

# Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score

Julia Hippisley-Cox,<sup>1,2</sup> Carol Coupland<sup>1,2</sup>

**To cite:** Hippisley-Cox J, Coupland C. Predicting risk of emergency admission to hospital using primary care data: derivation and validation of QAdmissions score. *BMJ Open* 2013;**3**:e003482. doi:10.1136/bmjopen-2013-003482

► Prepublication history and additional material for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2013-003482>).

Received 26 June 2013  
Accepted 5 July 2013

## ABSTRACT

**Objective:** To develop and externally validate a risk algorithm (QAdmissions) to estimate the risk of emergency hospital admission for patients aged 18–100 years in primary care.

**Design:** Prospective open cohort study using routinely collected data from general practice linked to hospital episode data during the 2-year study period 1 January 2010 to 31 December 2011.

**Setting:** 405 general practices in England contributing to the national QResearch database to develop the algorithm. Two validation cohorts to validate the algorithm (1) 202 different QResearch practices and (2) 343 practices in England contributing to the Clinical Practice Research DataLink (CPRD). All general practices had data linked to hospital episode statistics at the individual patient level.

**Participants:** We studied 2 849 381 patients aged 18–100 years in the derivation cohort with over 4.6 million person-years of follow-up. 265 573 of these patients had one or more emergency admissions during follow-up. For the QResearch validation cohort, we identified 1 340 622 patients aged 18–100 years with over 2.2 million person-years of follow-up. Of these patients, 132 723 had one or more emergency admissions during follow-up. The CPRD cohort included 2 475 360 patients aged 18–100 years with over 3.8 million person-years of follow-up. 234 204 of these patients had one or more emergency admissions during follow-up. We excluded patients without a valid NHS number and a valid Townsend score.

**Endpoint:** First (ie, incident) emergency admission to hospital in the next 2 years as recorded on the linked hospital episodes records.

**Risk factors:** Candidate variables recorded on the general practitioner computer system including (1) demographic variables (age, sex, strategic health authority, Townsend deprivation score, ethnicity); (2) lifestyle variables (smoking, alcohol intake); (3) chronic diseases; (4) prescribed medication; (5) clinical values (body mass index, systolic blood pressure); (6) laboratory test results (haemoglobin, platelets, erythrocyte sedimentation rate, ratio of total serum cholesterol to high density lipoprotein cholesterol concentrations, liver function tests). We also included the number of emergency admissions in

## ARTICLE SUMMARY

### Article focus

- Methods to identify patients at increased risk of emergency admission to hospital are needed to identify patients for whom interventions may be required to reduce risk of admission.
- Current risk scoring methods are expensive, unpublished or difficult to implement.

### Key messages

- We have developed and validated a new algorithm to quantify absolute risk of emergency admission to hospital, which includes established risk factors, and which is designed to work in primary care.
- The QAdmissions model provides a valid measure of absolute emergency admission risk in the general population of patients as shown by its performance in a separate validation cohort.
- Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using these algorithms in primary care.

### Strengths and limitations of this study

- The key strengths include use of linked data on hospital admissions, study size, representativeness, and lack of selection and recall bias.
- Limitations include potential for bias due to missing data.

the preceding year based on information recorded on the linked hospital episodes records.

**Results:** The final QAdmissions algorithm incorporated 30 variables. When applied to the QResearch validation cohort, it explained 41% of the variation in women and 43% of that in men. The D statistic for QAdmissions was 1.7 in women and 1.8 in men. The receiver operating curve statistic was 0.78 for men and 0.77 for women. QAdmissions had good performance on all measures of discrimination and calibration. The positive predictive value for emergency admissions for the top tenth of patients at highest risk was 42% and the sensitivity was 39%. The results for the CPRD validation cohort were similar.

<sup>1</sup>Division of Primary Care, University of Nottingham, Nottingham, UK.

<sup>2</sup>ClinRisk Ltd, White Rose House Leeds, UK

### Correspondence to

Professor Julia Hippisley-Cox; [Julia.hippisley-cox@nottingham.ac.uk](mailto:Julia.hippisley-cox@nottingham.ac.uk)

**Conclusions:** The QAdmissions model provided a valid measure of absolute risk of emergency admission to hospital in the general population as shown by its performance in a separate validation cohort. Further research is needed to evaluate the cost-effectiveness of using these algorithms in primary care.

## INTRODUCTION

Unplanned admissions account for an estimated 11 billion pounds a year in England, which is a considerable proportion of the National Health Service (NHS) budget.<sup>1</sup> Such admissions are not only costly but also potentially distressing to individuals. Successive governments have tried to implement approaches to prevent the rise in emergency admissions including identifying patients at high risk of emergency admission so that these patients can be targeted before preventable or avoidable costs have been incurred.

In Spring 2013, the NHS commissioning Board (now NHS England) announced a new Enhanced Service Specification to reward general practitioner (GP) practices for the identification and case management of patients identified as seriously ill or at risk of an emergency admission.<sup>2</sup> As part of this, GP practices need to undertake risk profiling and risk stratification of their registered patients on at least a quarterly basis.

Central to any risk stratification and case identification programme is the accuracy and utility of the algorithm used to undertake the risk assessment. In general, a risk stratification algorithm needs to be developed using data from the setting where it will subsequently be used (eg, primary care in England). It needs to be able to distinguish between patients who do or do not experience the event of interest (discrimination) and accurately quantify the level of risk (calibration). It should predict the outcome of interest (eg, emergency admission) for the population of interest (eg, all adult patients registered with the general practitioner). It needs to apply over the relevant time period (eg, 1–2 years) assuming that sufficient time is needed for interventions to have an effect. It needs to include predictors with good clinical face validity and, ideally, include some clinically relevant factors which are amenable to change (ie, help reduce risk of emergency admission). It should preferably incorporate measures of socioeconomic deprivation and ethnicity not only in recognition of the role these factors have as predictors of major diseases, but also to prevent widening health inequalities which can occur when new programmes are introduced. The risk algorithm needs to have the potential to be updated or recalibrated, and its performance should be tested in a population of patients separate from that used to develop the tool to demonstrate that it can reliably identify the target population. Finally, the tool needs to be suitable for implementation in clinical practice.

While a number of emergency admission risk assessment tools have been developed, they are generally

designed for use in hospital to identify patients at risk of readmission.<sup>3–5</sup> Other current tools focus on specific populations or have not been published or validated. For example, there are a number of American algorithms based on patients enrolled in health maintenance organisations with questionable generalisability.<sup>6–8</sup> There are several tools which have been intended for use in primary care. The Emergency Admission Risk Likelihood Index is a six item questionnaire which was developed using data from patients aged 75+ from 17 general practices in the North of England.<sup>9</sup> Hence, it only applies to elderly patients and may not be sufficiently representative for wider use. The Predicting Emergency Admissions Over the Next Year (PEONY) score was designed for use in Scottish primary care patients aged 40–65 years.<sup>10</sup> However, it does not include morbidity data from primary care, and currently the underlying algorithm is not published or independently validated. Finally, the combined predictive model<sup>11</sup> (CPM), developed using data from two Primary Care Trusts, had been designed to work on primary care data linked to three secondary care data sources (inpatient, outpatient, accident and emergency). However, the Department of Health announced in August 2011 that the tools were outdated and in urgent need of a refresh.<sup>12</sup>

One problem which has beset all the existing risk algorithms is the practical difficulty in implementing them into primary care since they have not been designed to run off routinely collected data that are already in GP computer systems or have not been validated in that setting. While it is possible to extract the primary care data from GP clinical systems into a data warehouse for linkage, processing and feeding back to the practice, this is a complex technical process to achieve in real time. It also has significant information governance challenges given the necessary controls around the processing of personal confidential data by third parties without patient consent.

Therefore, we decided to develop and validate a new risk prediction algorithm to predict the absolute risk of emergency admissions to hospital (QAdmissions) which could meet the above requirements. We were interested to develop an algorithm which incorporates ethnicity and clinical diagnoses, medications and abnormal laboratory results which the healthcare professional in practice can then follow-up. In addition, we decided to develop a tool which could be automatically populated using data solely from GP computer systems and so provide an expedient practical alternative where primary care data are not routinely linked to secondary care data.

