

Michael Noble, Annie Burden, Susan Stirling, Allan B Clark, Stanley Musgrave, Mohammad A Alsallakh, David Price, Gwyneth A Davies, Hilary Pinnock, Martin Pond, Aziz Sheikh, Erika J Sims, Samantha Walker and Andrew M Wilson

## Predicting asthma-related crisis events using routine electronic healthcare data:

a quantitative database analysis study

### Abstract

#### Background

There is no published algorithm predicting asthma crisis events (accident and emergency [A&E] attendance, hospitalisation, or death) using routinely available electronic health record (EHR) data.

#### Aim

To develop an algorithm to identify individuals at high risk of an asthma crisis event.

#### Design and setting

Database analysis from primary care EHRs of people with asthma across England and Scotland.

#### Method

Multivariable logistic regression was applied to a dataset of 61 861 people with asthma from England and Scotland using the Clinical Practice Research Datalink. External validation was performed using the Secure Anonymised Information Linkage Databank of 174 240 patients from Wales. Outcomes were  $\geq 1$  hospitalisation (development dataset) and asthma-related hospitalisation, A&E attendance, or death (validation dataset) within a 12-month period.

#### Results

Risk factors for asthma-related crisis events included previous hospitalisation, older age, underweight, smoking, and blood eosinophilia. The prediction algorithm had acceptable predictive ability with a receiver operating characteristic of 0.71 [95% confidence interval (CI) = 0.70 to 0.72] in the validation dataset. Using a cut-point based on the 7% of the population at greatest risk results in a positive predictive value of 5.7% [95% CI = 5.3% to 6.1%] and a negative predictive value of 98.9% [95% CI = 98.9% to 99.0%], with sensitivity of 28.5% [95% CI = 26.7% to 30.3%] and specificity of 93.3% [95% CI = 93.2% to 93.4%]; those individuals had an event risk of 6.0% compared with 1.1% for the remaining population. In total, 18 people would need to be followed to identify one admission.

#### Conclusion

This externally validated algorithm has acceptable predictive ability for identifying patients at high risk of asthma-related crisis events and excluding those not at high risk.

#### Keywords

algorithms; asthma; asthma attack; general practice; prediction; risk.

### INTRODUCTION

The challenge of reducing unplanned hospital admissions and avoidable deaths for common chronic conditions, such as asthma, remains unresolved. Despite effective treatments, evidence-based guidelines,<sup>1</sup> and financially incentivised community-based chronic disease management (via the Quality and Outcomes Framework<sup>2</sup>), each year in the UK an average of 1500 people die<sup>3</sup> (on average, 3 a day) and 93 000 are hospitalised due to asthma.<sup>4</sup> A total of 5.4 million people in the UK are currently receiving treatment for asthma: 1.1 million children (1 in 11) and 4.3 million adults (1 in 12).<sup>3</sup> Identification of those at increased risk of these events is beneficial both at an individual level to tailor disease management, and at a population level to inform and modify processes of care.

Many risk factors for poor asthma outcomes have been identified,<sup>5–8</sup> some of which have been combined into risk algorithms, including: Asthma UK's Asthma Attack Risk Checker tool;<sup>9</sup> the Asthma Disease Activity Score;<sup>10</sup> and wheeze frequency, admissions, reliever use, and step on British Thoracic Society medication

guidelines (WARS) score.<sup>11</sup> Recently, an algorithm has also been developed to identify children at risk of life-threatening asthma.<sup>12</sup> These have been derived from small datasets, including those from clinical trials, or the variables used in the prediction tools have required up-to-date personal characteristics, including psychosocial characteristics or adherence to medication for which comprehensive data are difficult to obtain in large populations.<sup>13</sup> An algorithm to identify patients at greatest risk of poor outcomes using electronic healthcare data would overcome this problem and enable a register of patients at high risk to be generated efficiently.

Most prediction algorithms define a severe asthma attack as one that requires oral corticosteroid therapy or hospital attendance/admission;<sup>14</sup> however, this composite scoring includes variables that are not necessarily colinear. Early treatment with prednisolone may stop the deterioration and prevent an accident and emergency (A&E) attendance and, as such, this composite definition may mask the benefits of prompt management of an attack, with increased prednisolone treatment and reduced hospitalisations;<sup>13</sup> as

**M Noble**, MBBChir, GP, Acle Medical Centre, Acle, Norfolk, UK. **A Burden**, MSc, statistician, Observational and Pragmatic Research Institute, Singapore. **S Stirling**, MSc, statistician; **AB Clark**, PhD, statistician; **S Musgrave**, MD, trial manager; **M Pond**, MSc, data manager; **EJ Sims**, PhD, operations manager, Norwich Medical School, University of East Anglia, Norwich, UK. **MA Alsallakh**, PhD, clinical data scientist; **GA Davies**, MD, professor of respiratory medicine, Swansea University Medical School, Swansea, UK. **D Price**, FRCGP, professor of primary care respiratory medicine, Observational & Pragmatic Research Institute, 883 North Bridge Road, #02-05, Southbank, Singapore. **H Pinnock**, MD, professor of primary care respiratory medicine; **A Sheikh**, MD, professor of primary care, Asthma UK Centre for Applied Research, University of Edinburgh, Edinburgh, UK. **S Walker**, PhD, public

and patient representative, Asthma UK, London, UK. **AM Wilson**, MD, professor of respiratory medicine, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK.

#### Address for correspondence

Andrew M Wilson, Floor 2, Bob Champion Research and Education Building, James Watson Road, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.

**Email:** a.m.wilson@uea.ac.uk

**Submitted:** 28 November 2020; **Editor's response:** 12 January 2021; **final acceptance:** 11 June 2021.

#### ©The Authors

This is the full-length article (published online 19 Oct 2021) of an abridged version published in print. Cite this version as: **Br J Gen Pract 2021**; DOI: <https://doi.org/10.3399/BJGP.2020.1042>

## How this fits in

Risk stratification is commonly undertaken in primary care but there are no validated prediction algorithms for people with asthma using routine data. An algorithm was developed using a primary care dataset and externally validated showing acceptable predictive ability with a receiver operating characteristic of 0.71 (95% confidence interval = 0.70 to 0.72). The 7% of the population most at risk had an event rate of 6.0%, compared with 1.1% for the remaining population. This algorithm can be used to identify individuals at high risk of an asthma-related crisis event from primary care electronic health records.

such, it is important to develop algorithms that identify these two risks separately.

