# BMJ Open

# Validating the recording of exacerbations of asthma in electronic health records: a systematic review protocol

Elizabeth Moore [ID] , Zakariah Z Gassasse [ID] , Jennifer K Quint [ID]

School of Public Health, Imperial College London, London, UK

**Correspondence to**
Elizabeth Moore;
liz.moore@imperial.ac.uk

## ABSTRACT

**Introduction** Asthma exacerbations or 'attacks' can vary in severity from mild worsening of symptoms to life-threatening changes that require urgent hospital care. Understanding these exacerbations is crucial to improving treatment and support for patients. Electronic health records (EHR) using anonymised data from people with asthma in primary and secondary care can be used to understand exacerbations and outcomes. However, previous studies found significant heterogeneity in the algorithms used to define asthma exacerbations. Validating definitions of asthma exacerbations in EHR will lead to more robust and comparable evidence in future research.

**Methods and analysis** Medline and Embase will be searched for the key concepts relating to asthma exacerbations, EHR and validation. All studies that validate exacerbations of asthma in EHR and administrative claims databases published before 30 May 2024 and written in English will be considered. Validated algorithms for asthma exacerbations or attacks must be compared against a reference or gold standard definition, and a measure of validity must be included. Articles will be screened for inclusion by two independent reviewers with any disagreements resolved by consensus or arbitration by a third reviewer. Study details will be extracted, and the risk of bias will be assessed using a QUADAS-2 tailored to this review.

**Ethics & dissemination** No ethical approval is required as this is a review of previously published literature. Results will be disseminated in a peer-reviewed journal with the aim of being used in future research to help identify asthma exacerbation in EHR.

**PROSPERO registration number** CRD42024545081

## STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This review will report the methods and findings of studies that validate definitions of asthma exacerbations in electronic health records (EHR).
⇒ Medical databases will be searched for the key concepts relating to asthma exacerbations, EHR and validation.
⇒ Study details will be extracted, and the risk of bias will be assessed using a QUADAS-2 tailored to this review.
⇒ Validated algorithms for asthma exacerbations or attacks will be compared against a reference or gold standard definition, and measures of validity will be reported.

## INTRODUCTION

Asthma is a common chronic lung disease that affects people of all ages and is characterised by inflammation and muscle tightening around the airways resulting in symptoms such as cough, wheeze, shortness of breath and chest tightness.[1] Exacerbations of asthma are defined in the Global Initiative for Asthma as events that involve a progressive increase in symptoms and a progressive decrease in lung function that are 'sufficient to require a change in treatment'.[2] Exacerbations of asthma or 'attacks' can vary in severity from mild worsening of symptoms that only require the use of inhalers to severe, life-threatening changes that require urgent treatment in hospital.[3] Severe asthma attacks have accounted for over 90,000 UK hospital admissions per annum.[4] There is convincing evidence of heterogeneity in acute asthma, as there is in stable asthma, and validating criteria to define asthma attacks may improve treatment and outcomes.[5]

Electronic health records (EHR) can be used in observational studies to understand diseases, treatments and outcomes. Researchers can access anonymised data collected from primary and secondary care, providing them with large study samples that are often more generalisable to wider populations. Information regarding diagnoses and clinical events is stored in EHR as clinical codes and/or associated values. Health records can be retrieved using coding (either a single code or an algorithm consisting of multiple codes or associated values), and researchers can apply additional restrictions if desired (eg, age or exclusion of other diseases). Some authors have also used natural language

processing and machine learning techniques to identify asthma diagnoses in large databases through automated algorithm generation.[6–8] This methodology follows a set of rules that may make codelist generation 'Blackbox' and are not always clear.

Previous studies have shown the importance of validating definitions in EHR to ensure robust, comparable study findings.[9] Comparing EHR data with a gold standard is the most common method to assess validity of algorithms, and these gold standards can include paper records, verification by a treating physician or through patient questionnaires, data review and alternative data sources such as linkable datasets.[10]

