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

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003482
Article Type:	Research
Date Submitted by the Author:	26-Jun-2013
Complete List of Authors:	Hippisley-Cox, Julia; University of Nottingham, ; ClinRisk Ltd, Coupland, Carol; University of Nottingham, Division of Primary Care
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Health informatics
Keywords:	PRIMARY CARE, PREVENTIVE MEDICINE, EPIDEMIOLOGY

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

Paper Presenting Original Research

Submitted to the BMJ Open June 2013

(original reference BMJ.2013.012731)

Authors

Julia Hippisley-Cox Professor of Clinical Epidemiology & General Practice[1]

Carol Coupland Associate Professor and Reader in Medical Statistics[1]

Institutions

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

Guarantor & Author for correspondence - Julia Hippisley-Cox

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

Telephone: 0115 8466915

Fax: 0115 8466904

ABSTRACT

Objective To develop and externally validate a risk algorithm (QAdmissions) to estimate 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 two year study period 01 January 2010 to 31 Dec 2011.

Setting 405 general practices in England contributing to the national QResearch database to develop the algorithm. Two validation cohorts to validate the algorithm (a) 202 different QResearch practices and (b) 343 practices in England contributing to the Clinical Practice Research DataLink (CPRD). All general practices had data linked to hospital episode statistics at 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. 132,723 of these patients had one or more emergency admissions during follow-up. For the CPRD cohort identified 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 both a valid NHS number and a valid Townsend score.

Endpoint First (i.e. incident) emergency admission to hospital in the next two years as recorded on the linked hospital episodes records.

Risk factors Candidate variables recorded on the GP computer system including (a) demographic variables (age, sex, strategic health authority, Townsend deprivation score, ethnicity); (b) lifestyle variables (smoking, alcohol intake); (c) chronic diseases; (d) prescribed medication; (e) clinical values (body mass index, systolic blood pressure); (f) laboratory test results (haemoglobin, platelets, erythrocyte sedimentation rate (ESR), ratio of total serum cholesterol to high density lipoprotein cholesterol concentrations, liver function tests). We also included the number of emergency admissions in 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% 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.

Conclusion. 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.

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.

Web calculator

Here is a simple web calculator to implement the QAdmissions algorithm which will be publically available alongside the paper. It also has the open source software for download.

URL	http://qadmissions.org
Username	reviewer
Password	TrueFirstHorses

1 Introduction

Unplanned admissions account for an estimated 11 billion pounds a year in England which is a considerable proportion of the NHS budget[1]. Not only are such admissions 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 GP practices for the identification and case management of patients identified as seriously ill or at risk of an emergency admission[2]. 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 program, 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 (e.g. 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 (e.g. emergency admission) for the population of interest (e.g. all adult patients registered with the general practitioner). It needs to apply over the relevant time period (e.g. 1-2 years) assuming 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 (i.e. help reduce risk of emergency admission). It should preferably incorporate measures of socio-economic deprivation and ethnicity 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 programs are introduced. The risk algorithm needs to have the potential to be updated or recalibrated and its performance should be tested in a separate population of patients from that used to develop the tool to demonstrate that it can reliably identify the target population. Lastly, the tool needs to be suitable for implementation in clinical practice.

Whilst 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 re-admission [3-5]. 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 generalizability [6-8]. There are several tools which have been intended for use in primary care. The Emergency Admission Risk Likelihood Index (EARLI) is a six item questionnaire which was developed using data from patients aged 75+ from 17 general practices in the North of England [9]. Hence it only applies to elderly patients and may not be sufficiently representative for wider use. The PEONY score was designed for use in Scottish primary care patients aged 40-65 years [10]. However, it does not include morbidity data from primary care and currently the underlying algorithm is not published or independently validated. Lastly the Combined Predictive Model[11] (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 tools were outdated and in urgent need of a refresh[12].

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 already in GP computer systems or been validated in that setting. Whilst it's 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 health care professional in the 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.

2 Methods

2.1 Study design and data source

We conducted a prospective cohort study studying a large UK primary care population using a similar method to our analyses for other risk prediction scores such as QRISK2 [13]. Version 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[13]. Practices and patients contained on the database are nationally representative[14] and similar to those on other primary care databases using other clinical software systems[15]. 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 individual patient level via pseudonymised NHS number. We also assembled a second validation cohort using 343 English practices contributing to the Clinical Research Data Link which had linked HES data (August 2012 download).

2.2 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 01 January 2010 and 31 Dec 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 01 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 one year, date on which the practice computer system was installed plus one year, and the beginning of the study period (01 January 2010). Patients were censored at the earliest date of: 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 Dec 2011).