The authors aimed to develop and validate a prediction tool to identify individuals at high risk of an asthma-related crisis event (A&E attendance, hospital admission, or death due to asthma) during the following 12 months, calculated from routinely captured electronic health record (EHR) data.

## METHOD

### Data sources

**Derivation dataset.** An analytical dataset was used from a published cohort study<sup>15</sup> that used a database of people aged 12–80 years registered at one of 650 primary care practices in the UK with physician-diagnosed and recorded asthma (with no subsequent code for asthma resolved), measurement of full blood count (FBC) at any time in the past, and 2 years' continuous data. The dataset comprised data from the Clinical Practice Research Datalink (CPRD)<sup>16</sup> between 2001 and 2012. Although the CPRD database contains record-linked primary and secondary care data, including reason for admission to hospital, only data from primary care were used to derive the algorithm because EHRs in UK primary care do not consistently code secondary care events. However, both primary and secondary care data were used when assessing the outcome.

**Validation dataset.** A separate dataset of patients from the Secure Anonymised Information Linkage (SAIL) Databank<sup>17,18</sup> who were registered at 340 general practices in Wales was used to validate the algorithm. Record-linked data from primary and secondary care were available for individual patients and included reason for admission

to hospital. Data on asthma outcomes, healthcare interactions (including GP consultations), and prescribed medications were obtained from the SAIL Databank.

### Eligibility

Patients included in the existing analytical dataset for the derivation of the at-risk algorithm comprised those with:

- active asthma (that is, with a coded diagnosis of asthma and a prescription for asthma treatment in the previous 12 months<sup>19</sup>);
- no diagnosis of any other chronic respiratory disease;
- a valid blood eosinophil count ( $\leq 5000$  blood eosinophils/microlitre [ $\mu\text{L}$ ]); and
- complete data for the baseline and outcome years (the year prior to, and the year following, the last eosinophil count, respectively).

Patients included in the SAIL Databank validation dataset comprised those with at least one asthma diagnosis code before 31 December 2011, no 'asthma resolved' codes between 1 January 2010 and 31 December 2011, and at least one asthma prescription (bronchodilator, corticosteroid, or leukotriene receptor antagonist) code between 1 January 2010 and 31 December 2010. Patients were continuously registered at one general practice between 1 January 2010 and 31 December 2010 (baseline data-collection year) and continually registered (or died) between 1 January 2011 and 31 December 2011 (outcome year).

### Predictors

Details of all variables considered as potential predictors for the at-risk algorithm are shown in Supplementary Table S1. These included age, sex, smoking history, comorbidities, respiratory-related medication, healthcare contacts, and blood eosinophil count. For diagnostic variables (for example, ischaemic heart disease and diabetes), Read codes were queried any time up to the end of the baseline year (that is, 31 December 2010) from the validation and derivation databases. Similarly, for blood eosinophil count, body mass index (BMI), and smoking status, the most recent codes any time before 31 December 2010 were used. For the rest of the variables — prescriptions for asthma, allergic rhinitis, diabetes, anxiety and depression, paracetamol use (which is positively associated with asthma<sup>20</sup>), lower respiratory

tract infection (LRTI) consultations, allergic rhinitis diagnosis — the codes were queried between 1 January 2010 and 31 December 2010.

### Outcome

For the development of the algorithm, the outcome was defined as  $\geq 1$  hospitalisation(s) within 12 months; for the validation of the algorithm, it was defined as a crisis event that comprised an asthma-related hospitalisation, A&E attendance, or death within a 12-month period.

### Statistical analysis

Univariate logistic regression models were used to identify baseline measures of disease severity, patient demographics, and comorbidities predictive of  $\geq 1$  future event(s). Variables showing an association ( $P < 0.05$ ) with an asthma exacerbation resulting in hospital admission in univariable analyses were entered into a multivariable model, which was reduced using backward elimination to produce a final list of predictors of hospital admission. No model updating was undertaken.

The final model was used to create at-risk scores, indicating the risk of an asthma-related crisis event for each patient in the dataset. To do this, coefficients for those factors present in each patient were summed, along with the intercept, to obtain the risk score ( $x$ ), which is the logit of the probability of asthma-related attendance at A&E or hospital admission; the probability is given by  $e^x/(1+e^x)$ . Internal validation was not investigated, as a separate dataset was used to perform external validation. The calibration slope coefficient was estimated by splitting the predicted risk into 10 groups, based on deciles and calculating the percentage of people in those with the outcome, estimating a linear regression model with the predicted risk group against the actual risk.

Discrimination (the ability to distinguish between those who do, and do not, experience the outcome) was assessed by calculating the receiver operating characteristic (ROC) for the risk scores. In addition, the specificity, sensitivity, positive predictive values (PPVs), and negative predictive values (NPVs) were calculated for five different at-risk cut-offs (top 1%, 2%, 5%, 7%, and 10%) for the risk scores for both the derivation and the validation datasets. The overall goodness of fit of the score was assessed by estimating the pseudo  $R^2$  from the logistic regression model. Assuming an asthma prevalence of 6%–7%, a 7% cut-off would, on average,

identify the most at risk 42–49 individuals from a practice of 10 000 patients. A sensitivity analysis was undertaken for the validation cohort including only data related to hospitalisation.

## RESULTS

### Participants

The derivation and validation datasets comprised 58 619 and 174 240 people, respectively (Figure 1). The mean age of participants was 50 years in the derivation dataset and 44 years in the validation dataset, with more females than males in both datasets (Table 1). There were proportionally more people receiving Global Initiative for Asthma (GINA) treatment step 4 or 5 (medium- or high-dose inhaled corticosteroid and long-acting beta-agonist/muscarinic antagonist +/- add on therapies) and more with a diagnosis of, or treatment for, rhinitis in the derivation dataset than in the validation dataset (Table 1). There were differences between the datasets in terms of smoking status, BMI, anxiety and depression, and paracetamol usage. The outcome was present in 1.65% of individuals in the derivation dataset and 1.40% in the validation dataset (Table 1).

The results of the logistic regression are presented in Table 2, which gives the estimated weight of each variable and describes the algorithm used to predict asthma crisis events.

