Preventing unscheduled hospitalisations from asthma: a retrospective cohort study using routine primary and secondary care data in the UK (The PUSH-Asthma Study)—protocol paper

Nikita Simms-Williams,1,2 Prasad Nagakumar,2,3 Rasiah Thayakaran,1 Nicola Adderley,1,4 Richard Hotham,1 Adel Mansur,3,4 Krishnarajah Nirantharakumar,1 Shamil Haroon1

ABSTRACT

Introduction Asthma is the most common chronic respiratory disease in children and adults. Asthma results in significant disease-related morbidity, healthcare costs and, in some cases, death. Despite efforts through implementation of national guidelines to improve asthma care, the UK has one of the highest asthma-related morbidity and mortality rates in the western world. New approaches are necessary to prevent asthma attacks in children and adults. The objectives of this study are to assess the association between demographic and clinical factors and asthma-related hospital admissions in children and adults, describe the epidemiology of asthma phenotypes among hospital attenders, and externally validate existing asthma risk prediction models.

Methods and analysis This is a retrospective cohort study of children and adults with asthma. Data will be extracted from the Clinical Practice Research Datalink (CPRD) Aurum database, which holds anonymised primary care data for over 13 million actively registered patients and covers approximately 19% of the UK population. The primary outcome will be asthma-related hospital admissions. The secondary outcomes will be prescriptions of short courses of oral corticosteroids (as a surrogate measure for asthma exacerbations), a composite outcome measure including hospital admissions and prescriptions of short courses of oral corticosteroids and delivery of asthma care management following hospital discharge. The primary analysis will use a Poisson regression model to assess the association between demographic and clinical risk factors and the primary and secondary outcomes. Latent class analysis will be used to identify distinct subgroups, which will further our knowledge on potential phenotypes of asthma among patients at high risk of asthma-related hospital admissions. A Concordance statistic (C-statistic) and logistic regression model will also be used to externally validate existing risk prediction models for asthma-related hospitalisations to allow for the optimal model to be identified and evaluated provide evidence for potential use of the optimal performing risk prediction model in primary care.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Clinical Practice Research Datalink Aurum will be used, which has over one million registered patients with a diagnosis of asthma, providing ample statistical power for both the main and subgroup analyses.
⇒ The study will include both adults and children, allowing for age-stratified analyses of asthma-related hospital admissions.
⇒ The study will not be able to analyse emergency department attendances due to the unavailability of linked emergency department data.
⇒ Asthma is often both underdiagnosed and misdiagnosed in routine clinical care, which risks introducing misclassification bias in our analyses.

INTRODUCTION

Context

Asthma is a chronic respiratory disease that affects both adults and children. It is one of the most common chronic diseases and results in significant morbidity and mortality.1 Despite asthma being a manageable disease, the UK has one of the highest asthma-related morbidity and mortality rates in the western world, accounting for three deaths per day and approximately 60,000 emergency hospital admissions/year.2,3 This has resulted in a significant and preventable health and economic burden that costs the UK National Health Service £1.1 billion per year.1
Acute asthma attacks requiring hospital attendance often indicate poor symptom control and inadequate management of risk factors. Possible risk factors for the high incidence of asthma attacks include severe disease, poor adherence to prescribed therapies, inadequate clinical care and disease monitoring, presence of comorbidities, tobacco use, socioeconomic deprivation, previous asthma attacks, and air pollution.6

Despite efforts to improve the management of asthma through implementation of national clinical guidelines published by the British Thoracic Society and the National Institute for Health and Care excellence (NICE), there has been little reduction in asthma attacks across all age groups.2

**Current knowledge**

Previous studies have found that female sex,3–12 previous exacerbations requiring hospital admission,9 11 15 increased disease severity,9 14 older age (≥45 years),9 11 12 younger age (<18),10 15 ethnic minorities (African-American, Black and Indian),10 15 16 history of inhaled corticosteroids (ICS) (defined as ICS use in the preceding year),15 over-use of short-acting beta agonists16 and comorbidities such as chronic rhinitis and gastroesophageal reflux disease (GORD)11 are associated with asthma exacerbations including those requiring hospitalisation in children and adults. Previous studies also suggest that continued care and follow-up after hospitalisation reduce future asthma-related hospitalisations.17–19