2.3 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 Hospital Episode Statistics (HES) data which includes all hospital trusts in England. The hospital episode data was linked at individual patient level to the QResearch database via pseudonymised NHS number. Emergency admissions were identified by selecting the standard codes to represent all emergency admission 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 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.

2.4 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[10] or re-admission[4 7]. We also included predictors used in other risk algorithms where the outcome is likely to require emergency admission (for example as thrombosis[16] or cardiovascular disease[17 18]). 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 similar way to the implementation of similar risk prediction

algorithms developed using the QResearch database [4 11-14]. The full list of candidate variables is shown in Table 1 and is summarized here:

(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, 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, ESR, total serum cholesterol/HDL 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 (a) asthma and chronic obstructive airways disease; (b) schizophrenia and manic depression. We defined abnormal liver function tests as a single variable which denoted either a high gamma GT, AST or bilirubin where a high value was at least three times the upper limit of normal.

2.5 Model derivation and development

As in previous studies[17], we used the Cox proportional hazards model in the derivation dataset to estimate the coefficients and hazard ratios 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[19 20]. We tested for interactions between each variable and age and included significant interactions in the final model where they improved model fit. Continuous variables were centered 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[20] [21]. We took the logarithm of the hazard ratio 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 year 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 modeling in the derivation cohort[22].

2.6 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 Research Data Link (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[13] and compared these by tenth of predicted risk for each score. The observed risk at 2 years was obtained using the 2 year Kaplan-Meier estimate. We calculated the ROC statistic, D statistic (a measure of discrimination where higher values indicate better discrimination)[23] and an R squared statistic (which is a measure of explained variation for survival data where higher values indicate more variation is explained)[24].

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 two 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 was 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 appendix. This was then used alongside the other GP data derived predictor variables to calculate the risk scores. This was done to evaluate the performance of the algorithm in a primary care setting where GP-HES linked data is not available (GP-HES is not routinely available in all primary care settings).

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

3 Results

3.1 Practices and patients

Overall, 607 QResearch practices in England met our inclusion criteria and had been using their current computer system for at least one 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 4,518 (0.16%) had an invalid NHS number and 3,577 (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 1,254 (0.09%) had an invalid NHS number and 1,398 (0.10%) had a missing Townsend score leaving 1,340,622 eligible patients for analysis.

Error! Reference source not found. 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 recording of ethnicity was higher in the two QResearch cohorts (75% and 76%) than in CPRD (53%).

3.2 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 in each cohort. Overall in the derivation cohort, we identified 265,573 patients (9.3% of 2,857,476) 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 (32.6%) 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; 10,741 (4.0%) had two and 7,783 (2.9%) had 3 or more. 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 men and rose steeply with age. The age-sex standardized emergency admission rates varied between Strategic Health Authorities with highest rates in the 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.

3.3 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, 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 age and atrial fibrillation and cardiovascular disease. The interactions with age indicated higher hazard ratios 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 didn't have ethnicity recorded. For men, Indian, Bangladeshi, Chinese and the other Asian group had significantly lower risks compared with men who were white or who didn't 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 3 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 than 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 co-morbidities and medications in the table were significantly associated with increased risks in men and women. Patients with a haemoglobin value of <11g/dl, those with raised platelets and those with at least one abnormal liver function test also had increased risk of emergency admission.

3.4 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% 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 2 displays the predicted and observed risks of emergency admission at 2 years across each tenth of 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 highest risk based on the GP-HES data (i.e. those with a score of 23% or higher), then 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.

Box one shows a clinical example of applying the QAdmissions score to two individual patients.

4 Discussion

4.1 Summary of key findings

We have developed and externally validated a new algorithm (QAdmissions) to identify patients at high risk of 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 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 similar way to how other risk prediction tools such as QRISK2[17], QDiabetes[25] and QFracture[26] have been implemented. Alternatively, a standalone version is available at the publically available website www.qadmissions.org. This can be used for the assessment of individual patients

QAdmissions provides an estimate of absolute risk of admission either at one year or two 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 strategic health authority level to help take account of local differences in configuration of services. Like the Combined Predictive Model[11], it can be applied across the general population to help health organisations to design and implement interventions across the risk spectrum ranging from: 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[11].

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 within GP computer systems based on 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[10], QAdmissions includes age, deprivation, prior emergency admission and medications (e.g. antidepressants, antipsychotics, 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 modeled 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 both 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 0.79 reported in the validation cohort of PEONY and significantly higher than the 0.69 reported by the authors of the PARR score[4] and the 0.70 for PARR-30[27]. 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 re-admission to hospital which is a different outcome to the outcome in our study[28].

We have not provided definite comment on what threshold of absolute risk 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 patient 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.