The overall ability of the algorithm to discriminate between patients who subsequently had an asthma-related crisis event and those who did not was acceptable, and similar in the derivation dataset (ROC 0.72, 95% CI = 0.71 to 0.74) and the validation dataset (ROC 0.71, 95% CI = 0.70 to 0.72) (Table 3). Using a cut-point based on the 7% of the population at greatest risk results in a PPV of 5.7% (95% CI = 5.3% to 6.1%) and an NPV of 98.9% (95% CI = 98.9% to 99.0%), with sensitivity and specificity of 28.5% (95% CI = 26.7% to 30.3%) and 93.3% (95% CI = 93.2% to 93.4%), respectively (Table 3). The discriminative ability of the algorithm was similar in the validation cohort when the outcome was confined to hospitalisation only (see Supplementary Table S2). These individuals had a risk of event of 5.68% (Table 4) and 3.31% when considering hospitalisation only (see Supplementary Table S3). The at-risk algorithm showed acceptable prognostic performance in the validation data with a 5.4-fold higher asthma-related crisis event rate in the high-risk group (6.0%) versus the rest of the population (1.1%) at the 7% cut-off (Table 5) or an absolute difference of 4.9%.

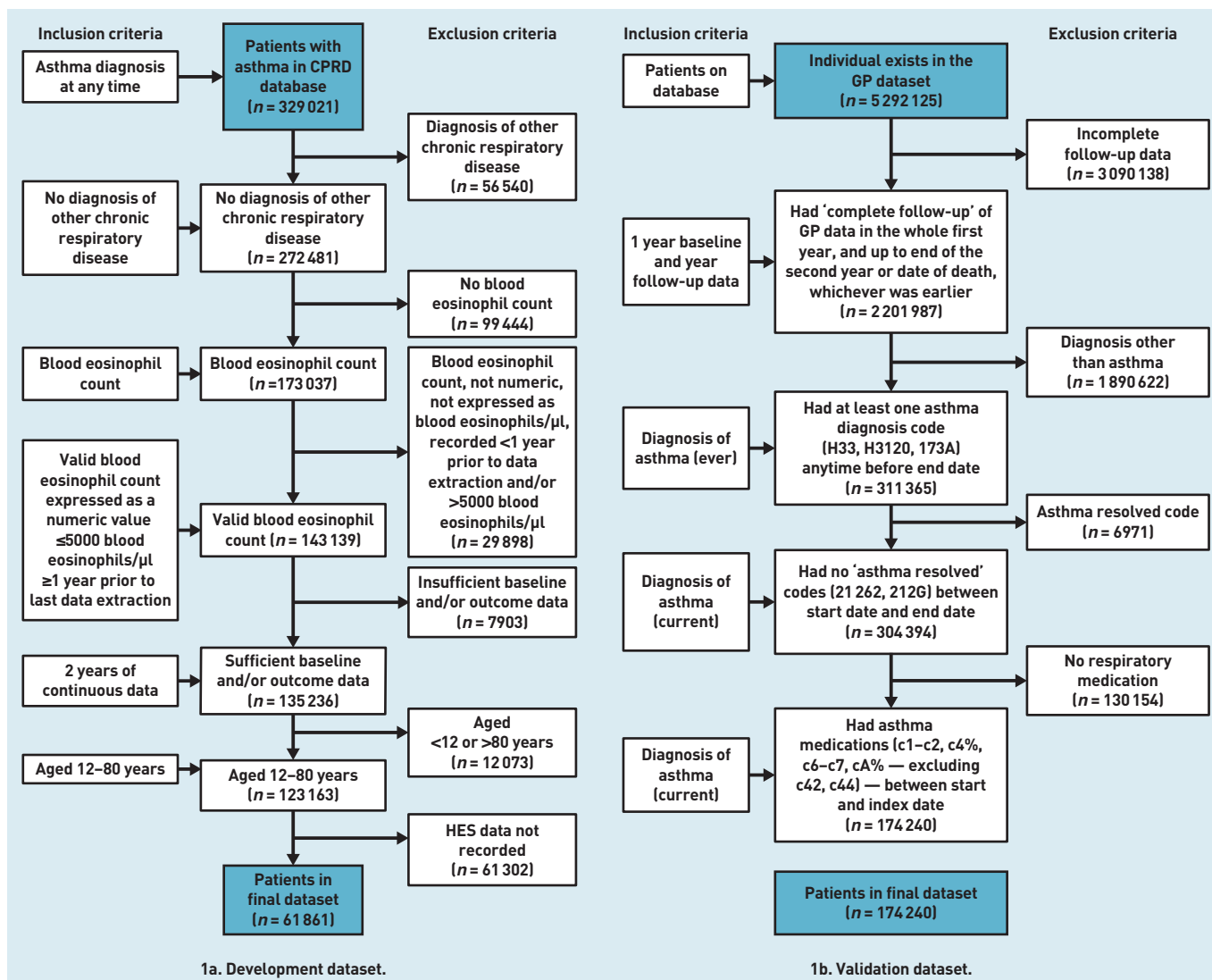


Figure 1. Flow of participants through the study. CPRD = Clinical Practice Research Datalink. HES = Hospital Episode Statistics.

The calibration slopes showed acceptable agreement between deciles of mean risk score and proportions of people experiencing asthma-related crisis events in each decile group, with data points close to the line of equality. The slope coefficient for the derivation dataset was 0.99 [95% CI = 0.92 to 1.05], while that for the validation was 0.85 [95% CI = 0.75 to 0.96] (data not shown).

## DISCUSSION

### Summary

Using data that are routinely available in UK primary care EHRs, the authors derived and externally validated an algorithm containing hospitalisation, older age, underweight, smoking, and blood eosinophilia variables to identify individuals at increased risk of experiencing an asthma-related crisis event. This had acceptable overall characteristics with an ROC of 0.72 in the derivation and

0.71 in the validation cohorts, respectively. Using the top 7% of the score as a cut-off, the algorithm correctly identified 28.5% of the asthma population most at risk and 93.3% of those not at risk. A practice can expect a crisis event to occur in 6.0% of the group that is at risk compared with 1.1% of the rest of the population with asthma. Eighteen people would need to be followed to identify one admission. The algorithm can identify people who are at a five-fold increased risk (absolute difference of 5%) of an asthma-related crisis event compared with those not at risk.