## METHODS

### Study design and data source

We conducted a prospective cohort study in a large UK primary care population using a method similar to our

analyses for other risk prediction scores such as QRISK2.<sup>13</sup> V.35 of the QResearch database was used for this study (<http://www.qresearch.org>). This is a large validated primary care electronic database containing the health records of 13 million patients registered from 660 general practices using the Egton Medical Information System (EMIS) computer system.<sup>13</sup> Practices and patients contained on the database are nationally representative<sup>14</sup> and similar to those on other primary care databases using other clinical software systems.<sup>15</sup> We included all QResearch practices in England once they had been using their current EMIS system for at least a year (to ensure completeness of recording of morbidity and prescribing data), randomly allocating two-thirds of practices to the derivation dataset with one-third to the validation dataset. The analysis was conducted on QResearch practices in England in order to incorporate hospital episode data linked at the individual patient level via a pseudonymised NHS number. We also assembled a second validation cohort using 343 English practices contributing to the Clinical Practice Research Datalink (CPRD) which had linked hospital episode statistics (HES) data (August 2012 download).

### Cohort selection

We identified three open cohorts of patients aged 18–100 at the study entry date, drawn from patients registered with eligible practices between 1 January 2010 and 31 December 2011. We used an open cohort design, rather than a closed cohort design, as this allows patients to enter the population throughout the whole study period rather than require registration on 1 January 2010, thus better reflecting the realities of routine general practice. We excluded registered patients without a valid pseudonymised NHS number as this was needed to link the primary and secondary care data together. We also excluded patients without a valid postcode-related Townsend deprivation score.

For each patient, we determined an entry date to the cohort, which was the latest of the following dates: 18th birthday, date of registration with the practice plus 1 year, date on which the practice computer system was installed plus 1 year, and the beginning of the study period (1 January 2010). Patients were censored at the earliest date of the following: the first emergency hospital admission in the study period, death, deregistration with the practice, last upload of computerised data or the study end date (31 December 2011).

### Emergency hospital admission outcomes

The primary outcome measure of interest was the first recorded emergency admission to hospital in the study period. We identified emergency hospital admissions from the HES data, which includes all hospital trusts in England. The HES data was linked at the individual patient level to the QResearch database via a pseudonymised NHS number. Emergency admissions were identified by selecting the standard codes to represent all

emergency admissions in England. This information is derived from the method of admission field recorded for each admission. The following codes were included—coded as 21 (accident and emergency); 22 (GP direct to hospital); 23 (GP via a bed bureau); 24 (consultant clinic); 25 (mental health crisis resolution team); 28 (other means). We only included emergency admissions where the admission date and discharge date were both recorded and where the admission date was on or before the discharge date.

### Risk factors for emergency admission

We identified a list of candidate variables, focusing on variables which have previously been established to increase risk of emergency admission<sup>10</sup> or readmission.<sup>4–7</sup> We also included predictors used in other risk algorithms where the outcome is likely to require emergency admission (eg, as thrombosis<sup>16</sup> or cardiovascular disease<sup>17–18</sup>). We decided to focus on variables which are recorded in the primary care electronic record in order to ensure that the resulting algorithm could be implemented into existing GP computer systems in a way similar to the implementation of similar risk prediction algorithms developed using the QResearch database.<sup>4–11–14</sup> The full list of candidate variables is shown in [table 1](#) and is summarised as follows:

- A. Demographic variables: age, sex, Strategic Health Authority, Townsend deprivation score, ethnicity.
- B. Lifestyle variables: smoking status, alcohol intake.
- C. Chronic diseases.
- D. Medication for statins, non-steroidal anti-inflammatory drugs (NSAIDs), anticoagulants, corticosteroids, antidepressants and antipsychotics at study entry date.
- E. Clinical values: body mass index, systolic blood pressure.
- F. Laboratory test results: haemoglobin, platelets, erythrocyte sedimentation rate, total serum cholesterol/high-density lipoprotein ratio, liver function tests.
- G. Emergency admissions in the year before study entry date (none, 1, 2, 3 or more).

All the above variables were derived from the patients' primary care record except for the number of emergency admissions in the year before the study entry date where we used the HES-linked data. We restricted all values of these candidate predictor variables to those recorded in the person's electronic healthcare record before baseline, except for ethnicity where we used the most recently recorded value in the study period before the patient had the outcome or was censored.

We imputed missing values where necessary as described below. Given the large number of candidate variables, we combined factors where appropriate. For example, we combined (1) asthma and chronic obstructive airways disease and (2) schizophrenia and manic depression. We defined abnormal liver function tests as

**Table 1** Baseline characteristics of patients in the QResearch derivation cohort, the QResearch validation cohort and the CPRD validation cohort

	QResearch derivation (n=2 849 381)	QResearch validation (n=1 340 622)	CPRD validation (n=2 475 360)
Female	1 446 784 (50.8)	677 897 (50.6)	1 260 015 (50.9)
Male	1 402 597 (49.2)	662 725 (49.4)	1 215 345 (49.1)
Mean age (SD)	46.3 (18.9)	47.8 (18.6)	48.2 (18.6)
Strategic Health Authority			
East Midlands SHA	225 092 (7.9)	165 734 (12.4)	70 695 (2.9)
Yorkshire & Humberside SHA	220 560 (7.7)	75 976 (5.7)	287 374 (11.6)
East of England SHA	197 453 (6.9)	158 962 (11.9)	390 573 (15.8)
London SHA	560 544 (19.7)	234 346 (17.5)	52 618 (2.1)
North East SHA	141 974 (5.0)	103 200 (7.7)	398 889 (16.1)
North West SHA	268 958 (9.4)	264 508 (19.7)	317 867 (12.8)
South Central SHA	310 830 (10.9)	74 588 (5.6)	274 296 (11.1)
South East SHA	253 288 (8.9)	63 455 (4.7)	314 779 (12.7)
South West SHA	421 052 (14.8)	92 822 (6.9)	275 566 (11.1)
West Midlands SHA	249 630 (8.8)	107 031 (8.0)	92 703 (3.7)
Ethnicity			
Ethnicity recorded	2 129 124 (74.7)	1 015 630 (75.8)	1 301 115 (52.6)
White/not recorded	2 554 557 (89.7)	1 212 057 (90.4)	2 320 487 (93.7)
Indian	49 360 (1.7)	22 888 (1.7)	31 800 (1.3)
Pakistani	23 947 (0.8)	15 243 (1.1)	13 739 (0.6)
Bangladeshi	22 309 (0.8)	11 076 (0.8)	4482 (0.2)
Other Asian	38 463 (1.3)	14 870 (1.1)	22 394 (0.9)
Caribbean	23 704 (0.8)	9038 (0.7)	11 086 (0.4)
Black African	43 471 (1.5)	22 355 (1.7)	26 533 (1.1)
Chinese	28 803 (1.0)	8086 (0.6)	7514 (0.3)
Other	64 767 (2.3)	25 009 (1.9)	37 325 (1.5)
Smoking status			
Smoking status recorded	2 766 234 (97.1)	1 300 728 (97.0)	2 388 744 (96.5)
Non-smoker	1 568 956 (55.1)	731 480 (54.6)	1 220 054 (49.3)
Ex-smoker	612 156 (21.5)	288 031 (21.5)	642 110 (25.9)
Light smoker (1–9/day)	353 026 (12.4)	165 471 (12.3)	161 185 (6.5)
Moderate smoker (10–19/day)	152 631 (5.4)	75 157 (5.6)	210 441 (8.5)
Heavy smoker (20+/day)	79 465 (2.8)	40 589 (3.0)	120 768 (4.9)
Smoker amount not recorded	n/a	n/a	34 186 (1.4)
Alcohol intake			
Alcohol status recorded	2 340 360 (82.1)	1 097 278 (81.8)	1 968 156 (79.5)
Non-drinker	746 788 (26.2)	354 328 (26.4)	393 692 (15.9)
Trivial <1 unit/day	792 730 (27.8)	368 465 (27.5)	878 965 (35.5)
Light 1–2 units/day	365 897 (12.8)	166 881 (12.4)	508 687 (20.6)
Moderate 3–6 units/day	387 161 (13.6)	183 738 (13.7)	150 466 (6.1)
Heavy 7–9 units/day	27 501 (1.0)	13 579 (1.0)	17 695 (0.7)
Very Heavy >9 units/day	16 260 (0.6)	8112 (0.6)	18 651 (0.8)
Drinker—amount not recorded	4023 (0.1)	2175 (0.2)	0 (0)
Emergency admissions in the past year (HES record)			
No emergency admission (HES record)	2 695 651 (94.6)	1 264 555 (94.3)	2 334 640 (94.3)
1 emergency admission (HES record)	118 002 (4.1)	58 078 (4.3)	107 182 (4.3)
2 emergency admissions (HES record)	23 301 (0.8)	11 687 (0.9)	21 802 (0.9)
3+ emergency admissions (HES record)	12 427 (0.4)	6302 (0.5)	11 736 (0.5)
Emergency admissions in the past year (GP record)			
No emergency admission (GP record)	2 731 533 (95.9)	1 283 422 (95.7)	2 261 885 (91.4)
1 emergency admission (GP record)	89 457 (3.1)	44 263 (3.3)	158 723 (6.4)