4.2 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 [16 17 25 26 29]. The strengths and limitations of the approach have already been discussed in detail [15 16 25 30-32] including information on multiple imputation of missing data. In summary, 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,[33] [34]. We think our study has good face validity since it has been conducted in the setting where the majority of patients in the UK are assessed, treated and followed up. Limitations include lack of formally adjudicated outcomes, information bias, 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 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.

5 Other information

5.1 Acknowledgements

We acknowledge the contribution of EMIS practices who contribute to QResearch® and to the University of Nottingham and EMIS for expertise in establishing, developing and supporting the database. We acknowledge the contribution of the NHS Information Centre for pseudonymising the Hospital Episodes Statistics dataset so that data could be linked to patients in the QResearch database.

5.2 Approvals:

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)

5.3 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. CC contributed to the design, analysis, interpretation and drafting of the paper. All authors have read and approved the final manuscript.

5.4 Funding

There was no funding initially for this work. 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 Research Data Link which was used for the external validation of QAdmissions.

5.5 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.

5.6 Copyright

5.7 Data Sharing statement

The patient level data from the QResearch are specifically licensed according to its governance framework. See www.qresearch.org for further details. The QAdmissions algorithm will be published as open source software under the GNU Lesser Public License.

6 References

1. Lewis G, Curry N, Bardsley M. Choosing a predictive risk model: a guide for commissioners in England. In: Trust N, ed.: Nuffield Trust, 2011:20.
2. 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.
3. 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**(8):406-14 doi: 10.1258/jrsm.99.8.406[published Online First: Epub Date]|.
4. 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**(7563):327 doi: 10.1136/bmj.38870.657917.AE[published Online First: Epub Date]|.
5. 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.
6. 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? Journal of the American Geriatrics Society 1998;**46**(4):419-25
7. 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. The American journal of medicine 1999;**107**(1):13-7
8. Reuben DB, Keeler E, Seeman TE, et al. Development of a method to identify seniors at high risk for high hospital utilization. Medical care 2002;**40**(9):782-93 doi: 10.1097/01.MLR.0000024611.65466.AE[published Online First: Epub Date]|.

9. 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**(2):158-67 doi: 10.1093/fampra/cml069[published Online First: Epub Date]|.
10. 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**(13):1416-22 doi: 10.1001/archinte.168.13.1416[published Online First: Epub Date]|.
11. Wennberg D, Siegel MB, Darin B, et al. Combined predictive model - final report. London: The Kings Fund, 2006.
12. 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.pdf.
13. 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 doi: 10.1136/bmj.39261.471806.55[published Online First: Epub Date]|.
14. 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. Report to the Health and Social Care Information Centre.: University of Nottingham, 2005.
15. 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 doi: 10.1136/hrt.2007.134890[published Online First: Epub Date]|.
16. 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 doi: 10.1136/bmj.d4656[published Online First: Epub Date]|.
17. 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 doi: 10.1136/bmj.39609.449676.25[published Online First: Epub Date]|.
18. Hippisley-Cox J, Coupland C, Brindle P. Predicting risk of ischaemic stroke in primary care: derivation and validation of QStroke and a comparison with CHADS2 and CHA2DS2VASC. BMJ 2013:(in press)
19. 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
20. 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
21. Royston P. Multiple imputation of missing values. Stata Journal 2004;**4**(3):227-41
22. Steyerberg E. *Clinical Prediction Models*: Springer, 2009.

23. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. Stat Med 2004;**23**:723-48
24. Royston P. Explained variation for survival models. Stata J 2006;**6**:1-14
25. 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- doi: 10.1136/bmj.b880[published Online First: Epub Date]|.
26. 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 doi: 10.1136/bmj.e3427[published Online First: Epub Date]|.
27. 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**(4) doi: 10.1136/bmjopen-2012-001667[published Online First: Epub Date]|.
28. Donzé J ADWDSJL. Potentially avoidable 30-day hospital readmissions in medical patients: Derivation and validation of a prediction model. JAMA Internal Medicine 2013;**173**(8):632-38 doi: 10.1001/jamainternmed.2013.3023[published Online First: Epub Date]|.
29. 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
30. 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- doi: 10.1136/bmj.b4229[published Online First: Epub Date]|.
31. 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
32. 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 doi: 10.1111/j.1464-5491.2011.03237.x[published Online First: Epub Date]|.
33. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. British Medical Journal 1991;**302**:766-68
34. Majeed A. Sources, uses, strengths and limitations of data collected in primary care in England. Health Statistics Quarterly 2004(21):5-14

Table 1 Baseline characteristics of patients in the QResearch derivation, the QResearch validation cohorts and the CPRD validation cohort. Values are numbers (percentages of total number in cohort) unless stated otherwise.