### Strengths and limitations

The main strength of this study is that it used two separate large databases capturing people from different geographical areas with record linkage between primary and secondary care data. The generalisability of the algorithm is illustrated by its similar

**Table 1. Participant characteristics in the derivation ( $n = 58\,619$ ) and validation ( $n = 174\,240$ ) datasets<sup>a</sup>**

Characteristic	Derivation dataset, $n(\%)^b$	Validation dataset, $n(\%)^b$
<b>Sex</b>		
Male	19 684 (33.58)	78 437 (45.02)
Female	38 935 (66.42)	95 803 (54.98)
<b>Age, years, mean (SD)</b>		
	49.70 (16.86)	44.47 (22.57)
<b>Outcome, derivation data: HES hospital admissions for asthma</b>		
≥1 admission	969 (1.65)	—
No admissions	57 650 (98.35)	—
<b>Outcome, validation data: asthma-related crisis event: hospital admission, A&amp;E attendance, or death</b>		
Hospital admission	—	1434 (0.82)
A&E attendance	—	75 (0.04)
Death	—	1235 (0.7)
≥1 asthma events	—	2439 (1.40) <sup>c</sup>
No asthma events	—	171 801 (98.60)
<b>Age group, years</b>		
12–60	40 809 (69.62)	125 802 (72.20)
61–80	17 810 (30.38)	48 438 (27.80)
<b>Smoking status</b>		
Current smoker	10 498 (17.91)	33 880 (19.44)
Ex-smoker	15 564 (26.55) <sup>d</sup>	70 544 (40.49) <sup>e</sup>
<b>BMI</b>		
Underweight: <18.5 kg/m <sup>2</sup>	1332 (2.27)	5160 (2.96)
Normal: 18.5–<25 kg/m <sup>2</sup>	18 403 (31.39)	82 221 (47.19)
Overweight: 25–<30 kg/m <sup>2</sup>	19 182 (32.72)	42 725 (24.52)
Obese: ≥30 kg/m <sup>2</sup>	19 702 (36.61)	44 134 (25.33)
<b>Count of blood eosinophils at baseline</b>		
≤400 $\mu$ l	49 172 (83.88)	155 922 (89.49)
>400 $\mu$ l	9447 (16.12)	18 318 (10.51)
<b>IHD</b>	3549 (6.05)	11 120 (6.30)
<b>≥1 LTRA prescriptions</b>	2871 (4.90)	11 258 (6.46)
<b>Diabetes diagnosis and/or therapy</b>	15 210 (25.95)	12 895 (7.40)
<b>Paracetamol</b>	18 482 (31.53)	28 018 (16.08)
<b>Rhinitis diagnosis and/or drugs in baseline</b>	27 845 (47.50)	27 127 (15.57)
<b>Number of courses of acute oral steroids</b>		
1	6444 (10.99)	17 390 (9.98)
2	2272 (3.88)	7083 (4.07)
≥3	2754 (4.70)	9056 (5.20)
<b>Previous hospitalisations</b>		
≥1	959 (1.64)	1057 (0.61)
<b>GP consultations for LRTI</b>		
1	7346 (12.53)	20 764 (11.92)
≥2	2550 (4.35) <sup>f</sup>	6990 (4.01) <sup>g</sup>
<b>Number of SABA prescriptions</b>		
0–2	30 368 (51.81)	73 688 (42.29)
3–6	17 912 (30.56)	51 725 (29.69)
7–12	7945 (13.55)	33 235 (19.07)
≥13	2394 (4.08)	15 592 (8.95)
<b>Anxiety and/or depression</b>	24 222 (41.32)	39 664 (22.76)
<b>History of anaphylaxis</b>	360 (0.61)	850 (0.49)

... continued

### Funding

The dataset and statistical analyses for the derivation of the algorithm were funded and delivered by the Observational and Pragmatic Research Institute. This article presents independent research funded by the National Institute for Health Research (NIHR) under its Health Technology Assessment programme (grant reference number: 13/34/70). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

### Ethical approval

The study protocol was approved by the Clinical Practice Research Database Independent Scientific Advisory Committee (ISAC) (ISAC approval number: 10\_087).

### Provenance

Freely submitted; externally peer reviewed.

### Competing interests

David Price has: board membership with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan, Mundipharma, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, Teva Pharmaceutical Industries, and Thermo Fisher Scientific; consultancy agreements with Amgen, AstraZeneca, Boehringer Ingelheim, Chiesi, GlaxoSmithKline, Mylan, Mundipharma, Novartis, Pfizer, Teva Pharmaceutical Industries, and Theravance Biopharma; grants and unrestricted funding for investigator-initiated studies (conducted through the Observational and Pragmatic Research Institute) from AstraZeneca, Boehringer Ingelheim, Chiesi, Circassia, Mylan, Mundipharma, Novartis, Pfizer, Regeneron Pharmaceuticals, Respiratory Effectiveness Group, Sanofi Genzyme, Teva Pharmaceutical Industries, Theravance, and the UK National Health Service; payment for lectures/speaking engagements from AstraZeneca, Boehringer Ingelheim, Chiesi, Cipla, GlaxoSmithKline, KYORIN Pharmaceutical, Mylan, Mundipharma, Novartis, Regeneron Pharmaceuticals, Sanofi Genzyme, and Teva Pharmaceutical Industries; payment for the development of educational materials from Mundipharma and Novartis; payment for travel/accommodation/meeting expenses from AstraZeneca, Boehringer Ingelheim, Mundipharma, Mylan, Novartis, Thermo Fisher Scientific; funding for patient enrolment or completion of research from Novartis; stock/stock options from AKL Research and Development, which produces phytopharmaceuticals; ownership of 74% of the social enterprise Optimum Patient Care (Australia and UK) and 74% of Observational and Pragmatic Research Institute (Singapore); and 5% shareholding in Timestamp, which develops adherence monitoring technology. He is also a peer reviewer for grant committees of the NIHR's Efficacy and Mechanism Evaluation programme and Health Technology Assessment, and was an expert witness for GlaxoSmithKline.

**Table 1 Continued. Participant characteristics in the derivation (n = 58 619) and validation (n = 174 240) datasets<sup>a</sup>**

Characteristic	Derivation dataset, n(%) <sup>b</sup>	Validation dataset, n(%) <sup>b</sup>
<b>GINA management steps, regrouped</b>		
No therapy	5180 (8.84)	519 (0.30)
Step 1–2	31 923 (54.46)	73 470 (42.17)
Step 3	7990 (13.63)	85 645 (49.15)
Step 4–5	13 526 (23.07)	14 606 (8.38)

<sup>a</sup>This table contains frequencies of each variable for those individuals for whom all variables included in the algorithm were available. <sup>b</sup>Unless otherwise stated. <sup>c</sup>Individuals can have ≥1 event, so the total is less than the sum of the individual occurrences. <sup>d</sup>Missing 32 557 were non-smokers. <sup>e</sup>Missing 69 816 were non-smokers. <sup>f</sup>The remaining 48 723 are 'none'. <sup>g</sup>The remaining 146 486 are 'none'. A&E = accident and emergency. BMI = body mass index. GINA = Global Initiative for Asthma. HES = Hospital Episode Statistics. IHD = ischaemic heart disease. LRTA = leukotriene receptor antagonist. LTRI = lower respiratory tract infection. SABA = short-acting beta-agonist. SD = standard deviation.

behaviour in two different datasets. The data on cause (asthma related or not) for hospital admission when deriving the algorithm were deliberately ignored as this information, although predictive of future events, is not routinely available in primary care datasets. However, by linking primary care data with that from secondary care for the purposes of assessing the outcome, it was possible to confirm that the algorithm identifies people at risk of an asthma-related crisis event.