Most of the existing evidence on the association between asthma and hospitalisations is based on patient data from the USA and Asia, and the evidence from these studies is unlikely to be generalisable to a UK population, partly due to significant differences in healthcare provision. Risk factors for exacerbations are likely to differ across age groups, but there is a paucity of studies describing risk factors for asthma exacerbations stratified by age.

A UK-based study describing the characteristics of 3776 children aged 5–12 years in primary care record databases (Clinical Practice Research Datalink (CPRD) and Optimum Patient Care Research Database) showed that previous history of asthma attacks was the strongest predictor for future asthma attacks.20 A similar study using data from over 200,000 patients in CPRD showed that the risk of subsequent asthma attacks requiring hospitalisation increased with disease severity.14 A study of 460 children aged 5–16 years from 10 UK general practices showed that recent asthma attacks, high fractional exhaled nitric oxide, socioeconomic deprivation and poor symptom control in the previous 6 months was associated with a higher risk of future asthma attacks.21 22

Two previous studies have used UK data from CPRD to assess asthma outcomes; however, one study included only children20 and the other investigated the association with disease severity alone.14 There is currently no evidence published using UK primary care data in a population of adults and children stratified by important demographic characteristics, including age and sex. A UK representative population-based cohort study using primary care data is needed to understand the multiple risk factors present for future asthma exacerbations requiring hospital admission, in addition to understanding the primary care management of patients after hospitalisation in the UK.

A recent systematic review summarised a number of risk prediction models that have been developed to estimate the risk of asthma exacerbations requiring hospitalisation.23–25 The main parameters used in these risk prediction models included age, sex, smoking history, comorbidities, course of systemic corticosteroids, emergency department visits and previous hospitalisation for asthma exacerbations. A more recent study used data from a subset of CPRD to generate a risk prediction model for asthma exacerbations.25 The model included patients aged 12–80 years and did not demonstrate if the risk prediction model was able to accurately predict the risk of asthma exacerbation-related hospitalisations in children below the age of 12 years.

External validation of existing risk prediction models is needed in a common data set to identify the model with optimal performance in a UK primary care population to accurately predict asthma exacerbation risk. There is a need to assess both the discrimination and calibration of these models.26 This could help guideline developers and clinicians to determine whether the models are appropriate for use in UK primary care. This would also help to identify the most promising models to prospectively evaluate in randomised controlled impact trials.27

**Study aims and objectives**

We aim to describe detailed characteristics of patients with asthma and their association with asthma-related hospital admissions to identify a comprehensive range of risk factors for high-risk patients. The aim is to enable primary care services to use these findings on risk factors to identify and flag high-risk patients for the delivery of targeted interventions to prevent future hospitalisations. We also aim to identify distinct risk factor clusters to describe the clinical heterogeneity in high-risk patients with asthma and delineate potential phenotypes that may help better target therapies. The treatment and care delivered following hospitalisation will be analysed to aid the identification of the gaps and discontinuity between primary and secondary care services, which could then be targeted for intervention and service improvement. We also aim to externally validate existing risk prediction models for asthma exacerbations in a UK-based primary population.

The specific objectives for this study are to (1) define and characterise a cohort of adults and children with asthma, (2) assess the incidence of asthma-related hospitalisations, (3) identify risk factors for asthma-related hospitalisations, (4) phenotype asthma patients at high risk of asthma-related hospitalisations, (5) externally validate existing risk prediction models for hospitalisations for asthma, (6) assess the primary care management of
METHODS AND ANALYSIS

Study design

A retrospective open cohort study using routine primary and secondary healthcare data, consisting of patients with asthma between 1 January 2011 and 1 January 2021.