	QResearch Derivation (n=2,849,381)	QResearch Validation (n=1,340,622)	CPRD Validation (n=2,475,360)
female	1446784 (50.8)	677897 (50.6)	1260015 (50.9)
male	1402597 (49.2)	662725 (49.4)	1215345 (49.1)
Mean age (SD)	46.3 (18.9)	47.8 (18.6)	48.2 (18.6)
Strategic Health Authority			
East Midlands SHA	225092 (7.9)	165734 (12.4)	70695 (2.9)
Yorkshire & Humberside SHA	220560 (7.7)	75976 (5.7)	287374 (11.6)
East of England SHA	197453 (6.9)	158962 (11.9)	390573 (15.8)
London SHA	560544 (19.7)	234346 (17.5)	52618 (2.1)
North East SHA	141974 (5.0)	103200 (7.7)	398889 (16.1)
North West SHA	268958 (9.4)	264508 (19.7)	317867 (12.8)
South Central SHA	310830 (10.9)	74588 (5.6)	274296 (11.1)
South East SHA	253288 (8.9)	63455 (4.7)	314779 (12.7)
South West SHA	421052 (14.8)	92822 (6.9)	275566 (11.1)
West Midlands SHA	249630 (8.8)	107031 (8.0)	92703 (3.7)
Ethnicity			
ethnicity recorded	2129124 (74.7)	1015630 (75.8)	1301115 (52.6)
White/not recorded	2554557 (89.7)	1212057 (90.4)	2320487 (93.7)
Indian	49360 (1.7)	22888 (1.7)	31800 (1.3)
Pakistani	23947 (0.8)	15243 (1.1)	13739 (0.6)
Bangladeshi	22309 (0.8)	11076 (0.8)	4482 (0.2)
Other Asian	38463 (1.3)	14870 (1.1)	22394 (0.9)
Caribbean	23704 (0.8)	9038 (0.7)	11086 (0.4)
Black African	43471 (1.5)	22355 (1.7)	26533 (1.1)
Chinese	28803 (1.0)	8086 (0.6)	7514 (0.3)
Other	64767 (2.3)	25009 (1.9)	37325 (1.5)
Smoking status			
smoking status recorded	2766234 (97.1)	1300728 (97.0)	2388744 (96.5)
non smoker	1568956 (55.1)	731480 (54.6)	1220054 (49.3)
Ex-smoker	612156 (21.5)	288031 (21.5)	642110 (25.9)
light smoker (1-9/day)	353026 (12.4)	165471 (12.3)	161185 (6.5)
moderate smoker (10-19/day)	152631 (5.4)	75157 (5.6)	210441 (8.5)
heavy smoker (20+/day)	79465 (2.8)	40589 (3.0)	120768 (4.9)
smoker amount not recorded	n/a	n/a	34,186 (1.4)
Alcohol intake			
Alcohol status recorded	2340360 (82.1)	1097278 (81.8)	1968156 (79.5)
non drinker	746788 (26.2)	354328 (26.4)	393692 (15.9)
Trivial <1 unit/day	792730 (27.8)	368465 (27.5)	878965 (35.5)
Light 1-2 units/day	365897 (12.8)	166881 (12.4)	508687 (20.6)
Moderate 3-6 units/day	387161 (13.6)	183738 (13.7)	150466 (6.1)

