2.002 to 2.075)	2.13 (2.10 to 2.15)	1.61 (1.56 to 1.66)

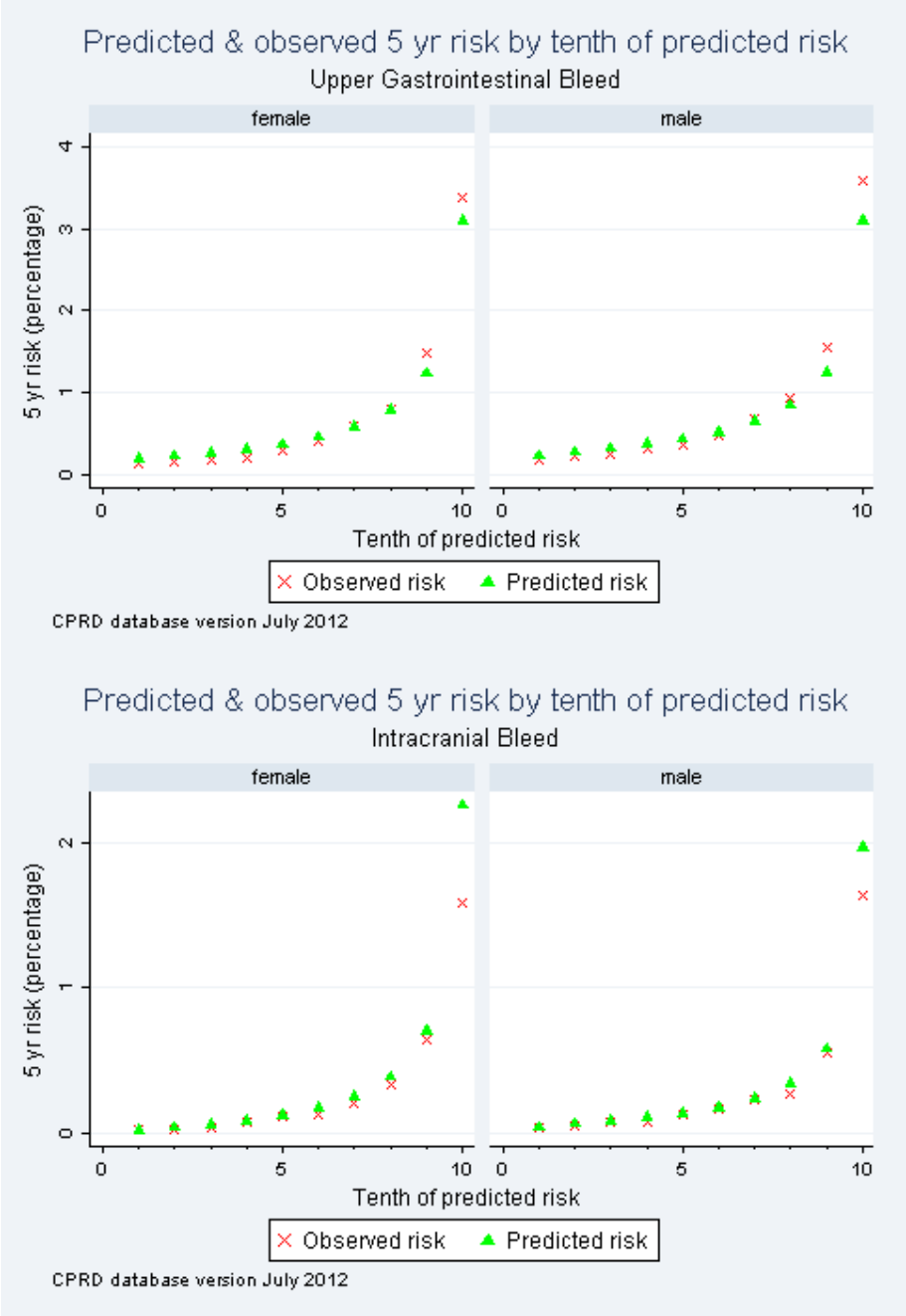
Notes on understanding validation statistics:

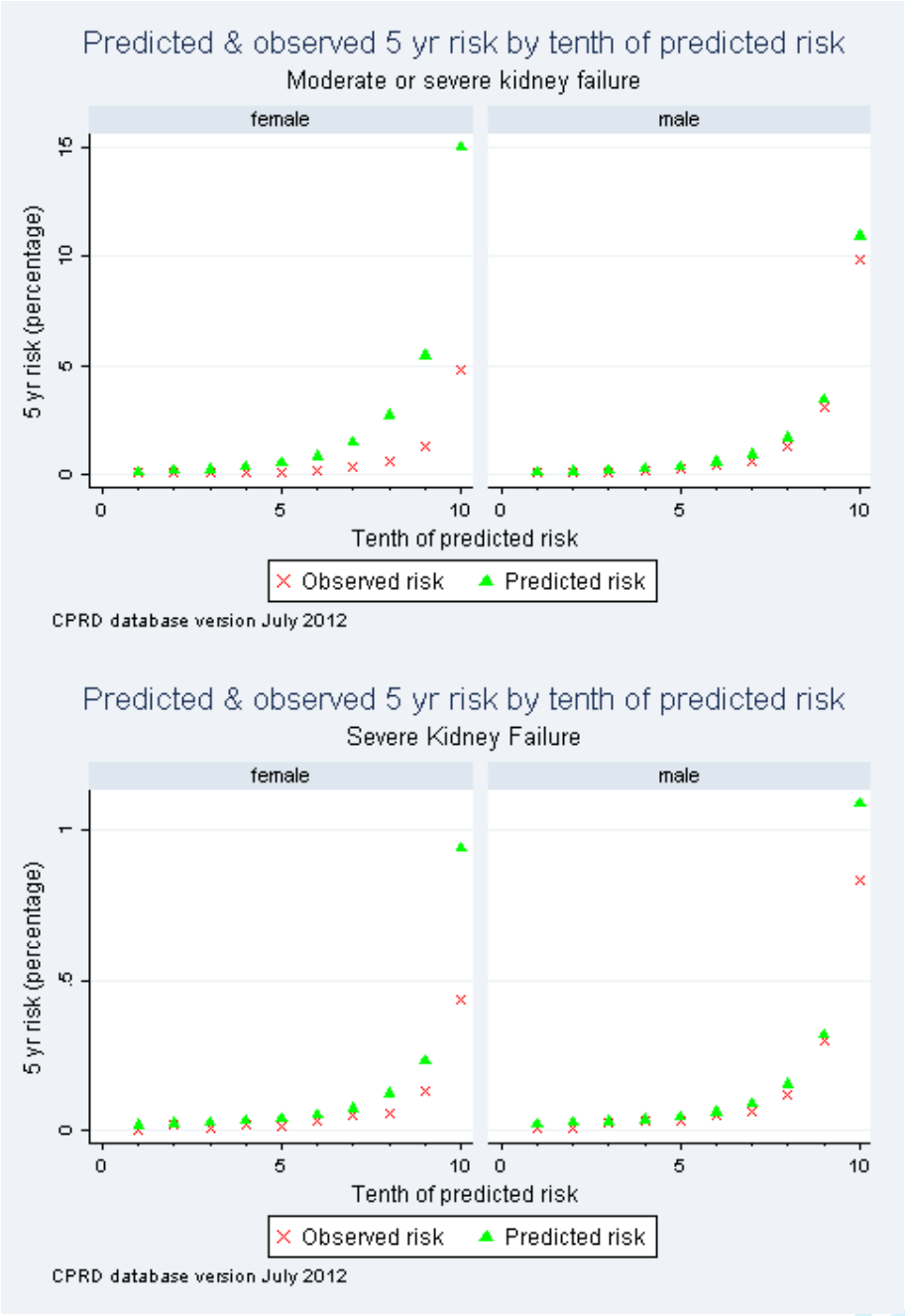
D statistic is a measure of discrimination - higher values indicate better discrimination
ROC statistic is a measure of discrimination - higher values indicate better discrimination
The R^2 statistic is a measure of explained variation - higher values indicate more variation is explained

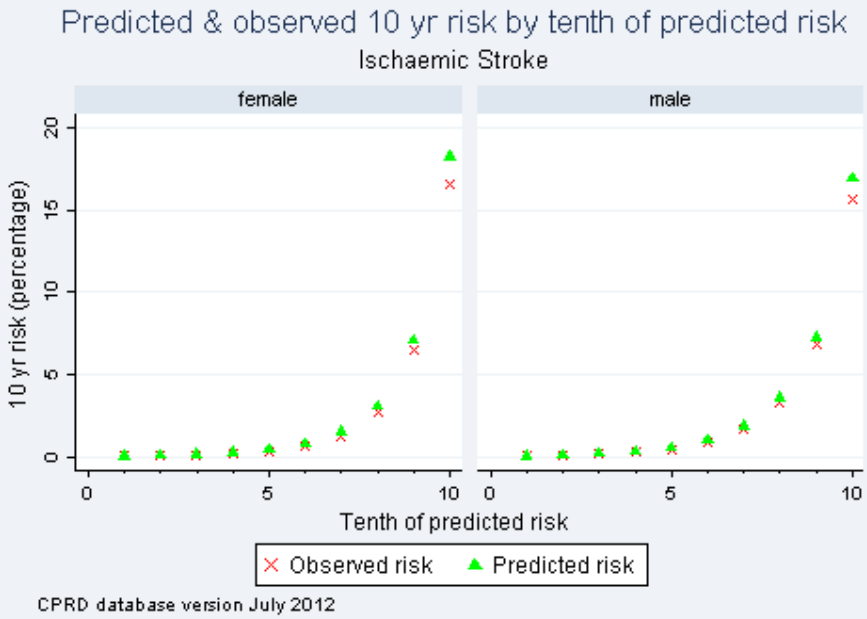
Table 8 Performance of each score for predicting the relevant outcome in the CPRD validation cohort. The cut off is the threshold of predicted risk for the top decile in the CPRD cohort.

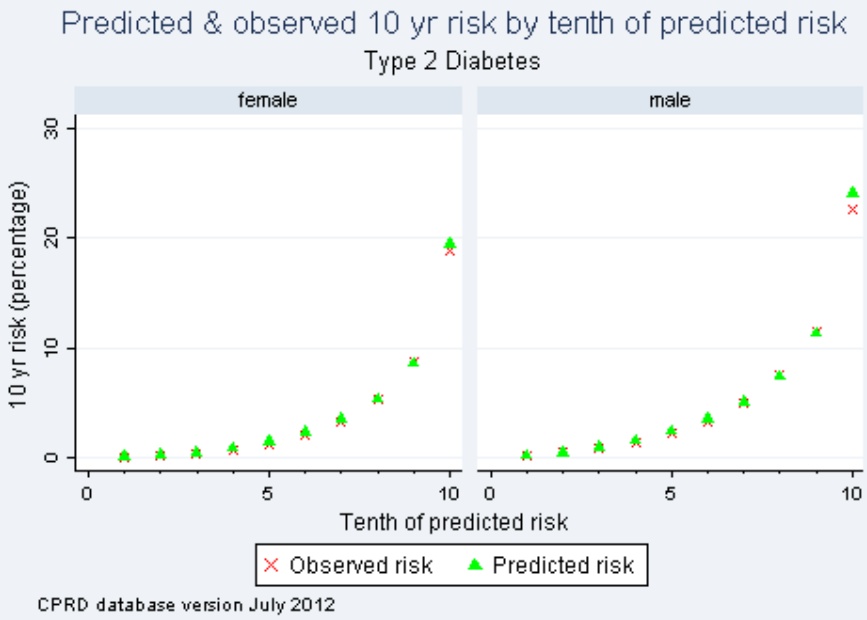
score	*outcome*	*duration*	*cut off (%) for top decile predicted risk*	*Sensitivity (%)*	*Specificity (%)*	*observed risk (%)*
QDiabetes	Type 2 diabetes	10 yr risk	13.0	44.8	91.0	20.8
QStroke	Ischaemic stroke	10 yr risk	10.5	54.7	90.8	16.1
QRISK2	cardiovascular disease	10 yr risk	20.7	49.9	91.9	31.8
QThrombosis	venous thromboembolism	5 yr risk	1.5	36.2	90.1	2.6
QKidney	moderate-severe kidney failure	5 yr risk	6.3	59.1	90.5	6.9
QKidney	severe kidney failure	5 yr risk	0.4	58.5	90.0	0.7
QBleed	upper GI bleed	5 yr risk	1.6	38.0	90.2	3.5
QBleed	intracranial bleed	5 yr risk	0.9	44.2	90.1	1.6
QFracture	fractured neck of femur	10 yr risk	3.7	66.5	90.4	9.4
QFracture	osteoporotic fracture	10 yr risk	7.8	49.6	90.5	13.1

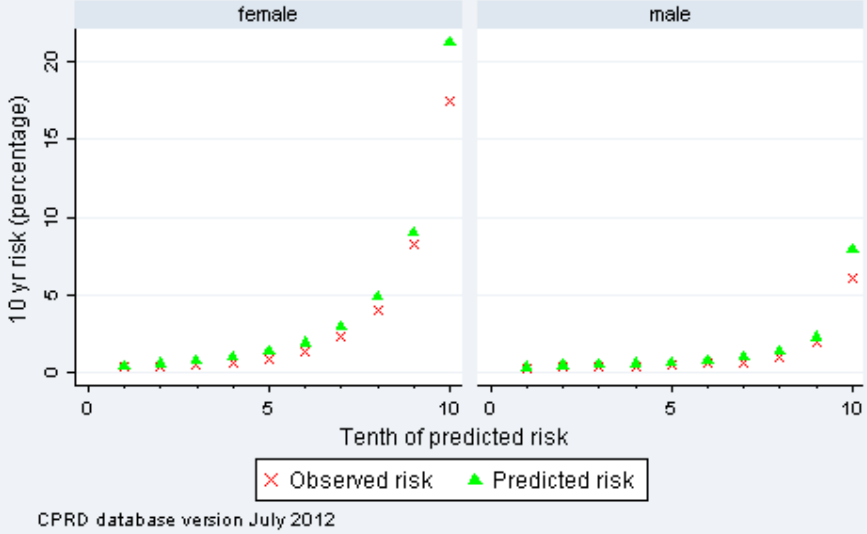
Figure 1 Calibration of each QPrediction score comparing the mean predicted risks with the observed risks in the CPRD cohort.

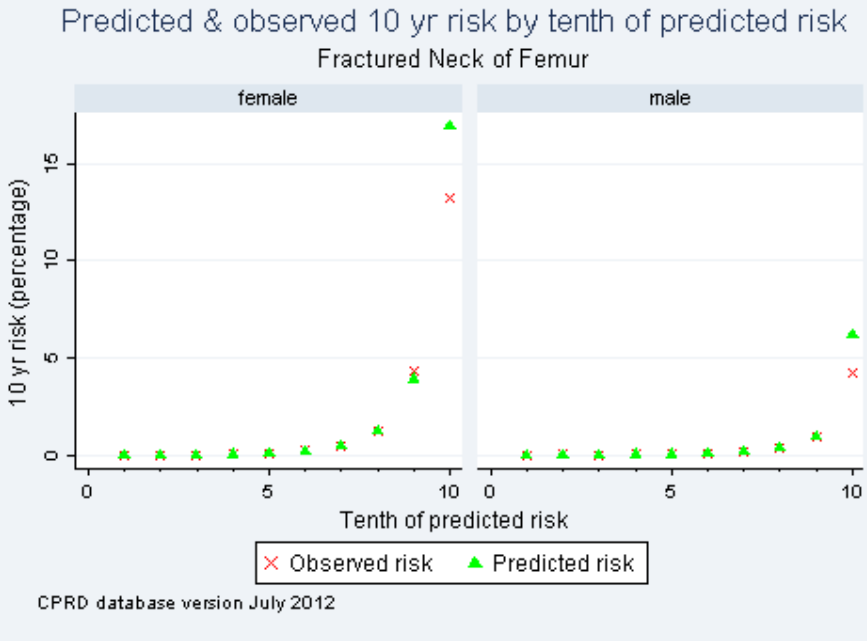


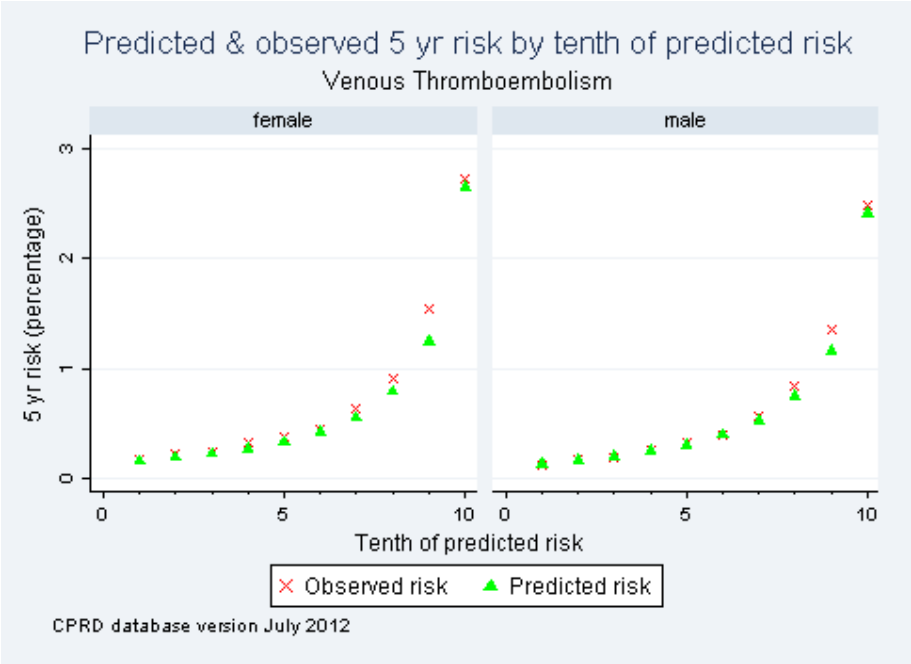












Web extra table 1. Numbers of patients in CPRD and QResearch by geographical area

	CPRD	*Col %*	*QResearch*	*Col %*
East Midlands	109,428	3.3	500,970	7.4
East of England	397,008	12.1	528,379	7.8
London	563,353	17.2	1,711,956	25.3
North East	59,558	1.8	269,695	4.0
North West	474,457	14.5	830,047	12.3
South Central	411,571	12.6	696,070	10.3
South East	362,319	11.1	545,811	8.1
South West	397,735	12.2	700,041	10.4
West Midlands	348,614	10.7	589,548	8.7
Yorks & Humber	147,469	4.5	386,132	5.7
Total	**3,271,512**	**100.00**	**6,758,649**	**100.00**

.

Web table 2 Recording of ethnicity and family history of coronary heart disease (FH CHD) by geographical area

	CPRD linked data					QResearch					Ratio recording QResearch:CPRD	
	total patients	ethnicity recorded		FH CHD recorded		total patients	ethnicity recorded		FH CHD recorded			
	count	count	Row %	count	Row %	count	count	Row %	count	Row %	ethnicity	FH
East Midlands	109,428	23,636	21.6	7,912	7.2	500,970	284,194	56.7	60,278	12.0	2.6	1.7
East of England	397,008	127,427	32.1	17,597	4.4	528,379	319,742	60.5	54,252	10.3	1.9	2.3
London	563,353	308,285	54.7	21,034	3.7	1,711,956	1,095,835	64.0	162,282	9.5	1.2	2.5
North East	59,558	17,140	28.8	3,326	5.6	269,695	195,127	72.4	43,168	16.0	2.5	2.9
North West	474,457	196,987	41.5	25,915	5.5	830,047	460,640	55.5	112,718	13.6	1.3	2.5
South Central	411,571	140,448	34.1	16,989	4.1	696,070	380,908	54.7	73,051	10.5	1.6	2.5
South East	362,319	90,633	25.0	12,618	3.5	545,811	262,861	48.2	45,765	8.4	1.9	2.4
South West	397,735	137,806	34.6	17,829	4.5	700,041	375,155	53.6	75,091	10.7	1.5	2.4
West Midlands	348,614	148,012	42.5	20,375	5.8	589,548	347,479	58.9	63,412	10.8	1.4	1.8
Yorks & Humber	147,469	56,340	38.2	6,195	4.2	386,132	214,702	55.6	54,515	14.1	1.5	3.4
Total	**3,271,512**	**1,246,714**	**38.1**	**149,790**	**4.6**	**6,758,649**	**3,936,643**	**58.2**	**744,532**	**11.0**	**1.5**	**2.4**

Web extra table 3 comparison of baseline characteristics for the CPRD cohort with linked data used for the validation cohort in this study and the CPRD cohort without linked data (which was not used for the validation of the QPrediction scores).

	CPRD linked data	CPRD unlinked data
female	1682709 (51.4)	1166103 (51.3)
male	1588803 (48.6)	1108235 (48.7)
25-34 years	895167 (27.4)	665645 (29.3)
35-44 years	760830 (23.3)	503091 (22.1)
45-54 years	571937 (17.5)	384686 (16.9)
55-64 years	424453 (13.0)	293826 (12.9)
65-74 years	312352 (9.5)	216937 (9.5)
75+ years	306773 (9.4)	210153 (9.2)
Ethnicity recorded	1246714 (38.1)	645829 (28.4)
White or not recorded	3117325 (95.3)	2209396 (97.1)
Indian	32467 (1.0)	11751 (0.5)
Pakistani	12752 (0.4)	6358 (0.3)
Bangladeshi	4107 (0.1)	2682 (0.1)
Other Asian	22668 (0.7)	8854 (0.4)
Caribbean	11414 (0.3)	4812 (0.2)
Black African	27654 (0.8)	9751 (0.4)
Chinese	7090 (0.2)	3416 (0.2)
Other ethnic group	36035 (1.1)	17318 (0.8)
Smoking status recorded	3037626 (92.9)	2066777 (90.9)
Non smoker	1448554 (44.3)	1006511 (44.3)
Ex-smoker	475488 (14.5)	306460 (13.5)
Light smoker (1-9/day)	214330 (6.6)	127487 (5.6)
Moderate smoker (10-19/day)	362391 (11.1)	277693 (12.2)
Heavy smoker (20+/day)	229912 (7.0)	159718 (7.0)
Smoker amount not recorded	306951 (9.4)	188908 (8.3)
Alcohol status recorded	2617112 (80.0)	1759758 (77.4)
Non drinker	482513 (14.7)	361566 (15.9)
Trivial <1u/day	1186942 (36.3)	754470 (33.2)
Light 1-2u/day	701808 (21.5)	479407 (21.1)
Moderate 3-6/day	203091 (6.2)	135130 (5.9)
Heavy 7-9u/day	22465 (0.7)	15645 (0.7)
Very Heavy >/day	20293 (0.6)	13540 (0.6)
family history		
family history of CHD	149790 (4.6)	97401 (4.3)
family history of diabetes	229200 (7.0)	144713 (6.4)
family history osteoporosis	10942 (0.3)	6164 (0.3)
family history kidney disease	2839 (0.1)	1592 (0.1)

prescribed medication		
antidepressants	337350 (10.3)	232657 (10.2)
anticoagulants	29032 (0.9)	20338 (0.9)
antipsychotics	113398 (3.5)	76819 (3.4)
oral NSAIDs	592931 (18.1)	425739 (18.7)
tamoxifen	9343 (0.3)	6107 (0.3)
antiplatelets	190291 (5.8)	137793 (6.1)
oestrogen only HRT	119413 (3.7)	75448 (3.3)
corticosteroids	116949 (3.6)	70793 (3.1)
oral contraceptive pill	174288 (5.3)	126218 (5.5)
recorded diagnoses		
congestive cardiac failure	35543 (1.1)	25823 (1.1)
atrial fibrillation	40227 (1.2)	27032 (1.2)
coronary heart disease	138080 (4.2)	102493 (4.5)
cardiovascular disease	184597 (5.6)	134650 (5.9)
rheumatoid arthritis	24711 (0.8)	17427 (0.8)
chronic renal disease	8050 (0.2)	5774 (0.3)
type 1 diabetes	11162 (0.3)	7778 (0.3)
type 2 diabetes	94905 (2.9)	63240 (2.8)
venous thromboembolism	38162 (1.2)	23593 (1.0)
varicose veins	65991 (2.0)	44717 (2.0)
moderate/severe kidney failure	23607 (0.7)	15072 (0.7)
severe kidney failure	2728 (0.1)	1839 (0.1)
oesophageal varices	809 (0.0)	674 (0.0)
inflammatory bowel disease	19170 (0.6)	13095 (0.6)
SLE	1984 (0.1)	1273 (0.1)
peripheral vascular disease	30130 (0.9)	23066 (1.0)
dementia	23320 (0.7)	15858 (0.7)
Parkinson's disease	9222 (0.3)	5854 (0.3)
cancer	70774 (2.2)	45637 (2.0)
liver disease	6852 (0.2)	5041 (0.2)
malabsorption	16718 (0.5)	12007 (0.5)
endocrine diseases	17179 (0.5)	12479 (0.5)
COPD	44766 (1.4)	33190 (1.5)
chronic liver disease or pancreatitis	9572 (0.3)	6899 (0.3)
renal stones	19858 (0.6)	14935 (0.7)
care home resident	4873 (0.1)	2859 (0.1)
falls	90783 (2.8)	53221 (2.3)
prior fracture	70017 (2.1)	50346 (2.2)
asthma or COPD	312477 (9.6)	207765 (9.1)
treated hypertension	285293 (8.7)	190707 (8.4)
platelets < 150 or > 480	24333 (0.7)	12651 (0.6)
emergency admission or hip op	8749 (0.3)	6468 (0.3)
prior haemorrhage	177327 (5.4)	122024 (5.4)

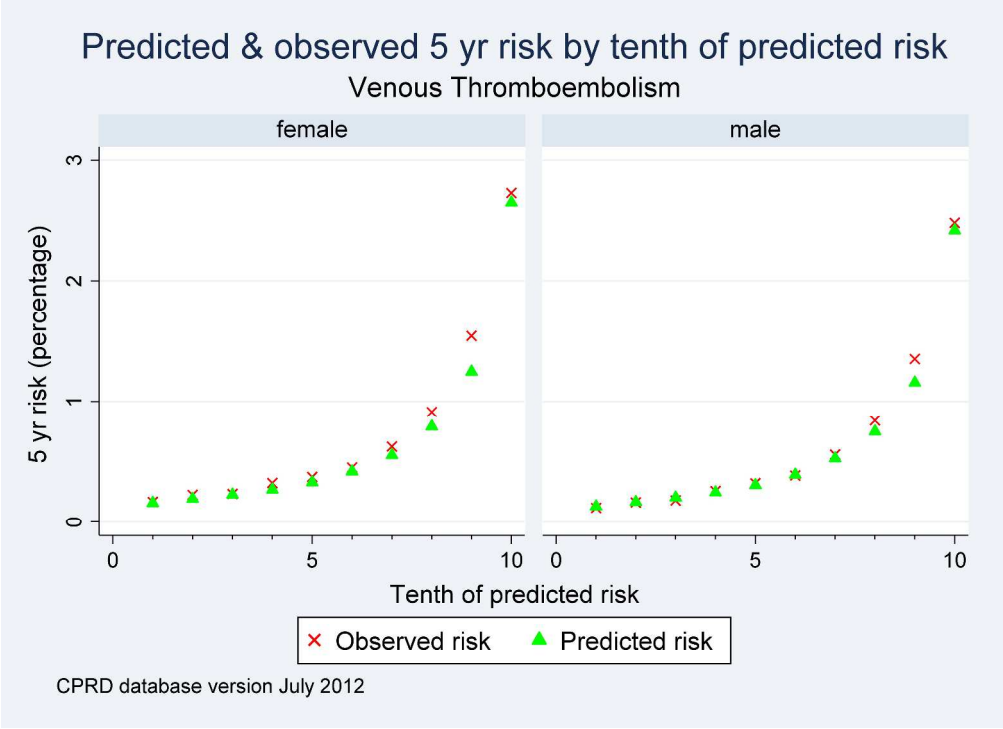
Recorded values		
BMI recorded	2750153 (84.1)	1864134 (82.0)
SBP reccorded	2949786 (90.2)	2010003 (88.4)
cholesterol/HDL ratio recorded	1193900 (36.5)	761573 (33.5)
platelets recorded	606260 (18.5)	302478 (13.3)
mean age (SD)	47.9 (17.0)	47.4 (17.2)
mean townsend score (SD)	-.5 (3.2)	.1 (3.7)
mean BMI (SD)	28.9 (6.9)	29.2 (7.1)
mean cholesterol raito (SD)	4.1 (1.4)	4.1 (1.4)
mean SBP(SD)	130.7 (23.2)	130.1 (151.3)

Web extra table 4 Number of cases of upper gastrointestinal bleed and intracranial bleed on CPRD and QResearch (one third sample database). Incidence rates per 1000 pyrs have been standardised to the QResearch population in 5 year bands.