Data source

The data used to conduct this study will be UK primary care records extracted from the Clinical Practice Research Datalink —Aurum (CPRD Aurum) database with linked Hospital Episode Statistics (HES). Over 1000 general practices contribute to CPRD Aurum, covering approximately 19% of the UK population. The database holds data for over 39 million patients, approximately 15 million of whom are current patients. Over 90% of patient data contributing to CPRD Aurum have linked HES data. The general practices that contribute to CPRD Aurum use a patient records software called EMIS, which codes clinical data using Read Code clinical classification (V.3) and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) codes, and codes for drug prescriptions using medical drug codes, which are linked to the British National Formulary. The linked HES data uses ICD-10 codes to code diagnoses.

Population

General practices that have contributed to CPRD on or after the index date (1 January 2011) will be eligible for inclusion in the study. Data will be extracted for adults and children aged 5 years and over with diagnosis of asthma who have been registered to an eligible general practice after the index date (1 January 2011) will be eligible for inclusion in the study. This included atopic dermatitis, allergic rhinitis, allergies, GORD, cardiovascular disease, chronic obstructive pulmonary disease (COPD), obstructive sleep apnoea, non-cystic fibrosis bronchiectasis, interstitial lung disease, chronic rhinosinusitis, anxiety, depression, obesity and diabetes mellitus. Asthma-related prescriptions including short-acting bronchodilators, ICS and leukotriene receptor antagonists and other medical prescriptions unrelated to asthma diagnosis such as beta blockers and non-steroidal anti-inflammatory drugs will also be reported due to their likely association with asthma hospitalisations. All covariates will be captured before the index date at baseline from CPRD data.

Follow-up period

The recruitment/study period will be from 1 January 2011 to 1 January 2021. The start of follow-up for patients will be the latest of 12 months after registration with general practice, study start date or the date of patient diagnosis of asthma. The end of follow-up for patients will be the earliest of at least one study outcome reached, death, exit from the database or study end date (1 January 2021).

Missing data

Missing data are not likely to be missing at random. Therefore, multiple imputation will not be used. A separate ‘missing’ category will be created for missing data for each variable. Continuous variables will be converted into categorical variables to allow for the addition of the ‘missing’ category. Data will likely be missing for ethnic group, BMI and smoking status. The absence of clinical codes for diagnoses (eg, comorbidities) or drug treatment codes will indicate the absence of the condition or drug.

Data analysis

The data will be extracted using the Data extraction for epidemiological research (DExtER) tool and the analysis will be undertaken using STATA SE V.16. The asthma cohort will be stratified into the following age-groups: 5–11, for asthma exacerbations, delivery of clinical care for asthma after discharge from hospital (including prescription of ICS, smoking cessation counselling, provision of an asthma management plan, demonstration of inhaler technique and provision of an asthma review) and a composite outcome measure including hospital admissions or prescriptions of short courses of oral corticosteroids. The outcomes will be defined by the relevant clinical codes (ICD-10) or drug/SNOMED-CT codes.

Predictor variables

The baseline patient demographics and characteristics to be extracted include age (at index date), sex, ethnic group, socioeconomic status, smoking status (smoker, ex-smoker, never smoked), body mass index (BMI) and alcohol consumption. Current comorbidities will also be extracted. A literature search was conducted in addition to obtaining expert opinions to identify the most relevant comorbidities to be included in the study. This included atopic dermatitis, allergic rhinitis, allergies, GORD, cardiovascular disease, chronic obstructive pulmonary disease (COPD), obstructive sleep apnoea, non-cystic fibrosis bronchiectasis, interstitial lung disease, chronic rhinosinusitis, anxiety, depression, obesity and diabetes mellitus. Asthma-related prescriptions including short-acting bronchodilators, ICS and leukotriene receptor antagonists and other medical prescriptions unrelated to asthma diagnosis such as beta blockers and non-steroidal anti-inflammatory drugs will also be reported due to their likely association with asthma hospitalisations. All covariates will be captured before the index date at baseline from CPRD data.