Previous studies have sought to examine the validity of asthma diagnoses in EHR. A scoping review by Al Sallakh et al[11] found a lack of consensus in approaches to defining asthma or assessing asthma outcomes in EHR with significant heterogeneity in the algorithms used to define exacerbations of asthma. It highlighted the fundamental need to reach a consensus on the definitions of asthma exacerbations in EHR. Nissen et al[12] carried out a systematic review of studies looking at asthma recording in EHR. Their review found that definitions and methods of asthma diagnosis validation vary widely across different EHR databases, and asthma symptoms present differently depending on the setting (eg, primary care, secondary care and urgent care). Sharifi et al[13] conducted a systematic review of validated methods to capture acute bronchospasm, which is a hallmark of asthma, using administrative or claims data. They found a paucity of studies using rigorous methods to validate algorithms for the identification of acute asthma or bronchospasms in general populations, with only three studies reporting any validation, and all were among paediatric populations. A similar review to validate acute exacerbations of chronic obstructive pulmonary disease (COPD) is currently being undertaken[14]; however, to our knowledge, there has not been another systematic review of studies that validate definitions of asthma exacerbations in EHR.

## OBJECTIVE

This review will report the methods and findings of studies that validate definitions of asthma exacerbations in EHR. The target population are people with asthma, the intervention measured (index test) will be the detection algorithms for exacerbations of asthma, the comparison will be the reference standard used to confirm exacerbations of asthma and the outcome will be the validity of the detection algorithms. Studies will be included from any country, in any EHR database and using any clinical coding. In our review, we will specifically explore the following:

► The database and type of EHR used.
► The algorithm (codelists) used to detect asthma exacerbations.
► The reference standard used to validate asthma exacerbations.

► The estimated validity of the detection algorithm for asthma exacerbations.

## METHODS AND ANALYSIS

Medline and Embase will be searched via Ovid for the key concepts of 'asthma exacerbations', 'electronic health records' or 'administrative claims database' and 'validation'. The full search strategy is described in online supplemental file 1. In order to detect the validation terms, we will use the same search strategy used by Stone et al[14] in a similar review validating COPD exacerbations in EHR. This was based on search methodology by Benchimol et al[15] and strategies used in similar reviews of validation studies in EHR databases.[12 16–18]

### Inclusion criteria

► Studies written in English and published before 30 May 2024 will be considered.
► Data must be from an EHR or administrative claims database.
► Adult and paediatric studies will be included and where appropriate, different treatment regimens and outcomes will be taken into consideration.
► Validated algorithms for asthma exacerbations or attacks must be compared against a reference or gold standard definition (eg, records review by treating physicians).
► A measure of validity must be included (eg, positive predictive value (PPV), negative predictive value (NPV), and sensitivity and specificity) or can be calculable from information provided in the study.

### Exclusion criteria

Studies will be excluded if they only look at asthma diagnosis rather than asthma attacks/exacerbations.

## DATA MANAGEMENT AND SYNTHESIS

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist by Moher et al[19] will be followed, and the PRISMA flow diagram for this protocol is shown in figure 1 in online supplemental file 2. Articles found in the literature search will be stored on EndNote 21 (Clarivate Analytics, Philadelphia, Pennsylvania, USA), and duplicates will be removed. All unique articles will be screened by two reviewers, and if the inclusion criteria are met, a full text review will be carried out. Any disagreement regarding the inclusion/exclusion of articles will be resolved by consensus or arbitration by a third reviewer, and reasons for exclusion from the review will be recorded. Both reviewers will read the full texts and will independently extract study details. Data will be tabulated and stored in Microsoft Excel (Microsoft, Redmond, Washington, USA), and the following information from each study will be recorded:

► Study details (including title, first author and year of publication).
► Study aim/research question.

- ► EHR database used.
- ► Population (location, time period).
- ► Type of algorithm(s) used to detect asthma exacerbations (eg, clinical coding scheme).
- ► Reference/gold standard used to compare the algorithm(s) against.
- ► Measure(s) of validity (eg, PPV and NPV).
- ► Results of validated measures.
- ► Prevalence of asthma exacerbations.
- ► Information to calculate validity (where available: true positives, false positives, true negatives, false negatives).

The primary outcome measure sought will be the validity of the asthma exacerbation detection algorithm.

Risk of bias (ROB) will be assessed using a quality assessment tool for diagnostic accuracy studies known as QUADAS-2.[20] We will tailor the QUADAS-2 to this specific review as done in other similar reviews[14 21] using a recommended reporting checklist by Benchimol et al[15] for use in validation studies of health administrative data. Our tailored QUADAS-2 can be found in online supplemental file 3. If there are multiple validations reported, we will complete a ROB assessment for each validation. Results will be presented in the text and in tables to summarise study details, the algorithms used to validate exacerbations of asthma in EHR, the reference standard used to validate the algorithm, the validity of the algorithms and the ROB in the studies.