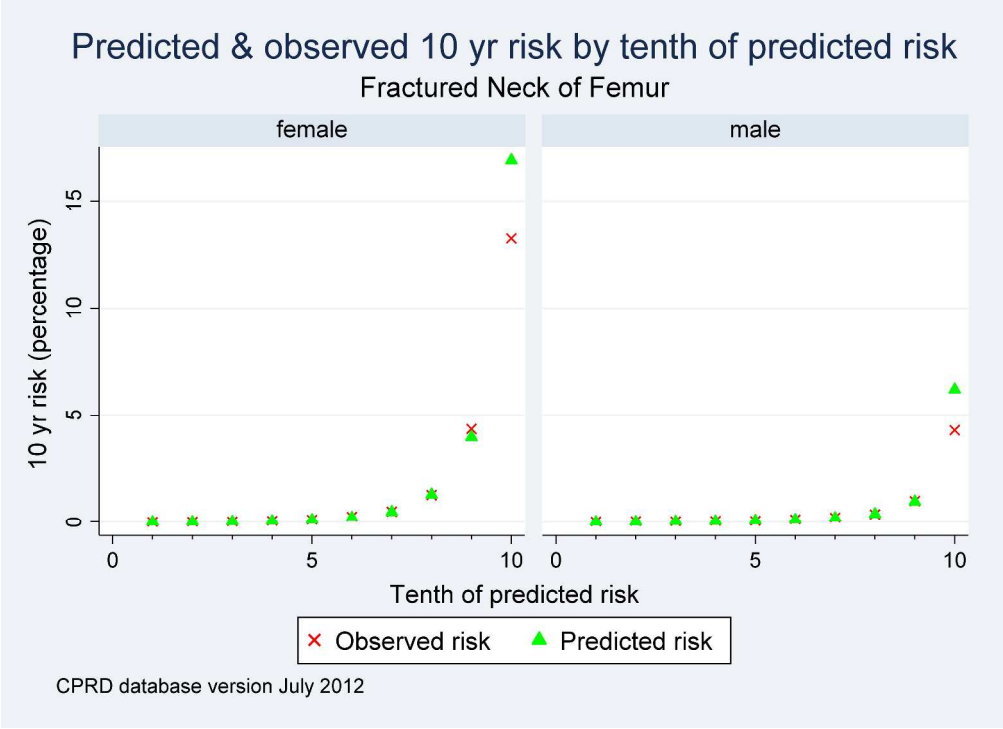
	CPRD validation		*QResearch validation*	
	cases on	**age standardised Incidence rate per 1000pyrs**	**cases**	**Age standardised Incidence rate per 1000pyrs**
no anticoagulants	13,314	1.41 (1.39 to 1.43)	6,447	1.33 (1.30 to 1.36)
anticoagulants	359	6.70 (4.06 to 9.34)	153	6.10 (3.20 to 8.98)
Intracranial bleed				
no anticoagulants	5,190	0.53 (0.51 to 0.54)	2,716	0.56 (0.54 to 0.58)
anticoagulants	233	2.45 (1.23 to 3.68)	104	2.87 (1.11 to 4.39)

Web extra table 5 Performance of QPrediction scores on the CPRD validation cohort, restricted to patients with complete data for relevant laboratory and clinical values

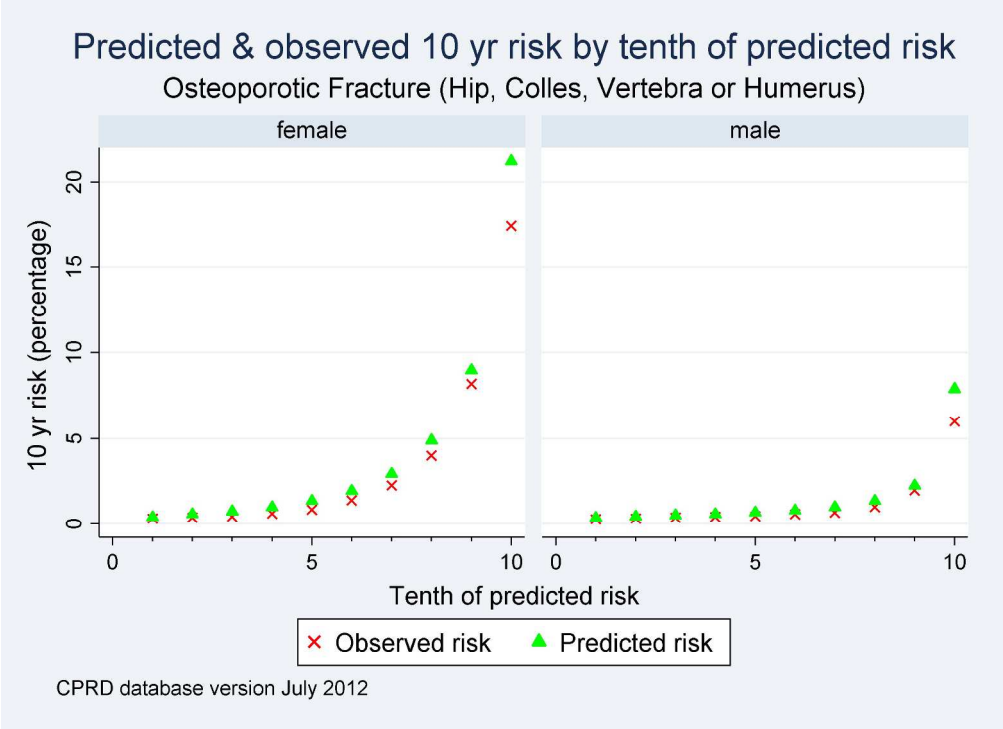
		CPRD	*CPRD*
		women	**men**
	statistic	**mean (95%CI)**	**mean (95%CI)**
QDiabetes-2013	ROC	0.849 (0.847 to 0.85	0.814 (0.813 to 0.816
(type 2 diabetes)	R^2 (%)	49.8 (49.4 to 50.2)	44.4 (44 to 44.8)
	D statistic	2.04 (2.024 to 2.056)	1.828 (1.814 to 1.842)
QKidney-2010	ROC	0.847 (0.842 to 0.852	0.839 (0.835 to 0.842
(moderate or severe kidney failure)	R^2 (%)	53.4 (52.8 to 54)	49.7 (49.3 to 50.1)
	D statistic	2.19 (2.165 to 2.215)	2.036 (2.018 to 2.054)
QKidney -2010	ROC	0.816 (0.798 to 0.834	0.808 (0.795 to 0.822
(severe kidney failure)	R^2 (%)	47.3 (45.2 to 49.4)	46.1 (44.5 to 47.7)
	D statistic	1.938 (1.856 to 2.02)	1.895 (1.834 to 1.956)
QRISK2-2014	ROC	0.791 (0.787 to 0.796	0.757 (0.753 to 0.761
(cardiovascular disease)	R^2 (%)	40.9 (40 to 41.8)	31.8 (30.9 to 32.7)
	D statistic	1.704 (1.673 to 1.735)	1.398 (1.371 to 1.425)
QStroke-2013	ROC	0.794 (0.79 to 0.797	0.771 (0.768 to 0.774
(ischaemic stroke or TIA)	R^2 (%)	42.1 (41.4 to 42.8)	36.6 (35.9 to 37.3)
	D statistic	1.747 (1.723 to 1.771)	1.557 (1.535 to 1.579)
QThrombosis-2010	ROC	0.755 (0.75 to 0.76	0.762 (0.756 to 0.767
(venous thromboembolism)	R^2 (%)	34.6 (33.8 to 35.4)	32.6 (31.7 to 33.5)
	D statistic	1.487 (1.462 to 1.512)	1.424 (1.395 to 1.453)
QBleed-20141	ROC statistic	0.773 (0.766 to 0.779	0.751 (0.744 to 0.758
(upper GI bleed)	R^2 (%)	43.6 (42.1 to 45.1)	39.6 (38.1 to 41.1)
	D statistic	1.799 (1.744 to 1.854)	1.658 (1.605 to 1.711)
QBleed-2014	ROC statistic	0.812 (0.803 to 0.822	0.791 (0.78 to 0.802
(intracranial bleed)	R^2 (%)	53 (51 to 55)	50.2 (48 to 52.4)
	D statistic	2.172 (2.086 to 2.258)	2.057 (1.967 to 2.147)
QFracture-2012	ROC	0.899 (0.896 to 0.901	0.866 (0.86 to 0.872
(fracture neck of femur)	R^2 (%)	70.4 (69.9 to 70.9)	67.1 (66.2 to 68)
	D statistic	3.159 (3.124 to 3.194)	2.922 (2.861 to 2.983)
QFracture -2012	ROC	0.819 (0.816 to 0.821	0.757 (0.751 to 0.763
(osteoporotic fracture)	R^2 (%)	53.9 (53.4 to 54.4)	46.1 (45 to 47.2)
	D statistic	2.213 (2.189 to 2.237)	1.893 (1.852 to 1.934)



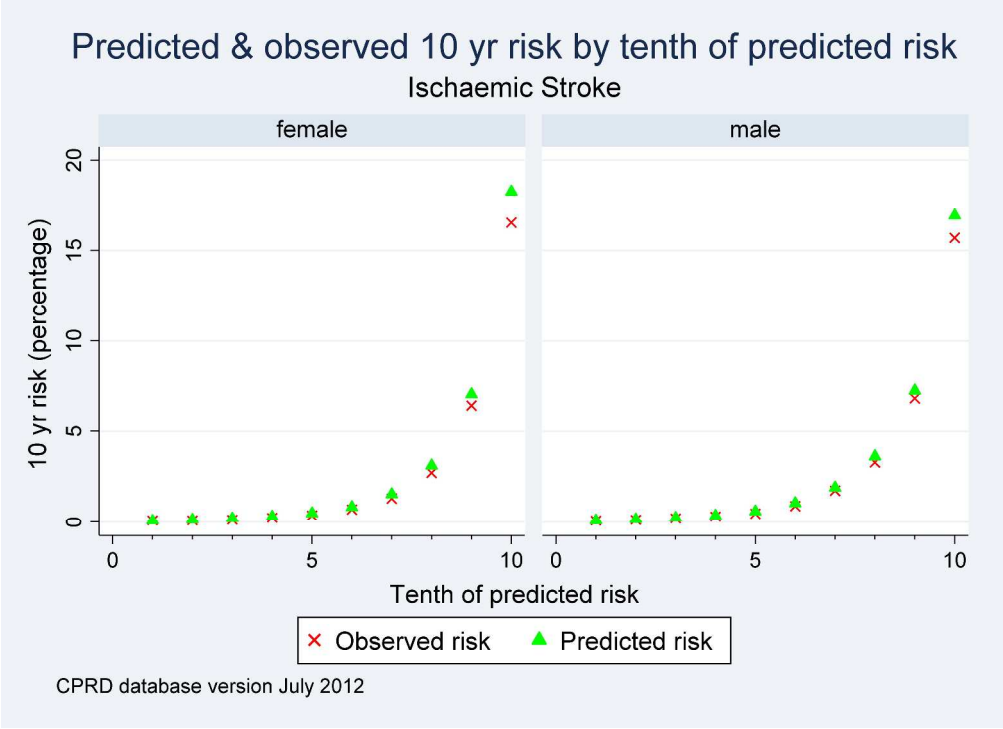
QThrombosis (venous thromboembolism)



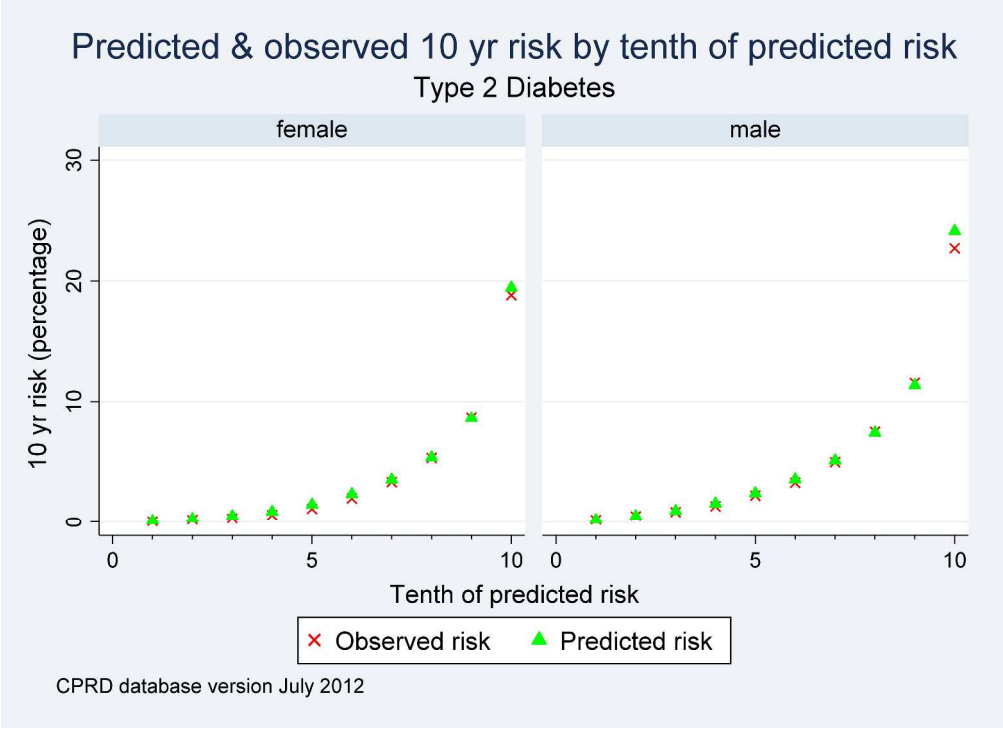
QFracture (hip)



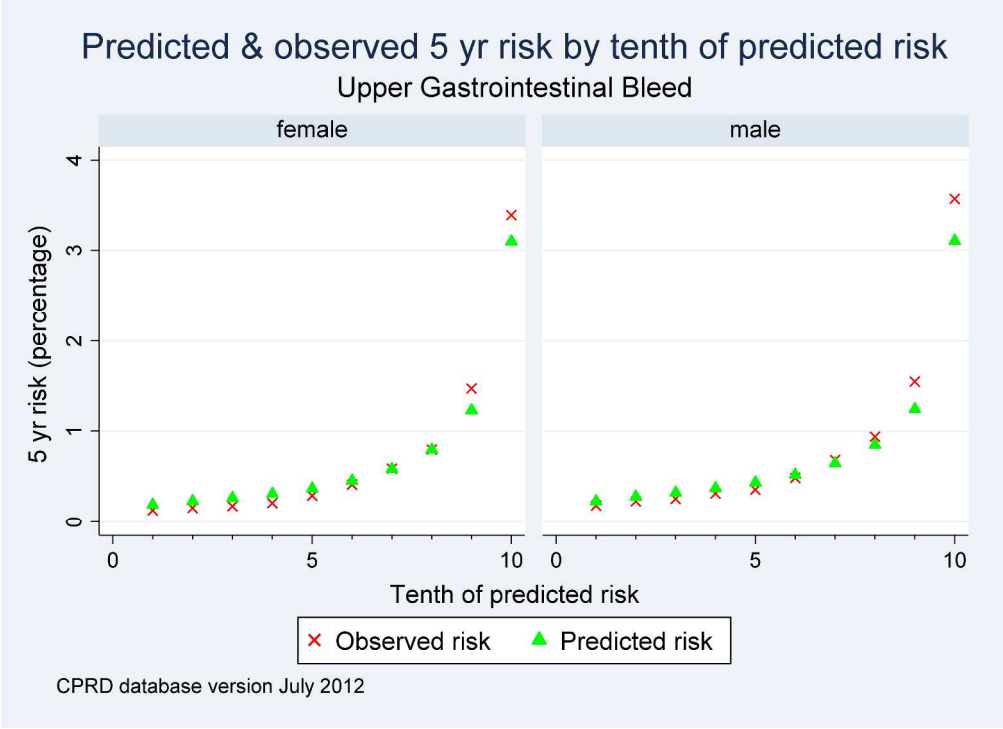
QFracture (hip, colles, spine, shoulder)



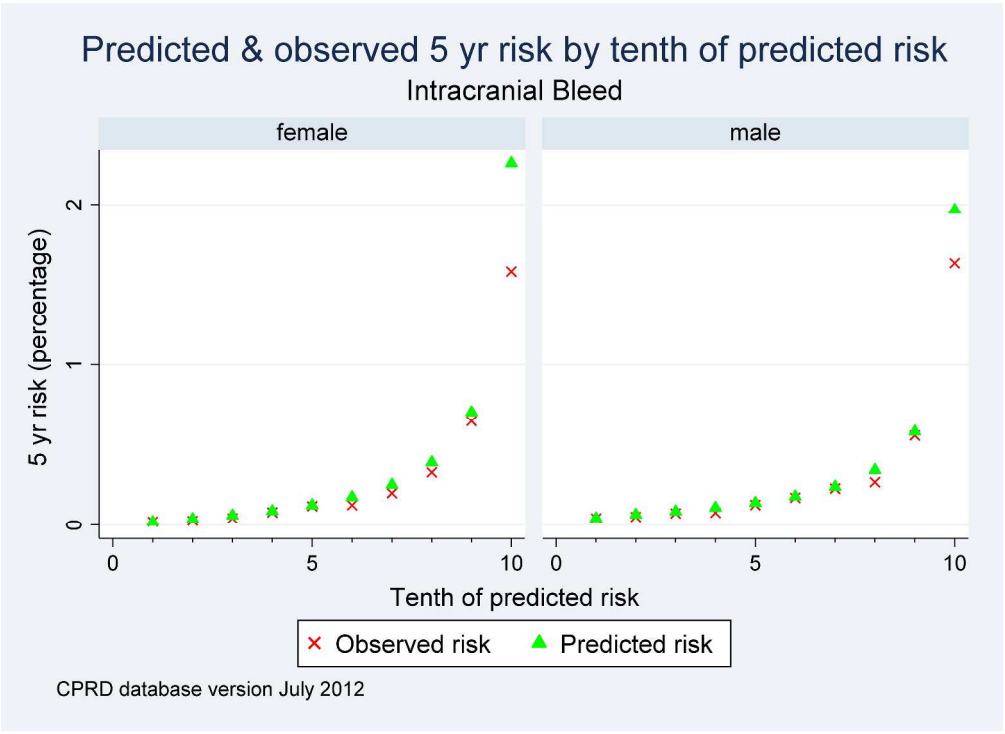
QStroke (ischaemic stroke)



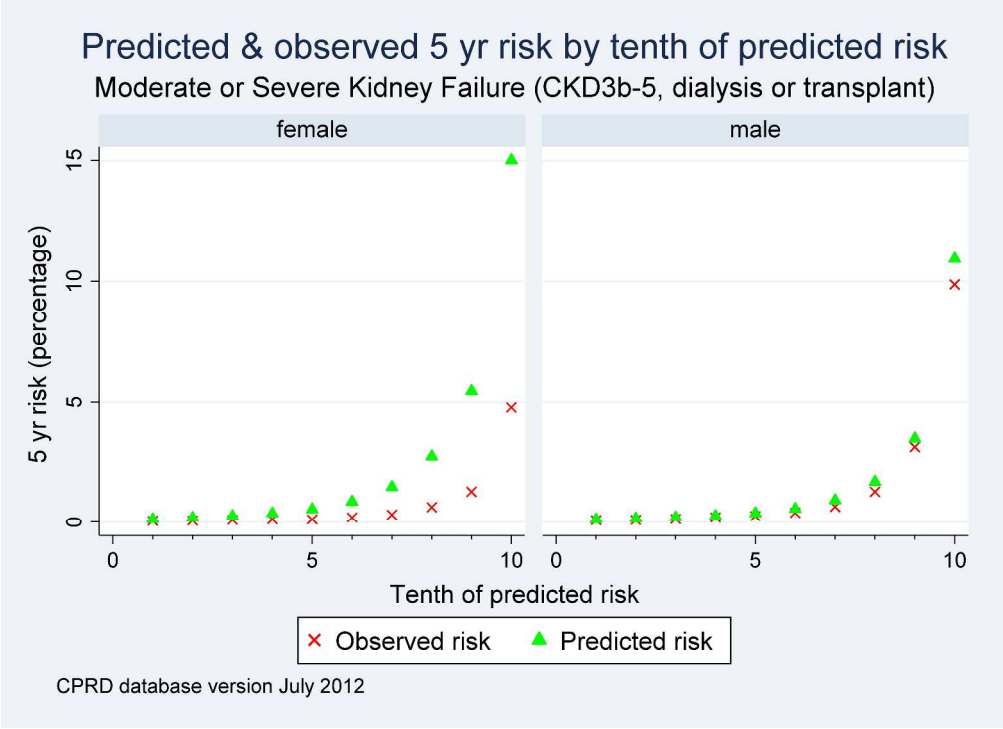
QDiabetes (type 2 diabetes)



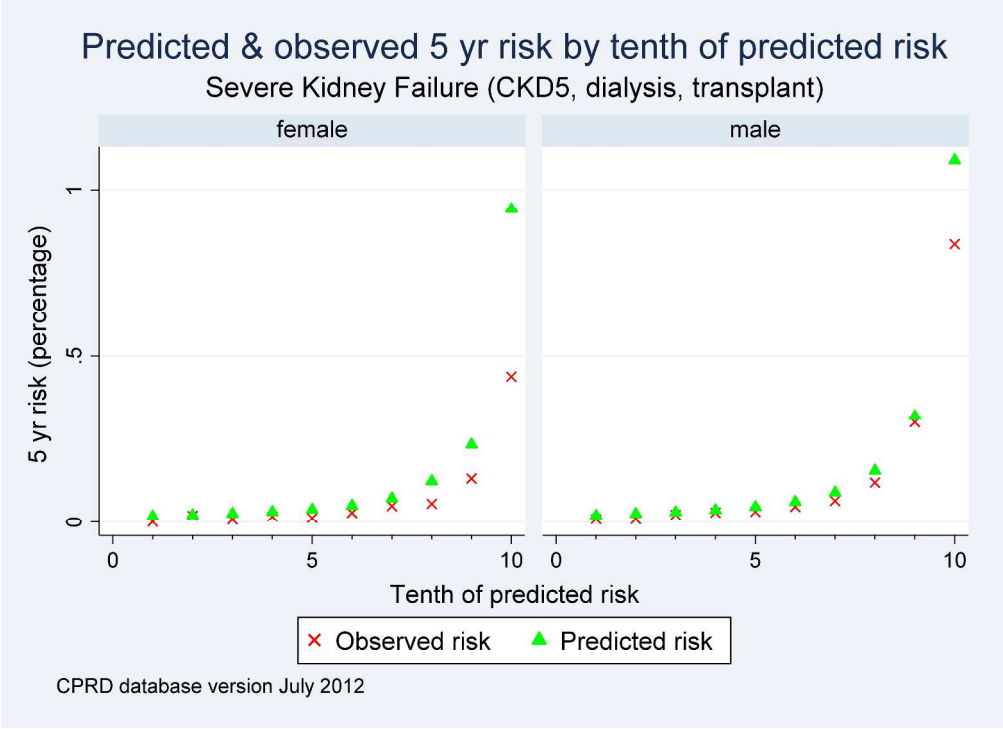
QBleed (upper gastrointestinal haemorrhage)



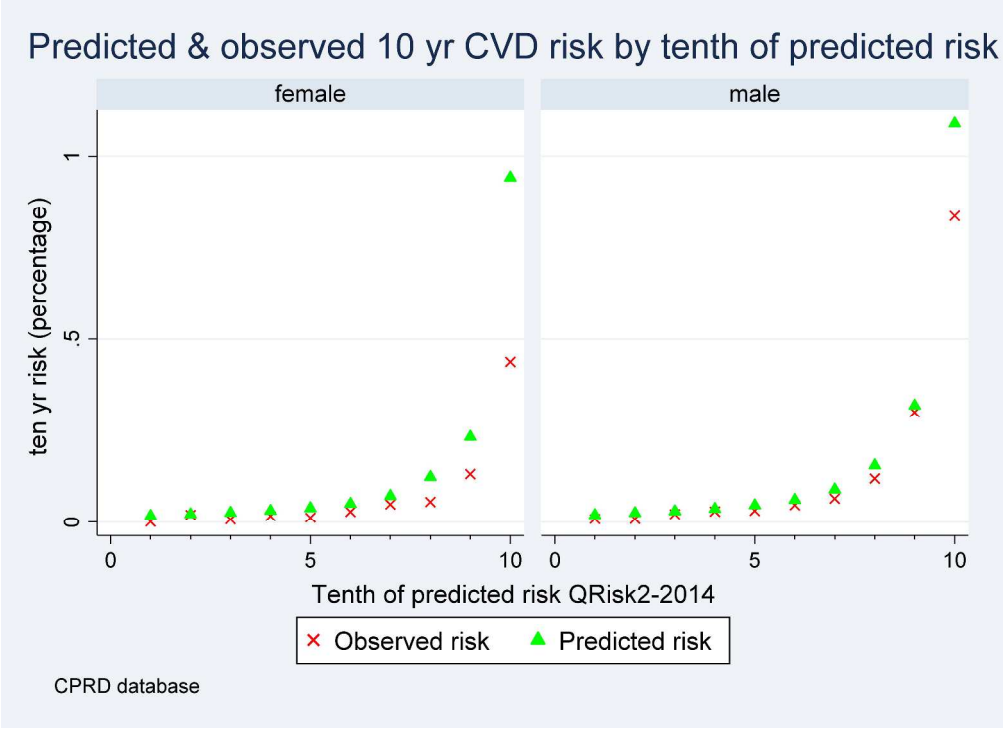
QBleed (intracranial haemorrhage)



QKidney (moderate or severe kidney failure)



QKidney (severe kidney failure)



QRisk2 (cardiovascular disease)

Web extra table 1. Numbers of patients in CPRD and QResearch by geographical area

	CPRD	*Col %*	*QResearch*	*Col %*
East Midlands	109,428	3.3	500,970	7.4
East of England	397,008	12.1	528,379	7.8
London	563,353	17.2	1,711,956	25.3
North East	59,558	1.8	269,695	4.0
North West	474,457	14.5	830,047	12.3
South Central	411,571	12.6	696,070	10.3
South East	362,319	11.1	545,811	8.1
South West	397,735	12.2	700,041	10.4
West Midlands	348,614	10.7	589,548	8.7
Yorks & Humber	147,469	4.5	386,132	5.7
Total	**3,271,512**	**100.00**	**6,758,649**	**100.00**

.