Study variables

Exposure variables

Asthma will be defined as the presence of a Read code or SNOMED-CT code for asthma. An extensive list of relevant clinical codes for asthma has been developed to identify eligible patients eligible for inclusion (see online supplemental file 1).

Outcome variables

The primary outcome will be the number of hospital admissions for asthma, assessed with linked HES Admitted Patient Care (APC) data (defined by the ICD-10 asthma diagnosis codes J45 and J46). The secondary outcomes will be prescriptions of short courses of oral corticosteroids for asthma exacerbations, delivery of clinical care for asthma after discharge from hospital (including prescription of ICS, smoking cessation counselling, provision of an asthma management plan, demonstration of inhaler technique and provision of an asthma review) and a composite outcome measure including hospital admissions or prescriptions of short courses of oral corticosteroids. The outcomes will be defined by the relevant clinical codes (ICD-10) or drug/SNOMED-CT codes.

Predictor variables

The baseline patient demographics and characteristics to be extracted include age (at index date), sex, ethnic group, socioeconomic status, smoking status (smoker, ex-smoker, never smoked), body mass index (BMI) and alcohol consumption. Current comorbidities will also be extracted. A literature search was conducted in addition to obtaining expert opinions to identify the most relevant comorbidities to be included in the study. This included atopic dermatitis, allergic rhinitis, allergies, GORD, cardiovascular disease, chronic obstructive pulmonary disease (COPD), obstructive sleep apnoea, non-cystic fibrosis bronchiectasis, interstitial lung disease, chronic rhinosinusitis, anxiety, depression, obesity and diabetes mellitus. Asthma-related prescriptions including short-acting bronchodilators, ICS and leukotriene receptor antagonists and other medical prescriptions unrelated to asthma diagnosis such as beta blockers and non-steroidal anti-inflammatory drugs will also be reported due to their likely association with asthma hospitalisations. All covariates will be captured before the index date at baseline from CPRD data.

Follow-up period

The recruitment/study period will be from 1 January 2011 to 1 January 2021. The start of follow-up for patients will be the latest of 12 months after registration with general practice, study start date or the date of patient diagnosis of asthma. The end of follow-up for patients will be the earliest of at least one study outcome reached, death, exit from the database or study end date (1 January 2021).

Missing data

Missing data are not likely to be missing at random. Therefore, multiple imputation will not be used. A separate ‘missing’ category will be created for missing data for each variable. Continuous variables will be converted into categorical variables to allow for the addition of the ‘missing’ category. Data will likely be missing for ethnic group, BMI and smoking status. The absence of clinical codes for diagnoses (eg, comorbidities) or drug treatment codes will indicate the absence of the condition or drug.

Data analysis

The data will be extracted using the Data extraction for epidemiological research (DExtER) tool and the analysis will be undertaken using STATA SE V.16. The asthma cohort will be stratified into the following age-groups: 5–11, for asthma exacerbations, delivery of clinical care for asthma after discharge from hospital (including prescription of ICS, smoking cessation counselling, provision of an asthma management plan, demonstration of inhaler technique and provision of an asthma review) and a composite outcome measure including hospital admissions or prescriptions of short courses of oral corticosteroids. The outcomes will be defined by the relevant clinical codes (ICD-10) or drug/SNOMED-CT codes.
12–17, 18+ years. The primary analysis will use descriptivestatistics to describe the characteristics of the cohortincluding age, sex, ethnic group, socioeconomic status,smoking status (smoker, ex-smoker, never smoked), BMI,comorbidities (including allergies and atopic diseases),prescriptions of short-acting bronchodilators, ICS, oralcorticosteroids and blood eosinophil levels (counts andpercentage where these are available). Categorical data atbaseline will be described using proportions, and continuoundata will be described using means with SD, where thedata follow a normal distribution or median with IQR,where the distribution is skewed.

