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

Protocol for Original Research

Authors

Julia Hippisley-Cox Professor of Clinical Epidemiology & General Practice
Carol Coupland Associate Professor and Reader in Medical Statistics

Institutions

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

Contents

1 Introduction

Unplanned admissions account for an estimated 11 billion pounds a year in England which is a considerable portion 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 the 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 following an holistic approach embracing physical and mental health problems; work with a local multidisciplinary team to identify those at risk and co-ordinate the management of these patients. The intended benefits for patients being improved quality of life and care and fewer emergency admissions to hospital". In return for this, practices will then receive remuneration of 0.74 pence per registered patient per practices which represents an estimated 37 million pounds in respect of the population of England in the next financial year.

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 algorithms need to be developed using data from the setting where it will subsequently be used (e.g. primary care in England). It needs to distinguish between different patients according to their level of risk (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 (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 lower risk of emergency admission). It needs to incorporate measures of socio-economic deprivation and ethnicity (a) in recognition of the role these factors as predictors of major diseases but also (b) to prevent widening health inequalities which can occur when new programs are introduced.

Once developed, the risk algorithm needs to be scientifically reviewed and published. It needs to have the potential to be updated or recalibrated over time to reflect changes in demography, burden of disease, service deliver, data quality, data coding and NHS policy. Its performance needs to 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. Ideally this validation should ideally be done an independent team. Lastly, the tool needs to be suitable for implementation in clinical practice. This is likely to mean that it can run off data which is already present in routinely used clinical systems used at the point of care or which can be easily be assimilated. This will not only help ensure the tool is practical to use but also clinically safe since the underlying data must be accurate and up to date if it is to be used to inform clinical decisions.

Whilst a number of emergency admission risk assessment tool have been developed, they are generally designed for use in hospital to identify patients at risk of re-admission [3-5]. Hence they are not applicable in a primary care setting or to patients who have not already had a recent admission. Other current tools focus on specific populations or have not been published or validated. For example, there are a number of American algorithms are based on patients enrolled in health maintenance organisation with questionable generalizability [6-8]. Version 3 of SPARRA [5] is designed for use in Scotland for patients who have contact with secondary care services or repeat prescriptions. 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. A questionnaire approach is also likely to be difficult to systematically implement in everyday practice especially if the patients do not respond because they are elderly or infirm. The PEONY score was designed for use in Scottish primary care patients age 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 both the CPM and the PARR tools were outdated and in urgent need of a refresh [12]. The weightings of the predictors and the Health Resource Groups were incompatible with current NHS data [12]. The Department of Health withdrew funding and the Kings Fund subsequently withdrew its associated implementation, documentation and support.

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 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 complex technical process to achieve in real time. It also has significant information governance challenge 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 risk of emergency admission to hospital (QAdmissions) which could meet the above requirements. We were interested to develop an algorithm which incorporates ethnicity and primary diagnoses, medication and abnormal laboratory results which the clinician can then follow up. In addition, we decided to develop a tool which could be automatically populated using data from GP computer system 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 will undertake a prospective cohort study studying a large UK primary care population using a similar method to our original analysis for other risk prediction scores such as QRISK2 [13]. Version 35 of the QResearch database will be 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 will include 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 will be conducted on QResearch practices in England in order to incorporate hospital episode data linked at individual patient level via pseudonymised NHS number.

2.2 Cohort selection

We will identify an open cohort 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 will use 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 will exclude patients without a valid postcode related Townsend deprivation score. We will also exclude registered patients without a valid pseudonymised NHS number as this is 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 will determine an entry date to the cohort, which is 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 will be censored at the earliest date of the date of 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 is the first recorded diagnosis of emergency admission to hospital in the study period. We will identify emergency hospital admissions

using the Hospital Episode Statistics (HES) data which were linked at individual patient level to the QResearch database via pseudonymised NHS number. Emergency admissions will be identified by selecting the relevant codes from the method of admission field: 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 will include events 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 will utilise 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 will include predictors used in other risk algorithms where the outcome is likely to require emergency admission (for example as thrombosis[16], osteoporotic fracture[17] or cardiovascular disease[18 19]). We will 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].

(a) demographic variables: age, sex, Strategic Health Authority, Townsend deprivation score, ethnicity
(b) Lifestyle variables: smoking, alcohol intake
(c) Chronic diseases:
(d) Current medication for statins, NSAIDS, anticoagulants, corticosteroids, antidepressants and antipsychotics.
(e) Clinical values: body mass index, systolic blood pressure
(f) Laboratory test results: haemoglobin, platelets, ESR, cholesterol/HDL ratio, liver function tests.

All the above variables will be derived from the patients' primary care record. In addition, we will include the number of emergency admissions in the preceding year as recorded on the HES-GP linked data. This will be coded into four groups - none; one only; two only; three or more. For the validation cohort, we will extract information on admissions recorded in the primary care record. We will include all admissions and then removed those which were coded as routine.

We will restrict all values of these variables to those recorded in the person's electronic healthcare record before baseline, except for ethnicity where we will use the most recently recorded value in the study period before the patient had the outcome or was censored. We will impute missing values where necessary as described below. Given the large number of candidate variables, we may factors where appropriate.

2.5 Model derivation and development

As in previous studies[18], we will use 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 will use fractional polynomials to model non-linear risk relationships with age and body mass index where appropriate[2021]. We will test for interactions between each variable and age and include significant interactions in the final model where they improve model fit. Continuous variables will be centered for analysis. Our main analyses will use multiple imputation to replace missing values for systolic blood pressure, cholesterol, smoking status, and body mass index.

Our final model, will be fitted based on five multiply imputed datasets using Rubin's rules to combine effect estimates and standard errors to allow for the uncertainty due to imputing missing data[21] [22]. We will take the log of the hazard ratio for each variable from the final model and use these as weights for the risk equations. We will combine 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 will be at least 100 events per variable considered in the prediction modeling for the outcome in the derivation cohort [23].

2.6 Model Validation

We will test the performance of the final model (QAdmissions) in the validation cohort. We will calculate the 2 year estimated risk of emergency admission for each patient in the validation dataset using multiple imputation to replace missing values as in the derivation dataset.

We will calculate the mean predicted and observed risk at 2 years[13] and compare these by tenth of predicted risk for each score. The observed risk at 2 years will be obtained using the 2 year Kaplan-Meier estimate. We will calculate the ROC statistic, D statistic (a measure of discrimination where higher values indicate better discrimination)[24] and an R squared statistic (which is a measure of explained variation for survival data where higher values indicate more variation is explained)[25].

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

For the main analyses, we will base the calculation of risk of admission on data recorded in the GP record except for prior emergency admission which will be derived from the HES-GP linked data. We will repeat the analyses by calculating risk of admission based hospital admissions recorded on the GP record using Read codes instead of the HES linked data. This will be done to test the likely performance in a clinical setting where GP-HES linked data is not available (GP-HES is not routinely available in all primary care settings). We will examine

the clinical codes used to identify hospital admissions on the GP record and selected admissions were coded either as emergency admissions or where the urgency of the admission is not unspecified. Analyses will be conducted using STATA (version 12).

3 Other information

3.1 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.

3.2 Funding

None. Later stages of this work were supported by North East London Commissioning Support Group.

3.3 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.