Web table 2 Recording of ethnicity and family history of coronary heart disease (FH CHD) by geographical area

	CPRD linked data					QResearch					Ratio recording QResearch:CPRD	
	total patients	ethnicity recorded		FH CHD recorded		total patients	ethnicity recorded		FH CHD recorded			
	count	count	Row %	count	Row %	count	count	Row %	count	Row %	ethnicity	FH
East Midlands	109,428	23,636	21.6	7,912	7.2	500,970	284,194	56.7	60,278	12.0	2.6	1.7
East of England	397,008	127,427	32.1	17,597	4.4	528,379	319,742	60.5	54,252	10.3	1.9	2.3
London	563,353	308,285	54.7	21,034	3.7	1,711,956	1,095,835	64.0	162,282	9.5	1.2	2.5
North East	59,558	17,140	28.8	3,326	5.6	269,695	195,127	72.4	43,168	16.0	2.5	2.9
North West	474,457	196,987	41.5	25,915	5.5	830,047	460,640	55.5	112,718	13.6	1.3	2.5
South Central	411,571	140,448	34.1	16,989	4.1	696,070	380,908	54.7	73,051	10.5	1.6	2.5
South East	362,319	90,633	25.0	12,618	3.5	545,811	262,861	48.2	45,765	8.4	1.9	2.4
South West	397,735	137,806	34.6	17,829	4.5	700,041	375,155	53.6	75,091	10.7	1.5	2.4
West Midlands	348,614	148,012	42.5	20,375	5.8	589,548	347,479	58.9	63,412	10.8	1.4	1.8
Yorks & Humber	147,469	56,340	38.2	6,195	4.2	386,132	214,702	55.6	54,515	14.1	1.5	3.4
Total	**3,271,512**	**1,246,714**	**38.1**	**149,790**	**4.6**	**6,758,649**	**3,936,643**	**58.2**	**744,532**	**11.0**	**1.5**	**2.4**

Web extra table 3 comparison of baseline characteristics for the CPRD cohort with linked data used for the validation cohort in this study and the CPRD cohort without linked data (which was not used for the validation of the QPrediction scores).

	CPRD linked data	CPRD unlinked data
female	1682709 (51.4)	1166103 (51.3)
male	1588803 (48.6)	1108235 (48.7)
25-34 years	895167 (27.4)	665645 (29.3)
35-44 years	760830 (23.3)	503091 (22.1)
45-54 years	571937 (17.5)	384686 (16.9)
55-64 years	424453 (13.0)	293826 (12.9)
65-74 years	312352 (9.5)	216937 (9.5)
75+ years	306773 (9.4)	210153 (9.2)
Ethnicity recorded	1246714 (38.1)	645829 (28.4)
White or not recorded	3117325 (95.3)	2209396 (97.1)
Indian	32467 (1.0)	11751 (0.5)
Pakistani	12752 (0.4)	6358 (0.3)
Bangladeshi	4107 (0.1)	2682 (0.1)
Other Asian	22668 (0.7)	8854 (0.4)
Caribbean	11414 (0.3)	4812 (0.2)
Black African	27654 (0.8)	9751 (0.4)
Chinese	7090 (0.2)	3416 (0.2)
Other ethnic group	36035 (1.1)	17318 (0.8)
Smoking status recorded	3037626 (92.9)	2066777 (90.9)
Non smoker	1448554 (44.3)	1006511 (44.3)
Ex-smoker	475488 (14.5)	306460 (13.5)
Light smoker (1-9/day)	214330 (6.6)	127487 (5.6)
Moderate smoker (10-19/day)	362391 (11.1)	277693 (12.2)
Heavy smoker (20+/day)	229912 (7.0)	159718 (7.0)
Smoker amount not recorded	306951 (9.4)	188908 (8.3)
Alcohol status recorded	2617112 (80.0)	1759758 (77.4)
Non drinker	482513 (14.7)	361566 (15.9)
Trivial <1u/day	1186942 (36.3)	754470 (33.2)
Light 1-2u/day	701808 (21.5)	479407 (21.1)
Moderate 3-6/day	203091 (6.2)	135130 (5.9)
Heavy 7-9u/day	22465 (0.7)	15645 (0.7)
Very Heavy >/day	20293 (0.6)	13540 (0.6)
family history		
family history of CHD	149790 (4.6)	97401 (4.3)
family history of diabetes	229200 (7.0)	144713 (6.4)
family history osteoporosis	10942 (0.3)	6164 (0.3)
family history kidney disease	2839 (0.1)	1592 (0.1)

prescribed medication		
antidepressants	337350 (10.3)	232657 (10.2)
anticoagulants	29032 (0.9)	20338 (0.9)
antipsychotics	113398 (3.5)	76819 (3.4)
oral NSAIDs	592931 (18.1)	425739 (18.7)
tamoxifen	9343 (0.3)	6107 (0.3)
antiplatelets	190291 (5.8)	137793 (6.1)
oestrogen only HRT	119413 (3.7)	75448 (3.3)
corticosteroids	116949 (3.6)	70793 (3.1)
oral contraceptive pill	174288 (5.3)	126218 (5.5)
recorded diagnoses		
congestive cardiac failure	35543 (1.1)	25823 (1.1)
atrial fibrillation	40227 (1.2)	27032 (1.2)
coronary heart disease	138080 (4.2)	102493 (4.5)
cardiovascular disease	184597 (5.6)	134650 (5.9)
rheumatoid arthritis	24711 (0.8)	17427 (0.8)
chronic renal disease	8050 (0.2)	5774 (0.3)
type 1 diabetes	11162 (0.3)	7778 (0.3)
type 2 diabetes	94905 (2.9)	63240 (2.8)
venous thromboembolism	38162 (1.2)	23593 (1.0)
varicose veins	65991 (2.0)	44717 (2.0)
moderate/severe kidney failure	23607 (0.7)	15072 (0.7)
severe kidney failure	2728 (0.1)	1839 (0.1)
oesophageal varices	809 (0.0)	674 (0.0)
inflammatory bowel disease	19170 (0.6)	13095 (0.6)
SLE	1984 (0.1)	1273 (0.1)
peripheral vascular disease	30130 (0.9)	23066 (1.0)
dementia	23320 (0.7)	15858 (0.7)
Parkinson's disease	9222 (0.3)	5854 (0.3)
cancer	70774 (2.2)	45637 (2.0)
liver disease	6852 (0.2)	5041 (0.2)
malabsorption	16718 (0.5)	12007 (0.5)
endocrine diseases	17179 (0.5)	12479 (0.5)
COPD	44766 (1.4)	33190 (1.5)
chronic liver disease or pancreatitis	9572 (0.3)	6899 (0.3)
renal stones	19858 (0.6)	14935 (0.7)
care home resident	4873 (0.1)	2859 (0.1)
falls	90783 (2.8)	53221 (2.3)
prior fracture	70017 (2.1)	50346 (2.2)
asthma or COPD	312477 (9.6)	207765 (9.1)
treated hypertension	285293 (8.7)	190707 (8.4)
platelets < 150 or > 480	24333 (0.7)	12651 (0.6)
emergency admission or hip op	8749 (0.3)	6468 (0.3)
prior haemorrhage	177327 (5.4)	122024 (5.4)

Recorded values		
BMI recorded	2750153 (84.1)	1864134 (82.0)
SBP reccorded	2949786 (90.2)	2010003 (88.4)
cholesterol/HDL ratio recorded	1193900 (36.5)	761573 (33.5)
platelets recorded	606260 (18.5)	302478 (13.3)
mean age (SD)	47.9 (17.0)	47.4 (17.2)
mean townsend score (SD)	-.5 (3.2)	.1 (3.7)
mean BMI (SD)	28.9 (6.9)	29.2 (7.1)
mean cholesterol raito (SD)	4.1 (1.4)	4.1 (1.4)
mean SBP(SD)	130.7 (23.2)	130.1 (151.3)

Web extra table 4 Number of cases of upper gastrointestinal bleed and intracranial bleed on CPRD and QResearch (one third sample database). Incidence rates per 1000 pyrs have been standardised to the QResearch population in 5 year bands.

	CPRD validation		*QResearch validation*	
	cases on	**age standardised Incidence rate per 1000pyrs**	**cases**	**Age standardised Incidence rate per 1000pyrs**
no anticoagulants	13,314	1.41 (1.39 to 1.43)	6,447	1.33 (1.30 to 1.36)
anticoagulants	359	6.70 (4.06 to 9.34)	153	6.10 (3.20 to 8.98)
Intracranial bleed				
no anticoagulants	5,190	0.53 (0.51 to 0.54)	2,716	0.56 (0.54 to 0.58)
anticoagulants	233	2.45 (1.23 to 3.68)	104	2.87 (1.11 to 4.39)

Web extra table 5 Performance of QPrediction scores on the CPRD validation cohort, restricted to patients with complete data for relevant laboratory and clinical values

		CPRD	*CPRD*
		women	**men**
	statistic	**mean (95%CI)**	**mean (95%CI)**
QDiabetes-2013	ROC	0.849 (0.847 to 0.85	0.814 (0.813 to 0.816
(type 2 diabetes)	R^2 (%)	49.8 (49.4 to 50.2)	44.4 (44 to 44.8)
	D statistic	2.04 (2.024 to 2.056)	1.828 (1.814 to 1.842)
QKidney-2010	ROC	0.847 (0.842 to 0.852	0.839 (0.835 to 0.842
(moderate or severe kidney failure)	R^2 (%)	53.4 (52.8 to 54)	49.7 (49.3 to 50.1)
	D statistic	2.19 (2.165 to 2.215)	2.036 (2.018 to 2.054)
QKidney -2010	ROC	0.816 (0.798 to 0.834	0.808 (0.795 to 0.822
(severe kidney failure)	R^2 (%)	47.3 (45.2 to 49.4)	46.1 (44.5 to 47.7)
	D statistic	1.938 (1.856 to 2.02)	1.895 (1.834 to 1.956)
QRISK2-2014	ROC	0.791 (0.787 to 0.796	0.757 (0.753 to 0.761
(cardiovascular disease)	R^2 (%)	40.9 (40 to 41.8)	31.8 (30.9 to 32.7)
	D statistic	1.704 (1.673 to 1.735)	1.398 (1.371 to 1.425)
QStroke-2013	ROC	0.794 (0.79 to 0.797	0.771 (0.768 to 0.774
(ischaemic stroke or TIA)	R^2 (%)	42.1 (41.4 to 42.8)	36.6 (35.9 to 37.3)
	D statistic	1.747 (1.723 to 1.771)	1.557 (1.535 to 1.579)
QThrombosis-2010	ROC	0.755 (0.75 to 0.76	0.762 (0.756 to 0.767
(venous thromboembolism)	R^2 (%)	34.6 (33.8 to 35.4)	32.6 (31.7 to 33.5)
	D statistic	1.487 (1.462 to 1.512)	1.424 (1.395 to 1.453)
QBleed-20141	ROC statistic	0.773 (0.766 to 0.779	0.751 (0.744 to 0.758
(upper GI bleed)	R^2 (%)	43.6 (42.1 to 45.1)	39.6 (38.1 to 41.1)
	D statistic	1.799 (1.744 to 1.854)	1.658 (1.605 to 1.711)
QBleed-2014	ROC statistic	0.812 (0.803 to 0.822	0.791 (0.78 to 0.802
(intracranial bleed)	R^2 (%)	53 (51 to 55)	50.2 (48 to 52.4)
	D statistic	2.172 (2.086 to 2.258)	2.057 (1.967 to 2.147)
QFracture-2012	ROC	0.899 (0.896 to 0.901	0.866 (0.86 to 0.872
(fracture neck of femur)	R^2 (%)	70.4 (69.9 to 70.9)	67.1 (66.2 to 68)
	D statistic	3.159 (3.124 to 3.194)	2.922 (2.861 to 2.983)
QFracture -2012	ROC	0.819 (0.816 to 0.821	0.757 (0.751 to 0.763
(osteoporotic fracture)	R^2 (%)	53.9 (53.4 to 54.4)	46.1 (45 to 47.2)
	D statistic	2.213 (2.189 to 2.237)	1.893 (1.852 to 1.934)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page in manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2 abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 2, page 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 2, 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5 & 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*Pages 6*
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Page 8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Pages 6, 7, 8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Pages 7, 8
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6
		(*c*) Explain how missing data were addressed	Page 7
		(*d*) If applicable, explain how loss to follow-up was addressed	
		(*e*) Describe any sensitivity analyses	Page 7,8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 9
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Page 9 tables 2, 3
		(b) Indicate number of participants with missing data for each variable of interest	table 2,3
		(c) Summarise follow-up time (eg, average and total amount)	Table5
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9 table 5
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were	Pages 10, 11 Table 3, table 5

		included	
		(*b*) Report category boundaries when continuous variables were categorized	Page 6
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Page 11
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Page 11
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 12- 13,
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 13
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 15

BMJ Open

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2014-005809.R2
Article Type:	Research
Date Submitted by the Author:	28-Jul-2014
Complete List of Authors:	Hippisley-Cox, Julia; University of Nottingham, ; ClinRisk Ltd, Coupland, Carol; University of Nottingham, Division of Primary Care Brindle, Peter; Bristol Clinical Commissioning Group,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology, General practice / Family practice, Health informatics
Keywords:	qresearch, cprd, Epidemiology < TROPICAL MEDICINE, qrisk2, prognosis, validation

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Revised Research paper submitted to BMJ Open
Reference-2014-005809
July 2014

Authors

Julia Hippisley-Cox	Professor of Clinical Epidemiology & General Practice[1]
Carol Coupland	Associate Professor and Reader in Medical Statistics[1]
Peter Brindle	Research and Development Programme Director[2]

Institutions

[1]Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham, NG2 7RD.

[2]Avon Primary Care Research Collaborative, Bristol Clinical Commissioning Group, South Plaza, Marlborough Street, Bristol BS1 3NX.

Guarantor & Author for correspondence - Julia Hippisley-Cox

Email: Julia.hippisley-cox@nottingham.ac.uk

Telephone: 0115 8466915

Fax: 0115 8466904

Abstract

Objectives To validate the performance of a set of risk prediction algorithms developed using the QResearch database, in an independent sample from general practices contributing to the Clinical Research Data Link (CPRD).

Setting Prospective open cohort study using practices contributing to the CPRD database and practices contributing to the QResearch database.

Participants The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices between 01 Jan 1998 and 31 July 2012. The validation statistics for QResearch were obtained from the original published papers which used a one third sample of practices separate to those used to derive the score. A cohort from QResearch was used to compare incidence rates and baseline characteristics and consisted of 6.8 million patients from 753 practices registered between 01 Jan 1998 and until 31 July 2013.

Outcome measures Incident events relating to seven different risk prediction scores: QRISK2 (cardiovascular disease); QStroke (ischaemic stroke); QDiabetes (type 2 diabetes); QFracture (osteoporotic fracture and hip fracture); QKidney (moderate and severe kidney failure); QThrombosis (venous thromboembolism); QBleed (intracranial bleed and upper gastrointestinal haemorrhage). Measures of discrimination and calibration were calculated.

Results Overall, the baseline characteristics of the CPRD and QResearch cohorts were similar though QResearch had higher recording levels for ethnicity and family history. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with the published results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch and the ROC value was 0.85 on both databases. The scores were well calibrated in CPRD.

Conclusion Each of the algorithms performed practically as well in the external independent CPRD validation cohorts as they had in the original published QResearch validation cohorts.

Strengths and limitations of this study

- This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system from that used to derive the algorithms.
- The discrimination and calibration statistics for each score were very similar in CPRD to those published from validation cohorts from QResearch. This supports their potential utility in the general population of patients in primary care.
- A strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice.
- The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible.
- Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using these algorithms in primary care.

1 Introduction

In the last 7 years, we have developed a series of risk prediction algorithms using the QResearch database. QResearch is a large research database containing pseudonymised individual level data from over 700 general practices using the EMIS clinical system. The QResearch database consists of data collected from primary care (coded information on socio-demographic characteristics, diagnoses, symptoms, smoking/alcohol, clinical measurements, laboratory values, prescriptions and referrals) which has been linked to cause of death, hospital episodes and cancer registrations at individual patient level.

The algorithms predict outcomes such as cardiovascular disease(www.qrisk.org)[1], stroke (www.qstroke.org)[2], type 2 diabetes (www.qdiabetes.org)[3], osteoporotic fracture (www.qfracture.org)[4], moderate or severe kidney disease (www.qkidney.org)[5], venous thrombo-embolism (www.qthrombosis.org)[6], and emergency hospital admission (www.qadmissions.org)[7]. Generally, the "QPrediction" algorithms have been designed to systematically identify patients in primary care at high risk of a serious clinical outcome for whom further intervention to lower risk of that outcome might be possible. They are also designed to quantify absolute risk of serious outcomes in a way which patients can understand and which might help guide lifestyle and management decisions. A number of these algorithms are now integrated into GP clinical computer systems, included in national guidelines[1 4] and are in daily use across the NHS [1 3 8].

The algorithms were originally developed using a random two thirds sample of practices contributing to the QResearch database and validated on the remaining third. Whilst this represents a physically discrete population of patients and practices for validation, the practices all use the same clinical computer system (EMIS), which is in use in 53% of UK practices. A more stringent test of performance is to validate the algorithms on a fully external database derived from practices using a different but commonly used primary care computer system. This would help determine whether the predictions from the algorithms are likely to generalise to the whole population in England. Whilst some of the algorithms have been validated by an independent team using the THIN primary care database[9-12], there are currently no published validations of the algorithms using a primary care database which is routinely linked to mortality data in the same way as QResearch.

We therefore decided to validate the various QPrediction Scores using another database known as the Clinical Research Data Link (CPRD). The General Practice Research database (GPRD) was originally set up in 1988 and is of similar nature to QResearch although it is derived from practices using a different clinical computer system (Vision, which is used by 20% of GPs). It was extended to include linked mortality data and data from secondary care and was renamed the Clinical Research Data Link (CPRD) in 2012. Our secondary objective was to compare the ascertainment of incident clinical events recorded in GP data alone with that

recorded in either GP data or the linked mortality data in both the CPRD and QResearch.

2 Methods

2.1 CPRD Study population

For the validation using CPRD, we identified an open cohort of patients aged 25-99 years at entry to the cohort and followed this cohort up until 31st July 2012 (the latest date for which linked data were available at the time of analysis). We restricted the CPRD cohort to 357 practices in England which had linked ONS mortality and hospital admissions data. For each patient we determined an entry date to the cohort, which was the latest of the following dates: 25th birthday, date of registration with the practice plus one year, date on which the practice computer system was installed plus one year, and the beginning of the study period (01 January 1998). Patients were censored at the earliest date of the relevant outcome, de-registration with the practice, last upload of computerised data or the study end date (31 July 2012).

For the assessment of the two QBleed outcomes (intracranial bleed and upper gastrointestinal haemorrhage) we used a later cohort entry date of 01.01.2007 for comparability with the equivalent study period for the derivation of the algorithm on QResearch[13].

2.2 QResearch study population

For comparison of the validation statistics (ROC, D and R2 statistics), we extracted the original published values from the papers which had been calculated using a one third sample of practices from QResearch which were independent from the two thirds of practices used to derive the scores.

For comparison of the baseline characteristics, incidence rates and ascertainment rates we used the latest version of the QResearch database which is currently available (QResearch 38, 31st Dec 2013). We identified an open cohort in the same way as for CPRD, using all of the QResearch practices in England, and with follow-up until 31 July 2013.

2.3 Inclusion and exclusion criteria

For both databases, we excluded patients without a Townsend score (an area based measure of material deprivation derived from the post code) and temporary residents. For each score we then identified patients who were eligible to have the score calculated according to the relevant inclusion and exclusion criteria as summarised in Table 4

2.4 Risk scores included in validation

We validated the following risk prediction scores on CPRD:

1. QDiabetes - 10 year risk of type 2 diabetes[3]
2. QRISK2 – 10 year risk of cardiovascular disease[1]
3. QStroke – 10 year risk of stroke or transient ischaemic attack (TIA)[5]
4. QFracture - 10 year risk of hip or osteoporotic fracture[4]
5. QThrombosis – 5 year risk of venous thrombo-embolism (VTE)[6]
6. QBleed – 5 year risk of upper gastrointestinal haemorrhage and intracranial haemorrhage[13]
7. QKidney – 5 year risk of moderate-severe kidney disease[5]

2.5 Clinical outcomes

We identified the relevant clinical outcome using the same definition as had been applied in the original derivation of the risk scores using QResearch. The data sources used to identify the clinical outcomes had varied over the six years during which the original studies had been undertaken due to the changing availability of linked hospital and mortality data over that time. In 2008, the QResearch database was linked to mortality records for 1997 onwards. In 2013, the QResearch database was linked to hospital admissions records with data for patients from 1998 onwards. For the latest updated version of QRISK2 (QRISK2-2014), the outcome was identified by the presence of the relevant Read code on the GP record or an ICD10 code recorded on the linked mortality record or on the linked hospital admissions record. For QStroke, QDiabetes, QFracture and QThrombosis, the outcome was identified either by the presence of the relevant Read code recorded on the GP record or an ICD10 code recorded on the linked mortality record. For QKidney, the outcome was identified solely from information recorded in the GP record as in the original study as it required blood test values which were only present in the GP record. For QBleed, the outcome was identified in CPRD from events recorded either on the linked hospital admissions database or the linked mortality record in order to identify the events most likely to have serious clinical consequences for the patient.

We determined case ascertainment for each clinical outcome on both databases, by calculating the proportion of cases recorded on the GP record out of the total number of cases recorded on either the GP record or linked mortality record. We calculated the age standardised incidence rates of each outcome based on outcomes recorded on (1) the GP record alone and on (2) the GP record or linked mortality; (3) GP or linked mortality or hospital records. We standardised CPRD rates to the age distribution of the QResearch population in five year bands to ensure comparability.

2.6 Risk factors and missing values

We extracted data from CPRD for all the predictor variables included in one or more of the different algorithms using the same definitions as those used in the original

QResearch studies to enable a direct comparison of the results. We developed a mapping between the Read and medication reference tables to identify the equivalent code in each database. This included the following variables recorded at entry to the cohort:

- **demographics** – age (continuous), sex, ethnicity (9 categories – white, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean ,Black African, Chinese, Other ethnic group), resident in care home, material deprivation (as measured by the Townsend score)
- **clinical values** - smoking status (non-smoker, ex-smoker, light smoker [1-9 cigarettes/day], moderate smoker [10-19 cigarettes/day], heavy smoker [20+ cigarettes/day];body mass index, systolic blood pressure, alcohol consumption-none drinker, trivial (<1u/day), light (1-2u/day), moderate (3-6u/day), heavy (7-9u/day), very heavy (>9 day).
- **laboratory results** –cholesterol/HDL ratio, platelets
- **family history**- family history of osteoporosis or hip fracture in a first degree relative, coronary heart disease in first degree relative under the age of 60 years, diabetes in a first degree relative.
- **chronic diseases** – congestive cardiac failure, atrial fibrillation, coronary heart disease, cardiovascular disease, periperal vascular disease, venous thromboembolism, diabetes, rheumatoid arthritis, SLE, hypertension, renal disease, renal stones, inflammatory bowel disease, dementia, Parkinson's disease, epilepsy, cancer, chronic liver disease or pancreatitis, oesophageal varices, prior haemorrhage, malabsorption endocrine diseases, asthma or COPD, history of falls, prior osteoporotic fracture, varicose vein surgery, emergency admissions or hip surgery in last 6 months.
- **prescribed medication**- anticoagulants, antidepressants, antipsychotics, antiplatelets, oral NSAIDs, tamoxifen, oestrogen containing hormone replacement therapy (BNF chapter 6.4.1.1), systemic corticosteroids, combined oral contraceptive.

The combination of predictor variables required for each risk score varied with the score being validated as shown in Table 1. We used the clinical value recorded closest to the date on which the patient entered the study for body mass index, systolic blood pressure, smoking status, platelets, and total and HDL cholesterol. Patients were considered to be exposed to medication at entry to the cohort if they had at least 2 prescriptions for the relevant medication prescribed prior to the study entry date with the most recent one occurring within 28 days of the study entry date.

2.7 Townsend scores

We used the Townsend score evaluated at output area as a proxy for material deprivation. The CPRD dataset differs from the QResearch dataset in that each patient in the CPRD dataset is allocated to a tenth of deprivation (as measured by the Townsend score) and only the category number is provided. In contrast, each

patient in the QResearch dataset is allocated the individual Townsend score corresponding to their output area of residence (i.e. continuous data). In order to calculate risk scores in the CPRD cohort, we used the median value for each tenth as supplied by CPRD. Patients with missing Townsend scores were excluded from the cohorts.

2.8 Discrimination and calibration statistics

We used chained equations with the ICE procedure in STATA[14] to perform multiple imputation to replace missing values for body mass index, systolic blood pressure, smoking status, alcohol, and total and HDL cholesterol. We created five multiply imputed datasets and used Rubin's rules to combine effect estimates and standard errors to allow for the uncertainty due to imputing missing data[15 16].

We applied the algorithm for each score to eligible patients in the CPRD study cohort to obtain predicted risks for each of the relevant clinical outcomes. We calculated the estimated risk for eligible patients in the CPRD validation dataset over 5 years or 10 years depending on which score was used. We then tested the performance of each score in the CPRD cohort and compared it with the published results from the original QResearch validation cohorts.

In order to assess calibration (i.e. degree of similarity between predicted and observed risks), we calculated the mean predicted risk and the observed risk [17]obtained using the Kaplan-Meier estimate and compared the ratio of the mean predicted risk to the observed risk for patients in the validation cohort in each decile of predicted risk. We calculated the area under the Receiver Operator Curve (ROC) statistic to assess discrimination (i.e. ability of a risk prediction equation to distinguish between those who do and do not have an event during the follow-up period). We also calculated the D statistic[18] and an R squared statistic derived from the D statistic[19] which are measures of discrimination and explained variation appropriate for survival models. The D statistic has been developed as a new measure of discrimination specifically for censored survival data, higher values indicate improved discrimination, and an increase in the D statistic of at least 0.1 indicates an important difference in prognostic separation between different risk classification schemes. The R^2 statistic derived from the D statistic is a measure specific to censored survival data– it measures explained variation in time to the outcome event and higher values indicate more variation is explained[20]. We also repeated the assessment of discrimination by restricting the analysis for each score to patients without missing data for relevant clinical or laboratory measures used in the risk score (ie those with complete data for all predictor variables in the risk score).

We identified the proportion of patients in the CPRD validation cohort who were in the top decile of predicted risk and used this to calculate the sensitivity, specificity and observed risk at this threshold. We used the top decile for comparability across the scores and with previous studies though the choice of threshold for use in clinical practice will depend on the context and cost-effectiveness of relevant interventions. Analyses were conducted using Stata (version 13.1).

2.9 Sample size estimation

There is currently no clear guidance on sample size requirements for studies evaluating the performance (validation) of a multivariable risk score, but a commonly used rule-of-thumb is that it is desirable to seek a dataset with at least 100 patients with the outcome of interest. We used all the available data on the CPRD to maximize the power of the study.

3 Results

3.1 Study populations

The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices with linked data between 01 Jan 1998 and 31 July 2012. The QResearch cohort consisted of 6.8 million patients from 753 practices with linked data, registered between 01 Jan 1998 and until 31 July 2013. The numbers of patients in each geographical region are shown in Web Extra Table 1.

3.2 Baseline characteristics

Table 2 shows a comparison of the demographic characteristics for the CPRD and QResearch cohorts.

The QResearch population was marginally younger with 34.2% of women and 32.8% of men aged 24-34 years compared with 27.8% and 26.9% for CPRD.

3.2.1 Recording of ethnicity

QResearch had a higher proportion of patients with self-assigned ethnicity recorded compared with CPRD both overall (58.2% vs 38.1%) and in each of the 10 geographical areas within England (web extra table 2). We repeated the analysis restricting information on QResearch to that recorded prior to 31 July 2012 (for comparability with the calendar time available on CPRD). Of the 6,758,649 patients in the QResearch cohort, 3,856,244 (57.1%) had ethnicity recorded prior to this date.

3.2.2 Recording of family history

Recording of a positive family history of coronary heart disease and diabetes was more than twice as high in QResearch compared with CPRD. For example, for family history of coronary heart disease, 11.0% of patients had a value recorded for QResearch compared with 4.6% for CPRD (web extra table 2). Restricting information to that recorded prior to July 2012 for QResearch, then 6,758,649 (10.7%) had a positive family history of coronary heart disease recorded.

3.2.3 Recording of alcohol and smoking levels

Recording of alcohol levels was very similar in both QResearch and CPRD. For example, 82.1% of women had alcohol level recorded in both databases. Recording of smoking status was marginally higher in women compared with men in QResearch (93.2% vs 89.1%) and also CPRD (94.8% vs 90.8%).

3.2.4 Recording of clinical values

Recording of cholesterol/HDL ratio was marginally higher on QResearch compared with CPRD (40.1% vs 36.0%). Recording of body mass index and systolic blood pressure tended to be marginally higher on CPRD than QResearch. However the mean values for the various clinical values (BMI, systolic blood pressure, serum creatinine and cholesterol/HDL ratio) were extremely similar.