Heavy 7-9 units/day	27501 (1.0)	13579 (1.0)	17695 (0.7)
Very Heavy >9 units/day	16260 (0.6)	8112 (0.6)	18651 (0.8)
Drinker - amount not recorded	4023 (0.1)	2175 (0.2)	0 (0)
emergency admissions in past year (HES record)			
no emergency admission (HES record)	2695651 (94.6)	1264555 (94.3)	2334640 (94.3)
1 emergency admission (HES record)	118002 (4.1)	58078 (4.3)	107182 (4.3)
2 emergency admission (HES record)	23301 (0.8)	11687 (0.9)	21802 (0.9)
3+ emergency admissions (HES record)	12427 (0.4)	6302 (0.5)	11736 (0.5)
Emergency admissions in past year (GP record)			
no emergency admission (GP record)	2731533 (95.9)	1283422 (95.7)	2261885 (91.4)
1 emergency admission (GP record)	89457 (3.1)	44263 (3.3)	158723 (6.4)
2 emergency admission (GP record)	19581 (0.7)	8812 (0.7)	36567 (1.5)
3+ emergency admissions (GP record)	8810 (0.3)	4125 (0.3)	18185 (0.7)
Clinical values, family history and deprivation			
Body mass index recorded	2281550 (80.1)	1083278 (80.8)	1980327 (80.0)
Mean body mass index (SD)	26.1 (4.9)	26.4 (4.9)	26.4 (5.0)
systolic blood pressure recorded*	2437745 (85.6)	1186261 (88.5)	n/a
Mean systolic blood pressure (SD)	127.0 (16.4)	127.3 (16.5)	n/a
cholesterol/HDL recorded*	824938 (29.0)	413117 (30.8)	n/a
Mean cholesterol/HDL ratio	3.8 (1.2)	3.8 (1.2)	n/a
family history CHD*	327668 (11.5)	169286 (12.6)	n/a
mean Townsend score (SD)	0.1 (3.6)	0.1 (3.5)	-0.7 (3.1)
Haemoglobin recorded	1645857 (57.8)	816261 (60.9)	1512841 (61.1)
Haemoglobin < 11g/dl	56293 (2.0)	28113 (2.1)	49339 (2.0)
Platelets recorded	1632357 (57.3)	810551 (60.5)	1505945 (60.8)
Platelets > 480	16501 (0.6)	8434 (0.6)	14127 (0.6)
Liver function test recorded	1225813 (43.0)	628439 (46.9)	1148893 (46.4)
Abnormal liver function tests	34260 (1.2)	19112 (1.4)	32230 (1.3)
Erythrocyte sedimentation rate (ESR) recorded	755536 (26.5)	409183 (30.5)	n/a
Abnormal Erythrocyte sedimentation rate (ESR)	5989 (0.2)	3306 (0.2)	n/a
comorbidity			
Type 1 diabetes	11000 (0.4)	5445 (0.4)	9854 (0.4)
type 2 diabetes	125374 (4.4)	63461 (4.7)	117754 (4.8)
atrial fibrillation	52603 (1.8)	26285 (2.0)	48490 (2.0)
cardiovascular disease	154825 (5.4)	79116 (5.9)	150108 (6.1)
congestive cardiac failure	27404 (1.0)	14304 (1.1)	22685 (0.9)
venous thromboembolism	42870 (1.5)	21298 (1.6)	37925 (1.5)

cancer	97279 (3.4)	48370 (3.6)	82513 (3.3)
asthma or COPD	378048 (13.3)	179635 (13.4)	342371 (13.8)
epilepsy	36615 (1.3)	17904 (1.3)	34607 (1.4)
falls	124248 (4.4)	64299 (4.8)	172555 (7.0)
manic depression or schizophrenia	21277 (0.7)	10155 (0.8)	16792 (0.7)
chronic renal disease	9841 (0.3)	4700 (0.4)	9476 (0.4)
Conditions leading to malabsorption	29206 (1.0)	14432 (1.1)	19078 (0.8)
chronic liver disease or pancreatitis	15811 (0.6)	7669 (0.6)	10895 (0.4)
valvular heart disease*	30924 (1.1)	15960 (1.2)	n/a
treated hypertension*	371503 (13.0)	188901 (14.1)	n/a
rheumatoid arthritis or SLE*	45966 (1.6)	23020 (1.7)	n/a
depression (QOF definition)*	372341 (13.1)	176638 (13.2)	n/a
current prescribed medication			
statins*	341765 (12.0)	174252 (13.0)	
NSAIDs	416749 (14.6)	208936 (15.6)	365927 (14.8)
anticoagulants	38790 (1.4)	19764 (1.5)	36166 (1.5)
corticosteroids	101067 (3.5)	49683 (3.7)	109847 (4.4)
antidepressants	341194 (12.0)	168305 (12.6)	302457 (12.2)
antipsychotics	74039 (2.6)	38324 (2.9)	69498 (2.8)

*denotes variables which were considered but which didn't meet the criteria for inclusion in the final model. These variables were therefore not needed from CPRD for the external validation so have been reported as not applicable.

Predicting Risk of Emergency Admission - QAdmissions

Table 2 Incidence rates of first emergency admissions to hospital during follow-up for men and women in the QResearch derivation , the QResearch validation cohort and the CPRD validation cohort. Rates are per 100,000 person years. Adjusted rates have been directly standardized by age and sex using 5 year ageband.

	QResearch Derivation cohort			QResearch Validation cohort			CPRD validation cohort		
	cases	pyrs	Crude rate per 1000(95% CI)	cases	pyrs	Crude rate per 1000(95% CI)	cases	pyrs	Crude rate per 1000(95% CI)
total	265,573	4,597,543	**57.8 (57.5 to 58.0)**	**132723**	**2222285**	**59.7 (59.4 to 60.0)**	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)
Ageband									
18-24 years	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 years	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 years	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 years	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 years	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 years	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+ years	84,274	406,859	207 (206 to 209)	42,072	203,280	207 (205 to 209)	74,695	363,477	206 (204 to 207)
SHA	**cases**	**pyrs**	**Age/sex standardised rate per 1000 (95% CI)**	**cases**	**pyrs**	**Age/sex standardised rate per 1000 (95% CI)**	**cases**	**pyrs**	**Age/sex standardised rate per 1000 (95% CI)**
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)