The limitations were that patients in the derivation, but not the validation, cohort needed to have had a valid FBC to be entered into the database (although specific

values, such as eosinophil counts, were not required). This may have resulted in differences in some of the characteristics (for example, age, sex, asthma severity, and number of comorbidities); however, the authors do not believe there is any difference in the diagnosis or management of people with asthma between Wales and England as both countries follow national guidelines.<sup>1</sup>

The databases contained data that are now a decade old (validation 2001–2012, validation 2011–2012) and asthma guidelines have been updated in this time. Modifications made have included the use of high-dose inhaled corticosteroids

**Table 2. Adjusted estimated regression coefficients in algorithm derivation**

Coefficient	β-coefficient (SE)	P-value
<b>Constant</b>	-5.013	—
<b>Factors considered for algorithm</b>		
<b>Age group, years, reference group ≤60 years</b>		
61–80	0.192 (0.076)	0.01
<b>BMI, kg/m<sup>2</sup>, reference group 18.5–&lt;25 kg/m<sup>2</sup></b>		
Underweight: <18.5 kg/m <sup>2</sup>	0.597 (0.179)	0.001
Overweight: 25–<30 kg/m <sup>2</sup>	-0.210 (0.087)	0.02
Obese: ≥30 kg/m <sup>2</sup>	-0.103 (0.083)	0.21
<b>Smoking status, reference group non-smoker</b>		
Current smoker	0.251 (0.089)	0.005
Ex-smoker	0.093 (0.078)	0.24
<b>Blood eosinophil counts, reference group ≤400/μl</b>		
>400/μl	0.237 (0.084)	0.005
<b>Rhinitis diagnosis and/or drugs</b>	-0.212 (0.068)	0.002
<b>Diabetes diagnosis and/or therapy</b>	0.378 (0.071)	<0.001
<b>IHD diagnosis</b>	0.243 (0.116)	0.036

... continued



**Table 2 Continued. Adjusted estimated regression coefficients in algorithm derivation**

Coefficient	$\beta$ -coefficient (SE)	P-value
Anxiety and/or depression	0.192 (0.068)	0.005
History of anaphylaxis	0.790 (0.275)	0.004
GINA management step, reference group step 1–2		
No therapy	0.568 (0.121)	<0.001
Step 3	0.101 (0.108)	0.35
Step 4–5	0.461 (0.078)	<0.001
GP consultations for LRTIs		
1	0.313 (0.086)	<0.001
$\geq 2$	0.206 (0.122)	0.09
Acute oral steroids		
1	0.551 (0.095)	<0.001
2	0.975 (0.120)	<0.001
$\geq 3$	1.141 (0.107)	<0.001
Paracetamol	0.204 (0.071)	0.004
Hospitalisation, HES data		
$\geq 1$	1.877 (0.105)	<0.001

The pseudo  $R^2$  for the equation was 0.0846. Example: for a 24-year-old non-smoker with a BMI of 28 kg/m<sup>2</sup>, a history of rhinitis, and anxiety, who is receiving Step 3 GINA therapy and has had one course of prednisolone for asthma in the previous year, the at-risk probability =  $e^x / (1 + e^x) = e^{-2.615} / (1 + e^{-2.615}) = 0.01$ , where  $x = -4.591$ , calculated from  $-5.013$  (constant)  $-0.210$  (BMI)  $-0.212$  (rhinitis)  $+ 0.192$  (anxiety)  $+ 0.101$  (asthma treatment)  $+ 0.551$  (asthma attacks). BMI = body mass index. GINA = Global Initiative for Asthma. HES = Hospital Episode Statistics. IHD = ischaemic heart disease. LTRI = lower respiratory tract infection. SE = standard error.

to abort an asthma attack,<sup>21</sup> vitamin D monitoring and therapy,<sup>22</sup> and the use of monoclonal antibody therapies.<sup>23</sup> However, there have been no marked changes to the understanding of the aetiology of asthma crises or deaths since the data were collected, and the software systems and determinants of coding decisions in day-to-day practice remain comparable. The authors did not, however, have access to information on medication adherence or social circumstances.

Socioeconomic status has been shown to be a risk factor for hospitalisation<sup>24</sup> and an independent predictor for life-threatening asthma in children.<sup>12</sup> Unfortunately, routine data do not contain this information, although algorithms have been developed for assessing prescription uptake<sup>25</sup> and socioeconomic status is available from postcode data,<sup>26</sup> both of which may be applied to future algorithms.

In addition, the authors did not have death or A&E data in the derivation cohort,

**Table 3. Measures of the prognostic performance of the algorithm in the derivation dataset ( $n = 58\ 619$ ) and validation dataset ( $n = 174\ 240$ ) (hospitalisation, death, or A&E attendance) for risk score thresholds of the top 1%, 2%, 5%, 7%, and 10% of patients with asthma**

Quintile	Derivation dataset, ROC 0.72 (95% CI = 0.71 to 0.74)					Validation dataset, ROC 0.71 (95% CI = 0.70 to 0.72)				
	Risk score threshold	PPV, %	NPV, %	Sensitivity, %	Specificity, %	Risk score threshold	PPV, %	NPV, %	Sensitivity, %	Specificity, %
Top 1%	0.104	19.3	98.5	11.7	99.2	0.073	8.9	98.7	6.4	99.1
Top 2%	0.067	13.6	98.6	16.5	98.2	0.056	7.6	98.7	11.0	98.1
Top 5%	0.041	8.29	98.7	25.1	95.3	0.038	6.2	98.9	22.2	95.2
Top 7%	0.035	6.99	98.7	29.6	93.4	0.031	5.7	98.9	28.5	93.3
Top 10%	0.028	5.89	98.8	35.6	90.4	0.025	4.9	99.0	35.1	90.4

A&E = accident and emergency. NPV = negative predictive value. PPV = positive predictive value. ROC = receiver operating characteristic.