The incidence rates for asthma-related hospitalisationswill be calculated for the entire cohort and stratified bydemographic and clinical characteristics. Poissonregression models will be used to assess the associationbetween the prespecified covariates and the primary andsecondary outcomes. The models will be fitted with theprespecified covariates as the independent variables andasthma-related hospital admissions (primary outcome)asthe dependent variable. The secondary outcomes,prescriptions of short courses of oral corticosteroids andthe composite outcome of hospital admissions or shortcourses of oral corticosteroids, will also be analysed usingseparate Poisson regression models.

A latent class analysis will be used to identify clusters ofcharacteristics based on demographic and clinical factorsthat differentiate patients with asthma. This will allow us tophenotype subgroups of patients with asthma at highrisk of asthma-related hospital admissions.

To quantify the performance of existing risk predictionmodels, we will conduct an external validation onmodels derived from primary care where there has been sufficientreporting of model parameters. The models will beidentified from the literature and recent systematicreviews. The variables used in the models that are alsoavailable in CPRD include age, sex, smoking history,comorbidities, prescriptions of systemic corticosteroids andprevious hospitalisation for asthma exacerbations. A novelrisk prediction model that has been developed by theUniversity of Edinburgh will also be externally validated inCPRD Aurum.34

The discrimination and calibration of the predictionmodels will be assessed. The concordance statistic (C-statistic) will be used to calculate discrimination, which willassess how well the model differentiates between patientswho experience the outcome and patients who do not. Calibration will be used to assess how closely the predic-tions of the model match the observed outcomes in thedata. Calibration will be assessed by fitting a logisticregression model with the outcome variable set as theobserved outcomes and the independent variable set as the log odd-transformed model predictors. We will assesstheseseasures using both the primary and secondaryoutcomes.

An analysis of hospitalised patients with asthma willbe conducted to assess the primary care managementof patients following discharge from hospital after anasthma exacerbation. We will report the frequency andpercentages of asthma-related drug prescriptions and anyother interventions delivered before and after dischargefrom hospital. We will use $\chi^2$ tests to test for statisticallysignificant differences between the frequency of drugprescriptions and other interventions before and afterhospitalisation. An ordinal logistic regression model willbe used to assess the association between patient riskfactors and the delivery of clinical care after dischargefrom hospital. The items of clinical care management tobe included are prescriptions of ICS, smoking cessationcounselling, provision of an asthma management plan,demonstration of inhaler technique and provision of anasthma review. A composite management score will be derived as the sum of the number of clinical care manage-ment items recorded.

A sensitivity analysis will be conducted to assess theeffect of the COVID-19 pandemic on the results of theprimary analysis. The events of the COVID-19 pandemicare likely to have had an impact on the frequency ofasthma-related exacerbations and hospital admissionsdue to self-isolations and national lockdowns beingimplemented over the past year. A recent study foundthat asthma exacerbations significantly decreased duringthe pandemic across all age groups, both sexes and acrossmost regions in England.35 The sensitivity analysis will beconducted in two parts. The first part includes restrictingthe analysis to all patients eligible for inclusion prior to 31January 2020 (patients eligible for inclusion after this datewill be excluded). Therefore, the study start date for theanalysis will be 1 January 2011 and the study end date willbe 31 January 2020. The second part includes restrictingthe analysis to all patients eligible for inclusion after 31January 2020 to the latest available date. The sensitivityanalysis will allow any differences in outcome resultsattributed to COVID-19 to be identified and compared withthat of the primary analyses.

In older adults, COPD can be misdiagnosed as asthma.Smoking is a known key risk factor for COPD. We willtherefore, undertake a subgroup analysis by smokingstatus and COPD status to assess whether the observedassociations differ between these two groups. Additionally,we will conduct a subgroup analysis focusing on patientswith eosinophilia on their blood count (peripheral bloodeosinophil count >3%, or absolute count >300 cells/μl)aseosinophilia can be used as an indicator for potentialcorticosteroid responsiveness in airway diseases; this ismainly in the form of asthma and also includes a propor-tion of patients with COPD.