Continued



Table 1 Continued

	QResearch derivation (n=2 849 381)	QResearch validation (n=1 340 622)	CPRD validation (n=2 475 360)
2 emergency admissions (GP record)	19 581 (0.7)	8812 (0.7)	36 567 (1.5)
3+ emergency admissions (GP record)	8810 (0.3)	4125 (0.3)	18 185 (0.7)
Clinical values, family history and deprivation			
Body mass index recorded	2 281 550 (80.1)	1 083 278 (80.8)	1 980 327 (80.0)
Mean body mass index (SD)	26.1 (4.9)	26.4 (4.9)	26.4 (5.0)
Systolic blood pressure recorded*	2 437 745 (85.6)	1 186 261 (88.5)	n/a
Mean systolic blood pressure (SD)	127.0 (16.4)	127.3 (16.5)	n/a
Cholesterol/HDL recorded*	824 938 (29.0)	413 117 (30.8)	n/a
Mean cholesterol/HDL ratio	3.8 (1.2)	3.8 (1.2)	n/a
Family history CHD*	327 668 (11.5)	169 286 (12.6)	n/a
Mean Townsend score (SD)	0.1 (3.6)	0.1 (3.5)	-0.7 (3.1)
Haemoglobin recorded	1 645 857 (57.8)	816 261 (60.9)	1 512 841 (61.1)
Haemoglobin < 11 g/dl	56 293 (2.0)	28 113 (2.1)	49 339 (2.0)
Platelets recorded	1 632 357 (57.3)	810 551 (60.5)	1 505 945 (60.8)
Platelets > 480	16 501 (0.6)	8434 (0.6)	14 127 (0.6)
Liver function test recorded	1 225 813 (43.0)	628 439 (46.9)	1 148 893 (46.4)
Abnormal liver function tests	34 260 (1.2)	19 112 (1.4)	32 230 (1.3)
ESR recorded	755 536 (26.5)	409 183 (30.5)	n/a
Abnormal ESR	5989 (0.2)	3306 (0.2)	n/a
Comorbidity			
Type 1 diabetes	11 000 (0.4)	5445 (0.4)	9854 (0.4)
Type 2 diabetes	125 374 (4.4)	63 461 (4.7)	117 754 (4.8)
Atrial fibrillation	52 603 (1.8)	26 285 (2.0)	48 490 (2.0)
Cardiovascular disease	154 825 (5.4)	79 116 (5.9)	150 108 (6.1)
Congestive cardiac failure	27 404 (1.0)	14 304 (1.1)	22 685 (0.9)
Venous thromboembolism	42 870 (1.5)	21 298 (1.6)	37 925 (1.5)
Cancer	97 279 (3.4)	48 370 (3.6)	82 513 (3.3)
Asthma or COPD	378 048 (13.3)	179 635 (13.4)	342 371 (13.8)
Epilepsy	36 615 (1.3)	17 904 (1.3)	34 607 (1.4)
Falls	124 248 (4.4)	64 299 (4.8)	172 555 (7.0)
Manic depression or schizophrenia	21 277 (0.7)	10 155 (0.8)	16 792 (0.7)
Chronic renal disease	9841 (0.3)	4700 (0.4)	9476 (0.4)
Conditions leading to malabsorption	29 206 (1.0)	14 432 (1.1)	19 078 (0.8)
Chronic liver disease or pancreatitis	15 811 (0.6)	7669 (0.6)	10 895 (0.4)
Valvular heart disease*	30 924 (1.1)	15 960 (1.2)	n/a
Treated hypertension*	371 503 (13.0)	188 901 (14.1)	n/a
Rheumatoid arthritis or SLE*	45 966 (1.6)	23 020 (1.7)	n/a
Depression (QOF definition)*	372 341 (13.1)	176 638 (13.2)	n/a
Current prescribed medication			
Statins*	341 765 (12.0)	174 252 (13.0)	
NSAIDs	416 749 (14.6)	208 936 (15.6)	365 927 (14.8)
Anticoagulants	38 790 (1.4)	19 764 (1.5)	36 166 (1.5)
Corticosteroids	101 067 (3.5)	49 683 (3.7)	109 847 (4.4)
Antidepressants	341 194 (12.0)	168 305 (12.6)	302 457 (12.2)
Antipsychotics	74 039 (2.6)	38 324 (2.9)	69 498 (2.8)

Values are numbers (percentages of total number in cohort) unless stated otherwise.

CPRD, Clinical Practice Research DataLink; COPD, chronic obstructive pulmonary disease; CHD, coronary heart disease; ESR, erythrocyte sedimentation rate; GP, general practitioner; HES, hospital episode statistics; HDL, high-density lipoprotein; NSAIDs, non-steroidal anti-inflammatory drugs; SHA, Strategic Health Authority; SLE, systemic lupus erythematosus

\*Variables which were considered but did not meet the criteria for inclusion in the final model. These variables were therefore not needed from CPRD for the external validation, so they have been reported as not applicable.

a single variable which denoted either a high  $\gamma$ -GT, aspartate aminotransferase or bilirubin where a high value was at least three times the upper limit of normal.

### Model derivation and development

As in previous studies,<sup>17</sup> we used the Cox proportional hazards model in the derivation dataset to estimate the coefficients and hazard ratios (HRs) associated with each potential risk factor for the first recorded emergency admission to hospital for males and females separately. We used fractional polynomials to model non-linear risk relationships with age and body mass index where appropriate.<sup>19</sup> We tested for interactions between each variable and age and included significant interactions in the final model where they improved the model fit. Continuous variables were centred for analysis. Our main analyses used multiple imputation to replace missing values for systolic blood pressure, cholesterol, smoking status, alcohol status and body mass index.

Our final model was fitted based on five multiply imputed datasets using Rubin's rules to combine estimates and standard errors to allow for the uncertainty due to imputing missing data.<sup>20</sup> We took the logarithm of HR for each variable from the final model and used these as weights for the risk equations. We combined these weights with the baseline survivor function evaluated at 1 and 2 years to derive a risk equation which could be applied for each time period. There were at least 100 outcome events per variable considered in the prediction model in the derivation cohort.<sup>21</sup>

### Model validation

We tested the performance of the final model (QAdmissions) in the QResearch validation cohort and also in a cohort of practices and patients derived from the Clinical Practice Research Datalink (CPRD). We calculated the 2-year estimated risk of emergency admission for each patient in the validation datasets using multiple imputation to replace missing values as in the derivation dataset.

We calculated the mean predicted and observed risks at 2 years<sup>13</sup> and compared these by 10th of the predicted risk for each score. The observed risk at 2 years was obtained using the 2-year Kaplan-Meier estimate. We calculated the receiver operating characteristic (ROC) statistic, the D statistic (a measure of discrimination where higher values indicate better discrimination)<sup>22</sup> and an  $R^2$  statistic (which is a measure of explained variation for survival data where higher values indicate more variation is explained).<sup>23</sup>

Since there is no currently accepted threshold for classifying a high risk of emergency admission based on an absolute risk estimate, we examined the distribution of predicted risk values for QAdmissions and calculated a series of centile values. For each centile threshold, we calculated the sensitivity and the observed risk of

admission (as an estimate of the positive predictive value) over the 2-year follow-up.

For the main validation analyses, we estimated the risk of emergency admission using predictor variables derived from data recorded in the GP record, except for prior emergency admissions which were derived from the HES-GP-linked data.

We repeated the analyses using data on hospital admissions recorded on the GP record instead of the HES-linked data to derive the prior admissions variable. For this second analysis, we examined the clinical Read codes used to identify hospital admissions on the GP record and selected admissions which were coded either as emergency admissions or referral to accident and emergency. A list of the clinical codes used to identify prior hospital events on the GP data can be found in the first table of the online supplementary appendix. This was then used alongside the other GP data derived predictor variables to calculate the risk scores. This was performed to evaluate the performance of the algorithm in a primary care setting where the GP-HES-linked data are not available (GP-HES is not routinely available in all primary care settings).

All analyses were conducted on the QResearch and CPRD validation cohorts. We used STATA (V.12.1) for all analyses.

## RESULTS

### Practices and patients

Overall, 607 QResearch practices in England met our inclusion criteria and had been using their current computer system for at least 1 year. Of these, 405 were randomly assigned to the derivation dataset and 202 to the QResearch validation dataset. We identified 2 857 476 patients aged 18–100 years in the derivation cohort. Of these, 4518 (0.16%) had an invalid NHS number and 3577 (0.13%) had a missing Townsend score leaving 2 849 381 eligible patients for analysis. Similarly, we identified 1 343 274 patients in the QResearch validation cohort. Of these 1254 (0.09%) had an invalid NHS number and 1398 (0.10%) had a missing Townsend score leaving 1 340 622 eligible patients for analysis.