Predicting Risk of Emergency Admission - QAdmissions

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)
	cases	**pyrs**	**Age/sex standardised rate per 1000 (95% CI)**	**cases**	**pyrs**	**Age/sex standardised rate per 1000 (95% CI)**	**cases**	**pyrs**	**Age/sex standardised rate per 1000 (95% CI)**
Ethnicity									
White/not recorded	248,023	4,179,915	56.8 (56.6 to 57.0)	123918	2031918	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)

Table 3 Adjusted hazard ratios (95% CI) for emergency admission to hospital for the final QAdmissions model in the derivation cohort. Hazard ratios are adjusted for fractional polynomial terms for age and BMI. Final model included age interaction terms.

	Women adjusted hazard ratio[§] (95% CI)	Men adjusted hazard ratio[§] (95% CI)
Ethnicity		
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)
Strategic Health Authority		
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)
smoking status		
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)
alcohol status		
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)
Emergency admissions in last year		
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)

Predicting Risk of Emergency Admission - QAdmissions

Clinical values and deprivation		
Townsend Score (5 unit increase)	1.10 (1.09 to 1.11)	1.11 (1.10 to 1.12)
most recent Haemoglobin <11g/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)
co-morbidity		
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)
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)

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

§ hazard ratios 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. Hazard ratios for these variables in the table are evaluated at mean age in men and women.

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^2 (%)	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^2 (%)	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 1.750)	1.64 (1.63 to 1.65)

Notes on understanding validation statistics:

Discrimination is the ability of the risk prediction model to differentiate between patients who experience a 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^2 is another measure specific to censored survival data– it measures explained variation and higher values indicate more variation is explained.

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.

	QResearch validation cohort				CPRD validation cohort			
2 year risk score	cut off for 2 year predicted risk %	total classified as high risk	Sensitivity (%)	Observed risk of admission at 2 years value* (%)	cut off for 2 year predicted risk %	total classified as high risk	Sensitivity (%)	Observed risk of admission at 2 years value* (%)
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
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.

Figure legends:

Figure 1 Mean predicted risks and observed risk of emergency admission to hospital at 2 years by tenth of predicted risk applying the QAdmissions risk prediction scores to all patients in the QResearch validation cohort (results from CPRD available from the authors).

Figure 2 and 3: Clinical cases:

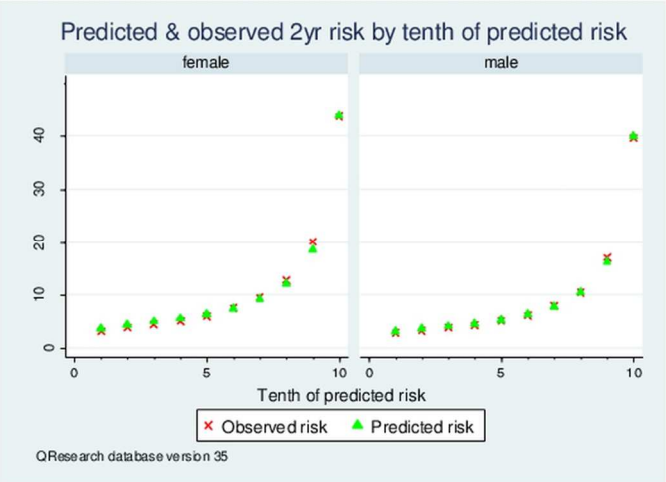
STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	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
		(c) Summarise follow-up time (eg, average and total amount)	Page 9 Table2
Outcome data	15*	Report numbers of outcome events or summary measures over time	Page 9
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

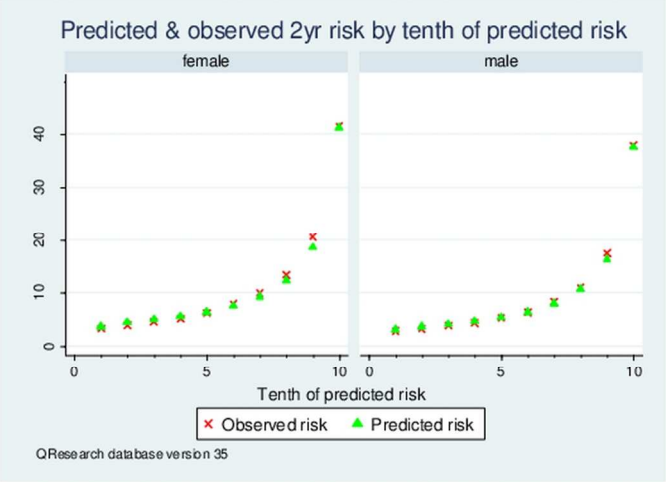
		included	
		(*b*) Report category boundaries when continuous variables were categorized	Page 6
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	Page 11
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, 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

Figure 1 Mean predicted risks and observed risk of emergency admission to hospital at 2 years by tenth of predicted risk applying the QAdmissions risk prediction scores to all patients in the QResearch validation cohort (results from CPRD available from the authors).