Sample size calculation
Based on the sample size calculation derived by Signoriniand Zhang and Yuan,36 37 we have determined that, forthe smallest outcome baseline rate considered (0.5%),in order to have 80% power at an alpha level of 0.05 to detectthe smallest effect size (incidence rate ratio 0.2),we will need a sample size of approximately 360 000.Based on data from NICE and Asthma UK, we expect a
cumulative incidence of asthma-related hospital admissions of approximately 0.75% in a single year. Therefore, we will have a sufficient sample size for this study. The population size of current patients within CPRD Aurum is 13,320,051. We estimate that approximately 1,000,000 patients will meet the criteria for inclusion in this study.

**Patient and public involvement**

We will be working closely with members of the Patient and Public Involvement groups of Asthma UK and the NIHR Children and Young Persons Steering Group for dissemination of the study findings. Patients and members of the public were not involved in the design of the study.

**STRENGTHS AND LIMITATIONS**

The study will have a large sample size and data will be generalisable to the UK population as CPRD contains data that covers approximately 20% of the UK population. In contrast with previous studies, we will stratify patients across three age groups to describe risk factors and adjust for a broader range of relevant risk factors and confounders in the analysis. This will enable us to identify at-risk patients across multiple age groups and provide a comprehensive evaluation of the impact of various comorbidities on asthma-related hospital admissions. The large sample size also provides sufficient statistical power to perform several relevant subgroup analyses. The HES APC data linkage to primary care records provides a unique opportunity to characterise patients admitted to hospital due to asthma exacerbations.

Limitations of the study include the potential inaccuracy of asthma diagnoses. We acknowledge that patients with asthma can often be misdiagnosed, which may result in the potential misclassification of asthma. The misclassification of asthma could result in the dilution of effect sizes and, therefore, the effect sizes in this study are likely to be relatively conservative, as will model performance.

A limitation of using data from primary care databases is that ethnicity is often poorly recorded with approximately 30% of individuals having missing ethnicity data. However, for those without an ethnicity recording, a separate category will be created for the missing data values. Another limitation is that oral corticosteroids may be prescribed for a number of alternative conditions such as other atopic diseases (eg, atopic dermatitis), COPD and rheumatoid arthritis, leading to potential misclassification bias of the secondary outcome. We also acknowledge that a prescription recorded in primary care records does not necessitate that the prescribed drug was actually consumed by the patient or reflect medication compliance. We will acknowledge these study limitations in our main study report.

A further limitation to the study includes the lack of accident and emergency (A&E) data use. Some asthmatic patients may require A&E but may not be admitted to hospital or visit their GP; therefore, data from these patients may be missed. However, this is likely to be a small proportion of patients with asthma and is unlikely to affect the results of this study.

**Author affiliations**

1 Institute of Applied Health Research, University of Birmingham, Birmingham, UK
2 Respiratory medicine, Birmingham Women’s and Children’s Hospitals NHS Foundation Trust, Birmingham, UK
3 Institute of inflammation and ageing, University of Birmingham, Birmingham, UK
4 Respiratory Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK

**Twitter** Shamil Haroon @ShamilHaroon

**Contributors** SH, PN and AM led the design of the study. NS-W, SH, PN, RH, AM and RT were involved in the design of the study. NSW drafted the manuscript. SH, PN, NA, AM, RT and KN provided feedback on the manuscript; all authors approved the final version.

**Funding** This research was funded by the National Institute of Health and Research West Midlands Improvement and Innovation Strategic Fund. Grant number: 1 002 281.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were involved in the design, conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

**Patient consent for publication** Not applicable.

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

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) 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, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.
37 Mansur AH. Clinical spectrum of patients assessed in primary care respiratory diagnostic hub; informing referral criteria. Thorax BMJ 2021;76.