Table 1 compares the characteristics of eligible patients in the QResearch derivation and validation cohort. It also includes the characteristics of the 2 475 360 patients from 343 CPRD practices which met the inclusion criteria and which constitute the second validation cohort. The baseline characteristics of all three cohorts were similar except that the recording of ethnicity was higher in the two QResearch cohorts (75% and 76%) than in CPRD (53%).

### Emergency admissions outcome

Table 2 shows the numbers of cases (patients with one or more admissions in follow-up) and incidence rates of first emergency admissions by age, sex, ethnicity and Strategic Health Authority (SHA) in each cohort.

**Table 2** Incidence rates of first emergency admissions to hospital during follow-up for men and women in the QResearch derivation cohort, the QResearch validation cohort and the CPRD validation cohort

	QResearch derivation cohort			QResearch validation cohort			CPRD validation cohort		
	Cases	person-years	Crude rate/1000 (95% CI)	Cases	Person-years	Crude rate/1000 (95% CI)	Cases	Person-years	Crude rate/1000 (95% CI)
<b>Total</b>	265 573	4 597 543	<b>57.8 (57.5 to 58.0)</b>	<b>132 723</b>	<b>2 222 285</b>	<b>59.7 (59.4 to 60.0)</b>	234 204	3 878 996	60.4 (60.1 to 60.6)
Female	143 524	2 307 505	62.2 (61.9 to 62.5)	71 700	1 116 041	64.2 (63.8 to 64.7)	126 630	1 962 447	64.5 (64.2 to 64.9)
Male	122 049	2 290 038	53.3 (53.0 to 53.6)	61 023	1 106 244	55.2 (54.7 to 55.6)	107 574	1 916 550	56.1 (55.8 to 56.5)
Age band (years)									
18–24	19 563	546 478	35.8 (35.3 to 36.3)	8 687	218 427	39.8 (38.9 to 40.6)	15 749	378 473	41.6 (41.0 to 42.3)
25–34	26 301	799 454	32.9 (32.5 to 33.3)	12 798	366 120	35.0 (34.4 to 35.6)	22 264	608 225	36.6 (36.1 to 37.1)
35–44	29 210	861 476	33.9 (33.5 to 34.3)	15 193	426 812	35.6 (35.0 to 36.2)	25 738	735 573	35.0 (34.6 to 35.4)
45–54	32 359	821 316	39.4 (39.0 to 39.8)	16 186	415 342	39.0 (38.4 to 39.6)	28 572	732 828	39.0 (38.5 to 39.4)
55–64	34 350	678 292	50.6 (50.1 to 51.2)	17 425	343 970	50.7 (49.9 to 51.4)	31 255	621 903	50.3 (49.7 to 50.8)
65–74	39 516	483 667	81.7 (80.9 to 82.5)	20 362	248 334	82.0 (80.9 to 83.1)	35 931	438 517	81.9 (81.1 to 82.8)
75+	84 274	406 859	207 (206 to 209)	42 072	203 280	207 (205 to 209)	74 695	363 477	206 (204 to 207)
<b>SHA</b>			<b>Age/sex standardised rate per 1000 (95% CI)</b>	<b>Cases</b>	<b>Person-years</b>	<b>Age/sex standardised rate per 1000 (95% CI)</b>	<b>Cases</b>	<b>Person-years</b>	<b>Age/sex standardised rate per 1000 (95% CI)</b>
East Midlands	18 226	353 210	53.3 (52.6 to 54.1)	16 269	283 709	54.9 (54.1 to 55.7)	5 185	76 158	69.0 (67.2 to 70.8)
Yorks & Humber	21 018	346 172	61.5 (60.7 to 62.3)	8 458	129 444	61.9 (60.6 to 63.2)	24 987	430 346	55.2 (54.5 to 55.8)
East of England	19 633	333 388	53.6 (52.8 to 54.3)	13 822	262 783	51.7 (50.8 to 52.5)	30 149	585 433	56.4 (55.8 to 57)
London	39 647	846 604	55.8 (55.3 to 56.3)	17 708	363 511	58.4 (57.5 to 59.2)	6 913	87 279	77.6 (75.9 to 79.3)
North East	17 144	229 358	74.6 (73.6 to 75.7)	13 791	175 554	75.2 (74.0 to 76.4)	45 946	656 831	69.0 (68.4 to 69.6)
North West	32 202	452 867	69.3 (68.5 to 70.0)	29 851	436 418	66.2 (65.5 to 66.9)	27 562	521 701	51.4 (50.8 to 52)
South Central	26 134	515 603	50.1 (49.5 to 50.7)	5 741	126 728	43.9 (42.8 to 45.0)	25 571	450 142	55.1 (54.5 to 55.8)
South East Coast	23 849	408 445	52.9 (52.3 to 53.6)	5 482	105 833	50.3 (49.0 to 51.6)	29 319	471 571	57.4 (56.7 to 58)
South West	40 724	691 067	54.0 (53.5 to 54.5)	10 114	156 593	57.6 (56.5 to 58.7)	28 495	450 503	60.7 (60.1 to 61.4)
West Midlands	26 996	420 830	59.6 (58.9 to 60.3)	11 487	181 712	59.2 (58.2 to 60.2)	10 077	149 034	63.5 (62.3 to 64.6)
Ethnicity									
White/not recorded	248 023	4 179 915	56.8 (56.6 to 57.0)	123 918	2 031 918	58.4 (58.1 to 58.7)	224 317	3 667 301	58.9 (58.6 to 59.1)
Indian	2 822	69 939	55.7 (53.5 to 58.0)	1 542	34 821	58.9 (55.8 to 62.1)	2 027	44 446	59.9 (57.2 to 62.7)
Pakistani	1 981	35 724	75.5 (71.4 to 79.5)	1 452	23 474	85.8 (80.4 to 91.2)	1 230	19 049	89.3 (82.9 to 95.6)
Bangladeshi	1 548	33 347	75.4 (70.3 to 80.5)	848	16 546	84.5 (77.0 to 92.0)	297	5 956	76.3 (65.9 to 86.7)
Other Asian	1 757	52 332	51.5 (48.4 to 54.5)	757	20 622	51.6 (46.8 to 56.4)	1 134	29 731	52.9 (49.1 to 56.7)
Caribbean	2 631	37 728	72.3 (69.6 to 75.1)	925	14 644	64.6 (60.5 to 68.7)	1 093	16 468	69.4 (65.3 to 73.6)
Black African	2 637	62 229	56.8 (53.4 to 60.1)	1 442	32 407	62.0 (56.8 to 67.2)	1 538	35 515	59.7 (53.8 to 65.7)
Chinese	499	35 304	34.2 (30.1 to 38.2)	254	11 556	37.5 (32.2 to 42.7)	233	9 838	37.9 (31.8 to 44.0)
Other	3 675	91 026	58.0 (55.7 to 60.3)	1 585	36 297	56.9 (53.4 to 60.3)	2 335	50 694	63.2 (59.9 to 66.4)

Rates are per 1000 person-years. Standardised rates have been directly standardised by age and sex using 5-year age bands.

Overall in the derivation cohort, we identified 265 573 patients (9.3% of 2 849 381) with an incident emergency admission arising from 4.6 million person-years of observation. Of these, 181 784 (68.5%) had one admission and 83 789 (31.6%) had more than one emergency admission in the study period. Of the 265 573 patients with an emergency admission, 212 803 (80.1%) had no emergency admissions in the previous 12 months; 34 246 (12.9%) had one admission; 10 741 (4.0%) had two admissions and 7783 (2.9%) had three or more admissions. The median duration of admission was 2 days (IQR 0–6 days).

In the QResearch validation cohort, we identified 132 723 patients (9.9% of 1 340 622) with an incident emergency admission arising from 2.2 million years of observation. Of these, 90 622 (68.3%) had one admission only and 42 101 (31.7%) had more than one admission. The median duration of admission was 2 days (IQR 0–6 days).

The crude incidence rate of emergency admission was higher in women than in men and rose steeply with age. The age-sex standardised emergency admission rates varied between SHAs with the highest rates in SHAs in the North East. The emergency admission rates for the CPRD validation cohort as recorded on the CPRD-HES-linked data are similar to those for both QResearch cohorts for age, sex and ethnicity.

### Model development

Table 3 shows the results of the Cox regression analysis for the final QAdmissions model. Details of the fractional polynomial terms for age and body mass index are shown in the footnote of the table. The final model included interactions between age and the following variables in men and women: prior admissions, type 2 diabetes, venous thromboembolism, epilepsy, manic depression/schizophrenia, chronic renal disease, malabsorption, chronic liver/pancreatic disease, NSAIDs, anticoagulants, antidepressants and antipsychotics. In addition for men, there were interactions between age and atrial fibrillation and cardiovascular disease. The interactions with age indicated higher HRs for these risk factors among younger patients compared with older patients.