Table 3 shows prescribed medication and clinical diagnoses recorded in patients on or prior to entry to the study cohort. Overall, the prevalence of clinical diagnoses were similar on the two databases with CPRD having marginally higher prescribing rates.

The inclusion and exclusion criteria for each risk score are shown in Table 4 along with the numbers of patients eligible for each analysis on CPRD. For example, there were 3,177,192 patients aged 25-84 years. Of these, 99,189 had existing diabetes at baseline leaving 3,078,003 for the validation of QDiabetes. Table 4 also shows the numbers and percentage out of those eligible for inclusion with complete data for risk factors necessary for calculation of the score which would otherwise need to be imputed (i.e. laboratory or clinical values). The amount of missing data varies substantially between the scores with scores requiring multiple laboratory or clinical values (such as QRISK2) having the lowest levels of completeness.

3.2.5 Comparison between CPRD linked and unlinked data

Web extra Table 3 shows characteristics for CPRD cohort with linked data with CPRD cohort without linked data. The CPRD cohort with linked data tended to have higher recorded of ethnicity compared with the CPRD cohort without linked data (38.1% vs 28.4%). Recording of smoking, alcohol, body mass index, systolic blood pressure, cholesterol and platelets were all higher on the CPRD cohort with linked data than those without linked data.

3.3 Incidence rates of clinical outcomes

Table 5 shows the number of incident events for each clinical outcome in women recorded on GP data and those recorded on either GP data or cause specific mortality data for both the CPRD and QResearch cohorts. It also shows the age standardized incidence rates per 1000 person years. Table 6 shows the comparable information for men.

For example, there were 35,617 incident ischaemic stroke or TIA events for women on CPRD. Of these, 32,283 had been identified on the GP record with an additional 3,334 events identified on the linked ONS mortality record. The ascertainment of events on the GP record was therefore 32283/35617 i.e. 90.6%. For QResearch, there were 70,477 incident stroke events recorded on either the GP or linked ONS mortality record of which 63,572 had been identified on the GP record. The ascertainment was therefore 90.2%.

For thromboembolism in women, 91.1% of events recorded on either the GP or linked ONS mortality record on CPRD were identified on the GP record compared with 90.6% for QResearch. Similar results were obtained for men with levels of ascertainment between the two databases being extremely close suggesting similar recording patterns between the two groups of GP practices contributing to each database.

The age standardized incidence rates of events on CPRD tended to be marginally lower than those on QResearch as shown by the ratio of the CPRD rates to those in QResearch (Table 5). For example, the rate ratio for fractured neck of femur in women was 0.94 indicating that CPRD had a 6% lower incidence rate compared with QResearch. The effect was more marked for moderate or severe kidney failure where the incidence rates for CPRD were approximately 25% lower than those for QResearch in women and 16% lower in men.

The age standardized incidence rates of upper gastrointestinal haemorrhage and intracranial haemorrhage among patients prescribed anticoagulants and those not prescribed anticoagulants are shown in Web extra table 4. The rates are similar for CPRD and QResearch.

3.4 Validation statistics

Table 7 shows the discrimination statistics for each score in CPRD in men and women and also the published values from previous validations using QResearch. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch. The D statistic for women was 2.03 within CPRD compared with 2.08 for QResearch. The ROC value for women was 0.85 on both databases.

Of all the scores, QFracture (fractured neck of femur) had the best performance in men in CPRD with a ROC value of 0.89, R^2 value of 71% and D statistic of 3.17. The corresponding figures for QResearch in men were 0.89, 72% and 3.26.

QThrombosis had the lowest values for women in CPRD with an ROC value of 0.77, R^2 of 34.5 and D statistic of 1.49. The corresponding figures for women in QResearch were 0.75, 33.5 and 1.45.

Figure 1(a-j) compares the mean predicted risks and observed risks for each score across each tenth of predicted risk (1 representing the lowest risk and 10 the highest risk) and demonstrates that the models are generally well calibrated for patients on CPRD.

The QKidney score (moderate or severe kidney failure) showed the observed risk was lower than the predicted risk. This might indicate a degree of over prediction of the score. Alternatively, it could be related to the lower incidence rate of kidney failure observed among women on the CPRD compared with QResearch.

Web extra table 5 presents the ROC, D and R^2 statistic for each score restricted to patients from CPRD with complete recording of laboratory and risk factor data for each score. The results were very similar to the results obtained using multiply imputed dataset for the majority of scores except for QRISK2 and QStroke where values were lower. For example, the results for QFracture (hip fracture) in women on CPRD using multiply imputed data were ROC of 0.89; R^2 of 70.6%; D statistic of 3.17. The corresponding results restricted to women on CPRD with complete data were ROC of 0.90; R^2 of 70.4%; D statistic of 3.16. For QRISK2, the results for women for imputed data on CPRD were ROC of 0.88; R2 of 56.4% ; D statistic of 2.33. The corresponding results for complete data were ROC of 0.79; R2 of 40.9%; D statistic of 1.70.

3.5 Performance for the top decile of risk.

Table 8 shows the sensitivity, specificity and observed risk for patients in the top decile of each score on CPRD. The observed risk is higher than the risk threshold value since this represents the observed risk within the top decile of predicted risk. For example, the cut off for the top tenth of risk for QFracture (fractured neck of femur) was a 10 year risk of 3.7%. At this threshold the sensitivity was 66.5%, specificity 90.4% and observed risk 9.4%. The results are similar to those obtained from QResearch (not shown).

4 Discussion

4.1 Summary of key findings

This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system (Vision system supplied by In Practice Systems) from the QResearch database (which is based on practices using EMIS clinical systems). Practices using the Vision system together with practices using EMIS make up approximately 75% of all the English general practices. The discrimination and calibration statistics for each score were remarkably similar in CPRD to those published from validation cohorts from QResearch. Our paper also provides updated information on a direct comparison between two of the world's largest general practice databases which have both been linked to mortality and second care data.

Before a clinical risk score can be reliably used in clinical practice, evidence is needed that it can successfully predict the intended outcome in groups of patients other than ones used to develop the score but similar to ones in whom the score might be used. Not all risk scores perform well in external samples – this can be due to deficiencies in the design or modelling methods used to derive the algorithm, if the model is over fitted or if there is an important predictor which is absent[21]. Other reasons for poor performance include differences between the setting of patients in the new and derivation samples, differences in how information is recorded and differences in patient characteristics[21]. It is for these reasons, that we have meticulously assembled the CPRD cohort using the same inclusion/exclusion criteria, definitions of predictor and outcome variables as in the original derivation studies. Any differences observed are therefore more likely to be due to capture of information and underlying population characteristics. In this study, we have found marginal differences in incidence rates between QResearch and CPRD and higher rates of recording of family history and ethnicity in QResearch though these have not been large enough to materially affect our results.

4.2 Strengths and limitations

One strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice. CPRD only had Townsend score for patients recorded for approximately half their practices (unlike QResearch where Townsend score is included for all practices) so we had to limit the validation cohort in CPRD for this analysis to those practices with linked Townsend scores. We undertook a comparison between patients registered with CPRD practices with and without linked data. We found marginally higher recording for ethnicity, smoking, alcohol, clinical values for the CPRD cohort with linked data compared with the unlinked data but similar characteristics for demographics, comorbidities,

medication and clinical values (results not shown) so we have no reason to believe this would have biased our results.

Another strength of general practice databases is the large volume of patients who tend to be representative of the general population. A limitation of routinely collected data is that not all patients will have all clinical and laboratory data recorded leading to missing data values in some of the parameters needed to calculate the risk scores. We have reported performance in all patients using multiple imputation to replace missing values and restricted to patients without missing values and found very similar results for the majority of algorithms tested. There was some degradation of performance for algorithms, particularly for QRISK2 and QStroke, where there were large amounts of missing data. However in clinical practice, the risk scores can be calculated using information recorded during consultation reducing the amount of missing data. Alternatively, the software which implements QPrediction scores includes algorithms which estimate body mass index, systolic blood pressure and cholesterol/HDL ratio. The estimated values can be used where the relevant data is not recorded in order to generate an estimated risk score. Effectively, the software emulates the multiple imputation used in our validation which then gives the results based on multiply imputed data reasonable face validity.

The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible. The CPRD database uses the same clinical coding system as QResearch for clinical values (it uses Read version 2). However, there is a third clinical system in use in England (SystmOne) which uses a different coding system known as Clinical terms version 3(CTV3). Whilst there is a mapping between Read codes and CTV3, we have not tested the algorithms on a database using CTV3 in this study so are unable to draw conclusions regarding the generalisability of the results of the validation to practices using this system.

The quality of information on CPRD is likely to be good since previous studies have validated similar outcomes and exposures and found levels of completeness and accuracy to be good[22 23].

4.3 Comparison with other studies

The aim of this study was to validate a collection of QPrediction tools. The details of the derivation and first validation of each prediction tool have been separately published in the peer reviewed literature including information on definitions of predictor variables with supplementary information available on the relevant websites. We haven't duplicated information in the present paper but have provided the relevant links and references.

Our validation results confirm earlier studies undertaken on the THIN database (another general practice database which is derived from the Vision system but which isn't linked to mortality data). These earlier studies include external validations of QRISK2[10 11 24], QDiabetes[12], QFracture[9] and QKidney[25] by an independent team who were not involved in the development of the algorithms. These independent validations have demonstrated similar performance compared with the validations performed by study authors using the QResearch database. This study builds on previous validations by providing new information on the performance of scores not previously validated on an external database (QBleed and QThrombosis) and by utilising the linked data which was not available on the THIN database. Together with the present study (which includes a number of scores not previously tested in an external population), the results provide consistent evidence that these QPrediction scores are likely to provide appropriate estimates of disease risk in contemporary primary care populations in England and to discriminate between patients at different levels of risk with reasonable reliability.

4.4 Comparison of QResearch and CPRD baseline characteristics

Overall, our results show a striking similarity between CPRD and QResearch cohorts for nearly all baseline characteristics. There are two notable exceptions. First, recording of ethnicity was higher in QResearch than CPRD. Second, fewer patients in the CPRD cohort had a recorded family history of diabetes and coronary heart disease in a first degree relative under the age of 60 years. Recording differences in ethnicity and family history were not explained by geographic differences or difference in data capture period between the two databases. Given the similarity for the other risk factors and treatments, it is likely that the difference in ethnicity and family history recording reflects a difference in recording patterns between the two clinical computer systems rather than a true difference between the two cohorts. A similar pattern for recording of ethnicity and family history was also reported in the validation of QRISK on the Health Improvement Network (THIN database) [11 26]. This was thought to be due to different usage of clinical templates in the clinical system, with EMIS practices having ethnicity and family history included more often thereby prompting the user to enter this information in a more systematic fashion.

4.5 Comparison of QResearch and CPRD incidence rates

The age standardised incidence rates for each condition were generally marginally higher on QResearch than CPRD although the proportions of events identified on GP data (out of all events recorded on either GP or linked mortality data) was very close. This suggests that patterns of recording of major clinical events are very similar between QResearch and CPRD although the absolute value varies by clinical condition. For example, 91% of ischaemic stroke events recorded on either GP or linked mortality data are identified on the GP record compared with 99% of hip fractures. We also note the lower levels of total cardiovascular events in the GP

clinical record which was between 13-15% lower than the total recorded on either the GP record, the linked mortality record or the linked hospital admissions record. Some of this will reflect new sudden events where the first presentation was a hospital admission or death whilst others may reflect some under-representation of existing cases not recorded in the GP record. Our study is unable to distinguish between these two scenarios, though the latter one potentially has clinical consequences if the patient is not identified as having cardiovascular disease as they may not be offered secondary prevention.

We think that the information on baseline characteristics and incidence rates will have a utility beyond the present study since it suggests that both databases are fundamentally similar in many aspects and likely to generate similar results for a range of epidemiological studies[27].

4.6 Summary

In summary, we have tested a set of QPrediction scores using an external independent cohort of practices contributing to the CPRD. The results demonstrate good performance, comparable to the results obtained from QResearch, meaning that the findings of studies performed in either database are likely to be applicable in England.

5 Supporting information

Approvals

The project was approved in accordance with the QResearch® agreement with Trent Research Ethics Committee (ref 03/04/021) and approved by the ISAC committee of the CPRD (ref 13_079).

FOOTNOTES

Acknowledgements
We thank EMIS and EMIS practices for their contribution to the QResearch database. We thank CPRD and Vision Practices for allowing access to the CPRD for this study.

Source of funding:
The validation of the QPrediction scores is funded by the National Institute for Health Research's School for Primary Care Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funding body did not play a role in the study design, writing of the manuscript or in the decision to submit the manuscript for publication.

Statement
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd and its Licensees to permit this article (if accepted) to be published in HEART editions and any other BMJPGL products to exploit all subsidiary rights.

Contributorship
JHC initiated the study, undertook the literature review, data extraction, data manipulation and primary data analysis and write the first draft of the paper. JHC is the guarantor. CC contributed to the design, analysis, interpretation and drafting of the paper. PB contributed to the development of core ideas, the analysis plan, interpretation of the results and the drafting of the paper.

Competing interests
JHC 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). JHC is also director of ClinRisk Ltd which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of Medical Statistics at the University of Nottingham and a consultant statistician for ClinRisk Ltd. PB has received financial

support for undertaking the validation work from the National School for Primary Care Research. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations. There are no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

No additional data are available.

6 References

1. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008:bmj.39609.449676.25.
2. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ 2013;**346**.
3. Hippisley-Cox J, Coupland C, Robson J, et al. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009;**338**:b880-.
4. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. BMJ 2009;**339**:b4229-.
5. Hippisley-Cox J, Coupland C. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney(R) Scores. BMC Family Practice 2010;**11**:49.
6. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. BMJ 2011;**343**:d4656.
7. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions BMJ 2013:(in press).
8. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.
9. Collins GS, Mallett S, Altman DG. Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. BMJ 2011;**342**:d3651.
10. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010;**340**:c2442-.
11. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;**339**:b2584-.
12. Collins GS, Altman DG. External validation of the QDScore for predicting the 10-year risk of developing Type 2 diabetes. Diabetic Medicine 2011;**28**:599-607.
13. Hippisley-C J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleeding with anticoagulants: cohort study to derive and validate the QBleed score. BMJ 2014;**(in press)**.
14. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. Journal of statistical software 2011;**45**:1-20.
15. Gray A, Clarke P, Farmer A, et al. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). BMJ 2002;**325**(7369):860.
16. Royston P. Multiple imputation of missing values. Stata Journal 2004;**4**(3):227-41.

17. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007:bmj.39261.471806.55.
18. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;**23**:723-48.
19. Royston P. Explained variation for survival models. Stata J 2006;**6**:1-14.
20. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;**13**:33.
21. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;**338**(may28_1):b605-.
22. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**:4-14.
23. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;**60**:e128-36.
24. Collins GS, Altman A. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;**344**.
25. Collins GS, Altman D. Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney scores using a primary care database. British Journal of General Practice 2012;**62**(597):e243-e50.
26. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 2008;**94**:34-39.
27. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. BMJ Open 2014;**4**(4).
28. Hippisley-Cox J, Coupland C. QRISK2-2014 Annual Update Information, 2014:5.
29. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ 2012;**344**(may22 1):e3427-e27.
30. Hippisley-Cox J CC. QDiabetes 2013 Annual Update Information Secondary QDiabetes 2013 Annual Update Information 2013. http://qdiabetes.org/QDiabetes-2013-Annual-Update-Information.pdf.

Table 1 Summary of QPrediction scores including outcome and predictor variables

Score	Weblink∞	Outcome	Predictors
QDiabetes[3]	www.qdiabetes.org	10 year risk of type 2 diabetes±	**In men and women:** Age, sex, ethnicity, deprivation, smoking, family history of diabetes, cardiovascular disease, treated hypertension, steroid tables, body mass index
QRISK2[28]	www.qrisk.org	10 year risk of CVD **	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation.
QStroke[2]	www.qstroke.org	10 year risk of ischaemic stroke or TIA±	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, congestive cardiac failure, valvular heart disease
QKidney[5]	www.qkidney.org	5 year risk of moderate or severe kidney failureμ	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, body mass index, family history of kidney disease, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, SLE, peripheral vascular disease, kidney stones, NSAIDs
QThrombosis	www.qthrombosis.org	5 year risk of venous thromboembolism±	**In men and women**: age, body mass index, smoking status, varicose veins, congestive cardiac failure, chronic renal disease, cancer, chronic obstructive pulmonary disease, inflammatory bowel disease, hospital admission in past six months, and current prescriptions for antipsychotic drugs. Additionally in women: combined oral contraceptives, tamoxifen, and hormone replacement therapy
QBleed[13]	www.qbleed.org	5 year risk of upper gastrointestinal bleed in patient starting anticoagulants vs others*	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; venous thromboembolism; congestive cardiac failure; treated hypertension; cancer; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QBleed[13]	www.qbleed.org	5 year risk of intracranial bleed in patient starting anticoagulants vs others *	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; treated hypertension; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QFracture[29]	www.qfracture.org	10 year risk of hip fracture± 10 year risk of	**In women:** HRT usage, age, body mass index, smoking status, recorded alcohol use, parental history of osteoporosis, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls, menopausal symptoms, chronic liver disease, gastrointestinal malabsorption and other endocrine disorders.

		osteoporotic fracture [µ]	**In men:** age, body mass index, smoking status, recorded alcohol use, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls and liver disease.

∞the web link has the relevant calculator, links to academic papers, additional information including links to the open source software

[±] recorded either on GP record or linked ONS mortality record;

µ recorded on the GP record.

*Recorded either on linked hospital admissions record or ONS mortality record

**Recorded either on linked hospital admissions record or ONS mortality or linked hospital admissions record

∞**th**∞

Table 2 Comparison of baseline characteristics of patients in CPRD validation cohort and QResearch comparison cohort

	CPRD	*CPRD*	*QResearch*	*QResearch*
	Men (%)	**Women (%)**	**Men (%)**	**Women (%)**
total	1,588,803	1,682,709	3,304,145	3,454,504
Ageband				
25-34 years	427975 (26.9)	467192 (27.8)	1083589 (32.8)	1179742 (34.2)
35-44 years	396680 (25.0)	364150 (21.6)	814988 (24.7)	731089 (21.2)
45-54 years	294274 (18.5)	277663 (16.5)	558553 (16.9)	516188 (14.9)
55-64 years	212817 (13.4)	211636 (12.6)	390229 (11.8)	389266 (11.3)
65-74 years	148180 (9.3)	164172 (9.8)	267997 (8.1)	298847 (8.7)
75+ years	108877 (6.9)	197896 (11.8)	188789 (5.7)	339372 (9.8)
mean Townsend score (SD)	-.5 (3.2)	-.5 (3.2)	.3 (3.6)	.2 (3.6)
Care home resident	1407 (0.1)	3466 (0.2)	2983 (0.1)	7411 (0.2)
Ethnicity recorded	587879 (37.0)	658835 (39.2)	1859462 (56.3)	2077181 (60.1)
White or not recorded	1515113 (95.4)	1602212 (95.2)	3010061 (91.1)	3149618 (91.2)
Indian	16442 (1.0)	16025 (1.0)	56156 (1.7)	50406 (1.5)
Pakistani	6606 (0.4)	6146 (0.4)	30632 (0.9)	23405 (0.7)
Bangladeshi	2419 (0.2)	1688 (0.1)	23017 (0.7)	17450 (0.5)
Other Asian	10795 (0.7)	11873 (0.7)	32513 (1.0)	36886 (1.1)
Caribbean	4989 (0.3)	6425 (0.4)	25782 (0.8)	32953 (1.0)
Black African	12883 (0.8)	14771 (0.9)	51980 (1.6)	56528 (1.6)
Chinese	2914 (0.2)	4176 (0.2)	16084 (0.5)	23043 (0.7)
Other ethnic group	16642 (1.0)	19393 (1.2)	57920 (1.8)	64215 (1.9)
Smoking status recorded	1442088 (90.8)	1595538 (94.8)	2943405 (89.1)	3219598 (93.2)
Non smoker	613833 (38.6)	834721 (49.6)	1449694 (43.9)	1973691 (57.1)
Ex-smoker	252873 (15.9)	222615 (13.2)	611837 (18.5)	545125 (15.8)
Light smoker (1-9/day)	104466 (6.6)	109864 (6.5)	472614 (14.3)	384482 (11.1)
Moderate smoker (10-19/day)	183000 (11.5)	179391 (10.7)	223631 (6.8)	202776 (5.9)
Heavy smoker (20+/day)	142438 (9.0)	87474 (5.2)	185629 (5.6)	113524 (3.3)
Smoker amount not recorded	145478 (9.2)	161473 (9.6)	0 (0.0)	0 (0.0)
Alcohol status recorded	1238110 (77.9)	1379002 (82.0)	2584335 (78.2)	2834426 (82.1)
Non drinker	163633 (10.3)	318880 (19.0)	583752 (17.7)	1035692 (30.0)
Trivial <1u/day	460091 (29.0)	726851 (43.2)	782985 (23.7)	1144469 (33.1)
Light 1-2u/day	411261 (25.9)	290547 (17.3)	481674 (14.6)	402750 (11.7)
Moderate 3-6/day	166328 (10.5)	36763 (2.2)	648549 (19.6)	237679 (6.9)
Heavy 7-9u/day	19612 (1.2)	2853 (0.2)	54083 (1.6)	7152 (0.2)
Very Heavy >/day	17185 (1.1)	3108 (0.2)	24468 (0.7)	5195 (0.2)
Family History				
family history of CHD	68805 (4.3)	80985 (4.8)	326995 (9.9)	417537 (12.1)
family history of diabetes	96810 (6.1)	132390 (7.9)	357109 (10.8)	487397 (14.1)
family history osteoporosis	880 (0.1)	10062 (0.6)	1655 (0.1)	17529 (0.5)
family history kidney disease	1253 (0.1)	1586 (0.1)	2034 (0.1)	2769 (0.1)

Clinical Values recorded				
BMI recorded	1268235 (79.8)	1481918 (88.1)	2553514 (77.3)	2857742 (82.7)
mean BMI (SD)	29.6 (6.8)	28.2 (7.0)	29.4 (6.8)	28.2 (7.0)
SBP recorded	1359560 (85.6)	1590226 (94.5)	2755733 (83.4)	3190390 (92.4)
mean SBP(SD)	133.1 (23.6)	128.6 (22.7)	132.2 (18.3)	127.1 (20.8)
cholesterol/HDL ratio recorded	587865 (37.0)	606035 (36.0)	1323503 (40.1)	1368180 (39.6)
mean cholesterol ratio (SD)	4.4 (1.4)	3.7 (1.2)	4.4 (1.4)	3.7 (1.2)
platelets recorded	223461 (14.1)	382799 (22.7)	478596 (14.5)	829702 (24.0)
platelets < 150 or > 480	11051 (0.7)	13282 (0.8)	23479 (0.7)	27009 (0.8)
Creatinine recorded	811779 (51.1)	997118 (59.3)	1714337 (51.9)	2053036 (59.4)
Mean creatinine (SD)	96.7 (32.5)	79.7 (24.1)	95.5 (30.7)	78 (23.7)

Table 3 Prescribed medication and clinical diagnoses recorded at baseline in CPRD validation cohort and QResearch comparison cohort

	CPRD men (%)	*CPRD women (%)*	*QResearch men (%)*	*QResearch women (%)*
Prescribed medication				
anticoagulants	15955 (1.0)	13077 (0.8)	27024 (0.8)	22178 (0.6)
antidepressants	101553 (6.4)	235797 (14.0)	178532 (5.4)	398018 (11.5)
antipsychotics	33884 (2.1)	79514 (4.7)	47464 (1.4)	92307 (2.7)
antiplatelets	97475 (6.1)	92816 (5.5)	160910 (4.9)	153405 (4.4)
oral NSAIDs	246515 (15.5)	346416 (20.6)	396026(12.0)	556644 (16.1)
tamoxifen	n/a	9231 (0.5)	n/a	18343 (0.5)
oestrogen only Hormone replacement therapy	n/a	119373 (7.1)	n/a	208333 (6.0)
oral corticosteroids	45597 (2.9)	71352 (4.2)	54354 (1.6)	88205 (2.6)
oral contraceptive pill	n/a	174287 (10.4)	n/a	332696 (9.6)
Recorded Diagnoses				
congestive cardiac failure	15836 (1.0)	19707 (1.2)	24965 (0.8)	28852 (0.8)
atrial fibrillation	20125 (1.3)	20102 (1.2)	33499 (1.0)	32580 (0.9)
coronary heart disease	80377 (5.1)	57703 (3.4)	130220 (3.9)	88606 (2.6)
cardiovascular disease	101430 (6.4)	83167 (4.9)	165495 (5.0)	130214 (3.8)
peripheral vascular disease	17029 (1.1)	13101 (0.8)	25004 (0.8)	17078 (0.5)
venous thromboembolism	15072 (0.9)	23090 (1.4)	27086 (0.8)	40813 (1.2)
rheumatoid or SLE	7455 (0.5)	19010 (1.1)	21453 (0.6)	48447 (1.4)
rheumatoid arthritis	7243 (0.5)	17468 (1.0)	21142 (0.6)	45542 (1.3)
SLE	228 (0.0)	1756 (0.1)	351 (0.0)	3374 (0.1)
type 1 diabetes	6238 (0.4)	4924 (0.3)	12029 (0.4)	9612 (0.3)
type 2 diabetes	51634 (3.2)	43271 (2.6)	95401 (2.9)	79654 (2.3)
treated hypertension	123584 (7.8)	161709 (9.6)	210516 (6.4)	267076 (7.7)
chronic renal disease	3968 (0.2)	4082 (0.2)	8550 (0.3)	8995 (0.3)
moderate/severe kidney failure	14107 (0.9)	9500 (0.6)	30407 (0.9)	21509 (0.6)
severe kidney failure	1603 (0.1)	1125 (0.1)	3641 (0.1)	2672 (0.1)
renal stones	13415 (0.8)	6443 (0.4)	37422 (1.1)	29204 (0.8)
inflammatory bowel disease	8962 (0.6)	10208 (0.6)	17762 (0.5)	19502 (0.6)
dementia	6686 (0.4)	16634 (1.0)	12872 (0.4)	30497 (0.9)
parkinsons disease	4546 (0.3)	4676 (0.3)	6830 (0.2)	6611 (0.2)
epilepsy or anticonvulsants	47170 (3.0)	71171 (4.2)	56516 (1.7)	61561 (1.8)
cancer	26866 (1.7)	43908 (2.6)	51649 (1.6)	79326 (2.3)
liver disease	3959 (0.2)	2893 (0.2)	9947 (0.3)	6410 (0.2)
chronic liver disease or pancreatitis	5521 (0.3)	4051 (0.2)	13069 (0.4)	8729 (0.3)
oesophageal varices	469 (0.0)	340 (0.0)	1626 (0.0)	1388 (0.0)
prior haemorrhage	97562 (6.1)	79765 (4.7)	203278 (6.2)	147533 (4.3)
malabsorption	7343 (0.5)	9375 (0.6)	21042 (0.6)	26002 (0.8)
endocrine diseases	3082 (0.2)	14097 (0.8)	6026 (0.2)	27731 (0.8)
COPD	24029 (1.5)	20737 (1.2)	41281 (1.2)	34785 (1.0)
asthma or COPD	142974 (9.0)	169503 (10.1)	273768 (8.3)	310027 (9.0)
history of falls	28878 (1.8)	61905 (3.7)	34584 (1.0)	67465 (2.0)
prior fracture	24265 (1.5)	45752 (2.7)	62092 (1.9)	89000 (2.6)

varicose vein surgery	18979	(1.2)	47012	(2.8)	35651	(1.1)	85602	(2.5)
emergency admissions or hip op	3483	(0.2)	5266	(0.3)	3335	(0.1)	5508	(0.2)

Table 4 Numbers of patients eligible for each score in the CPRD validation cohort and number of patients with complete risk factor recording not requiring multiple imputation.