**Table 4. Number of events by risk strata in the derivation dataset (n = 58 619) and validation dataset (n = 174 240) (hospitalisation, death, or A&E attendance) cohorts, with risk score thresholds of top 1%, 2%, 5%, 7%, and 10% of patients with asthma**

Quintile	Derivation dataset			Validation dataset		
	Risk score threshold	People, n(%)	Events, n(%)	Risk score threshold	People, n(%)	Events, n(%)
Top 1%	0.104	587 (1)	113 (19.3)	0.073	1751 (1)	155 (8.85)
Top 2%	0.067	1173 (2)	160 (13.6)	0.056	3494 (2)	267 (7.64)
Top 5%	0.041	2931 (5)	243 (8.3)	0.038	8727 (5)	541 (6.20)
Top 7%	0.035	4106 (7)	287 (7)	0.031	12 225 (7)	694 (5.68)
Top 10%	0.028	5862 (10)	345 (5.9)	0.025	17 427 (10)	857 (4.92)
Total	n/a	58 619 (100)	969 (1.65)	n/a	174 240 (100)	2439 (1.40)

A&E = accident and emergency; SAIL = Secure Anonymised Information Linkage.

although it was available for the validation cohort. Despite this, it has been shown that the performance of the prediction algorithm is similar when considering both hospitalisation alone or the composite of hospitalisation, A&E attendance, or death.

**Table 5. Selected characteristics of individuals identified as being at risk, according to suggested cut-point of top 7%**

	Derivation dataset		Validation dataset	
	At risk, n(%)	Not at risk, n(%)	At risk, n(%)	Not at risk, n(%)
<b>Total</b>	4106 (100.00)	54 513 (100.00)	10 042 (100.00)	164 198 (100.00)
<b>Age group, years</b>				
12–60	2250 (54.80)	38 559 (70.73)	3980 (39.63)	121 822 (74.19)
61–80	1856 (45.20)	15 954 (29.27)	6062 (60.37)	42 376 (25.81)
<b>Smoking status</b>				
Non-smoking	1879 (45.76)	30 678 (56.28)	1806 (17.98)	68 010 (41.42)
Current smoker	985 (23.99)	9513 (17.45)	2913 (29.01)	30 967 (18.86)
Ex-smoker	1242 (30.25)	14 322 (26.27)	5323 (53.01)	65 221 (39.72)
<b>IHD</b>				
No	3455 (84.15)	51 615 (94.68)	7868 (78.35)	155 252 (94.55)
Yes	651 (15.85)	2898 (5.32)	2174 (21.65)	8946 (5.45)
<b>History of anaphylaxis</b>				
No	3988 (97.13)	54 271 (99.56)	9797 (97.56)	163 593 (99.63)
Yes	118 (2.87)	242 (0.44)	245 (2.44)	605 (0.37)
<b>Diabetes and/or therapy</b>				
No	1957 (47.66)	41 452 (76.04)	7859 (78.26)	153 486 (93.48)
Yes	2149 (52.34)	13 061 (23.96)	2183 (21.74)	10 712 (6.52)
<b>Blood eosinophil count</b>				
≤400/μl	3113 (75.82)	46 059 (84.49)	8139 (81.0)	147 783 (90.0)
>400/μl	993 (24.18)	8454 (15.51)	1903 (19.0)	16 415 (10.0)
<b>BMI</b>				
Normal: 18.5–<25 kg/m <sup>2</sup>	1121 (27.3)	17 282 (31.7)	3943 (39.27)	78 278 (47.67)
Underweight: <18.5 kg/m <sup>2</sup>	221 (5.38)	1111 (2.04)	638 (6.35)	4522 (2.75)
Overweight: 25–<30 kg/m <sup>2</sup>	994 (24.21)	18 188 (33.36)	2094 (20.85)	40 631 (24.75)
Obese: ≥29 kg/m <sup>2</sup>	1770 (43.11)	17 932 (32.89)	3367 (33.53)	40 767 (24.83)

... continued

Although the number of short-acting beta-agonist (SABA) prescriptions were included in the list of potential variables, long-acting beta-agonist as monotherapy (which has been described as a risk factor in asthma deaths<sup>27</sup>) was not, as this regime is rarely prescribed.<sup>28</sup> This algorithm does not predict community-based asthma attacks requiring oral prednisolone.

#### Comparison with existing literature

The WARS score had an ROC of 0.83 for prednisolone use,<sup>11</sup> but the performance of the score in terms of crisis events is unknown; likewise, the performance measurements of the risk score developed by Bateman *et al*<sup>10</sup> for asthma attacks are not published. However, the Respiratory Effectiveness Group initiative published an algorithm to predict the risk of ≥2 attacks in the subsequent 2 years with an ROC of 0.79 [95% CI=0.78 to 0.79].<sup>29</sup> Recent evidence<sup>27</sup> suggests that disease severity is an unreliable measure of risk and, indeed, the results presented here confirmed that GINA treatment step 'no therapy' was as statistically significant a risk factor as steps 4–5.

In terms of non-respiratory hospitalisation prediction algorithms, the QRISK2 score — which is widely used in the NHS to predict cardiovascular events — has an *R*<sup>2</sup> of 43.5 and 38.4, and an ROC statistic of 0.82 and 0.79 for females and males, respectively.<sup>30</sup> A systematic review of risk prediction models to predict emergency admission in community-dwelling adults<sup>31</sup> identified 27 different such models and showed that those using clinical data (as in the algorithm presented here) outperformed those using self-reported data; C-statistics ranged from 0.63 to 0.83. The algorithm presented here, which utilised clinical data, had a comparable level of calibration (C-statistic 0.72) to other clinically useful algorithms.

Outcome data were collected as events over a 12-month period to avoid seasonal variations. The algorithm, therefore, predicts hospitalisation in the following year; however, an individual's risk status can change if, for example, they had a hospitalisation just within, or without, of a 365-day period. Different algorithms can show substantial variation in risk at the individual level<sup>32</sup> and should complement physician assessment based on knowledge about individuals.