Increasing material deprivation (as measured by the Townsend score) was associated with increasing risk of admission. Women in the Pakistani, Caribbean and Black African groups had significantly increased risks of emergency admission compared with women who were white or who did not have ethnicity recorded. Men in the Indian, Bangladeshi, Chinese and the other Asian groups had significantly lower risks compared with men who were white or who did not have ethnicity recorded.

Prior emergency admission to hospital was associated with increased risk of emergency admission in men and women. For example, compared with men with no emergency admissions in the previous 12 months, there was a 2.7-fold increased risk in men with one previous

admission; a 4.4-fold increased risk for two prior admissions and an 8.3-fold increased risk for those with three or more prior admissions. There was a similar pattern for women.

There was a 'dose-response' relationship for smoking with heavy smokers having higher risks than moderate smokers, light smokers or ex-smokers. There was a 'J-shaped' effect for alcohol with lower risks for those recorded as trivial, light or moderate drinkers and higher risks for those recorded as very heavy drinkers or non-drinkers. This was despite adjustment for a diagnosis of chronic liver/pancreatic disease and the presence of abnormal liver function tests.

All the other comorbidities and medications in the table were significantly associated with increased risks in men and women. Patients with a haemoglobin value of <11 g/dL, those with raised platelets and those with at least one abnormal liver function test also had increased risks of emergency admission.

### Calibration and discrimination in the validation cohort

In the QResearch validation cohort, the QAdmissions risk scores calculated using the GP-HES-linked data explained 41% of the variation in women and 43% of that in men (table 4). The D statistic was 1.7 in women and 1.8 in men. The ROC value was 0.77 for women and 0.78 for men.

Figure 1 displays the predicted and observed risks of emergency admission at 2 years across each 10th of the predicted risk (1 representing the lowest risk and 10 the highest risk). This shows that the QAdmissions algorithm was well calibrated.

Table 5 shows the performance statistics for QAdmissions at different thresholds in the QResearch validation cohort using the GP-HES-linked data and the GP data alone. For example, for the top 10% of men and women at the highest risk based on the GP-HES data (ie, those with a score of 23% or higher), QAdmissions had a sensitivity of 39% and a positive predictive value (based on the observed risk at 2 years) of 42%.

The performance of the QAdmissions score calculated using the GP-HES-linked data was marginally better than that using data from the GP record alone. For example, the ROC values for women were 0.77 using the GP-HES-linked data and 0.76 for the GP data alone (table 4). Calibration was similar.

The results for the validation statistics in the CPRD cohort were very similar to those for the QResearch validation cohort, as shown in tables 4 and 5.

Figures 2 and 3 show clinical examples of applying the QAdmissions score to two individual patients.

## DISCUSSION

### Summary of key findings

We have developed and externally validated a new algorithm (QAdmissions) to identify patients at high risk of



**Table 3** Adjusted HRs (95% CI) for emergency admission to hospital for the final QAdmissions model in the derivation cohort. HRs are adjusted for fractional polynomial terms for age and BMI

	Women adjusted HR* (95% CI)	Men adjusted HR* (95% CI)
<b>Ethnicity</b>		
White/not recorded	1.00	1.00
Indian	1.00 (0.95 to 1.06)	0.92 (0.87 to 0.97)
Pakistani	1.18 (1.11 to 1.26)	1.01 (0.94 to 1.08)
Bangladeshi	1.03 (0.96 to 1.11)	0.86 (0.79 to 0.92)
Other Asian	0.88 (0.83 to 0.94)	0.87 (0.81 to 0.93)
Caribbean	1.21 (1.15 to 1.28)	1.16 (1.10 to 1.24)
Black African	1.23 (1.17 to 1.29)	0.95 (0.89 to 1.01)
Chinese	0.48 (0.43 to 0.54)	0.43 (0.37 to 0.49)
Other	1.03 (0.99 to 1.08)	0.95 (0.90 to 1.00)
<b>Strategic Health Authority (SHA)</b>		
East Midlands SHA	1.00	1.00
Yorkshire & Humber SHA	1.09 (1.06 to 1.12)	1.10 (1.07 to 1.13)
East of England SHA	0.99 (0.96 to 1.02)	1.00 (0.97 to 1.03)
London SHA	0.97 (0.95 to 0.99)	0.91 (0.89 to 0.94)
North East SHA	1.19 (1.16 to 1.23)	1.16 (1.12 to 1.19)
North West SHA	1.15 (1.12 to 1.18)	1.16 (1.13 to 1.19)
South Central SHA	0.98 (0.96 to 1.01)	0.99 (0.96 to 1.02)
South East SHA	1.04 (1.01 to 1.07)	1.02 (0.99 to 1.05)
South West SHA	1.00 (0.97 to 1.02)	1.01 (0.98 to 1.04)
West Midlands SHA	1.08 (1.05 to 1.11)	1.07 (1.04 to 1.10)
<b>Smoking status</b>		
Non-smoker	1.00	1.00
Ex-smoker	1.13 (1.11 to 1.14)	1.14 (1.12 to 1.15)
Light smoker (1–9/day)	1.31 (1.29 to 1.33)	1.36 (1.34 to 1.39)
Moderate smoker (10–19/day)	1.31 (1.28 to 1.35)	1.40 (1.37 to 1.44)
Heavy smoker (20+/day)	1.41 (1.37 to 1.46)	1.54 (1.50 to 1.59)
<b>Alcohol status</b>		
Non-drinker	1.00	1.00
Trivial <1 unit/day	0.85 (0.84 to 0.86)	0.85 (0.83 to 0.86)
Light 1–2 units/day	0.80 (0.79 to 0.82)	0.81 (0.79 to 0.82)
Moderate 3–6 units/day	0.82 (0.80 to 0.84)	0.81 (0.79 to 0.82)
Heavy 7–9 units/day	1.27 (1.17 to 1.37)	0.94 (0.90 to 0.97)
Very heavy >9 units/day	1.28 (1.17 to 1.39)	1.16 (1.11 to 1.22)
<b>Emergency admissions in the last year</b>		
None	1.00	1.00
1 emergency admission	2.74 (2.68 to 2.81)	2.62 (2.55 to 2.69)
2 emergency admissions	4.44 (4.27 to 4.62)	4.43 (4.23 to 4.64)
3+ emergency admissions	7.48 (7.14 to 7.84)	8.27 (7.85 to 8.71)
<b>Clinical values and deprivation</b>		
Townsend Score (5 unit increase)	1.10 (1.09 to 1.11)	1.11 (1.10 to 1.12)
Most recent haemoglobin <11 g/dL†	1.30 (1.27 to 1.32)	1.60 (1.54 to 1.65)
Most recent platelet >480†	1.28 (1.23 to 1.33)	1.25 (1.18 to 1.32)
Most recent LFTs 3 times normal†	1.44 (1.39 to 1.49)	1.48 (1.44 to 1.53)
<b>Co-morbidity</b>		
Type 1 diabetes†	2.17 (2.04 to 2.30)	2.15 (2.03 to 2.29)
Type 2 diabetes†	1.37 (1.31 to 1.43)	1.33 (1.27 to 1.40)
Atrial fibrillation†	1.32 (1.28 to 1.35)	1.77 (1.62 to 1.93)
Cardiovascular disease†	1.36 (1.34 to 1.38)	1.80 (1.71 to 1.89)
Congestive cardiac failure†	1.19 (1.15 to 1.22)	1.27 (1.23 to 1.30)
Venous thromboembolism†	1.41 (1.34 to 1.47)	1.66 (1.56 to 1.76)
Cancer†	1.35 (1.32 to 1.37)	1.44 (1.41 to 1.47)
Asthma or COPD†	1.20 (1.18 to 1.22)	1.20 (1.18 to 1.22)
Epilepsy†	1.59 (1.52 to 1.66)	1.71 (1.64 to 1.79)
Falls†	1.27 (1.25 to 1.29)	1.36 (1.33 to 1.38)
Manic depression or schizophrenia†	1.37 (1.30 to 1.44)	1.39 (1.31 to 1.48)
Chronic renal disease†	2.10 (1.94 to 2.27)	1.86 (1.70 to 2.03)

Continued

Table 3 Continued

	Women adjusted HR* (95% CI)	Men adjusted HR* (95% CI)
Conditions causing malabsorption†	1.47 (1.40 to 1.55)	1.60 (1.51 to 1.69)
Liver disease or chronic pancreatitis†	1.54 (1.44 to 1.64)	1.91 (1.81 to 2.03)
Medications		
NSAIDs†	1.35 (1.33 to 1.38)	1.48 (1.45 to 1.51)
Anticoagulant†	1.69 (1.57 to 1.82)	1.61 (1.49 to 1.75)
Corticosteroids†	1.50 (1.47 to 1.52)	1.52 (1.49 to 1.55)
Antidepressant†	1.66 (1.64 to 1.69)	1.72 (1.68 to 1.75)
Antipsychotic†	1.68 (1.64 to 1.73)	1.60 (1.53 to 1.66)

Final model included age interaction terms.