Risk Score	*Clinical outcome*	*Eligible age range*	*exclusion criteria at study entry*	*total in age range*	*total with exclusions*	*total eligible for analysis*	*Total complete data*	*% complete data*
QDiabetes	Type 2 diabetes	25-84	type 1 or 2 diabetes at study entry	3,177,192	99,189	3,078,003	2,467,642	80.2
QStroke	ischaemic stroke	25-84	existing stroke or anticoagulants at study entry	3,177,192	70,961	3,106,231	1,032,184	33.2
QRISK2	cardiovascular disease	25-84	existing CVD or statins at study entry	3,177,192	232,722	2,944,470	906,781	30.8
QThrombosis	thromboembolism	25-84	existing VTE or anticoagulants at study entry	3,177,192	53,904	3,123,288	2,513,347	80.5
QFracture	fractured neck of femur	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QFracture	osteoporotic fracture	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QKidney	moderate or severe kidney failure	35-74	existing moderate or severe kidney failure	2,069,572	10,518	2,059,054	1,146,619	55.7
QKidney	severe kidney failure	35-74	existing severe kidney failure	2,069,572	1,930	2,067,642	1,153,979	55.8
QBleed	upper gastro-intestinal bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0
QBleed	intracranial bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0

*entry date was 01.01.1998 except for upper GI bleed and intracranial bleed where entry date was 01.01.2007

Table 5 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in women

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	48,143	99.88	4.13 (4.10 to 4.17)	102,544	99.93	4.33 (4.31 to 4.36)	0.95
	GP or ONS	48,203	n/a	4.13 (4.10 to 4.17)	102,618	n/a	4.34 (4.31 to 4.36)	0.95
Ischaemic stroke	GP data	32,283	90.64	2.45 (2.42 to 2.48)	63,582	90.22	2.45 (2.44 to 2.47)	1.00
	GP or ONS	35,617	n/a	2.62 (2.59 to 2.64)	70,477	n/a	2.70 (2.68 to 2.72)	0.97
Cardiovascular disease	GP data	55,833	85.71	5.41 (5.37 to 5.46)	107,412	84.96	4.32 (4.30 to 4.35)	1.25
	GP or ONS	65,143	n/a	6.32 (6.27 to 6.37)	126,433	n/a	5.03 (5.01 to 5.06)	1.26
	GP or ONS or HES	69,202	n/a	6.72 (6.67 to 6.77)	140,510	n/a	5.63 (5.60 to 5.66)	1.19
Thromboembolism	GP data	18,199	91.1	1.52 (1.49 to 1.54)	35,971	90.55	1.46 (1.44 to 1.47)	1.04
	GP or ONS	19,978	n/a	1.64 (1.62 to 1.67)	39,727	n/a	1.60 (1.58 to 1.62)	1.03
Fractured neck of femur	GP data	17,529	99.98	1.32 (1.30 to 1.34)	34,821	99.99	1.40 (1.39 to 1.42)	0.94
	GP or ONS	17,533	n/a	1.32 (1.30 to 1.34)	34,825	n/a	1.40 (1.39 to 1.42)	0.94
Osteoporotic fracture	GP data	34,528	n/a	2.89 (2.58 to 3.20)	81,334	n/a	3.63 (3.61 to 3.66)	0.80
mod /severe kidney failure	GP data	19,902	n/a	2.06 (1.76 to 2.36)	48,665	n/a	2.81 (2.78 to 2.83)	0.73
severe kidney failure	GP data	1,737	n/a	0.18 (0.09 to 0.27)	4,150	n/a	0.24 (0.24 to 0.25)	0.74

Table 6 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in men

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	60,731	99.92	5.84 (5.79 to 5.89)	128,234	99.94	5.97 (5.94 to 6.00)	0.98
	GP or ONS	60,782	n/a	5.84 (5.80 to 5.89)	128,317	n/a	5.98 (5.94 to 6.01)	0.98
ischaemic stroke	GP data	32,223	93.55	3.17 (3.14 to 3.20)	63,480	92.85	3.10 (3.08 to 3.13)	1.02
	GP or ONS	34,443	n/a	3.33 (3.30 to 3.37)	68,366	n/a	3.37 (3.34 to 3.40)	0.99
Cardiovascular disease	GP data	70,283	86.7	7.38 (7.33 to 7.44)	137,136	86.12	7.12 (7.08 to 7.16)	1.03
	GP or ONS	81,068	n/a	8.52 (8.46 to 8.58)	159,240	n/a	8.37 (8.33 to 8.41)	1.02
	GP or ONS or HES	84,620	n/a	8.90 (8.84 to 8.96)	174,405	n/a	9.17 (9.13 to 9.21)	0.97
thromboembolism	GP data	15,655	92.32	1.49 (1.46 to 1.51)	31,503	92.22	1.44 (1.43 to 1.46)	1.03
	GP or ONS	16,958	n/a	1.61 (1.59 to 1.63)	34,161	n/a	1.57 (1.56 to 1.59)	1.02
fractured neck of femur	GP data	5,706	99.98	0.65 (0.63 to 0.67)	12,435	99.98	0.71 (0.70 to 0.73)	0.91
	GP or ONS	5,707	n/a	0.65 (0.63 to 0.67)	12,438	n/a	0.71 (0.70 to 0.73)	0.91
osteoporotic fracture	GP data	11,169	n/a	1.29 (1.05 to 1.52)	28,555	n/a	1.54 (1.52 to 1.55)	0.84
Mod/severe kidney failure	GP data	37,597	n/a	4.88 (4.37 to 5.38)	86,649	n/a	5.82 (5.78 to 5.85)	0.84
severe kidney failure	GP data	3,472	n/a	0.54 (0.38 to 0.71)	7,372	n/a	0.47 (0.46 to 0.48)	1.15

Table 7 Performance of QPrediction scores on the CPRD validation cohort compared with published results for the QResearch validation cohort

		CPRD	*CPRD*	*QResearch*	*QResearch*
		women	**men**	**women**	**men**
	statistic	**mean (95%CI)**	**mean (95%CI)**	**mean (95%CI)**	**mean (95%CI)**
QDiabetes 2013 (type 2 diabetes)[30]	ROC	0.846 (0.844 to 0.848)	0.818 (0.816 to 0.82)	0.853 (0.851 to 0.856)	0.837 (0.835 to 0.840)
	R^2 (%)	49.6 (49.2 to 50.1)	45.7 (45.3 to 46.2)	50.8 (50.3 to 51.4)	48.1 (47.6 to 48.6)
	D statistic	2.032 (2.015 to 2.049)	1.879 (1.863 to 1.895)	2.081 (2.058 to 2.104)	1.971 (1.951 to 1.991)
QKidney -2010[5] (moderate or severe kidney failure)	ROC	0.875 (0.87 to 0.879)	0.88 (0.878 to 0.883)	0.877 (0.873 to 0.880)	0.878 (0.874 to 0.882)
	R^2 (%)	58.3 (57.8 to 58.7)	57.5 (57.1 to 57.8)	56.45 (55.40 to 57.50)	58.29 (55.31 to 61.26)
	D statistic	2.418 (2.394 to 2.442)	2.379 (2.361 to 2.397)	2.33 (2.28 to 2.40)	2.42 (2.28 to 2.56)
QKidney -2010 (severe kidney failure)[5]	ROC	0.839 (0.822 to 0.855)	0.851 (0.84 to 0.862)	0.843 (0.825 to 0.860)	0.846 (0.829 to 0.862)
	R^2 (%)	51.4 (49.5 to 53.2)	53.8 (52.6 to 55.1)	55.39 (52.59 to 58.18)	56.65 (53.94 to 59.35)
	D statistic	2.103 (2.025 to 2.182)	2.21 (2.154 to 2.266)	2.28 (2.15 to 2.41)	2.34 (2.21 to 2.47)
QRISK2-2014[28] (cardiovascular disease)	ROC	0.883 (0.882 to 0.884)	0.859 (0.858 to 0.861)	0.892 (0.892 to 0.895)	0.871 (0.869 to 0.873)
	R^2 (%)	56.4 (56.1 to 56.7)	50.9 (50.6 to 51.2)	58.8 (58.4 to 59.1)	53.3 (52.9 to 53.7)
	D statistic	2.328 (2.313 to 2.343)	2.085 (2.071 to 2.098)	2.443 (2.423 to 2.463)	2.188 (2.171 to 2.205)
QStroke-2013[2] (ischaemic stroke or TIA)	ROC	0.882 (0.88 to 0.883)	0.869 (0.867 to 0.87)	0.877 (0.875 to 0.879)	0.866 (0.864 to 0.868)
	R^2 (%)	58.4 (58.1 to 58.8)	55.3 (54.9 to 55.7)	57.3 (56.8 to 57.8)	55.1 (54.6 to 55.7)
	D statistic	2.427 (2.408 to 2.446)	2.278 (2.259 to 2.297)	2.37 (2.35 to 2.40)	2.27 (2.24 to 2.30)
QThrombosis-2010[6] (venous thromboembolism)	ROC	0.756 (0.751 to 0.761)	0.765 (0.760 to 0.770)	0.75 (0.74 to 0.76)	0.75 (0.74 to 0.76)
	R^2 (%)	35.3 (34.5 to 36.1)	34.5 (33.7 to 35.4)	32.78 (31.08 to 34.48)	33.51 (31.71 to 35.30)

	D statistic	1.512 (1.485 to 1.538)	1.486 (1.458 to 1.513)	1.43 (1.37 to 1.49)	1.45 (1.39 to 1.51)
QBleed-2014[13] (upper gastrointestinal bleed)	ROC statistic	0.775 (0.770 to 0.781)	0.759 (0.753 to 0.764)	0.766 (0.758 to 0.775)	0.747 (0.738 to 0.756)
	R^2 (%)	44.7 (43.6 to 45.9)	41.6 (40.5 to 42.8)	40.7 (38.9 to 42.6)	36.9 (35.1 to 38.7)
	D statistic	1.842 (1.798 to 1.885)	1.729 (1.687 to 1.771)	1.70 (1.63 to 1.76)	1.57 (1.51 to 1.63)
QBleed-2014[13] (intracranial bleed)	ROC statistic	0.808 (0.801 to 0.816)	0.789 (0.780 to 0.797)	0.847 (0.838 to 0.856)	0.812 (0.80 to 0.824)
	R^2 (%)	51.7 (50.1 to 53.3)	50.0 (48.3 to 51.7)	58.0 (56.0 to 60.0)	53.3 (51.1 to 55.4)
	D statistic	2.118 (2.051 to 2.186)	2.046 (1.977 to 2.116)	2.40 (2.30 to 2.50)	2.19 (2.09 to 2.28)
QFracture-2012[29] (fractured neck of femur)	ROC	0.89 (0.888 to 0.892)	0.872 (0.867 to 0.877)	0.893 (0.890 to 0.896)	0.875 (0.868 to 0.883)
	R^2 (%)	70.6 (70.2 to 71)	69.2 (68.5 to 70)	71.73 (71.10 to 72.30)	70.37 (69.25 to 71.49)
	D statistic	3.171 (3.139 to 3.203)	3.07 (3.016 to 3.124)	3.26 (3.21 to 3.31)	3.15 (3.06 to 3.24)
QFracture -2012[29] (osteoporotic fracture: hip, spine, wrist,humerus)	ROC	0.817 (0.814 to 0.819)	0.768 (0.763 to 0.773)	0.790 (0.787 to 0.793)	0.711 (0.703 to 0.719)
	R^2 (%)	56.3 (55.8 to 56.7)	49.8 (48.9 to 50.7)	51.9 (51.2 to 52.6)	38.20 (36.89 to 39.57)
	D statistic	2.322 (2.301 to 2.343)	2.038 (2.002 to 2.075)	2.13 (2.10 to 2.15)	1.61 (1.56 to 1.66)

Notes on understanding validation statistics:

D statistic is a measure of discrimination - higher values indicate better discrimination
ROC statistic is a measure of discrimination - higher values indicate better discrimination
The R^2 statistic is a measure of explained variation - higher values indicate more variation is explained

Table 8 Performance of each score for predicting the relevant outcome in the CPRD validation cohort. The cut off is the threshold of predicted risk for the top decile in the CPRD cohort.

score	*outcome*	*duration*	*cut off (%) for top decile predicted risk*	*Sensitivity (%)*	*Specificity (%)*	*observed risk (%)*
QDiabetes	Type 2 diabetes	10 yr risk	13.0	44.8	91.0	20.8
QStroke	Ischaemic stroke	10 yr risk	10.5	54.7	90.8	16.1
QRISK2	cardiovascular disease	10 yr risk	20.7	49.9	91.9	31.8
QThrombosis	venous thromboembolism	5 yr risk	1.5	36.2	90.1	2.6
QKidney	moderate-severe kidney failure	5 yr risk	6.3	59.1	90.5	6.9
QKidney	severe kidney failure	5 yr risk	0.4	58.5	90.0	0.7
QBleed	upper GI bleed	5 yr risk	1.6	38.0	90.2	3.5
QBleed	intracranial bleed	5 yr risk	0.9	44.2	90.1	1.6
QFracture	fractured neck of femur	10 yr risk	3.7	66.5	90.4	9.4
QFracture	osteoporotic fracture	10 yr risk	7.8	49.6	90.5	13.1

Figure 1 Calibration of each QPrediction score comparing the mean predicted risks with the observed risks in the CPRD cohort.

1a QThrombosis (venous thromboembolism)

1b QFracture (hip)

1c QFracture (hip, colles, spine, shoulder)

1d QStroke (ischaemic stroke)

1e QDiabetes (type 2 diabetes)

1f QBleed (upper gastrointestinal haemorrhage)

1g QBleed (intracranial haemorrhage)

1h QKidney (moderate or severe kidney failure)

1i QKidney(severe kidney failure)

1j QRisk2 (cardiovascular disease)

The performance of seven QPrediction risk scores in an independent external sample of patients from general practice: a validation study

Revised Research paper submitted to BMJ Open
Reference-2014-005809
July 2014

Authors

Julia Hippisley-Cox	Professor of Clinical Epidemiology & General Practice[1]
Carol Coupland	Associate Professor and Reader in Medical Statistics[1]
Peter Brindle	Research and Development Programme Director[2]

Institutions

[1]Division of Primary Care, 13th floor, Tower Building, University Park, Nottingham, NG2 7RD.

[2]Avon Primary Care Research Collaborative, Bristol Clinical Commissioning Group, South Plaza, Marlborough Street, Bristol BS1 3NX.

Guarantor & Author for correspondence - Julia Hippisley-Cox

Email: Julia.hippisley-cox@nottingham.ac.uk

Telephone: 0115 8466915

Fax: 0115 8466904

Abstract

Objectives To validate the performance of a set of risk prediction algorithms developed using the QResearch database, in an independent sample from general practices contributing to the Clinical Research Data Link (CPRD).

Setting Prospective open cohort study using practices contributing to the CPRD database and practices contributing to the QResearch database.

Participants The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices between 01 Jan 1998 and 31 July 2012. The validation statistics for QResearch were obtained from the original published papers which used a one third sample of practices separate to those used to derive the score. A cohort from QResearch was used to compare incidence rates and baseline characteristics and consisted of 6.8 million patients from 753 practices registered between 01 Jan 1998 and until 31 July 2013.

Outcome measures Incident events relating to seven different risk prediction scores: QRISK2 (cardiovascular disease); QStroke (ischaemic stroke); QDiabetes (type 2 diabetes); QFracture (osteoporotic fracture and hip fracture); QKidney (moderate and severe kidney failure); QThrombosis (venous thromboembolism); QBleed (intracranial bleed and upper gastrointestinal haemorrhage). Measures of discrimination and calibration were calculated.

Results Overall, the baseline characteristics of the CPRD and QResearch cohorts were similar though QResearch had higher recording levels for ethnicity and family history. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with the published results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch and the ROC value was 0.85 on both databases. The scores were well calibrated in CPRD.

Conclusion Each of the algorithms performed practically as well in the external independent CPRD validation cohorts as they had in the original published QResearch validation cohorts.

Strengths and limitations of this study

- This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system from that used to derive the algorithms.
- The discrimination and calibration statistics for each score were very similar in CPRD to those published from validation cohorts from QResearch. This supports their potential utility in the general population of patients in primary care.
- A strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice.
- The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible.
- Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using these algorithms in primary care.

1 Introduction

In the last 7 years, we have developed a series of risk prediction algorithms using the QResearch database. QResearch is a large research database containing pseudonymised individual level data from over 700 general practices using the EMIS clinical system. The QResearch database consists of data collected from primary care (coded information on socio-demographic characteristics, diagnoses, symptoms, smoking/alcohol, clinical measurements, laboratory values, prescriptions and referrals) which has been linked to cause of death, hospital episodes and cancer registrations at individual patient level.

The algorithms predict outcomes such as cardiovascular disease(www.qrisk.org)[1], stroke (www.qstroke.org)[2], type 2 diabetes (www.qdiabetes.org)[3], osteoporotic fracture (www.qfracture.org)[4], moderate or severe kidney disease (www.qkidney.org)[5], venous thrombo-embolism (www.qthrombosis.org)[6], and emergency hospital admission (www.qadmissions.org)[7]. Generally, the "QPrediction" algorithms have been designed to systematically identify patients in primary care at high risk of a serious clinical outcome for whom further intervention to lower risk of that outcome might be possible. They are also designed to quantify absolute risk of serious outcomes in a way which patients can understand and which might help guide lifestyle and management decisions. A number of these algorithms are now integrated into GP clinical computer systems, included in national guidelines[1 4] and are in daily use across the NHS [1 3 8].

The algorithms were originally developed using a random two thirds sample of practices contributing to the QResearch database and validated on the remaining third. Whilst this represents a physically discrete population of patients and practices for validation, the practices all use the same clinical computer system (EMIS), which is in use in 53% of UK practices. A more stringent test of performance is to validate the algorithms on a fully external database derived from practices using a different but commonly used primary care computer system. This would help determine whether the predictions from the algorithms are likely to generalise to the whole population in England. Whilst some of the algorithms have been validated by an independent team using the THIN primary care database[9-12], there are currently no published validations of the algorithms using a primary care database which is routinely linked to mortality data in the same way as QResearch.

We therefore decided to validate the various QPrediction Scores using another database known as the Clinical Research Data Link (CPRD). The General Practice Research database (GPRD) was originally set up in 1988 and is of similar nature to QResearch although it is derived from practices using a different clinical computer system (Vision, which is used by 20% of GPs). It was extended to include linked mortality data and data from secondary care and was renamed the Clinical Research Data Link (CPRD) in 2012. Our secondary objective was to compare the ascertainment of incident clinical events recorded in GP data alone with that

recorded in either GP data or the linked mortality data in both the CPRD and QResearch.

2 Methods

2.1 CPRD Study population

For the validation using CPRD, we identified an open cohort of patients aged 25-99 years at entry to the cohort and followed this cohort up until 31st July 2012 (the latest date for which linked data were available at the time of analysis). We restricted the CPRD cohort to 357 practices in England which had linked ONS mortality and hospital admissions data. For each patient we determined an entry date to the cohort, which was the latest of the following dates: 25th birthday, date of registration with the practice plus one year, date on which the practice computer system was installed plus one year, and the beginning of the study period (01 January 1998). Patients were censored at the earliest date of the relevant outcome, de-registration with the practice, last upload of computerised data or the study end date (31 July 2012).

For the assessment of the two QBleed outcomes (intracranial bleed and upper gastrointestinal haemorrhage) we used a later cohort entry date of 01.01.2007 for comparability with the equivalent study period for the derivation of the algorithm on QResearch[13].

2.2 QResearch study population

For comparison of the validation statistics (ROC, D and R2 statistics), we extracted the original published values from the papers which had been calculated using a one third sample of practices from QResearch which were independent from the two thirds of practices used to derive the scores.

For comparison of the baseline characteristics, incidence rates and ascertainment rates we used the latest version of the QResearch database which is currently available (QResearch 38, 31st Dec 2013). We identified an open cohort in the same way as for CPRD, using all of the QResearch practices in England, and with follow-up until 31 July 2013.

2.3 Inclusion and exclusion criteria

For both databases, we excluded patients without a Townsend score (an area based measure of material deprivation derived from the post code) and temporary residents. For each score we then identified patients who were eligible to have the score calculated according to the relevant inclusion and exclusion criteria as summarised in Table 4

2.4 Risk scores included in validation

We validated the following risk prediction scores on CPRD:

1. QDiabetes - 10 year risk of type 2 diabetes[3]
2. QRISK2 – 10 year risk of cardiovascular disease[1]
3. QStroke – 10 year risk of stroke or transient ischaemic attack (TIA)[5]
4. QFracture - 10 year risk of hip or osteoporotic fracture[4]
5. QThrombosis – 5 year risk of venous thrombo-embolism (VTE)[6]
6. QBleed – 5 year risk of upper gastrointestinal haemorrhage and intracranial haemorrhage[13]
7. QKidney – 5 year risk of moderate-severe kidney disease[5]

2.5 Clinical outcomes

We identified the relevant clinical outcome using the same definition as had been applied in the original derivation of the risk scores using QResearch. The data sources used to identify the clinical outcomes had varied over the six years during which the original studies had been undertaken due to the changing availability of linked hospital and mortality data over that time. In 2008, the QResearch database was linked to mortality records for 1997 onwards. In 2013, the QResearch database was linked to hospital admissions records with data for patients from 1998 onwards. For the latest updated version of QRISK2 (QRISK2-2014), the outcome was identified by the presence of the relevant Read code on the GP record or an ICD10 code recorded on the linked mortality record or on the linked hospital admissions record. For QStroke, QDiabetes, QFracture and QThrombosis, the outcome was identified either by the presence of the relevant Read code recorded on the GP record or an ICD10 code recorded on the linked mortality record. For QKidney, the outcome was identified solely from information recorded in the GP record as in the original study as it required blood test values which were only present in the GP record. For QBleed, the outcome was identified in CPRD from events recorded either on the linked hospital admissions database or the linked mortality record in order to identify the events most likely to have serious clinical consequences for the patient.

We determined case ascertainment for each clinical outcome on both databases, by calculating the proportion of cases recorded on the GP record out of the total number of cases recorded on either the GP record or linked mortality record. We calculated the age standardised incidence rates of each outcome based on outcomes recorded on (1) the GP record alone and on (2) the GP record or linked mortality; (3) GP or linked mortality or hospital records. We standardised CPRD rates to the age distribution of the QResearch population in five year bands to ensure comparability.

2.6 Risk factors and missing values

We extracted data from CPRD for all the predictor variables included in one or more of the different algorithms using the same definitions as those used in the original

QResearch studies to enable a direct comparison of the results. We developed a mapping between the Read and medication reference tables to identify the equivalent code in each database. This included the following variables recorded at entry to the cohort:

- **demographics** – age (continuous), sex, ethnicity (9 categories – white, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean ,Black African, Chinese, Other ethnic group), resident in care home, material deprivation (as measured by the Townsend score)
- **clinical values** - smoking status (non-smoker, ex-smoker, light smoker [1-9 cigarettes/day], moderate smoker [10-19 cigarettes/day], heavy smoker [20+ cigarettes/day];body mass index, systolic blood pressure, alcohol consumption-none drinker, trivial (<1u/day), light (1-2u/day), moderate (3-6u/day), heavy (7-9u/day), very heavy (>9 day).
- **laboratory results** –cholesterol/HDL ratio, platelets
- **family history**- family history of osteoporosis or hip fracture in a first degree relative, coronary heart disease in first degree relative under the age of 60 years, diabetes in a first degree relative.
- **chronic diseases** – congestive cardiac failure, atrial fibrillation, coronary heart disease, cardiovascular disease, periperal vascular disease, venous thromboembolism, diabetes, rheumatoid arthritis, SLE, hypertension, renal disease, renal stones, inflammatory bowel disease, dementia, Parkinson's disease, epilepsy, cancer, chronic liver disease or pancreatitis, oesophageal varices, prior haemorrhage, malabsorption endocrine diseases, asthma or COPD, history of falls, prior osteoporotic fracture, varicose vein surgery, emergency admissions or hip surgery in last 6 months.
- **prescribed medication**- anticoagulants, antidepressants, antipsychotics, antiplatelets, oral NSAIDs, tamoxifen, oestrogen containing hormone replacement therapy (BNF chapter 6.4.1.1), systemic corticosteroids, combined oral contraceptive.

The combination of predictor variables required for each risk score varied with the score being validated as shown in Table 1. We used the clinical value recorded closest to the date on which the patient entered the study for body mass index, systolic blood pressure, smoking status, platelets, and total and HDL cholesterol. Patients were considered to be exposed to medication at entry to the cohort if they had at least 2 prescriptions for the relevant medication prescribed prior to the study entry date with the most recent one occurring within 28 days of the study entry date.

2.7 Townsend scores

We used the Townsend score evaluated at output area as a proxy for material deprivation. The CPRD dataset differs from the QResearch dataset in that each patient in the CPRD dataset is allocated to a tenth of deprivation (as measured by the Townsend score) and only the category number is provided. In contrast, each

patient in the QResearch dataset is allocated the individual Townsend score corresponding to their output area of residence (i.e. continuous data). In order to calculate risk scores in the CPRD cohort, we used the median value for each tenth as supplied by CPRD. Patients with missing Townsend scores were excluded from the cohorts.

2.8 Discrimination and calibration statistics

We used chained equations with the ICE procedure in STATA[14] to perform multiple imputation to replace missing values for body mass index, systolic blood pressure, smoking status, alcohol, and total and HDL cholesterol. We created five multiply imputed datasets and used Rubin's rules to combine effect estimates and standard errors to allow for the uncertainty due to imputing missing data[15 16].

We applied the algorithm for each score to eligible patients in the CPRD study cohort to obtain predicted risks for each of the relevant clinical outcomes. We calculated the estimated risk for eligible patients in the CPRD validation dataset over 5 years or 10 years depending on which score was used. We then tested the performance of each score in the CPRD cohort and compared it with the published results from the original QResearch validation cohorts.

In order to assess calibration (i.e. degree of similarity between predicted and observed risks), we calculated the mean predicted risk and the observed risk [17]obtained using the Kaplan-Meier estimate and compared the ratio of the mean predicted risk to the observed risk for patients in the validation cohort in each decile of predicted risk. We calculated the area under the Receiver Operator Curve (ROC) statistic to assess discrimination (i.e. ability of a risk prediction equation to distinguish between those who do and do not have an event during the follow-up period). We also calculated the D statistic[18] and an R squared statistic derived from the D statistic[19] which are measures of discrimination and explained variation appropriate for survival models. The D statistic has been developed as a new measure of discrimination specifically for censored survival data, higher values indicate improved discrimination, and an increase in the D statistic of at least 0.1 indicates an important difference in prognostic separation between different risk classification schemes. The R^2 statistic derived from the D statistic is a measure specific to censored survival data– it measures explained variation in time to the outcome event and higher values indicate more variation is explained[20]. We also repeated the assessment of discrimination by restricting the analysis for each score to patients without missing data for relevant clinical or laboratory measures used in the risk score (ie those with complete data for all predictor variables in the risk score).

We identified the proportion of patients in the CPRD validation cohort who were in the top decile of predicted risk and used this to calculate the sensitivity, specificity and observed risk at this threshold. We used the top decile for comparability across the scores and with previous studies though the choice of threshold for use in clinical practice will depend on the context and cost-effectiveness of relevant interventions. Analyses were conducted using Stata (version 13.1).

2.9 Sample size estimation

There is currently no clear guidance on sample size requirements for studies evaluating the performance (validation) of a multivariable risk score, but a commonly used rule-of-thumb is that it is desirable to seek a dataset with at least 100 patients with the outcome of interest. We used all the available data on the CPRD to maximize the power of the study.

3 Results

3.1 Study populations

The CPRD validation cohort consisted of 3.3 million patients, aged 25-99 years registered at 357 general practices with linked data between 01 Jan 1998 and 31 July 2012. The QResearch cohort consisted of 6.8 million patients from 753 practices with linked data, registered between 01 Jan 1998 and until 31 July 2013. The numbers of patients in each geographical region are shown in Web Extra Table 1.

3.2 Baseline characteristics

Table 2 shows a comparison of the demographic characteristics for the CPRD and QResearch cohorts.

The QResearch population was marginally younger with 34.2% of women and 32.8% of men aged 24-34 years compared with 27.8% and 26.9% for CPRD.