(a) QAdmissions score calculated using the HES-GP linked data

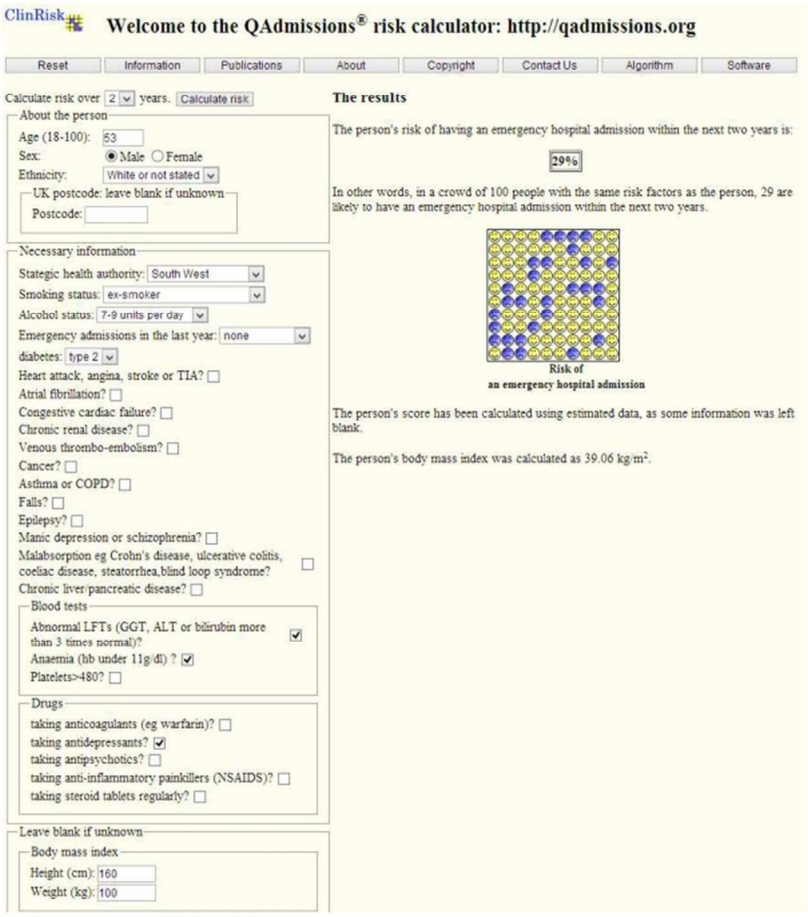


(b)QAdmissions score calculated based on GP data alone



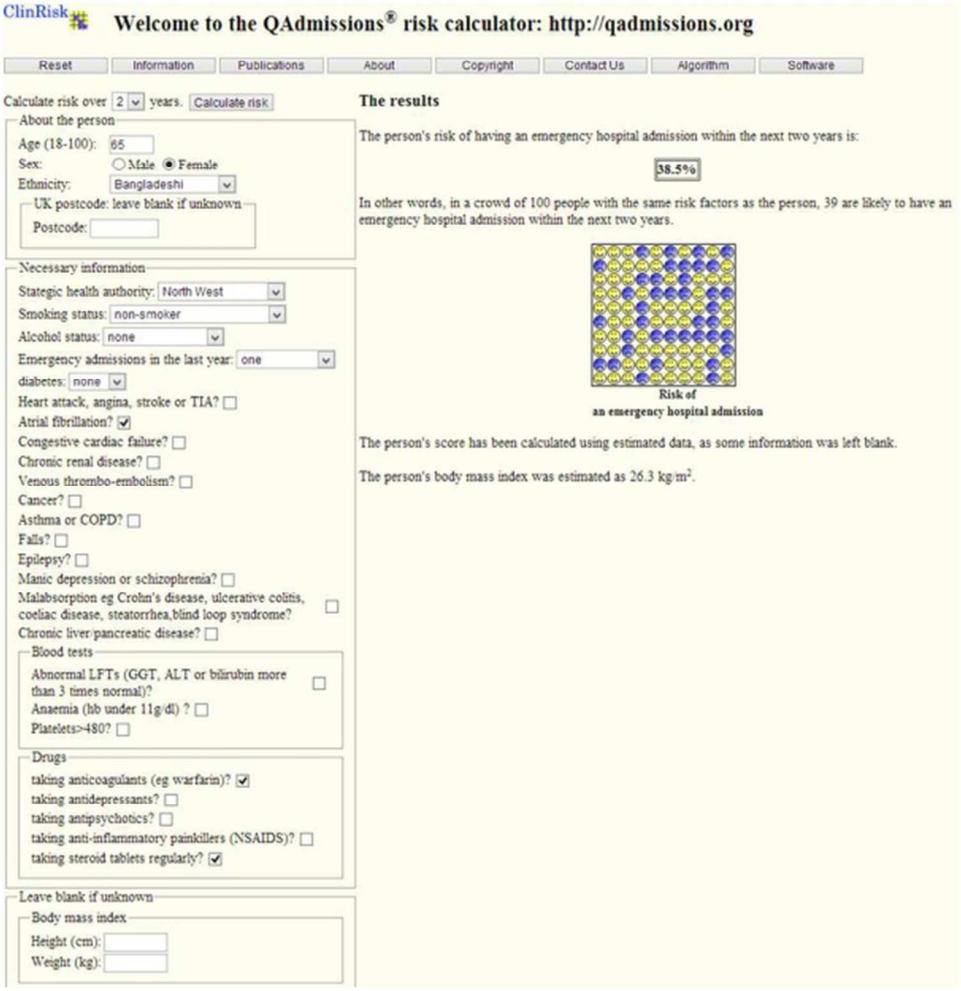
90x134mm (300 x 300 DPI)