Notes: Models also included fractional polynomial terms for age and body mass index.

\*HRs simultaneously adjusted for all the other variables shown in the table as well as fractional polynomial terms for age and body mass index.

†compared with patients without the condition/medication at baseline.

For women: fractional polynomial terms were  $(age/10)^{-2}$  and  $(age/10)^{-2} \ln(age)$ ;  $(bmi/10)^{-2}$  and  $(bmi/10)^{-2} \ln(bmi)$ .

For men: fractional polynomial terms were  $(age/10)^{-2}$  and  $(age/10)^{-2} \ln(age)$ ;  $(bmi/10)^{-2}$  and  $(bmi/10)^{-2} \ln(bmi)$ .

The models for men and women also included interactions between the age terms and prior admissions, type2 diabetes, venous thromboembolism, epilepsy, manic depression/schizophrenia, chronic renal disease, malabsorption, chronic liver/pancreatic disease, NSAIDs, anticoagulants, antidepressants and antipsychotics. In addition for men, there were interactions between the age terms and atrial fibrillation and cardiovascular disease. HRs for these variables in the table are evaluated at mean age in men and women.

BMI, body mass index; COPD, chronic obstructive pulmonary disease; LFTs, liver function tests; NSAIDs, non-steroidal anti-inflammatory drugs.

emergency admission to hospital using contemporaneous primary care data from the UK. The algorithm incorporates 30 predictor variables which are associated with increased risk of hospital admission including socio-demographic variables, lifestyle, morbidity, medication and laboratory results such as anaemia and abnormal liver function tests. The algorithm can be applied to any adult in a primary care setting regardless of whether they have had a prior emergency admission. The algorithm is intended to be used for regular batch processing of a dataset containing an entire population to generate a rank-ordered list of patients at high risk for further assessment and management. It can be integrated into GP clinical computer systems by the systems

suppliers in a way similar to how other risk prediction tools such as QRISK2,<sup>17</sup> QDiabetes<sup>24</sup> and QFracture<sup>25</sup> have been implemented. Alternatively, a stand-alone version is available at the publicly available website <http://www.qadmissions.org>. This can be used for the assessment of individual patients.

QAdmissions provides an estimate of absolute risk of admission either at 1 or 2 years—the latter being potentially useful for interventions which are likely to work over a more extended time period. It includes a weighting for geographical area at the SHA level to help take account of local differences in configuration of services. Like the CPM,<sup>11</sup> it can be applied across the general population to help health organisations to design and

**Table 4** Validation statistics for the QAdmissions prediction algorithm in the QResearch and CPRD validation cohorts using (a) the score calculated using the GP-HES-linked data and (b) the score calculated using the GP data alone

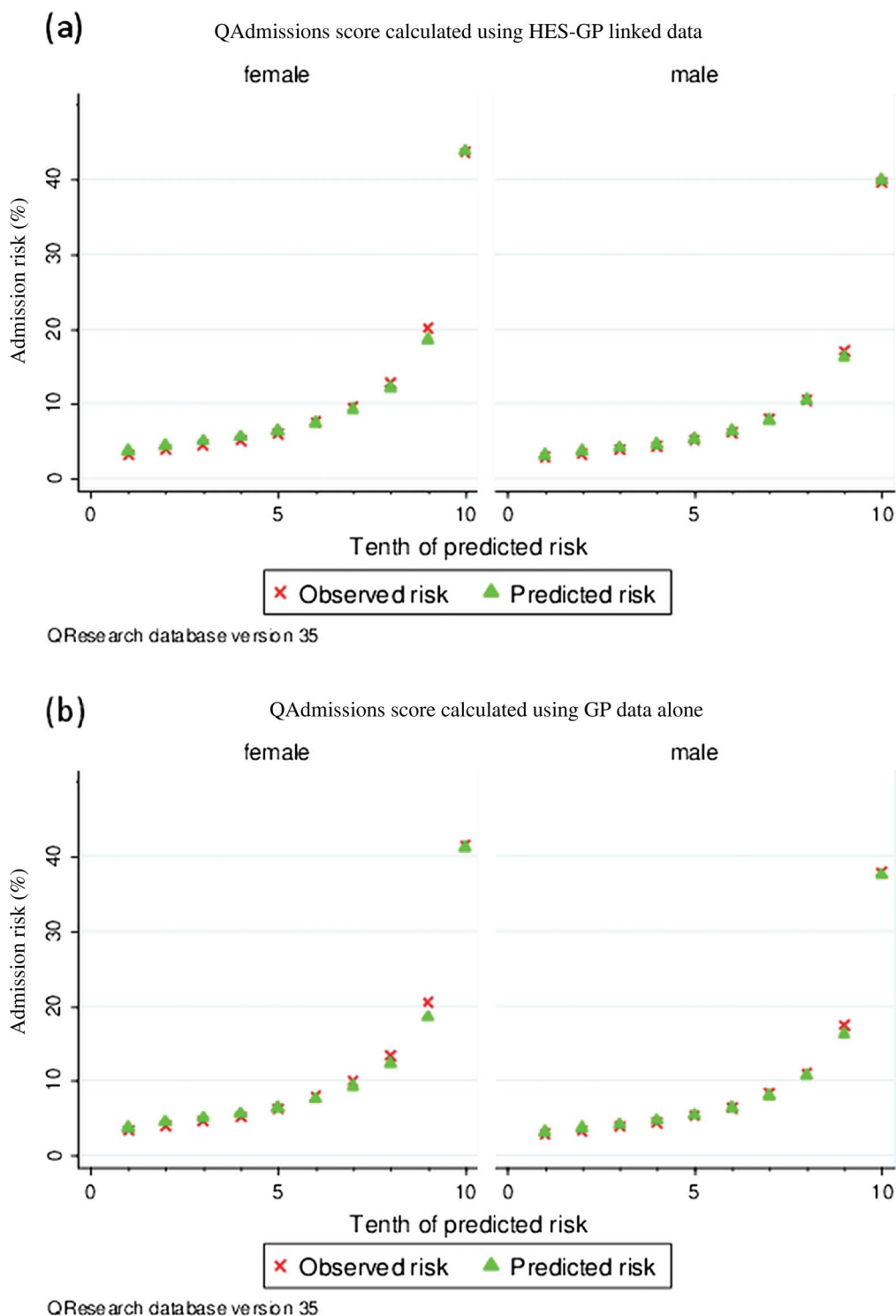
	QResearch validation cohort		CPRD validation cohort	
	HES-GP linked-data	GP data alone	HES-GP-linked data	GP data alone
Women				
ROC statistic	0.773 (0.771 to 0.774)	0.764 (0.762 to 0.766)	0.771 (0.770 to 0.773)	0.764 (0.763 to 0.766)
R <sup>2</sup> (%)	40.6 (40.2 to 40.9)	37.3 (37.0 to 37.8)	40.5 (40.2 to 40.8)	37.6 (37.3 to 37.9)
D statistic	1.69 (1.68 to 1.70)	1.58 (1.57 to 1.59)	1.69 (1.68 to 1.70)	1.59 (1.58 to 1.60)
Men				
ROC statistic	0.776 (0.774 to 0.778)	0.769 (0.767 to 0.771)	0.772 (0.771 to 0.774)	0.767 (0.765 to 0.768)
R <sup>2</sup> (%)	42.6 (42.2 to 42.9)	39.5 (39.1 to 39.9)	41.9 (41.6 to 42.2)	39.2 (38.9 to 39.5)
D statistic	1.76 (1.75 to 1.78)	1.65 (1.64 to 1.67)	1.74 (1.73 to 1.75)	1.64 (1.63 to 1.65)

CPRD, Clinical Practice Research DataLink; HES-GP, hospital episode statistics-general practitioner

Notes on understanding validation statistics: Discrimination is the ability of the risk prediction model to differentiate between patients who experience an admission event during the study and those who do not. This measure is quantified by calculating the area under the receiver operating characteristic curve (ROC) statistic, where a value of 1 represents perfect discrimination.

The D statistic is also a measure of discrimination which is specific to censored survival data. As with the ROC, higher values indicate better discrimination.

R<sup>2</sup> is another measure specific to censored survival data—it measures explained variation and higher values indicate more variation is explained.