3.2.1 Recording of ethnicity

QResearch had a higher proportion of patients with self-assigned ethnicity recorded compared with CPRD both overall (58.2% vs 38.1%) and in each of the 10 geographical areas within England (web extra table 2). We repeated the analysis restricting information on QResearch to that recorded prior to 31 July 2012 (for comparability with the calendar time available on CPRD). Of the 6,758,649 patients in the QResearch cohort, 3,856,244 (57.1%) had ethnicity recorded prior to this date.

3.2.2 Recording of family history

Recording of a positive family history of coronary heart disease and diabetes was more than twice as high in QResearch compared with CPRD. For example, for family history of coronary heart disease, 11.0% of patients had a value recorded for QResearch compared with 4.6% for CPRD (web extra table 2). Restricting information to that recorded prior to July 2012 for QResearch, then 6,758,649 (10.7%) had a positive family history of coronary heart disease recorded.

3.2.3 *Recording of alcohol and smoking levels*

Recording of alcohol levels was very similar in both QResearch and CPRD. For example, 82.1% of women had alcohol level recorded in both databases. Recording of smoking status was marginally higher in women compared with men in QResearch (93.2% vs 89.1%) and also CPRD (94.8% vs 90.8%).

3.2.4 *Recording of clinical values*

Recording of cholesterol/HDL ratio was marginally higher on QResearch compared with CPRD (40.1% vs 36.0%). Recording of body mass index and systolic blood pressure tended to be marginally higher on CPRD than QResearch. However the mean values for the various clinical values (BMI, systolic blood pressure, serum creatinine and cholesterol/HDL ratio) were extremely similar.

Table 3 shows prescribed medication and clinical diagnoses recorded in patients on or prior to entry to the study cohort. Overall, the prevalence of clinical diagnoses were similar on the two databases with CPRD having marginally higher prescribing rates.

The inclusion and exclusion criteria for each risk score are shown in Table 4 along with the numbers of patients eligible for each analysis on CPRD. For example, there were 3,177,192 patients aged 25-84 years. Of these, 99,189 had existing diabetes at baseline leaving 3,078,003 for the validation of QDiabetes. Table 4 also shows the numbers and percentage out of those eligible for inclusion with complete data for risk factors necessary for calculation of the score which would otherwise need to be imputed (i.e. laboratory or clinical values). The amount of missing data varies substantially between the scores with scores requiring multiple laboratory or clinical values (such as QRISK2) having the lowest levels of completeness.

3.2.5 *Comparison between CPRD linked and unlinked data*

Web extra Table 3 shows characteristics for CPRD cohort with linked data with CPRD cohort without linked data. The CPRD cohort with linked data tended to have higher recorded of ethnicity compared with the CPRD cohort without linked data (38.1% vs 28.4%). Recording of smoking, alcohol, body mass index, systolic blood pressure, cholesterol and platelets were all higher on the CPRD cohort with linked data than those without linked data.

3.3 Incidence rates of clinical outcomes

Table 5 shows the number of incident events for each clinical outcome in women recorded on GP data and those recorded on either GP data or cause specific mortality data for both the CPRD and QResearch cohorts. It also shows the age standardized incidence rates per 1000 person years. Table 6 shows the comparable information for men.

For example, there were 35,617 incident ischaemic stroke or TIA events for women on CPRD. Of these, 32,283 had been identified on the GP record with an additional 3,334 events identified on the linked ONS mortality record. The ascertainment of events on the GP record was therefore 32283/35617 i.e. 90.6%. For QResearch, there were 70,477 incident stroke events recorded on either the GP or linked ONS mortality record of which 63,572 had been identified on the GP record. The ascertainment was therefore 90.2%.

For thromboembolism in women, 91.1% of events recorded on either the GP or linked ONS mortality record on CPRD were identified on the GP record compared with 90.6% for QResearch. Similar results were obtained for men with levels of ascertainment between the two databases being extremely close suggesting similar recording patterns between the two groups of GP practices contributing to each database.

The age standardized incidence rates of events on CPRD tended to be marginally lower than those on QResearch as shown by the ratio of the CPRD rates to those in QResearch (Table 5). For example, the rate ratio for fractured neck of femur in women was 0.94 indicating that CPRD had a 6% lower incidence rate compared with QResearch. The effect was more marked for moderate or severe kidney failure where the incidence rates for CPRD were approximately 25% lower than those for QResearch in women and 16% lower in men.

The age standardized incidence rates of upper gastrointestinal haemorrhage and intracranial haemorrhage among patients prescribed anticoagulants and those not prescribed anticoagulants are shown in Web extra table 4. The rates are similar for CPRD and QResearch.

3.4 Validation statistics

Table 7 shows the discrimination statistics for each score in CPRD in men and women and also the published values from previous validations using QResearch. The validation statistics for each of the risk prediction scores were very similar in the CPRD cohort compared with results from QResearch validation cohorts. For example in women, the QDiabetes algorithm explained 50% of the variation within CPRD compared with 51% on QResearch. The D statistic for women was 2.03 within CPRD compared with 2.08 for QResearch. The ROC value for women was 0.85 on both databases.

Of all the scores, QFracture (fractured neck of femur) had the best performance in men in CPRD with a ROC value of 0.89, R^2 value of 71% and D statistic of 3.17. The corresponding figures for QResearch in men were 0.89, 72% and 3.26.

QThrombosis had the lowest values for women in CPRD with an ROC value of 0.77, R^2 of 34.5 and D statistic of 1.49. The corresponding figures for women in QResearch were 0.75, 33.5 and 1.45.

Figure 1(a-j) compares the mean predicted risks and observed risks for each score across each tenth of predicted risk (1 representing the lowest risk and 10 the highest risk) and demonstrates that the models are generally well calibrated for patients on CPRD.

The QKidney score (moderate or severe kidney failure) showed the observed risk was lower than the predicted risk. This might indicate a degree of over prediction of the score. Alternatively, it could be related to the lower incidence rate of kidney failure observed among women on the CPRD compared with QResearch.

Web extra table 5 presents the ROC, D and R^2 statistic for each score restricted to patients from CPRD with complete recording of laboratory and risk factor data for each score. The results were very similar to the results obtained using multiply imputed dataset for the majority of scores except for QRISK2 and QStroke where values were lower. For example, the results for QFracture (hip fracture) in women on CPRD using multiply imputed data were ROC of 0.89; R^2 of 70.6%; D statistic of 3.17. The corresponding results restricted to women on CPRD with complete data were ROC of 0.90; R^2 of 70.4%; D statistic of 3.16. For QRISK2, the results for women for imputed data on CPRD were ROC of 0.88; R2 of 56.4% ; D statistic of 2.33. The corresponding results for complete data were ROC of 0.79; R2 of 40.9%; D statistic of 1.70.

3.5 Performance for the top decile of risk.

Table 8 shows the sensitivity, specificity and observed risk for patients in the top decile of each score on CPRD. The observed risk is higher than the risk threshold value since this represents the observed risk within the top decile of predicted risk. For example, the cut off for the top tenth of risk for QFracture (fractured neck of femur) was a 10 year risk of 3.7%. At this threshold the sensitivity was 66.5%, specificity 90.4% and observed risk 9.4%. The results are similar to those obtained from QResearch (not shown).

4 Discussion

4.1 Summary of key findings

This is the first external validation of a set of QPrediction scores on the CPRD. It is important since CPRD represents a fully independent sample of patients registered with general practices using a different clinical computer system (Vision system supplied by In Practice Systems) from the QResearch database (which is based on practices using EMIS clinical systems). Practices using the Vision system together with practices using EMIS make up approximately 75% of all the English general practices. The discrimination and calibration statistics for each score were remarkably similar in CPRD to those published from validation cohorts from QResearch. Our paper also provides updated information on a direct comparison between two of the world's largest general practice databases which have both been linked to mortality and second care data.

Before a clinical risk score can be reliably used in clinical practice, evidence is needed that it can successfully predict the intended outcome in groups of patients other than ones used to develop the score but similar to ones in whom the score might be used. Not all risk scores perform well in external samples – this can be due to deficiencies in the design or modelling methods used to derive the algorithm, if the model is over fitted or if there is an important predictor which is absent[21]. Other reasons for poor performance include differences between the setting of patients in the new and derivation samples, differences in how information is recorded and differences in patient characteristics[21]. It is for these reasons, that we have meticulously assembled the CPRD cohort using the same inclusion/exclusion criteria, definitions of predictor and outcome variables as in the original derivation studies. Any differences observed are therefore more likely to be due to capture of information and underlying population characteristics. In this study, we have found marginal differences in incidence rates between QResearch and CPRD and higher rates of recording of family history and ethnicity in QResearch though these have not been large enough to materially affect our results.

4.2 Strengths and limitations

One strength of using CPRD for risk score validation is that the risk score can be assessed using data collected in a similar manner to the data that would be used when the risk score is used in clinical practice. CPRD only had Townsend score for patients recorded for approximately half their practices (unlike QResearch where Townsend score is included for all practices) so we had to limit the validation cohort in CPRD for this analysis to those practices with linked Townsend scores. We undertook a comparison between patients registered with CPRD practices with and without linked data. We found marginally higher recording for ethnicity, smoking, alcohol, clinical values for the CPRD cohort with linked data compared with the unlinked data but similar characteristics for demographics, comorbidities,

medication and clinical values (results not shown) so we have no reason to believe this would have biased our results.

Another strength of general practice databases is the large volume of patients who tend to be representative of the general population. A limitation of routinely collected data is that not all patients will have all clinical and laboratory data recorded leading to missing data values in some of the parameters needed to calculate the risk scores. We have reported performance in all patients using multiple imputation to replace missing values and restricted to patients without missing values and found very similar results for the majority of algorithms tested. There was some degradation of performance for algorithms, particularly for QRISK2 and QStroke, where there were large amounts of missing data. However in clinical practice, the risk scores can be calculated using information recorded during consultation reducing the amount of missing data. Alternatively, the software which implements QPrediction scores includes algorithms which estimate body mass index, systolic blood pressure and cholesterol/HDL ratio. The estimated values can be used where the relevant data is not recorded in order to generate an estimated risk score. Effectively, the software emulates the multiple imputation used in our validation which then gives the results based on multiply imputed data reasonable face validity.

The difficulty of obtaining a comprehensive code list for any given outcome or exposure is a limitation common to all research in primary care databases. We mitigated this by matching our code lists for the CPRD primary analysis to the code lists in the QResearch derivation data set wherever possible. The CPRD database uses the same clinical coding system as QResearch for clinical values (it uses Read version 2). However, there is a third clinical system in use in England (SystmOne) which uses a different coding system known as Clinical terms version 3(CTV3). Whilst there is a mapping between Read codes and CTV3, we have not tested the algorithms on a database using CTV3 in this study so are unable to draw conclusions regarding the generalisability of the results of the validation to practices using this system.

The quality of information on CPRD is likely to be good since previous studies have validated similar outcomes and exposures and found levels of completeness and accuracy to be good[22 23].

4.3 Comparison with other studies

The aim of this study was to validate a collection of QPrediction tools. The details of the derivation and first validation of each prediction tool have been separately published in the peer reviewed literature including information on definitions of predictor variables with supplementary information available on the relevant websites. We haven't duplicated information in the present paper but have provided the relevant links and references.

Our validation results confirm earlier studies undertaken on the THIN database (another general practice database which is derived from the Vision system but which isn't linked to mortality data). These earlier studies include external validations of QRISK2[10 11 24], QDiabetes[12], QFracture[9] and QKidney[25] by an independent team who were not involved in the development of the algorithms. These independent validations have demonstrated similar performance compared with the validations performed by study authors using the QResearch database. This study builds on previous validations by providing new information on the performance of scores not previously validated on an external database (QBleed and QThrombosis) and by utilising the linked data which was not available on the THIN database. Together with the present study (which includes a number of scores not previously tested in an external population), the results provide consistent evidence that these QPrediction scores are likely to provide appropriate estimates of disease risk in contemporary primary care populations in England and to discriminate between patients at different levels of risk with reasonable reliability.

4.4 Comparison of QResearch and CPRD baseline characteristics

Overall, our results show a striking similarity between CPRD and QResearch cohorts for nearly all baseline characteristics. There are two notable exceptions. First, recording of ethnicity was higher in QResearch than CPRD. Second, fewer patients in the CPRD cohort had a recorded family history of diabetes and coronary heart disease in a first degree relative under the age of 60 years. Recording differences in ethnicity and family history were not explained by geographic differences or difference in data capture period between the two databases. Given the similarity for the other risk factors and treatments, it is likely that the difference in ethnicity and family history recording reflects a difference in recording patterns between the two clinical computer systems rather than a true difference between the two cohorts. A similar pattern for recording of ethnicity and family history was also reported in the validation of QRISK on the Health Improvement Network (THIN database) [11 26]. This was thought to be due to different usage of clinical templates in the clinical system, with EMIS practices having ethnicity and family history included more often thereby prompting the user to enter this information in a more systematic fashion.

4.5 Comparison of QResearch and CPRD incidence rates

The age standardised incidence rates for each condition were generally marginally higher on QResearch than CPRD although the proportions of events identified on GP data (out of all events recorded on either GP or linked mortality data) was very close. This suggests that patterns of recording of major clinical events are very similar between QResearch and CPRD although the absolute value varies by clinical condition. For example, 91% of ischaemic stroke events recorded on either GP or linked mortality data are identified on the GP record compared with 99% of hip fractures. We also note the lower levels of total cardiovascular events in the GP

clinical record which was between 13-15% lower than the total recorded on either the GP record, the linked mortality record or the linked hospital admissions record. Some of this will reflect new sudden events where the first presentation was a hospital admission or death whilst others may reflect some under-representation of existing cases not recorded in the GP record. Our study is unable to distinguish between these two scenarios, though the latter one potentially has clinical consequences if the patient is not identified as having cardiovascular disease as they may not be offered secondary prevention.

We think that the information on baseline characteristics and incidence rates will have a utility beyond the present study since it suggests that both databases are fundamentally similar in many aspects and likely to generate similar results for a range of epidemiological studies[27].

4.6 Summary

In summary, we have tested a set of QPrediction scores using an external independent cohort of practices contributing to the CPRD. The results demonstrate good performance, comparable to the results obtained from QResearch, meaning that the findings of studies performed in either database are likely to be applicable in England.

5 Supporting information

Approvals
The project was approved in accordance with the QResearch® agreement with Trent Research Ethics Committee (ref 03/04/021) and approved by the ISAC committee of the CPRD (ref 13_079).

Acknowledgements
We thank EMIS and EMIS practices for their contribution to the QResearch database. We thank CPRD and Vision Practices for allowing access to the CPRD for this study.

Source of funding:
The validation of the QPrediction scores is funded by the National Institute for Health Research's School for Primary Care Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. The funding body did not play a role in the study design, writing of the manuscript or in the decision to submit the manuscript for publication.

Statement
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non-exclusive for government

Contributorship

JHC initiated the study, undertook the literature review, data extraction, data manipulation and primary data analysis and write the first draft of the paper. JHC is the guarantor. CC contributed to the design, analysis, interpretation and drafting of the paper. PB contributed to the development of core ideas, the analysis plan, interpretation of the results and the drafting of the paper.

Competing interests

JHC 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). JHC is also director of ClinRisk Ltd which produces open and closed source software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help improve patient care. CC is associate professor of Medical Statistics at the University of Nottingham and a consultant statistician for ClinRisk Ltd. PB has received financial support for undertaking the validation work from the National School for Primary Care Research. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations. There are no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

All the algorithms validated in this paper are published as open source software under the GNU Lesser Public License. No additional data are available.

Figure 1 Calibration of each QPrediction score comparing the mean predicted risks with the observed risks in the CPRD cohort.

1a QThrombosis (venous thromboembolism)

1b QFracture (hip)

1c QFracture (hip, colles, spine, shoulder)

1d QStroke (ischaemic stroke)

1e QDiabetes (type 2 diabetes)

1f QBleed (upper gastrointestinal haemorrhage)

1g QBleed (intracranial haemorrhage)

1h QKidney (moderate or severe kidney failure)

1i QKidney(severe kidney failure)

1j QRisk2 (cardiovascular disease)

6 References

1. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. BMJ 2008:bmj.39609.449676.25.
2. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. BMJ 2013;**346**.
3. Hippisley-Cox J, Coupland C, Robson J, et al. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. BMJ 2009;**338**:b880-.
4. Hippisley-Cox J, Coupland C. Predicting risk of osteoporotic fracture in men and women in England and Wales: prospective derivation and validation of QFractureScores. BMJ 2009;**339**:b4229-.
5. Hippisley-Cox J, Coupland C. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney(R) Scores. BMC Family Practice 2010;**11**:49.
6. Hippisley-Cox J, Coupland C. Development and validation of risk prediction algorithm (QThrombosis) to estimate future risk of venous thromboembolism: prospective cohort study. BMJ 2011;**343**:d4656.
7. Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions BMJ 2013:(in press).
8. Hippisley-Cox J, Coupland C, Robson J, et al. Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: cohort study using QResearch database. BMJ 2010;**341**:c6624.
9. Collins GS, Mallett S, Altman DG. Predicting risk of osteoporotic and hip fracture in the United Kingdom: prospective independent and external validation of QFractureScores. BMJ 2011;**342**:d3651.
10. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010;**340**:c2442-.
11. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;**339**:b2584-.
12. Collins GS, Altman DG. External validation of the QDScore for predicting the 10-year risk of developing Type 2 diabetes. Diabetic Medicine 2011;**28**:599-607.
13. Hippisley-C J, Coupland C. Predicting risk of upper gastrointestinal bleed and intracranial bleeding with anticoagulants: cohort study to derive and validate the QBleed score. BMJ 2014;**(in press)**.
14. Royston P, White I. Multiple Imputation by Chained Equations (MICE): Implementation in Stata. Journal of statistical software 2011;**45**:1-20.
15. Gray A, Clarke P, Farmer A, et al. Implementing intensive control of blood glucose concentration and blood pressure in type 2 diabetes in England: cost analysis (UKPDS 63). BMJ 2002;**325**(7369):860.
16. Royston P. Multiple imputation of missing values. Stata Journal 2004;**4**(3):227-41.

17. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ 2007:bmj.39261.471806.55.
18. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;**23**:723-48.
19. Royston P. Explained variation for survival models. Stata J 2006;**6**:1-14.
20. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol 2013;**13**:33.
21. Altman DG, Vergouwe Y, Royston P, et al. Prognosis and prognostic research: validating a prognostic model. BMJ 2009;**338**(may28_1):b605-.
22. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;**69**:4-14.
23. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010;**60**:e128-36.
24. Collins GS, Altman A. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;**344**.
25. Collins GS, Altman D. Predicting the risk of chronic kidney disease in the UK: an evaluation of QKidney scores using a primary care database. British Journal of General Practice 2012;**62**(597):e243-e50.
26. Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 2008;**94**:34-39.
27. Reeves D, Springate DA, Ashcroft DM, et al. Can analyses of electronic patient records be independently and externally validated? The effect of statins on the mortality of patients with ischaemic heart disease: a cohort study with nested case–control analysis. BMJ Open 2014;**4**(4).
28. Hippisley-Cox J, Coupland C. QRISK2-2014 Annual Update Information, 2014:5.
29. Hippisley-Cox J, Coupland C. Derivation and validation of updated QFracture algorithm to predict risk of osteoporotic fracture in primary care in the United Kingdom: prospective open cohort study. BMJ 2012;**344**(may22 1):e3427-e27.
30. Hippisley-Cox J CC. QDiabetes 2013 Annual Update Information Secondary QDiabetes 2013 Annual Update Information 2013. http://qdiabetes.org/QDiabetes-2013-Annual-Update-Information.pdf.

Table 1 Summary of QPrediction scores including outcome and predictor variables

Score	Weblink∞	Outcome	Predictors
QDiabetes[3]	www.qdiabetes.org	10 year risk of type 2 diabetes±	**In men and women:** Age, sex, ethnicity, deprivation, smoking, family history of diabetes, cardiovascular disease, treated hypertension, steroid tables, body mass index
QRISK2[28]	www.qrisk.org	10 year risk of CVD **	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation.
QStroke[2]	www.qstroke.org	10 year risk of ischaemic stroke or TIA±	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, cholesterol/HDL ratio, body mass index, family history of cardiovascular disease in first degree relative under 60 years, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, congestive cardiac failure, valvular heart disease
QKidney[5]	www.qkidney.org	5 year risk of moderate or severe kidney failureμ	**In men and women:** Age, smoking status, ethnic group (9 categories), systolic blood pressure, body mass index, family history of kidney disease, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease, type 2 diabetes, atrial fibrillation, SLE, peripheral vascular disease, kidney stones, NSAIDs
QThrombosis	www.qthrombosis.org	5 year risk of venous thromboembolism±	**In men and women**: age, body mass index, smoking status, varicose veins, congestive cardiac failure, chronic renal disease, cancer, chronic obstructive pulmonary disease, inflammatory bowel disease, hospital admission in past six months, and current prescriptions for antipsychotic drugs. Additionally in women: combined oral contraceptives, tamoxifen, and hormone replacement therapy
QBleed[13]	www.qbleed.org	5 year risk of upper gastrointestinal bleed in patient starting anticoagulants vs others*	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; venous thromboembolism; congestive cardiac failure; treated hypertension; cancer; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QBleed[13]	www.qbleed.org	5 year risk of intracranial bleed in patient starting anticoagulants vs others *	**In men and women** age; body mass index; Townsend score; smoking status; ethnicity; alcohol intake; prior bleed; oesophageal varices; chronic liver disease or pancreatitis; atrial fibrillation; treated hypertension; recent abnormal platelets (< 150μL or >480μL); new use of anticoagulants; current prescriptions for anti-platelets; NSAIDS; corticosteroids; antidepressants; anticonvulsants (phenytoin or carbamazepine)
QFracture[29]	www.qfracture.org	10 year risk of hip fracture± 10 year risk of	**In women:** HRT usage, age, body mass index, smoking status, recorded alcohol use, parental history of osteoporosis, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls, menopausal symptoms, chronic liver disease, gastrointestinal malabsorption and other endocrine disorders.

		osteoporotic fracture [µ]	**In men:** age, body mass index, smoking status, recorded alcohol use, rheumatoid arthritis, cardiovascular disease, type 2 diabetes, asthma, tricyclic antidepressants, corticosteroids, history of falls and liver disease.

∞the web link has the relevant calculator, links to academic papers, additional information including links to the open source software

[±] recorded either on GP record or linked ONS mortality record;

µ recorded on the GP record.

*Recorded either on linked hospital admissions record or ONS mortality record

**Recorded either on linked hospital admissions record or ONS mortality or linked hospital admissions record

Table 2 Comparison of baseline characteristics of patients in CPRD validation cohort and QResearch comparison cohort

	CPRD	*CPRD*	*QResearch*	*QResearch*
	Men (%)	**Women (%)**	**Men (%)**	**Women (%)**
total	1,588,803	1,682,709	3,304,145	3,454,504
Ageband				
25-34 years	427975 (26.9)	467192 (27.8)	1083589 (32.8)	1179742 (34.2)
35-44 years	396680 (25.0)	364150 (21.6)	814988 (24.7)	731089 (21.2)
45-54 years	294274 (18.5)	277663 (16.5)	558553 (16.9)	516188 (14.9)
55-64 years	212817 (13.4)	211636 (12.6)	390229 (11.8)	389266 (11.3)
65-74 years	148180 (9.3)	164172 (9.8)	267997 (8.1)	298847 (8.7)
75+ years	108877 (6.9)	197896 (11.8)	188789 (5.7)	339372 (9.8)
mean Townsend score (SD)	-.5 (3.2)	-.5 (3.2)	.3 (3.6)	.2 (3.6)
Care home resident	1407 (0.1)	3466 (0.2)	2983 (0.1)	7411 (0.2)
Ethnicity recorded	587879 (37.0)	658835 (39.2)	1859462 (56.3)	2077181 (60.1)
White or not recorded	1515113 (95.4)	1602212 (95.2)	3010061 (91.1)	3149618 (91.2)
Indian	16442 (1.0)	16025 (1.0)	56156 (1.7)	50406 (1.5)
Pakistani	6606 (0.4)	6146 (0.4)	30632 (0.9)	23405 (0.7)
Bangladeshi	2419 (0.2)	1688 (0.1)	23017 (0.7)	17450 (0.5)
Other Asian	10795 (0.7)	11873 (0.7)	32513 (1.0)	36886 (1.1)
Caribbean	4989 (0.3)	6425 (0.4)	25782 (0.8)	32953 (1.0)
Black African	12883 (0.8)	14771 (0.9)	51980 (1.6)	56528 (1.6)
Chinese	2914 (0.2)	4176 (0.2)	16084 (0.5)	23043 (0.7)
Other ethnic group	16642 (1.0)	19393 (1.2)	57920 (1.8)	64215 (1.9)
Smoking status recorded	1442088 (90.8)	1595538 (94.8)	2943405 (89.1)	3219598 (93.2)
Non smoker	613833 (38.6)	834721 (49.6)	1449694 (43.9)	1973691 (57.1)
Ex-smoker	252873 (15.9)	222615 (13.2)	611837 (18.5)	545125 (15.8)
Light smoker (1-9/day)	104466 (6.6)	109864 (6.5)	472614 (14.3)	384482 (11.1)
Moderate smoker (10-19/day)	183000 (11.5)	179391 (10.7)	223631 (6.8)	202776 (5.9)
Heavy smoker (20+/day)	142438 (9.0)	87474 (5.2)	185629 (5.6)	113524 (3.3)
Smoker amount not recorded	145478 (9.2)	161473 (9.6)	0 (0.0)	0 (0.0)
Alcohol status recorded	1238110 (77.9)	1379002 (82.0)	2584335 (78.2)	2834426 (82.1)
Non drinker	163633 (10.3)	318880 (19.0)	583752 (17.7)	1035692 (30.0)
Trivial <1u/day	460091 (29.0)	726851 (43.2)	782985 (23.7)	1144469 (33.1)
Light 1-2u/day	411261 (25.9)	290547 (17.3)	481674 (14.6)	402750 (11.7)
Moderate 3-6/day	166328 (10.5)	36763 (2.2)	648549 (19.6)	237679 (6.9)
Heavy 7-9u/day	19612 (1.2)	2853 (0.2)	54083 (1.6)	7152 (0.2)
Very Heavy >/day	17185 (1.1)	3108 (0.2)	24468 (0.7)	5195 (0.2)
Family History				
family history of CHD	68805 (4.3)	80985 (4.8)	326995 (9.9)	417537 (12.1)
family history of diabetes	96810 (6.1)	132390 (7.9)	357109 (10.8)	487397 (14.1)
family history osteoporosis	880 (0.1)	10062 (0.6)	1655 (0.1)	17529 (0.5)
family history kidney disease	1253 (0.1)	1586 (0.1)	2034 (0.1)	2769 (0.1)

Clinical Values recorded				
BMI recorded	1268235 (79.8)	1481918 (88.1)	2553514 (77.3)	2857742 (82.7)
mean BMI (SD)	29.6 (6.8)	28.2 (7.0)	29.4 (6.8)	28.2 (7.0)
SBP recorded	1359560 (85.6)	1590226 (94.5)	2755733 (83.4)	3190390 (92.4)
mean SBP(SD)	133.1 (23.6)	128.6 (22.7)	132.2 (18.3)	127.1 (20.8)
cholesterol/HDL ratio recorded	587865 (37.0)	606035 (36.0)	1323503 (40.1)	1368180 (39.6)
mean cholesterol ratio (SD)	4.4 (1.4)	3.7 (1.2)	4.4 (1.4)	3.7 (1.2)
platelets recorded	223461 (14.1)	382799 (22.7)	478596 (14.5)	829702 (24.0)
platelets < 150 or > 480	11051 (0.7)	13282 (0.8)	23479 (0.7)	27009 (0.8)
Creatinine recorded	811779 (51.1)	997118 (59.3)	1714337 (51.9)	2053036 (59.4)
Mean creatinine (SD)	96.7 (32.5)	79.7 (24.1)	95.5 (30.7)	78 (23.7)

Table 3 Prescribed medication and clinical diagnoses recorded at baseline in CPRD validation cohort and QResearch comparison cohort