Nevertheless, the growing workloads on primary care clinicians and the ongoing challenge of rising unplanned admissions and avoidable deaths makes accurate identification and targeting of the individuals



**Table 5 Continued. Selected characteristics of individuals identified as being at risk, according to suggested cut-point of top 7%**

	Derivation dataset		Validation dataset	
	At risk, n(%)	Not at risk, n(%)	At risk, n(%)	Not at risk, n(%)
<b>GP consults for LRTIs</b>				
0	1982 (48.27)	46 741 (85.74)	4014 (39.97)	142 472 (86.77)
1	1341 (32.66)	6005 (11.02)	3521 (35.06)	17 243 (10.50)
≥2	783 (19.07)	1767 (3.24)	2507 (24.97)	4483 (2.73)
<b>Rhinitis and/or therapy</b>				
No	2359 (57.45)	28 415 (52.13)	8616 (85.8)	138 497 (84.35)
Yes	1747 (42.55)	26 098 (47.87)	1426 (14.2)	25 701 (15.65)
<b>Anxiety and/or depression</b>				
No	1698 (41.35)	32 699 (59.98)	5534 (55.11)	129 042 (78.59)
Yes	2408 (58.65)	21 814 (40.02)	4508 (44.89)	35 156 (21.41)
<b>Acute oral steroids</b>				
0	643 (15.66)	46 706 (85.68)	543 (5.41)	140 168 (85.37)
1	817 (19.90)	5670 (10.40)	972 (9.68)	16 418 (10.0)
2	957 (23.31)	1336 (2.45)	2058 (20.49)	5025 (3.06)
≥3	1689 (41.13)	801 (1.47)	6469 (64.42)	2587 (1.58)
<b>Paracetamol</b>				
No	1729 (42.11)	38 408 (70.46)	6481 (64.54)	139 741 (85.11)
Yes	2377 (57.89)	16 105 (29.54)	3561 (35.46)	24 457 (14.89)
<b>GINA management step</b>				
No therapy	213 (5.19)	4967 (9.11)	17 (0.17)	502 (0.31)
Step 1–2	934 (22.75)	30 989 (56.85)	803 (8.0)	72 667 (44.26)
Step 3	389 (9.47)	7601 (13.94)	3404 (33.9)	82 241 (50.09)
Step 4–5	2570 (62.59)	10 956 (20.10)	5818 (57.94)	8788 (5.35)
<b>Baseline hospital admissions<sup>a</sup></b>				
No	3158 (76.91)	54 502 (99.98)	8989 (89.51)	164 194 (100)
Yes	948 (23.09)	11 (0.02)	1053 (10.49)	<5 (0)
<b>Outcome: asthma-related crisis event<sup>b</sup></b>				
No	3819 (93.01)	53 831 (98.75)	9438 (93.99)	162 363 (98.88)
Yes	287 (6.99)	682 (1.25)	604 (6.01)	1835 (1.12)

<sup>a</sup>In the validation dataset, actual values are masked due to small frequencies in one category. <sup>b</sup>Hospitalisation or A&E attendance in derivation dataset, and any of hospitalisation, A&E attendance, or death in validation data. A&E = accident and emergency; BMI = body mass index; GINA = Global Initiative for Asthma; IHD = ischaemic heart disease; LTRI = lower respiratory tract infection.

of adverse outcomes.<sup>27,35</sup> Current guidelines recommend that patients are assessed for risk of future attacks. The indicators recommended include a history of previous attacks, SABA use, and other markers of disease control, atopy, and environmental tobacco exposure in children; in adults, these include smoking, obesity, and depression.

In April 2020, Quality and Outcomes Framework indicators for disease control were changed<sup>36</sup> from 'Royal College of Physicians (RCP) 3 questions' (on asthma):<sup>37</sup>

*Have you had difficulty sleeping because of your asthma symptoms (including cough)? Have you had your usual asthma symptoms during the day (for example, cough, wheeze, chest tightness, or breathlessness)? Has your asthma interfered with your usual activities (for example, housework, work, school)?*

to the Asthma Control Test score plus the number of exacerbations in the previous 12 months. Achieving these new indicators requires more clinician time and greater participation from patients. Failure to attend appointments is, in itself, a risk factor for poor outcomes.<sup>35</sup>

The algorithm developed and presented here simplifies the collection and weights the statistical significance of multiple risk factors. It has the potential to save clinicians' time and provide accurate real-time assessments of patients' risk and, as it does not require patients to attend a consultation, also bypasses the dangers of inverse care associated with poor attendance at appointments. The algorithm also concurs with, and provides a mechanism to identify, important markers highlighted in the National Review of Asthma Deaths report,<sup>27</sup> such as patients on no treatment for their asthma. It can be used to generate alerts or prompts to identify patients at high risk of asthma crisis events (A&E attendance, hospitalisation, or death), when their EHRs are accessed so care can be targeted appropriately.

The algorithm is currently being used in a study<sup>38</sup> to validate the role of at-risk asthma registers in primary care. Further work is also needed to explore some of the unexpected indicators, such as low BMI, and to find a way to incorporate important social and behavioural determinants that are not currently captured in primary care EHRs.

### Contributors

Annie Burden and Susan Stirling contributed equally to the manuscript as statisticians.

### Acknowledgements

The authors would like to thank Derek Skinner of the Observational and Pragmatic Research Institute for his support in analytical dataset generation and statistical analyses. They would also like to acknowledge the support of the Asthma UK Centre for Applied Research for its help with this study.

### Open access

This article is Open Access: CC BY 4.0 licence (<http://creativecommons.org/licenses/by/4.0/>).

### Discuss this article

Contribute and read comments about this article: [bjgp.org/letters](http://bjgp.org/letters)

at highest risk an essential part of primary care strategy.

### Implications for practice

Primary care software systems routinely use prompts to alert clinicians to overdue asthma reviews and the overordering — and, by implication, overuse — of SABAs. Both are helpful markers of risk that are not always recognised as such,<sup>13,33,34</sup> but they do not reflect the range and complexity of factors found in patients who are most at risk