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/m2 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). 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%.



90x119mm (300 x 300 DPI)

A 65 year old Bangladeshi women from the North West, non-smoker, no-drinker, with a body mass index of 26.3 Kg/m2 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.



90x105mm (300 x 300 DPI)

Appendix Table A:

Clinical codes used to identify emergency hospital admission or casualty attendance on GP clinical record. This is for use as a predictor variable for when the algorithm is applied to GP data only.

Read or EMIS code	Description
7M300	Emergency operation NOC
8H1	Admit to intensive care unit
8H1-1	Admit to I.T.U.
8H11	Admit to cardiac ITU
8H12	Admit to respiratory ITU
8H13	Admit to neurological ITU
8H14	Admit to metabolic ITU
8H1Z	Admit to intensive c.u. NOS
8H2	Emergency hospital admission
8H21	Admit medical emergency unsp.
8H22	Admit surgical emergency unsp.
8H23	Admit psychiatric emergency
8H230	Emerg psychiatric admiss MHA
8H24	Admit geriatric emergency
8H25	Admit paediatric emergency
8H26	Admit gynaecological emergency
8H27	Admit obstetric emergency
8H28	Admit orthopaedic emergency
8H29	Admit ENT emergency
8H2A	Admit trauma emergency
8H2B	Admit ophthalmological emerg.
8H2C	Admit rheumatology emergency
8H2D	Admit dermatology emergency
8H2E	Admit neurology emergency
8H2F	Admit urology emergency
8H2G	Admit radiotherapy emergency
8H2H	Admit haematology emergency
8H2I	Admit plastic surgery emergenc
8H2J	Admit diabetic emergency
8H2K	Admit oral surgical emergency
8H2L	Admit psychogeriatric emergency
8H2M	Admit renal medicine emergency
8H2N	Admit neurosurgical emergency
8H2O	Admit cardiothoracic emergency
8H2P	Emergency admission, asthma

8H2Q	Admit cardiology emergency
8H2R	Admit COPD emergency
8H2S	Admit heart failure emergency
8H2T	Emergency voluntary psychiatric admission Mental Health Act
8H2V	Admit ischaemic heart disease emergency
8H2W	Admit vascular surgery emergency
8H2X	Emergency hospital admission from walk-in centre
8H2Y	Admit anticoagulation emergency
8H2Z	Admit hospital emergency NOS
8H63	Refer to casualty officer
8H64	Refer to house officer
8H65	Refer to hospital registrar
8Hb	Involuntary admission
8HC	Refer to hospital casualty
8HC1	Refer to A. & E. department
8HC2	Refer to hosp. eye casualty
8HC3	Refer to hosp. paeds casualty
8HCZ	Refer to hospital casualty NOS
8Hd1	Admission by accident and emergency doctor
8Hd3	Admission by out of hours service doctor
8Hd5	Admission to acute assessment unit
8Hd6	Admission to stroke unit
8HJA-1	Casualty self-referral
8HJZ	Self-referral to hospital NOS
9b0K	Hospital admission note
9N04	Seen in emergency clinic
9N19	Seen in hospital casualty
9Nk8	Seen in eye casualty department
EMISNQHO22	Hospital admission, emergency, indirect
EMISNQHO53	Hospital admission, emergency, from walk-in centre
EMISQEL1	Elderly psychiatric emergency admission