**Figure 1** Mean predicted risks and observed risk of emergency admission to hospital at 2 years by 10th of the predicted risk applying the QAdmissions risk prediction scores to all patients in the QResearch validation cohort (results from Clinical Practice Research DataLink available from the authors).

implement interventions across the risk spectrum as follows: prevention and wellness promotion for low-risk patients; supported self-care interventions for moderate risk patients; early intervention care management for patients with emerging risk and intensive case management for very high-risk patients.<sup>11</sup>

We undertook an additional validation by applying the final QAdmissions model to GP data alone and compared with the results using GP-HES-linked data. The results in both the QResearch and CPRD validation cohorts were comparable and hence provide evidence to support the implementation of QAdmissions

**Table 5** Performance of QAdmissions for predicting emergency admissions in the QResearch and CPRD validation cohorts based on (a) the score calculated using the GP-HES-linked data and (b) the score calculated using the GP data alone.

2 year risk score	QResearch validation cohort				CPRD validation cohort			
	Cut-off for 2 year predicted risk (%)	Total classified as high risk	Sensitivity (%)	Observed risk of admission at 2 years*	Cut-off for 2 year predicted risk (%)	Total classified as high risk	Sensitivity (%)	Observed risk of admission at 2 years*
HES-GP linked data								
Top 1%	69.2	13 406	6.6	72.5	67.5	24 753	6.7	72.7
Top 5%	35.9	67 031	24.6	53.0	35.1	123 768	24.9	53.3
Top 10%	23.0	134 062	39.3	41.8	22.4	247 536	39.4	41.8
Top 20%	13.4	268 124	56.9	30.0	13.1	495 072	56.8	29.9
GP data only								
Top 1%	56.7	13 406	6.0	65.9	65.6	24 753	6.1	66.0
Top 5%	30.9	67 031	23.4	50.0	35.9	123 768	23.2	49.6
Top 10%	20.6	134 062	37.7	39.8	23.8	247 536	37.4	39.7
Top 20%	12.6	268 124	55.5	29.1	14.2	495 072	55.1	29.1

\*Observed risk is an estimate of the positive predictive value.

within GP computer systems based solely on GP data. This potentially overcomes one of the main logistical difficulties in implementing other risk scores since they require real-time data linkage of primary data with secondary care data. Much of the apparent complexity relating to additional variables and interactions can be incorporated into the software using data already entered into the patient's electronic health record. The algorithm uses routinely collected data, which means it can be easily and regularly updated to reflect changes in populations, improvements in data quality or coding, advances in knowledge and evolving guidelines.

As with the PEONY algorithm,<sup>10</sup> QAdmissions includes age, deprivation, prior emergency admission and medications (eg, antidepressants, antipsychotics and analgesics) and these were all significantly associated with an increased risk of emergency admission. We found similar interactions between these variables and age with higher risks in younger patients, which diminished with increasing age. We have included many more emergency admissions in the derivation sample (265 573 events rather than 6793); more up-to-date data (2010–2011 rather than 1999–2004), which is important given the rise in emergency admission rates over the last 10 years. In contrast to PEONY, QAdmissions has been modelled using a more ethnically diverse population and includes morbidity in addition to prescribed medication. Apart from prior hospital admissions, all of the variables in the model are derived from the primary care record.

Although not directly comparable because of differences in the samples to which the algorithms can be applied and also the outcomes predicted, the positive predictive value for the top 1% of patients at highest risk was higher for QAdmissions (73%) than PEONY (59%), although the sensitivity was similar (7% vs 8%). Our

ROC value of 0.77 is comparable to the value of 0.79 reported in the validation cohort of PEONY and significantly higher than the 0.69 reported by the authors of the PARR score<sup>4</sup> and the 0.70 for PARR-30.<sup>26</sup> Our ROC value is also significantly higher than that reported by Donze *et al* (0.71), although their risk prediction model was designed to identify patients at high risk of 30 day readmission to hospital, which is an outcome different from the one in our study.<sup>27</sup>

We have not provided definite comment on the threshold of absolute risk that should be used for intervention, as that would require cost-effectiveness analyses which are outside the scope of this study. We have, however, provided analyses using a range of thresholds of risk, which can be used to help inform future analyses. Sensitivity is important as it is a measure of how well the algorithm performs in finding cases that might be suitable for intervention. If the risk threshold is set too high, then the sensitivity will be low and a large number of patients with emergency admission will be 'missed' by the algorithm. Conversely, a high-risk threshold is likely to result in a better positive predictive value, which means a higher proportion of those identified are likely to go on to have an emergency admission. So there is a balance to be struck between the sensitivity and positive predictive value of the score, which depends on the risk threshold selected, resources available and likely effectiveness of the interventions. For example, if the top 1% of patients at highest risk are targeted, then patients with an estimated absolute risk of admission of greater than 69% will be identified. This will have a good positive predictive value (73%) but a low sensitivity (7%). If the top 10% of patients at highest risk are identified, the sensitivity at this threshold will be 39% and the positive predictive value will be 42%. However, more patients will require assessment, so the costs of the intervention will be higher.



53 year old white man from the South West, ex-smoker, drinks 7-9 units/day, has type 2 diabetes, body mass index of 39.1 kg/m<sup>2</sup> currently prescribed antidepressants, with last haemoglobin of under 11g/dl and an abnormal liver function test, has a 29% risk of having an emergency admission within the next two years as shown in the screen shot below ([www.qadmissions.org](http://www.qadmissions.org)). This level of risk would put him in the top 10% of patients at risk of emergency admission. The same man without anaemia or abnormal LFTs would have a risk of 14%.

**ClinRisk** Welcome to the QAdmissions® risk calculator: <http://qadmissions.org>

Reset Information Publications About Copyright Contact Us Algorithm Software

Calculate risk over 2 years. Calculate risk

About the person

Age (18-100): 53  
 Sex:  Male  Female  
 Ethnicity: White or not stated  
 UK postcode: leave blank if unknown  
 Postcode:

Necessary information

Statagic health authority: South West  
 Smoking status: ex-smoker  
 Alcohol status: 7-9 units per day  
 Emergency admissions in the last year: none  
 diabetes: type 2  
 Heart attack, angina, stroke or TIA?   
 Atrial fibrillation?   
 Congestive cardiac failure?   
 Chronic renal disease?   
 Venous thrombo-embolism?   
 Cancer?   
 Asthma or COPD?   
 Falls?   
 Epilepsy?   
 Manic depression or schizophrenia?   
 Malabsorption eg Crohn's disease, ulcerative colitis, coeliac disease, steatorrhea, blind loop syndrome?   
 Chronic liver/pancreatic disease?   
 Blood tests  
 Abnormal LFTs (GGT, ALT or bilirubin more than 3 times normal)?   
 Anaemia (hb under 11g/dl)?   
 Platelets > 480?   
 Drugs  
 taking anticoagulants (eg warfarin)?   
 taking antidepressants?   
 taking antipsychotics?   
 taking anti-inflammatory painkillers (NSAIDs)?   
 taking steroid tablets regularly?   
 Leave blank if unknown  
 Body mass index  
 Height (cm): 160  
 Weight (kg): 100

**The results**

The person's risk of having an emergency hospital admission within the next two years is:

**29%**

In other words, in a crowd of 100 people with the same risk factors as the person, 29 are likely to have an emergency hospital admission within the next two years.

Risk of an emergency hospital admission

The person's score has been calculated using estimated data, as some information was left blank.

The person's body mass index was calculated as 39.06 kg/m<sup>2</sup>.

**Figure 2** Clinical case.

### Strengths and limitations of this study

The methods to derive and validate this model are the same as for a range of other clinical risk prediction tools derived from the QResearch database.<sup>16 17 24 25 28</sup> The strengths and limitations of the approach have already been discussed in detail<sup>15 16 24 29-31</sup> including information on multiple imputation of missing data. In summary, the key strengths include size, duration of

follow-up, representativeness and lack of selection, recall and respondent bias. UK general practices have good levels of accuracy and completeness in recording clinical diagnoses and prescribed medications.<sup>32 33</sup> We think our study has good face validity since it has been conducted in a setting where the majority of patients in the UK are assessed, treated and followed up. Limitations include lack of formally adjudicated outcomes, information bias

A 65 year old Bangladeshi woman from the North West, non-smoker, non-drinker, with a body mass index of 26.3 Kg/m<sup>2</sup> who has atrial fibrillation on anticoagulants and steroid, with one emergency admission in the last year, has a 39% risk of emergency admission in the next two years as shown in the screen shot below.