	CPRD men (%)	*CPRD women (%)*	*QResearch men (%)*	*QResearch women (%)*
Prescribed medication				
anticoagulants	15955 (1.0)	13077 (0.8)	27024 (0.8)	22178 (0.6)
antidepressants	101553 (6.4)	235797 (14.0)	178532 (5.4)	398018 (11.5)
antipsychotics	33884 (2.1)	79514 (4.7)	47464 (1.4)	92307 (2.7)
antiplatelets	97475 (6.1)	92816 (5.5)	160910 (4.9)	153405 (4.4)
oral NSAIDs	246515 (15.5)	346416 (20.6)	396026(12.0)	556644 (16.1)
tamoxifen	n/a	9231 (0.5)	n/a	18343 (0.5)
oestrogen only Hormone replacement therapy	n/a	119373 (7.1)	n/a	208333 (6.0)
oral corticosteroids	45597 (2.9)	71352 (4.2)	54354 (1.6)	88205 (2.6)
oral contraceptive pill	n/a	174287 (10.4)	n/a	332696 (9.6)
Recorded Diagnoses				
congestive cardiac failure	15836 (1.0)	19707 (1.2)	24965 (0.8)	28852 (0.8)
atrial fibrillation	20125 (1.3)	20102 (1.2)	33499 (1.0)	32580 (0.9)
coronary heart disease	80377 (5.1)	57703 (3.4)	130220 (3.9)	88606 (2.6)
cardiovascular disease	101430 (6.4)	83167 (4.9)	165495 (5.0)	130214 (3.8)
peripheral vascular disease	17029 (1.1)	13101 (0.8)	25004 (0.8)	17078 (0.5)
venous thromboembolism	15072 (0.9)	23090 (1.4)	27086 (0.8)	40813 (1.2)
rheumatoid or SLE	7455 (0.5)	19010 (1.1)	21453 (0.6)	48447 (1.4)
rheumatoid arthritis	7243 (0.5)	17468 (1.0)	21142 (0.6)	45542 (1.3)
SLE	228 (0.0)	1756 (0.1)	351 (0.0)	3374 (0.1)
type 1 diabetes	6238 (0.4)	4924 (0.3)	12029 (0.4)	9612 (0.3)
type 2 diabetes	51634 (3.2)	43271 (2.6)	95401 (2.9)	79654 (2.3)
treated hypertension	123584 (7.8)	161709 (9.6)	210516 (6.4)	267076 (7.7)
chronic renal disease	3968 (0.2)	4082 (0.2)	8550 (0.3)	8995 (0.3)
moderate/severe kidney failure	14107 (0.9)	9500 (0.6)	30407 (0.9)	21509 (0.6)
severe kidney failure	1603 (0.1)	1125 (0.1)	3641 (0.1)	2672 (0.1)
renal stones	13415 (0.8)	6443 (0.4)	37422 (1.1)	29204 (0.8)
inflammatory bowel disease	8962 (0.6)	10208 (0.6)	17762 (0.5)	19502 (0.6)
dementia	6686 (0.4)	16634 (1.0)	12872 (0.4)	30497 (0.9)
parkinsons disease	4546 (0.3)	4676 (0.3)	6830 (0.2)	6611 (0.2)
epilepsy or anticonvulsants	47170 (3.0)	71171 (4.2)	56516 (1.7)	61561 (1.8)
cancer	26866 (1.7)	43908 (2.6)	51649 (1.6)	79326 (2.3)
liver disease	3959 (0.2)	2893 (0.2)	9947 (0.3)	6410 (0.2)
chronic liver disease or pancreatitis	5521 (0.3)	4051 (0.2)	13069 (0.4)	8729 (0.3)
oesophageal varices	469 (0.0)	340 (0.0)	1626 (0.0)	1388 (0.0)
prior haemorrhage	97562 (6.1)	79765 (4.7)	203278 (6.2)	147533 (4.3)
malabsorption	7343 (0.5)	9375 (0.6)	21042 (0.6)	26002 (0.8)
endocrine diseases	3082 (0.2)	14097 (0.8)	6026 (0.2)	27731 (0.8)
COPD	24029 (1.5)	20737 (1.2)	41281 (1.2)	34785 (1.0)
asthma or COPD	142974 (9.0)	169503 (10.1)	273768 (8.3)	310027 (9.0)
history of falls	28878 (1.8)	61905 (3.7)	34584 (1.0)	67465 (2.0)
prior fracture	24265 (1.5)	45752 (2.7)	62092 (1.9)	89000 (2.6)

varicose vein surgery	18979	(1.2)	47012	(2.8)	35651	(1.1)	85602	(2.5)
emergency admissions or hip op	3483	(0.2)	5266	(0.3)	3335	(0.1)	5508	(0.2)

Table 4 Numbers of patients eligible for each score in the CPRD validation cohort and number of patients with complete risk factor recording not requiring multiple imputation.

Risk Score	*Clinical outcome*	*Eligible age range*	*exclusion criteria at study entry*	*total in age range*	*total with exclusions*	*total eligible for analysis*	*Total complete data*	*% complete data*
QDiabetes	Type 2 diabetes	25-84	type 1 or 2 diabetes at study entry	3,177,192	99,189	3,078,003	2,467,642	80.2
QStroke	ischaemic stroke	25-84	existing stroke or anticoagulants at study entry	3,177,192	70,961	3,106,231	1,032,184	33.2
QRISK2	cardiovascular disease	25-84	existing CVD or statins at study entry	3,177,192	232,722	2,944,470	906,781	30.8
QThrombosis	thromboembolism	25-84	existing VTE or anticoagulants at study entry	3,177,192	53,904	3,123,288	2,513,347	80.5
QFracture	fractured neck of femur	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QFracture	osteoporotic fracture	30-99	none except age	2,852,381	0	2,852,381	2,087,149	73.2
QKidney	moderate or severe kidney failure	35-74	existing moderate or severe kidney failure	2,069,572	10,518	2,059,054	1,146,619	55.7
QKidney	severe kidney failure	35-74	existing severe kidney failure	2,069,572	1,930	2,067,642	1,153,979	55.8
QBleed	upper gastro-intestinal bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0
QBleed	intracranial bleed*	25-99	anticoagulants in 180 days prior to study entry	2,429,696	35,283	2,394,413	1,890,804	79.0

*entry date was 01.01.1998 except for upper GI bleed and intracranial bleed where entry date was 01.01.2007

Table 5 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in women

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	48,143	99.88	4.13 (4.10 to 4.17)	102,544	99.93	4.33 (4.31 to 4.36)	0.95
	GP or ONS	48,203	n/a	4.13 (4.10 to 4.17)	102,618	n/a	4.34 (4.31 to 4.36)	0.95
Ischaemic stroke	GP data	32,283	90.64	2.45 (2.42 to 2.48)	63,582	90.22	2.45 (2.44 to 2.47)	1.00
	GP or ONS	35,617	n/a	2.62 (2.59 to 2.64)	70,477	n/a	2.70 (2.68 to 2.72)	0.97
Cardiovascular disease	GP data	55,833	85.71	5.41 (5.37 to 5.46)	107,412	84.96	4.32 (4.30 to 4.35)	1.25
	GP or ONS	65,143	n/a	6.32 (6.27 to 6.37)	126,433	n/a	5.03 (5.01 to 5.06)	1.26
	GP or ONS or HES	69,202	n/a	6.72 (6.67 to 6.77)	140,510	n/a	5.63 (5.60 to 5.66)	1.19
Thromboembolism	GP data	18,199	91.1	1.52 (1.49 to 1.54)	35,971	90.55	1.46 (1.44 to 1.47)	1.04
	GP or ONS	19,978	n/a	1.64 (1.62 to 1.67)	39,727	n/a	1.60 (1.58 to 1.62)	1.03
Fractured neck of femur	GP data	17,529	99.98	1.32 (1.30 to 1.34)	34,821	99.99	1.40 (1.39 to 1.42)	0.94
	GP or ONS	17,533	n/a	1.32 (1.30 to 1.34)	34,825	n/a	1.40 (1.39 to 1.42)	0.94
Osteoporotic fracture	GP data	34,528	n/a	2.89 (2.58 to 3.20)	81,334	n/a	3.63 (3.61 to 3.66)	0.80
mod /severe kidney failure	GP data	19,902	n/a	2.06 (1.76 to 2.36)	48,665	n/a	2.81 (2.78 to 2.83)	0.73
severe kidney failure	GP data	1,737	n/a	0.18 (0.09 to 0.27)	4,150	n/a	0.24 (0.24 to 0.25)	0.74

Table 6 comparison of age standardised incidence rates (95%CI) per 1000 person years for outcomes on CPRD vs QResearch database in men

	CPRD				*QResearch*			
outcome	**Source for case identification**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95% CI)**	**cases**	**% ascertainment**	standardised **rates per 1000 pyrs (95%CI)**	**ratio of CPRD to QR** standardised **rate**
Type 2 diabetes	GP data	60,731	99.92	5.84 (5.79 to 5.89)	128,234	99.94	5.97 (5.94 to 6.00)	0.98
	GP or ONS	60,782	n/a	5.84 (5.80 to 5.89)	128,317	n/a	5.98 (5.94 to 6.01)	0.98
ischaemic stroke	GP data	32,223	93.55	3.17 (3.14 to 3.20)	63,480	92.85	3.10 (3.08 to 3.13)	1.02
	GP or ONS	34,443	n/a	3.33 (3.30 to 3.37)	68,366	n/a	3.37 (3.34 to 3.40)	0.99
Cardiovascular disease	GP data	70,283	86.7	7.38 (7.33 to 7.44)	137,136	86.12	7.12 (7.08 to 7.16)	1.03
	GP or ONS	81,068	n/a	8.52 (8.46 to 8.58)	159,240	n/a	8.37 (8.33 to 8.41)	1.02
	GP or ONS or HES	84,620	n/a	8.90 (8.84 to 8.96)	174,405	n/a	9.17 (9.13 to 9.21)	0.97
thromboembolism	GP data	15,655	92.32	1.49 (1.46 to 1.51)	31,503	92.22	1.44 (1.43 to 1.46)	1.03
	GP or ONS	16,958	n/a	1.61 (1.59 to 1.63)	34,161	n/a	1.57 (1.56 to 1.59)	1.02
fractured neck of femur	GP data	5,706	99.98	0.65 (0.63 to 0.67)	12,435	99.98	0.71 (0.70 to 0.73)	0.91
	GP or ONS	5,707	n/a	0.65 (0.63 to 0.67)	12,438	n/a	0.71 (0.70 to 0.73)	0.91
osteoporotic fracture	GP data	11,169	n/a	1.29 (1.05 to 1.52)	28,555	n/a	1.54 (1.52 to 1.55)	0.84
Mod/severe kidney failure	GP data	37,597	n/a	4.88 (4.37 to 5.38)	86,649	n/a	5.82 (5.78 to 5.85)	0.84
severe kidney failure	GP data	3,472	n/a	0.54 (0.38 to 0.71)	7,372	n/a	0.47 (0.46 to 0.48)	1.15

Table 7 Performance of QPrediction scores on the CPRD validation cohort compared with published results for the QResearch validation cohort

		CPRD	*CPRD*	*QResearch*	*QResearch*
		women	**men**	**women**	**men**
	statistic	**mean (95%CI)**	**mean (95%CI)**	**mean (95%CI)**	**mean (95%CI)**
QDiabetes 2013 (type 2 diabetes)[30]	ROC	0.846 (0.844 to 0.848)	0.818 (0.816 to 0.82)	0.853 (0.851 to 0.856)	0.837 (0.835 to 0.840)
	R^2 (%)	49.6 (49.2 to 50.1)	45.7 (45.3 to 46.2)	50.8 (50.3 to 51.4)	48.1 (47.6 to 48.6)
	D statistic	2.032 (2.015 to 2.049)	1.879 (1.863 to 1.895)	2.081 (2.058 to 2.104)	1.971 (1.951 to 1.991)
QKidney -2010[5] (moderate or severe kidney failure)	ROC	0.875 (0.87 to 0.879)	0.88 (0.878 to 0.883)	0.877 (0.873 to 0.880)	0.878 (0.874 to 0.882)
	R^2 (%)	58.3 (57.8 to 58.7)	57.5 (57.1 to 57.8)	56.45 (55.40 to 57.50)	58.29 (55.31 to 61.26)
	D statistic	2.418 (2.394 to 2.442)	2.379 (2.361 to 2.397)	2.33 (2.28 to 2.40)	2.42 (2.28 to 2.56)
QKidney -2010 (severe kidney failure)[5]	ROC	0.839 (0.822 to 0.855)	0.851 (0.84 to 0.862)	0.843 (0.825 to 0.860)	0.846 (0.829 to 0.862)
	R^2 (%)	51.4 (49.5 to 53.2)	53.8 (52.6 to 55.1)	55.39 (52.59 to 58.18)	56.65 (53.94 to 59.35)
	D statistic	2.103 (2.025 to 2.182)	2.21 (2.154 to 2.266)	2.28 (2.15 to 2.41)	2.34 (2.21 to 2.47)
QRISK2-2014[28] (cardiovascular disease)	ROC	0.883 (0.882 to 0.884)	0.859 (0.858 to 0.861)	0.892 (0.892 to 0.895)	0.871 (0.869 to 0.873)
	R^2 (%)	56.4 (56.1 to 56.7)	50.9 (50.6 to 51.2)	58.8 (58.4 to 59.1)	53.3 (52.9 to 53.7)
	D statistic	2.328 (2.313 to 2.343)	2.085 (2.071 to 2.098)	2.443 (2.423 to 2.463)	2.188 (2.171 to 2.205)
QStroke-2013[2] (ischaemic stroke or TIA)	ROC	0.882 (0.88 to 0.883)	0.869 (0.867 to 0.87)	0.877 (0.875 to 0.879)	0.866 (0.864 to 0.868)
	R^2 (%)	58.4 (58.1 to 58.8)	55.3 (54.9 to 55.7)	57.3 (56.8 to 57.8)	55.1 (54.6 to 55.7)
	D statistic	2.427 (2.408 to 2.446)	2.278 (2.259 to 2.297)	2.37 (2.35 to 2.40)	2.27 (2.24 to 2.30)
QThrombosis-2010[6] (venous thromboembolism)	ROC	0.756 (0.751 to 0.761)	0.765 (0.760 to 0.770)	0.75 (0.74 to 0.76)	0.75 (0.74 to 0.76)
	R^2 (%)	35.3 (34.5 to 36.1)	34.5 (33.7 to 35.4)	32.78 (31.08 to 34.48)	33.51 (31.71 to 35.30)

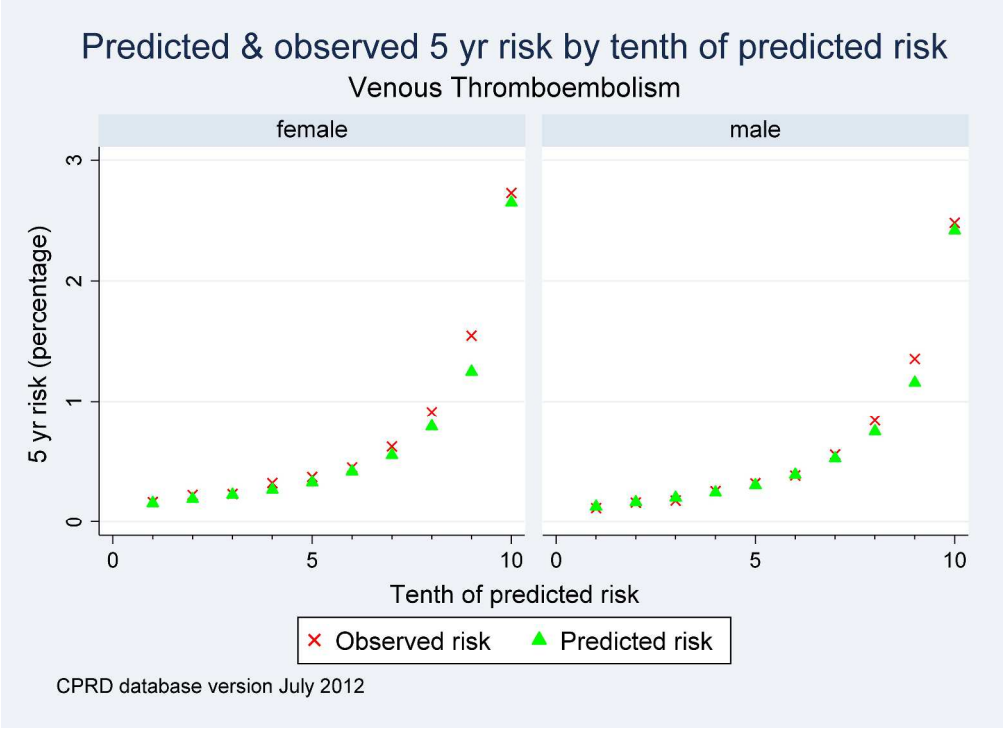
	D statistic	1.512 (1.485 to 1.538)	1.486 (1.458 to 1.513)	1.43 (1.37 to 1.49)	1.45 (1.39 to 1.51)
QBleed-2014[13] (upper gastrointestinal bleed)	ROC statistic	0.775 (0.770 to 0.781)	0.759 (0.753 to 0.764)	0.766 (0.758 to 0.775)	0.747 (0.738 to 0.756)
	R^2 (%)	44.7 (43.6 to 45.9)	41.6 (40.5 to 42.8)	40.7 (38.9 to 42.6)	36.9 (35.1 to 38.7)
	D statistic	1.842 (1.798 to 1.885)	1.729 (1.687 to 1.771)	1.70 (1.63 to 1.76)	1.57 (1.51 to 1.63)
QBleed-2014[13] (intracranial bleed)	ROC statistic	0.808 (0.801 to 0.816)	0.789 (0.780 to 0.797)	0.847 (0.838 to 0.856)	0.812 (0.80 to 0.824)
	R^2 (%)	51.7 (50.1 to 53.3)	50.0 (48.3 to 51.7)	58.0 (56.0 to 60.0)	53.3 (51.1 to 55.4)
	D statistic	2.118 (2.051 to 2.186)	2.046 (1.977 to 2.116)	2.40 (2.30 to 2.50)	2.19 (2.09 to 2.28)
QFracture-2012[29] (fractured neck of femur)	ROC	0.89 (0.888 to 0.892)	0.872 (0.867 to 0.877)	0.893 (0.890 to 0.896)	0.875 (0.868 to 0.883)
	R^2 (%)	70.6 (70.2 to 71)	69.2 (68.5 to 70)	71.73 (71.10 to 72.30)	70.37 (69.25 to 71.49)
	D statistic	3.171 (3.139 to 3.203)	3.07 (3.016 to 3.124)	3.26 (3.21 to 3.31)	3.15 (3.06 to 3.24)
QFracture -2012[29] (osteoporotic fracture: hip, spine, wrist,humerus)	ROC	0.817 (0.814 to 0.819)	0.768 (0.763 to 0.773)	0.790 (0.787 to 0.793)	0.711 (0.703 to 0.719)
	R^2 (%)	56.3 (55.8 to 56.7)	49.8 (48.9 to 50.7)	51.9 (51.2 to 52.6)	38.20 (36.89 to 39.57)
	D statistic	2.322 (2.301 to 2.343)	2.038 (2.002 to 2.075)	2.13 (2.10 to 2.15)	1.61 (1.56 to 1.66)

Notes on understanding validation statistics:

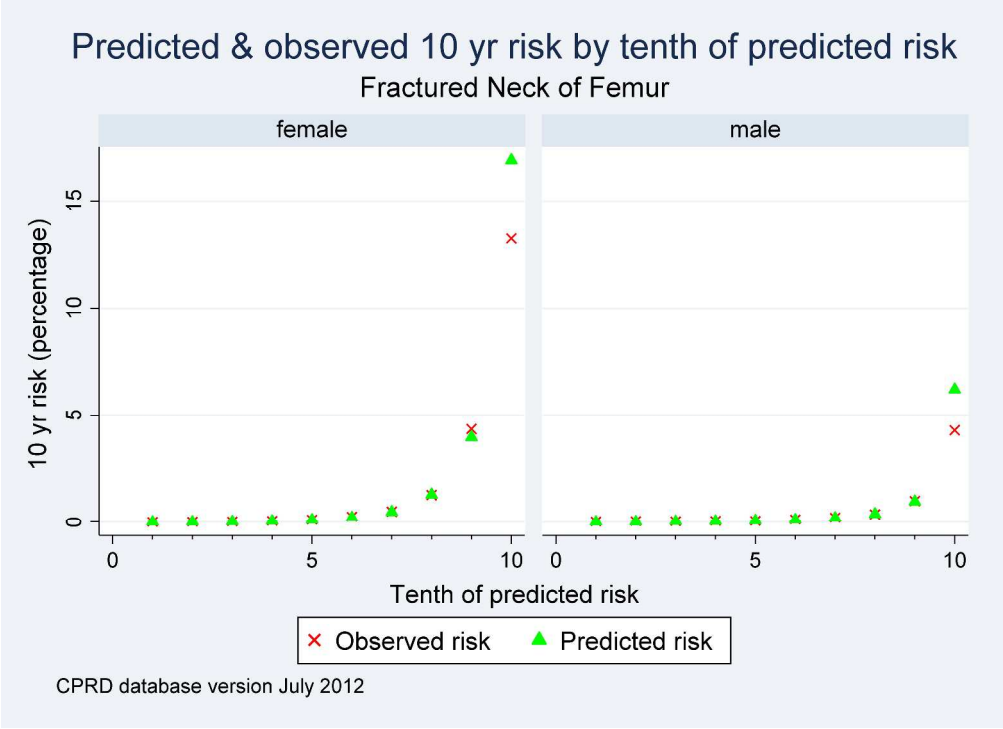
D statistic is a measure of discrimination - higher values indicate better discrimination
ROC statistic is a measure of discrimination - higher values indicate better discrimination
The R^2 statistic is a measure of explained variation - higher values indicate more variation is explained

Table 8 Performance of each score for predicting the relevant outcome in the CPRD validation cohort. The cut off is the threshold of predicted risk for the top decile in the CPRD cohort.

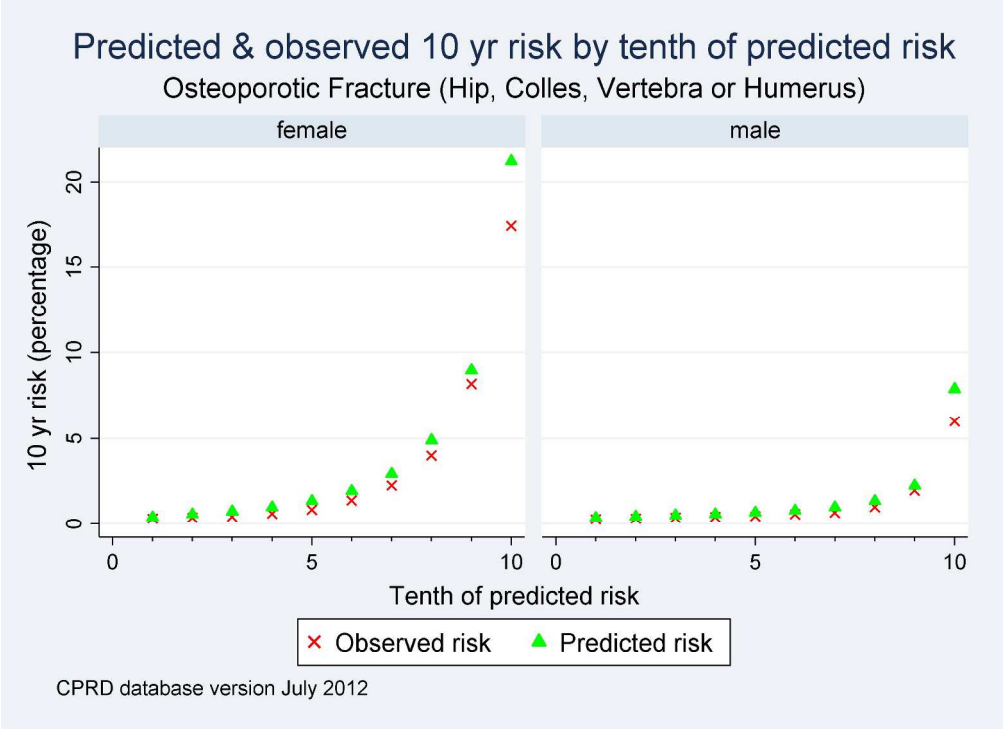
score	*outcome*	*duration*	*cut off (%) for top decile predicted risk*	*Sensitivity (%)*	*Specificity (%)*	*observed risk (%)*
QDiabetes	Type 2 diabetes	10 yr risk	13.0	44.8	91.0	20.8
QStroke	Ischaemic stroke	10 yr risk	10.5	54.7	90.8	16.1
QRISK2	cardiovascular disease	10 yr risk	20.7	49.9	91.9	31.8
QThrombosis	venous thromboembolism	5 yr risk	1.5	36.2	90.1	2.6
QKidney	moderate-severe kidney failure	5 yr risk	6.3	59.1	90.5	6.9
QKidney	severe kidney failure	5 yr risk	0.4	58.5	90.0	0.7
QBleed	upper GI bleed	5 yr risk	1.6	38.0	90.2	3.5
QBleed	intracranial bleed	5 yr risk	0.9	44.2	90.1	1.6
QFracture	fractured neck of femur	10 yr risk	3.7	66.5	90.4	9.4
QFracture	osteoporotic fracture	10 yr risk	7.8	49.6	90.5	13.1



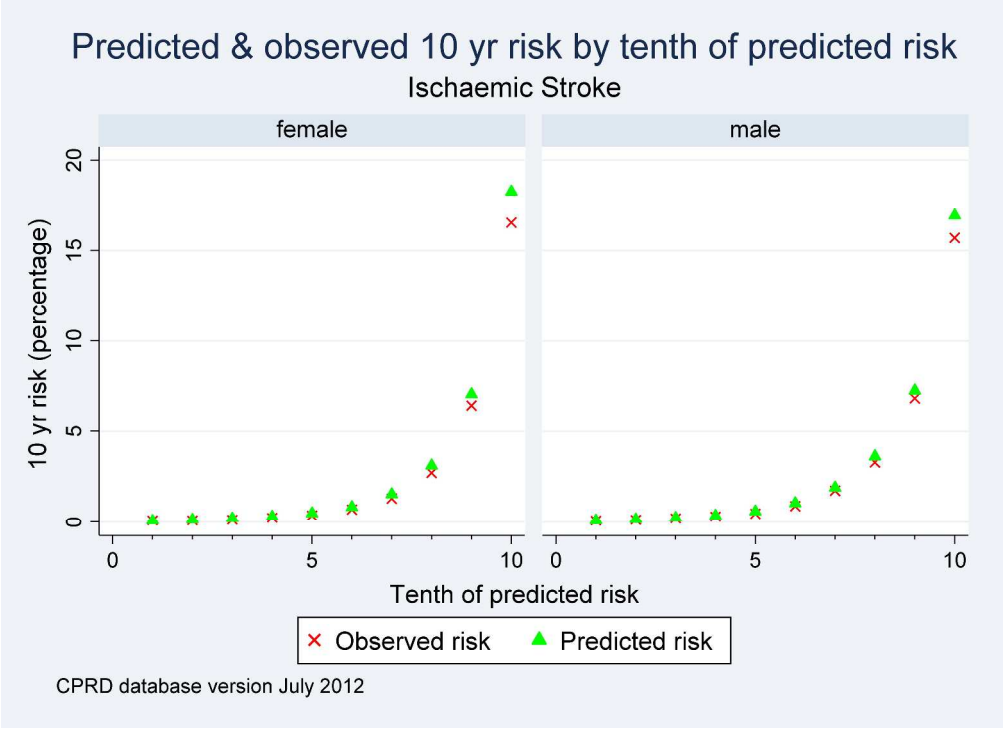
QThrombosis (venous thromboembolism)