## REFERENCES

1. British Thoracic Society and Scottish Intercollegiate Guidelines Network. *British guideline on the management of asthma: a national clinical guideline*. 2019. <https://www.brit-thoracic.org.uk/quality-improvement/guidelines/asthma> (accessed 8 Oct 2021).
2. NHS England. *Report of the Review of the Quality and Outcomes Framework in England*. <https://www.england.nhs.uk/wp-content/uploads/2018/07/quality-outcome-framework-report-of-the-review.pdf> (accessed 8 Oct 2021).
3. Asthma UK. Asthma facts and statistics. 2021. <https://www.asthma.org.uk/about/media/facts-and-statistics> (accessed 17 Sep 2021).
4. Mukherjee M, Stoddart A, Gupta RP, *et al*. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. *BMC Med* 2016; **14**(1): 113.
5. Hanson JR, Lee BR, Williams DD, *et al*. Developing a risk stratification model for predicting future health care use in asthmatic children. *Ann Allergy Asthma Immunol* 2016; **116**(1): 26–30.
6. Vollmer WM, Markson LE, O'Connor E, *et al*. Association of asthma control with health care utilization: a prospective evaluation. *Am J Respir Crit Care Med* 2002; **165**(2): 195–199.
7. Dougherty RH, Fahy JV. Acute exacerbations of asthma: epidemiology, biology and the exacerbation-prone phenotype. *Clin Exp Allergy* 2009; **39**(2): 193–202.
8. Buelo A, McLean S, Julious S, *et al*. At-risk children with asthma (ARC): a systematic review. *Thorax* 2018; **73**(9): 813–824.
9. Asthma UK. Asthma attack risk checker. <https://www.asthma.org.uk/advice/manage-your-asthma/risk> (accessed 8 Oct 2021).
10. Bateman ED, Buhl R, O'Byrne PM, *et al*. Development and validation of a novel risk score for asthma exacerbations: the risk score for exacerbations. *J Allergy Clin Immunol* 2015; **135**(6): 1457–1464.
11. Blakey JD, Obediat M, Pogson Z, *et al*. A simple asthma severity score predicts exacerbations. *Am J Respir Crit Care Med* 2011; **183**: A2248.
12. Lee M, Bogdanova Y, Chan M, *et al*. Development and validation of a risk score to identify children at risk of life-threatening asthma. *J Asthma* 2020; DOI: 10.1080/02770903.2020.1841224.
13. Smith JR, Noble MJ, Musgrave S, *et al*. The at-risk registers in severe asthma (ARRISA) study: a cluster-randomised controlled trial examining effectiveness and costs in primary care. *Thorax* 2012; **67**(12): 1052–1060.
14. Reddel HK, Taylor DR, Bateman ED, *et al*. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. *Am J Respir Crit Care Med* 2009; **180**(1): 59–99.
15. Price DB, Rigazio A, Campbell JD, *et al*. Blood eosinophil count and prospective annual asthma disease burden: a UK cohort study. *Lancet Respir Med* 2015; **3**(11): 849–858.
16. Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; **44**(3): 827–836.
17. Lyons RA, Jones KH, John G, *et al*. The SAIL databank: linking multiple health and social care datasets. *BMC Med Inform Decis Mak* 2009; **9**: 3.
18. Ford DV, Jones KH, Verplancke J-P, *et al*. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009; **9**: 157.
19. *New GMS Contract 2003: investing in general practice*. NHS Confederation, British Medical Association, 2003; London: BMA.
20. Shaheen SO, Sterne JA, Songhurst CE, Burney PG. Frequent paracetamol use and asthma in adults. *Thorax* 2000; **55**(4): 266–270.
21. Quon BS, Fitzgerald JM, Lemiere C, *et al*. Increased versus stable doses of inhaled corticosteroids for exacerbations of chronic asthma in adults and children. *Cochrane Database Syst Rev* 2010; **10**: CD007524.
22. Jolliffe DA, Greenberg L, Hooper RL, *et al*. Vitamin D supplementation to prevent asthma exacerbations: a systematic review and meta-analysis of individual participant data. *Lancet Respir Med* 2017; **5**(11): 881–890.
23. Bel EH, Wenzel SE, Thompson PJ, *et al*. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. *N Engl J Med* 2014; **371**(13): 1189–1197.
24. Wallar LE, De Prophetis E, Rosella LC. Socioeconomic inequalities in hospitalizations for chronic ambulatory care sensitive conditions: a systematic review of peer-reviewed literature, 1990–2018. *Int J Equity Health* 2020; **19**(1): 60.
25. Bryson CL, Au DH, Young B, *et al*. A refill adherence algorithm for multiple short intervals to estimate refill compliance (ReComp). *Med Care* 2007; **45**(6): 497–504.
26. Danesh J, Gault S, Semmence J, *et al*. Postcodes as useful markers of social class: population based study in 26 000 British households. *BMJ* 1999; **318**(7187): 843–844.
27. Royal College of Physicians. *National review of asthma deaths*. 2015. <https://www.rcplondon.ac.uk/projects/national-review-asthma-deaths> (accessed 8 Oct 2021).
28. Wasilevich EA, Clark SJ, Cohn LM, Dombkowski KJ. Long-acting  $\beta$ -agonist monotherapy among children and adults with asthma. *Am J Manag Care* 2011; **17**(4): e91–e95.
29. Blakey JD, Price DB, Pizzichini E, *et al*. Identifying risk of future asthma attacks using uk medical record data: a Respiratory Effectiveness Group initiative. *J Allergy Clin Immunol Pract* 2017; **5**(4): 1015–1024.
30. Hippisley-Cox J, Coupland C, Vinogradova Y, *et al*. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008; **336**(7659): 1475–1482.
31. Wallace E, Stuart E, Vaughan N, *et al*. Risk prediction models to predict emergency hospital admission in community-dwelling adults: a systematic review. *Med Care* 2014; **52**(8): 751–765.
32. Pate A, Emsley R, Ashcroft DM, *et al*. The uncertainty with using risk prediction models for individual decision making: an exemplar cohort study examining the prediction of cardiovascular disease in English primary care. *BMC Med* 2019; **17**(1): 134.
33. Nwaru BI, Ekström M, Hasvold P, *et al*. Overuse of short-acting  $\beta_2$ -agonists in asthma is associated with increased risk of exacerbation and mortality: a nationwide cohort study of the global SABINA programme. *Eur Respir J* 2020; **55**(4): 1901872.
34. McKibben S, Bush A, Thomas M, Griffiths C. "Tossing a coin": defining the excessive use of short-acting beta<sub>2</sub>-agonists in asthma — the views of general practitioners and asthma experts in primary and secondary care. *NPJ Prim Care Respir Med* 2018; **28**(1): 26.
35. Mohan G, Harrison BD, Badminton RM, *et al*. A confidential enquiry into deaths caused by asthma in an English health region: implications for general practice. *Br J Gen Pract* 1996; **46**(410): 529–532.
36. NHS England, British Medical Association. *2020/21 General Medical Services (GMS) contract Quality and Outcomes Framework (QOF): guidance for GMS contract 2020/21 in England*. 2020. <https://www.england.nhs.uk/coronavirus/wp-content/uploads/sites/52/2020/03/C0713-202021-General-Medical-Services-GMS-contract-Quality-and-Outcomes-Framework-QOF-Guidance.pdf> (accessed 8 Oct 2021).
37. Pearson MG, Bucknall CE, eds. *Measuring clinical outcome in asthma: a patient-focused approach*. London: Royal College of Physicians, 1999.
38. Smith JR, Musgrave S, Payerne E, *et al*. At-risk registers integrated into primary care to stop asthma crises in the UK (ARRISA-UK): study protocol for a pragmatic, cluster randomised trial with nested health economic and process evaluations. *Trials* 2018; **19**(1): 466.