**ClinRisk** Welcome to the QAdmissions® risk calculator: <http://qadmissions.org>

Reset Information Publications About Copyright Contact Us Algorithm Software

Calculate risk over 2 years. Calculate risk

**About the person**

Age (18-100): 65  
 Sex:  Male  Female  
 Ethnicity: Bangladeshi  
 UK postcode: leave blank if unknown  
 Postcode:

**Necessary information**

Statice health authority: North West  
 Smoking status: non-smoker  
 Alcohol status: none  
 Emergency admissions in the last year: one  
 diabetes: none  
 Heart attack, angina, stroke or TIA?   
 Atrial fibrillation?   
 Congestive cardiac failure?   
 Chronic renal disease?   
 Venous thrombo-embolism?   
 Cancer?   
 Asthma or COPD?   
 Falls?   
 Epilepsy?   
 Manic depression or schizophrenia?   
 Malabsorption eg Crohn's disease, ulcerative colitis, coeliac disease, steatorrhea, blind loop syndrome?   
 Chronic liver/pancreatic disease?

**Blood tests**

Abnormal LFTs (GGT, ALT or bilirubin more than 3 times normal)?   
 Anaemia (hb under 11g/dl)?   
 Platelets > 480?

**Drugs**

taking anticoagulants (eg warfarin)?   
 taking antidepressants?   
 taking antipsychotics?   
 taking anti-inflammatory painkillers (NSAIDs)?   
 taking steroid tablets regularly?

Leave blank if unknown

**Body mass index**

Height (cm):  
 Weight (kg):

**The results**

The person's risk of having an emergency hospital admission within the next two years is:

**38.5%**

In other words, in a crowd of 100 people with the same risk factors as the person, 39 are likely to have an emergency hospital admission within the next two years.

**Risk of an emergency hospital admission**

The person's score has been calculated using estimated data, as some information was left blank.

The person's body mass index was estimated as 26.3 kg/m<sup>2</sup>.

**Figure 3** Clinical case.

and potential for bias due to missing data. Our database has linked data for admission to hospital and is therefore likely to have picked up the majority of emergency admissions, thereby minimising ascertainment bias. There is scope for improvement in the recording of emergency admission on the GP clinical record as some codes are used which identify an admission has occurred but not the method or type of admission. An information standard for recording of hospital admissions on GP clinical records could help address this and is likely to improve the performance of the score when applied to GP data alone.

We excluded people without a valid NHS number as this was required to link the primary and secondary care

data for individual patients. We also excluded patients without a valid deprivation score since this group may represent a more transient population where follow-up could be unreliable or unrepresentative. Their deprivation scores are unlikely to be missing at random, so we did not think it would be appropriate to impute them.

The present validation has been done on two completely separate sets of practices and individuals to those which were used to develop the score. One of the validation cohorts was derived from the QResearch database, so the practices all use the same GP clinical computer system (EMIS—the computer system used by 55% of UK GPs). The favourable results from the validation which uses CPRD is a more stringent test since this is a fully

external set of practices which use a different computer system. Ideally, an additional validation should be undertaken using another external data source by an independent team not involving the study authors.

This QAdmissions model has been developed using data from general practices in England and includes a postcode-based deprivation score. It is therefore not likely to be applicable for clinical use in international settings without some modification of the English-specific risk factors, and validation in the setting in which it is intended to be used.

In summary, we have developed and validated a new algorithm to predict risk of emergency hospital admission. QAdmissions has some advantages compared with the current risk-scoring methods. QAdmissions also provides an accurate measure of absolute risk of emergency hospital admission in the general population as shown by its performance in a separate validation cohort. Further research is needed to evaluate the clinical outcomes and cost-effectiveness of using this algorithm in primary care.

**Acknowledgements** The authors would like to acknowledge the contribution of EMIS practices who contribute to QResearch and the University of Nottingham and EMIS for expertise in establishing, developing and supporting the database. They also acknowledge the contribution of the NHS Information Centre for pseudonymising the Hospital Episodes Statistics dataset so that the data could be linked to patients in the QResearch database.

**Contributors** JHC initiated the study, undertook the literature review, data extraction, data manipulation and primary data analysis and wrote the first draft of the paper. CC contributed to the design, analysis, interpretation and drafting of the paper. All authors have read and approved the final version of the manuscript.

**Funding** The North East London Commissioning support group provided limited funding to support the later stages of this work. The National School for Primary Care Research contributed to the license costs of the Clinical Practice Research Data Link which was used for the external validation of QAdmissions.

**Competing interests** JHC is the 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 the 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 the associate professor of Medical Statistics at the University of Nottingham and a consultant statistician for ClinRisk Ltd.

**Ethics approval** The project was approved in accordance with the QResearch agreement with Trent Multi-Centre Research Ethics Committee. The validation of QAdmissions on CPRD was approved by the Independent Scientific Advisory Group (Reference 13\_079).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** The patient level data from QResearch are specifically licensed according to its governance framework. See <http://www.qresearch.org> for further details. The QAdmissions algorithm will be published as open source software under the AGPLv3 Public License.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

## REFERENCES

- Lewis G, Curry N, Bardsley M. *Choosing a predictive risk model: a guide for commissioners in England*. Nuffield trust, 2011:20.
- NHS England. Enhanced service specification: Risk profiling and care management scheme. Secondary Enhanced service specification: risk profiling and care management scheme. 2013. <http://www.england.nhs.uk/wp-content/uploads/2013/03/ess-risk-profiling.pdf>
- Bottle A, Aylin P, Majeed A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. *J R Soc Med* 2006;99:406–14.
- Billings J, Dixon J, Mijanovich T, et al. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. *BMJ* 2006;333:327.
- ISD Scotland. Scottish Patients at Risk of Readmission and Admission (SPARRA)—a report on the development of SPARRA. Secondary Scottish Patients at Risk of Readmission and Admission (SPARRA)—a report on the development of SPARRA. 2011. <http://www.isdscotland.org/Health-Topics/Health-and-Social-Community-Care/SPARRA/2012-02-09-SPARRA-Version-3.pdf>
- Coleman EA, Wagner EH, Grothaus LC, et al. Predicting hospitalization and functional decline in older health plan enrollees: are administrative data as accurate as self-report? *J Am Geriatr Soc* 1998;46:419–25.
- Marcantonio ER, McKean S, Goldfinger M, et al. Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan. *Am J Med* 1999;107:13–17.
- Reuben DB, Keeler E, Seeman TE, et al. Development of a method to identify seniors at high risk for high hospital utilization. *Med Care* 2002;40:782–93.
- Lyon D, Lancaster GA, Taylor S, et al. Predicting the likelihood of emergency admission to hospital of older people: development and validation of the Emergency Admission Risk Likelihood Index (EARLI). *Fam Pract* 2007;24:158–67.
- Donnan PT, Dorward DWT, Mutch B, et al. Development and validation of a model for predicting emergency admissions over the next year (PEONY): A UK Historical Cohort Study. *Arch Intern Med* 2008;168:1416–22.
- Wennberg D, Siegel MB, Darin B, et al. *Combined predictive model—final report*. London: The Kings Fund, 2006.
- Department of Health. Risk Stratification and next steps with DH Risk Prediction tools—Patients at Risk of Re-hospitalisation and the Combined Predictive Model. Secondary Risk Stratification and next steps with DH Risk Prediction tools—Patients at Risk of Re-hospitalisation and the Combined Predictive Model. 2011. [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/147179/dh\\_129005.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/147179/dh_129005.pdf)
- 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;335:136.
- Hippisley-Cox J, Vinogradova Y, Coupland C, et al. *Comparison of key practice characteristics between general practices in England and Wales and general practices in the QRESEARCH data*. University of Nottingham; 2005, Report to the Health and Social Care Information Centre.
- 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;94:34–9.
- 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.
- Hippisley-Cox J, Coupland C, Vinogradova Y, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475. Published Online First: Epub Date. doi: 10.1136/bmj.39609.449676.25
- 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:f2573.
- Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;28:964–74.
- Royston P. Multiple imputation of missing values. *Stata J* 2004;4:227–41.
- Steyerberg E. *Clinical prediction models*. Springer, 2009.



22. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;23:723–48.
23. Royston P. Explained variation for survival models. *Stata J* 2006;6:1–14.
24. 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.
25. 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:e3427.
26. Billings J, Blunt I, Steventon A, *et al*. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). *BMJ Open* 2012;2:e001667.
27. Donzé J, Aujesky D, Williams D, *et al*. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA Intern Med* 2013;173:632–38.
28. 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 Scores. *BMC Fam Pract* 2010;11:49.
29. 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.
30. 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.
31. Collins GS, Altman DG. External validation of the QDScore for predicting the 10-year risk of developing Type 2 diabetes. *Diabetic Med* 2011;28:599–607.
32. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991;302:766–8.
33. Majeed A. Sources, uses, strengths and limitations of data collected in primary care in England. *Health Stat Q* 2004;21:5–14.