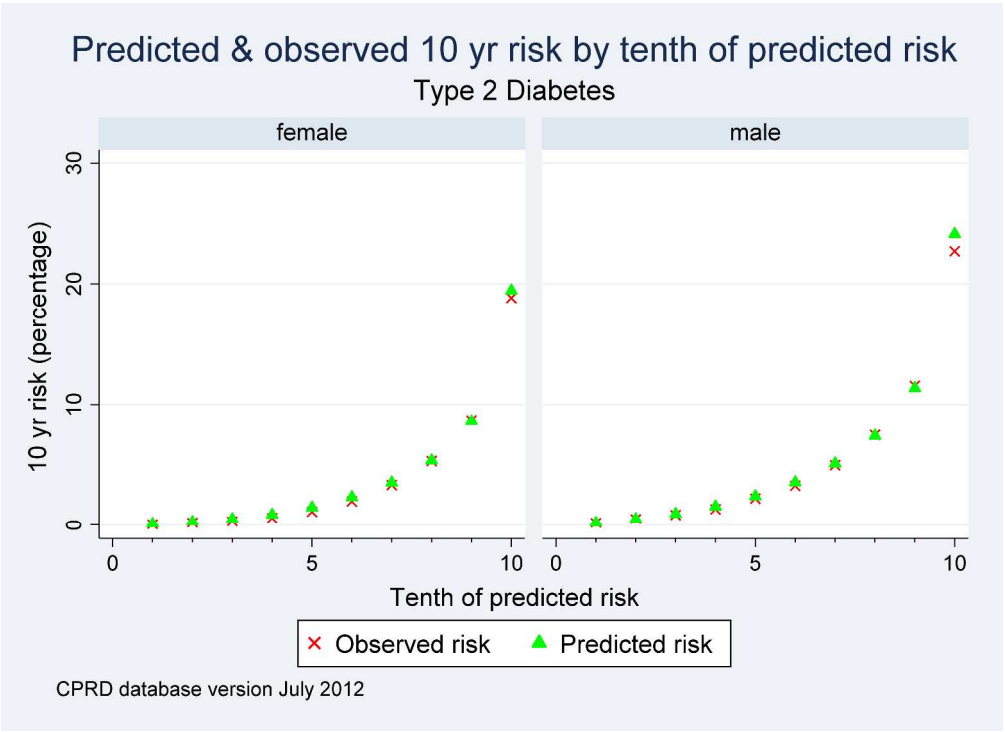
QFracture (hip)



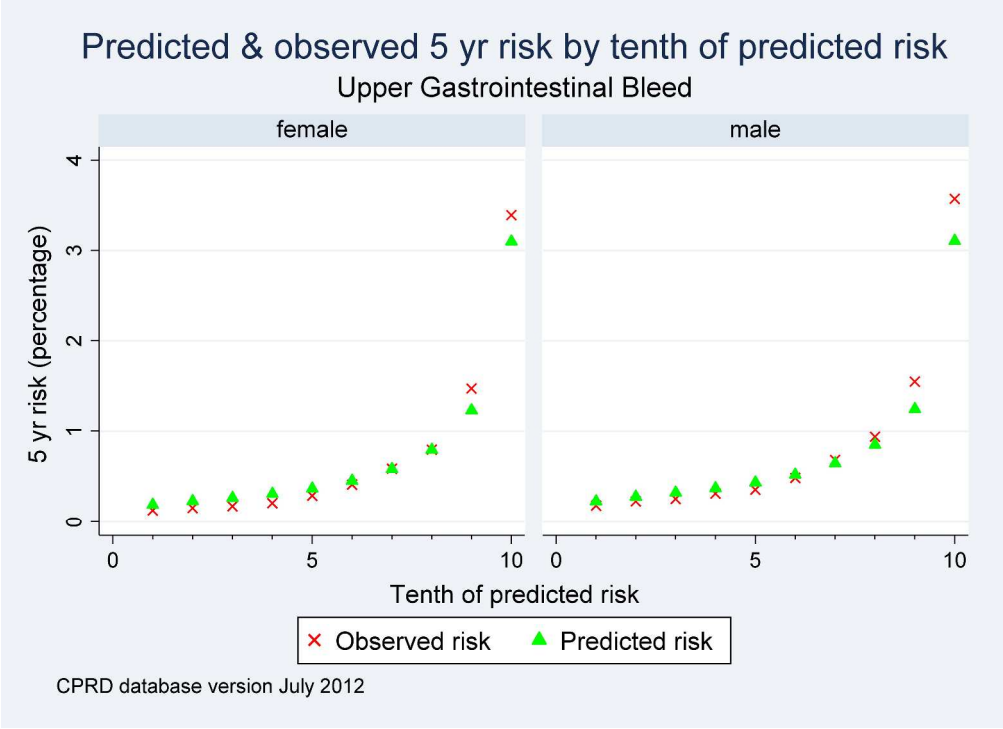
QFracture (hip, colles, spine, shoulder)



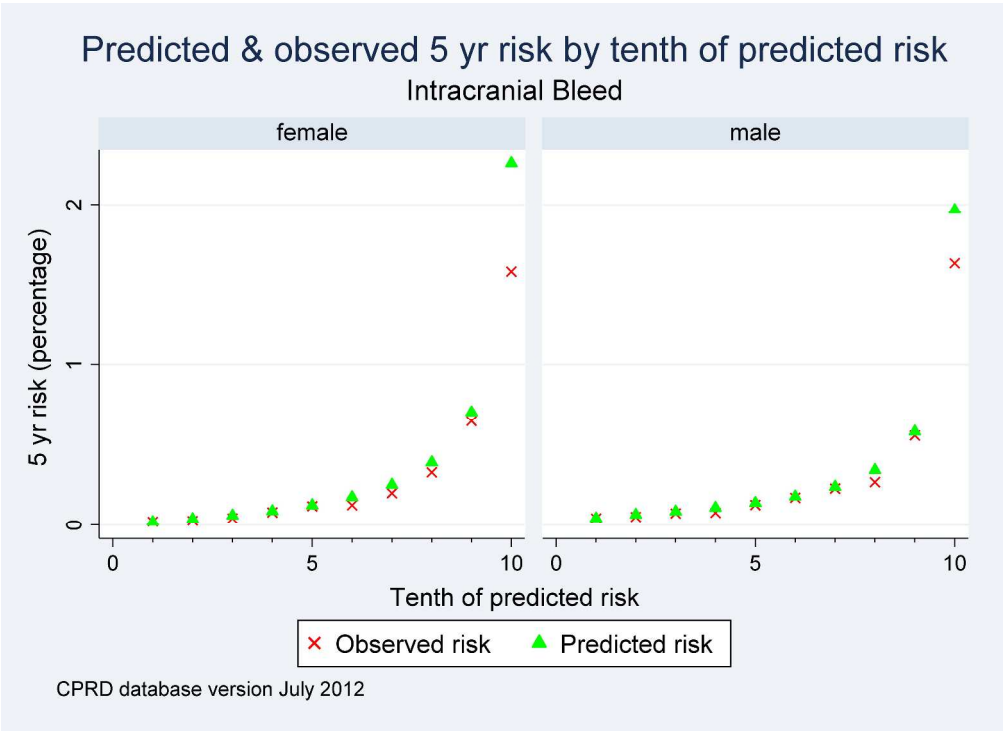
QStroke (ischaemic stroke)



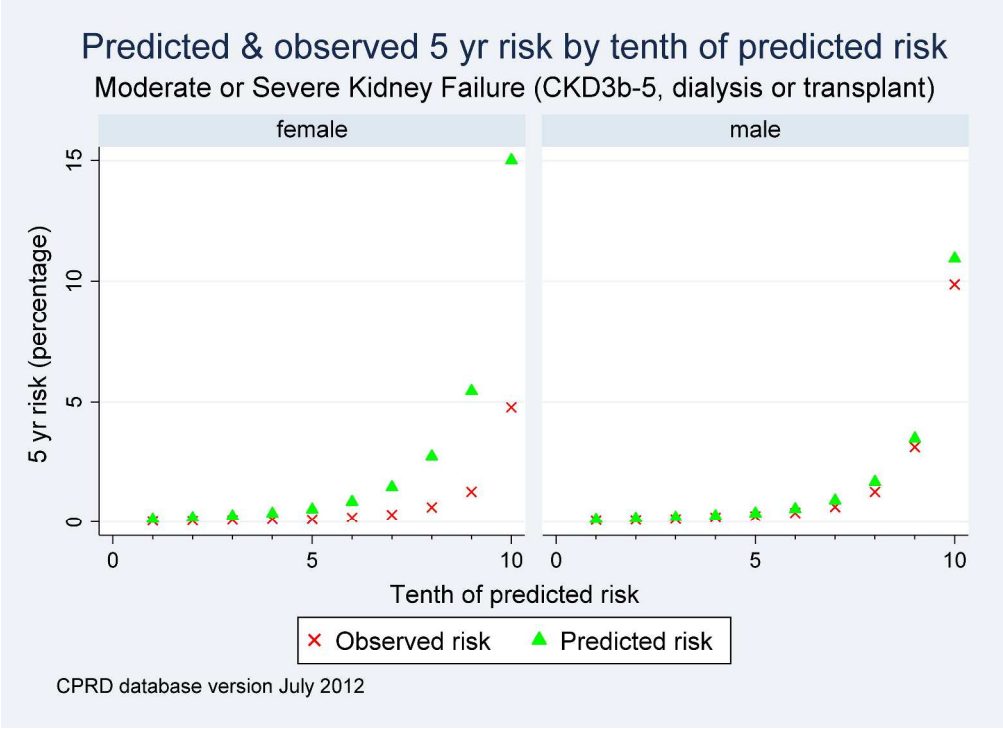
QDiabetes (type 2 diabetes)



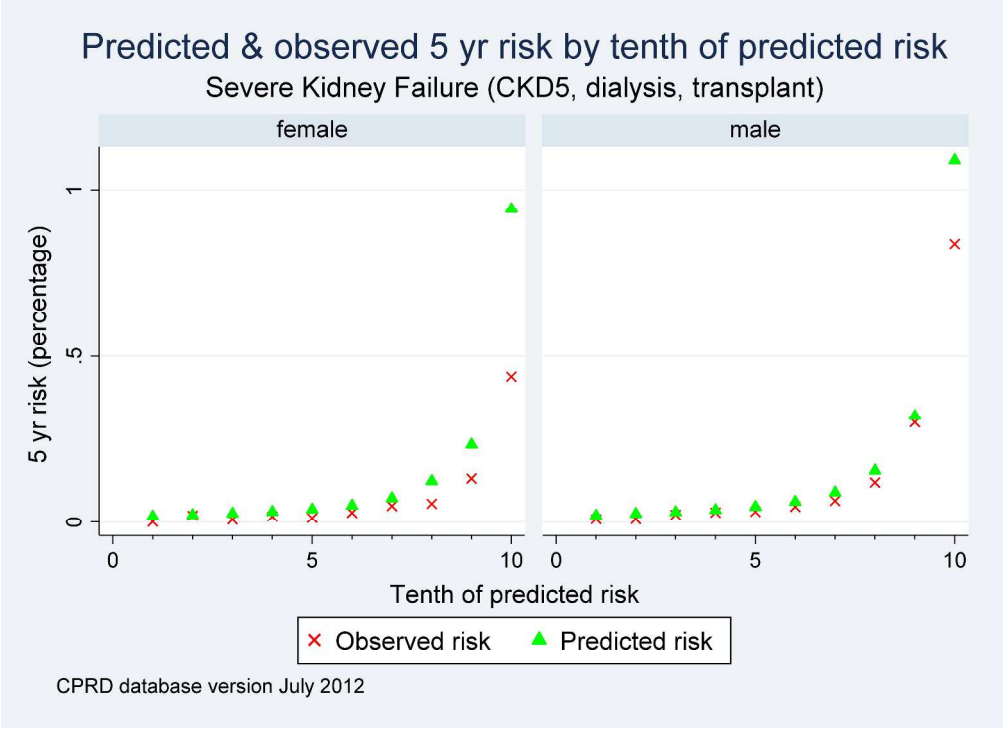
QBleed (upper gastrointestinal haemorrhage)



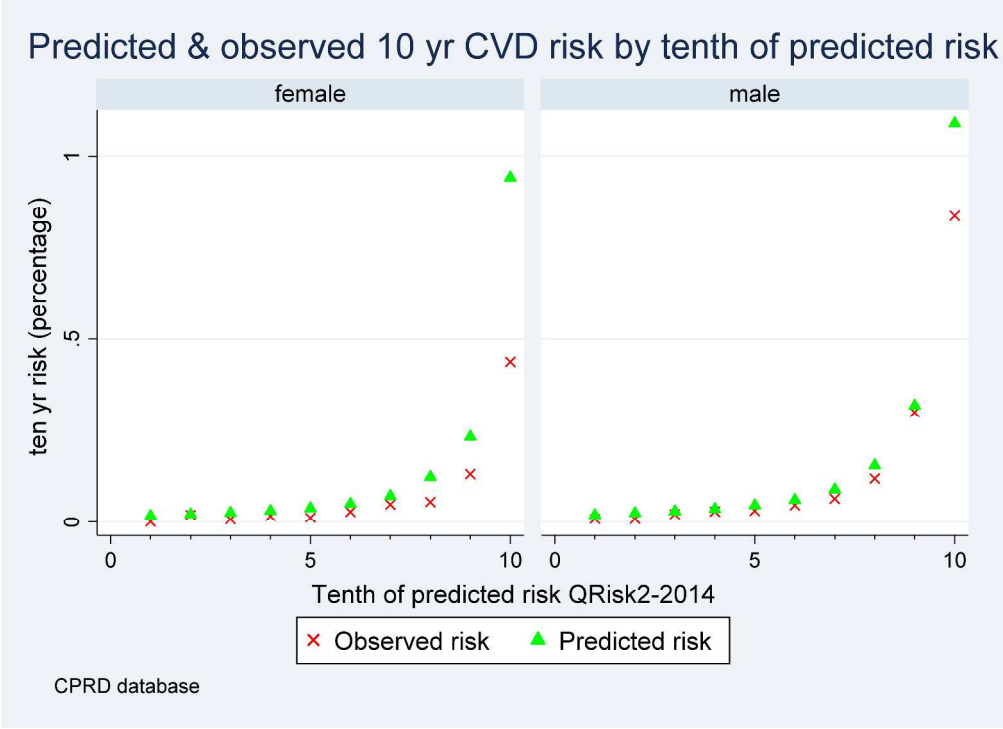
QBleed (intracranial haemorrhage)



QKidney (moderate or severe kidney failure)



QKidney (severe kidney failure)



QRisk2 (cardiovascular disease)

Web extra table 1. Numbers of patients in CPRD and QResearch by geographical area

	CPRD	*Col %*	*QResearch*	*Col %*
East Midlands	109,428	3.3	500,970	7.4
East of England	397,008	12.1	528,379	7.8
London	563,353	17.2	1,711,956	25.3
North East	59,558	1.8	269,695	4.0
North West	474,457	14.5	830,047	12.3
South Central	411,571	12.6	696,070	10.3
South East	362,319	11.1	545,811	8.1
South West	397,735	12.2	700,041	10.4
West Midlands	348,614	10.7	589,548	8.7
Yorks & Humber	147,469	4.5	386,132	5.7
Total	**3,271,512**	**100.00**	**6,758,649**	**100.00**

.

Web table 2 Recording of ethnicity and family history of coronary heart disease (FH CHD) by geographical area

	CPRD linked data					QResearch					Ratio recording QResearch:CPRD	
	total patients	ethnicity recorded		FH CHD recorded		total patients	ethnicity recorded		FH CHD recorded			
	count	count	Row %	count	Row %	count	count	Row %	count	Row %	ethnicity	FH
East Midlands	109,428	23,636	21.6	7,912	7.2	500,970	284,194	56.7	60,278	12.0	2.6	1.7
East of England	397,008	127,427	32.1	17,597	4.4	528,379	319,742	60.5	54,252	10.3	1.9	2.3
London	563,353	308,285	54.7	21,034	3.7	1,711,956	1,095,835	64.0	162,282	9.5	1.2	2.5
North East	59,558	17,140	28.8	3,326	5.6	269,695	195,127	72.4	43,168	16.0	2.5	2.9
North West	474,457	196,987	41.5	25,915	5.5	830,047	460,640	55.5	112,718	13.6	1.3	2.5
South Central	411,571	140,448	34.1	16,989	4.1	696,070	380,908	54.7	73,051	10.5	1.6	2.5
South East	362,319	90,633	25.0	12,618	3.5	545,811	262,861	48.2	45,765	8.4	1.9	2.4
South West	397,735	137,806	34.6	17,829	4.5	700,041	375,155	53.6	75,091	10.7	1.5	2.4
West Midlands	348,614	148,012	42.5	20,375	5.8	589,548	347,479	58.9	63,412	10.8	1.4	1.8
Yorks & Humber	147,469	56,340	38.2	6,195	4.2	386,132	214,702	55.6	54,515	14.1	1.5	3.4
Total	**3,271,512**	**1,246,714**	**38.1**	**149,790**	**4.6**	**6,758,649**	**3,936,643**	**58.2**	**744,532**	**11.0**	**1.5**	**2.4**

Web extra table 3 comparison of baseline characteristics for the CPRD cohort with linked data used for the validation cohort in this study and the CPRD cohort without linked data (which was not used for the validation of the QPrediction scores).

	CPRD linked data	CPRD unlinked data
female	1682709 (51.4)	1166103 (51.3)
male	1588803 (48.6)	1108235 (48.7)
25-34 years	895167 (27.4)	665645 (29.3)
35-44 years	760830 (23.3)	503091 (22.1)
45-54 years	571937 (17.5)	384686 (16.9)
55-64 years	424453 (13.0)	293826 (12.9)
65-74 years	312352 (9.5)	216937 (9.5)
75+ years	306773 (9.4)	210153 (9.2)
Ethnicity recorded	1246714 (38.1)	645829 (28.4)
White or not recorded	3117325 (95.3)	2209396 (97.1)
Indian	32467 (1.0)	11751 (0.5)
Pakistani	12752 (0.4)	6358 (0.3)
Bangladeshi	4107 (0.1)	2682 (0.1)
Other Asian	22668 (0.7)	8854 (0.4)
Caribbean	11414 (0.3)	4812 (0.2)
Black African	27654 (0.8)	9751 (0.4)
Chinese	7090 (0.2)	3416 (0.2)
Other ethnic group	36035 (1.1)	17318 (0.8)
Smoking status recorded	3037626 (92.9)	2066777 (90.9)
Non smoker	1448554 (44.3)	1006511 (44.3)
Ex-smoker	475488 (14.5)	306460 (13.5)
Light smoker (1-9/day)	214330 (6.6)	127487 (5.6)
Moderate smoker (10-19/day)	362391 (11.1)	277693 (12.2)
Heavy smoker (20+/day)	229912 (7.0)	159718 (7.0)
Smoker amount not recorded	306951 (9.4)	188908 (8.3)
Alcohol status recorded	2617112 (80.0)	1759758 (77.4)
Non drinker	482513 (14.7)	361566 (15.9)
Trivial <1u/day	1186942 (36.3)	754470 (33.2)
Light 1-2u/day	701808 (21.5)	479407 (21.1)
Moderate 3-6/day	203091 (6.2)	135130 (5.9)
Heavy 7-9u/day	22465 (0.7)	15645 (0.7)
Very Heavy >/day	20293 (0.6)	13540 (0.6)
family history		
family history of CHD	149790 (4.6)	97401 (4.3)
family history of diabetes	229200 (7.0)	144713 (6.4)
family history osteoporosis	10942 (0.3)	6164 (0.3)
family history kidney disease	2839 (0.1)	1592 (0.1)

prescribed medication		
antidepressants	337350 (10.3)	232657 (10.2)
anticoagulants	29032 (0.9)	20338 (0.9)
antipsychotics	113398 (3.5)	76819 (3.4)
oral NSAIDs	592931 (18.1)	425739 (18.7)
tamoxifen	9343 (0.3)	6107 (0.3)
antiplatelets	190291 (5.8)	137793 (6.1)
oestrogen only HRT	119413 (3.7)	75448 (3.3)
corticosteroids	116949 (3.6)	70793 (3.1)
oral contraceptive pill	174288 (5.3)	126218 (5.5)
recorded diagnoses		
congestive cardiac failure	35543 (1.1)	25823 (1.1)
atrial fibrillation	40227 (1.2)	27032 (1.2)
coronary heart disease	138080 (4.2)	102493 (4.5)
cardiovascular disease	184597 (5.6)	134650 (5.9)
rheumatoid arthritis	24711 (0.8)	17427 (0.8)
chronic renal disease	8050 (0.2)	5774 (0.3)
type 1 diabetes	11162 (0.3)	7778 (0.3)
type 2 diabetes	94905 (2.9)	63240 (2.8)
venous thromboembolism	38162 (1.2)	23593 (1.0)
varicose veins	65991 (2.0)	44717 (2.0)
moderate/severe kidney failure	23607 (0.7)	15072 (0.7)
severe kidney failure	2728 (0.1)	1839 (0.1)
oesophageal varices	809 (0.0)	674 (0.0)
inflammatory bowel disease	19170 (0.6)	13095 (0.6)
SLE	1984 (0.1)	1273 (0.1)
peripheral vascular disease	30130 (0.9)	23066 (1.0)
dementia	23320 (0.7)	15858 (0.7)
Parkinson's disease	9222 (0.3)	5854 (0.3)
cancer	70774 (2.2)	45637 (2.0)
liver disease	6852 (0.2)	5041 (0.2)
malabsorption	16718 (0.5)	12007 (0.5)
endocrine diseases	17179 (0.5)	12479 (0.5)
COPD	44766 (1.4)	33190 (1.5)
chronic liver disease or pancreatitis	9572 (0.3)	6899 (0.3)
renal stones	19858 (0.6)	14935 (0.7)
care home resident	4873 (0.1)	2859 (0.1)
falls	90783 (2.8)	53221 (2.3)
prior fracture	70017 (2.1)	50346 (2.2)
asthma or COPD	312477 (9.6)	207765 (9.1)
treated hypertension	285293 (8.7)	190707 (8.4)
platelets < 150 or > 480	24333 (0.7)	12651 (0.6)
emergency admission or hip op	8749 (0.3)	6468 (0.3)
prior haemorrhage	177327 (5.4)	122024 (5.4)

Recorded values		
BMI recorded	2750153 (84.1)	1864134 (82.0)
SBP reccorded	2949786 (90.2)	2010003 (88.4)
cholesterol/HDL ratio recorded	1193900 (36.5)	761573 (33.5)
platelets recorded	606260 (18.5)	302478 (13.3)
mean age (SD)	47.9 (17.0)	47.4 (17.2)
mean townsend score (SD)	-.5 (3.2)	.1 (3.7)
mean BMI (SD)	28.9 (6.9)	29.2 (7.1)
mean cholesterol raito (SD)	4.1 (1.4)	4.1 (1.4)
mean SBP(SD)	130.7 (23.2)	130.1 (151.3)

Web extra table 4 Number of cases of upper gastrointestinal bleed and intracranial bleed on CPRD and QResearch (one third sample database). Incidence rates per 1000 pyrs have been standardised to the QResearch population in 5 year bands.

	CPRD validation		*QResearch validation*	
	cases on	**age standardised Incidence rate per 1000pyrs**	**cases**	**Age standardised Incidence rate per 1000pyrs**
no anticoagulants	13,314	1.41 (1.39 to 1.43)	6,447	1.33 (1.30 to 1.36)
anticoagulants	359	6.70 (4.06 to 9.34)	153	6.10 (3.20 to 8.98)
Intracranial bleed				
no anticoagulants	5,190	0.53 (0.51 to 0.54)	2,716	0.56 (0.54 to 0.58)
anticoagulants	233	2.45 (1.23 to 3.68)	104	2.87 (1.11 to 4.39)

Web extra table 5 Performance of QPrediction scores on the CPRD validation cohort, restricted to patients with complete data for relevant laboratory and clinical values

		CPRD	*CPRD*
		women	**men**
	statistic	**mean (95%CI)**	**mean (95%CI)**
QDiabetes-2013	ROC	0.849 (0.847 to 0.85	0.814 (0.813 to 0.816
(type 2 diabetes)	R^2 (%)	49.8 (49.4 to 50.2)	44.4 (44 to 44.8)
	D statistic	2.04 (2.024 to 2.056)	1.828 (1.814 to 1.842)
QKidney-2010	ROC	0.847 (0.842 to 0.852	0.839 (0.835 to 0.842
(moderate or severe kidney failure)	R^2 (%)	53.4 (52.8 to 54)	49.7 (49.3 to 50.1)
	D statistic	2.19 (2.165 to 2.215)	2.036 (2.018 to 2.054)
QKidney -2010	ROC	0.816 (0.798 to 0.834	0.808 (0.795 to 0.822
(severe kidney failure)	R^2 (%)	47.3 (45.2 to 49.4)	46.1 (44.5 to 47.7)
	D statistic	1.938 (1.856 to 2.02)	1.895 (1.834 to 1.956)
QRISK2-2014	ROC	0.791 (0.787 to 0.796	0.757 (0.753 to 0.761
(cardiovascular disease)	R^2 (%)	40.9 (40 to 41.8)	31.8 (30.9 to 32.7)
	D statistic	1.704 (1.673 to 1.735)	1.398 (1.371 to 1.425)
QStroke-2013	ROC	0.794 (0.79 to 0.797	0.771 (0.768 to 0.774
(ischaemic stroke or TIA)	R^2 (%)	42.1 (41.4 to 42.8)	36.6 (35.9 to 37.3)
	D statistic	1.747 (1.723 to 1.771)	1.557 (1.535 to 1.579)
QThrombosis-2010	ROC	0.755 (0.75 to 0.76	0.762 (0.756 to 0.767
(venous thromboembolism)	R^2 (%)	34.6 (33.8 to 35.4)	32.6 (31.7 to 33.5)
	D statistic	1.487 (1.462 to 1.512)	1.424 (1.395 to 1.453)
QBleed-20141	ROC statistic	0.773 (0.766 to 0.779	0.751 (0.744 to 0.758
(upper GI bleed)	R^2 (%)	43.6 (42.1 to 45.1)	39.6 (38.1 to 41.1)
	D statistic	1.799 (1.744 to 1.854)	1.658 (1.605 to 1.711)
QBleed-2014	ROC statistic	0.812 (0.803 to 0.822	0.791 (0.78 to 0.802
(intracranial bleed)	R^2 (%)	53 (51 to 55)	50.2 (48 to 52.4)
	D statistic	2.172 (2.086 to 2.258)	2.057 (1.967 to 2.147)
QFracture-2012	ROC	0.899 (0.896 to 0.901	0.866 (0.86 to 0.872
(fracture neck of femur)	R^2 (%)	70.4 (69.9 to 70.9)	67.1 (66.2 to 68)
	D statistic	3.159 (3.124 to 3.194)	2.922 (2.861 to 2.983)
QFracture -2012	ROC	0.819 (0.816 to 0.821	0.757 (0.751 to 0.763
(osteoporotic fracture)	R^2 (%)	53.9 (53.4 to 54.4)	46.1 (45 to 47.2)
	D statistic	2.213 (2.189 to 2.237)	1.893 (1.852 to 1.934)

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page in manuscript
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	Page 2 abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	Page 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	Page4
Objectives	3	State specific objectives, including any prespecified hypotheses	Page 4
Methods			
Study design	4	Present key elements of study design early in the paper	Page 2, page 5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	Page 2, 5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up	Page 5
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed	n/a
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	Page 5 & 6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	*Pages 6*
Bias	9	Describe any efforts to address potential sources of bias	Page 6
Study size	10	Explain how the study size was arrived at	Page 8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Pages 6, 7, 8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	Pages 7, 8
		(*b*) Describe any methods used to examine subgroups and interactions	Page 6
		(*c*) Explain how missing data were addressed	Page 7
		(*d*) If applicable, explain how loss to follow-up was addressed	
		(*e*) Describe any sensitivity analyses	Page 7,8
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	Page 9
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Page 9 tables 2, 3
		(b) Indicate number of participants with missing data for each variable of interest	table 2,3
		(c) Summarise follow-up time (eg, average and total amount)	Table5
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9 table 5
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were	Pages 10, 11 Table 3, table 5

		included	
		(*b*) Report category boundaries when continuous variables were categorized	Page 6
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Page 11
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	Page 11
Discussion			
Key results	18	Summarise key results with reference to study objectives	Page 12
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	Page 12- 13,
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	Page 13
Generalisability	21	Discuss the generalisability (external validity) of the study results	Page 13
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	Page 15