This review will identify and assess the best algorithm to use in future asthma research when using particular clinical terminology by comparing the methods and results of the validated algorithms from similar databases that use the same clinical coding. If studies are sufficiently comparable in that they have been carried out in similar populations and using similar reference standards, we will use bivariate random effects regression to calculate the summary measure of sensitivity and specificity[22] or PPV and NPV[23] (where no sensitivity and specificity values are provided), as described in the protocol for COPD exacerbations by Stone et al.[14]

## PATIENT AND PUBLIC INVOLVEMENT

Patients and the public will be involved in determining our final consensus algorithm as part of the larger project but have not been included at this stage as we are collating what already exists in the literature.

## ETHICS AND DISSEMINATION

This is a review of previously published literature that is publicly available and therefore does not require ethical approval. This protocol has been registered on PROSPERO: International Prospective Register of Systematic Reviews (registration number: CRD42024545081). The findings of the review will be disseminated via presentation at relevant scientific conferences and publication in a peer-reviewed journal. The BMJ Open instructions for reviewers of study protocols can be found in online supplemental file 4.

**X** Zakariah Z Gassasse @zgassasse

**ORCID iDs**
Elizabeth Moore http://orcid.org/0000-0003-1660-7598
Zakariah Z Gassasse http://orcid.org/0000-0003-3079-1563
Jennifer K Quint http://orcid.org/0000-0003-0149-4869

## REFERENCES

1. World Health Organization. Asthma key facts. 2023. Available: https://www.who.int/news-room/fact-sheets/detail/asthma
2. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention 2019.
3. Jin J. Asthma Attacks. *JAMA* 2016;315:832.
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;14:113.
5. Martin MJ, Beasley R, Harrison TW. Towards a personalised treatment approach for asthma attacks. *Thorax* 2020;75:1119–29.
6. Sanders DL, Gregg W, Aronsky D. Identifying asthma exacerbations in a pediatric emergency department: a feasibility study. *Int J Med Inform* 2007;76:557–64.
7. Afzal Z, Engelkes M, Verhamme KMC, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. *Pharmacoepidemiol Drug Saf* 2013;22:826–33.
8. Wu ST, Juhn YJ, Sohn S, et al. Patient-level temporal aggregation for text-based asthma status ascertainment. *J Am Med Inform Assoc* 2014;21:876–84.
9. Whittaker H, Rothnie KJ, Quint JK. Exploring the impact of varying definitions of exacerbations of chronic obstructive pulmonary disease in routinely collected electronic medical records. *PLoS ONE* 2023;18:e0292876.
10. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013;20:144–51.

11 Al Sallakh MA, Vasileiou E, Rodgers SE, *et al*. Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. *Eur Respir J* 2017;49:1700204.

12 Nissen F, Quint JK, Wilkinson S, *et al*. Validation of asthma recording in electronic health records: a systematic review. *Clin Epidemiol* 2017;9:643–56.

13 Sharifi M, Krishanswami S, McPheeters ML. A systematic review of validated methods to capture acute bronchospasm using administrative or claims data. *Vaccine (Auckl)* 2013;31 Suppl 10:K12–20.

14 Stone P, Sood N, Feary J, *et al*. Validation of acute exacerbation of chronic obstructive pulmonary disease (COPD) recording in electronic health records: a systematic review protocol. *BMJ Open* 2020;10:e032467.

15 Benchimol EI, Manuel DG, To T, *et al*. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011;64:821–9.

16 Jayatunga W, Stone P, Aldridge RW, *et al*. Code sets for respiratory symptoms in electronic health records research: a systematic review protocol. *BMJ Open* 2019;9:e025965.

17 Leong A, Dasgupta K, Bernatsky S, *et al*. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PLoS ONE* 2013;8:e75256.

18 Shiff NJ, Jama S, Boden C, *et al*. Validation of administrative health data for the pediatric population: a scoping review. *BMC Health Serv Res* 2014;14:236.

19 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.

20 Whiting PF, Rutjes AWS, Westwood ME, *et al*. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.

21 Fawkner-Corbett D, Hayward G, Alkhmees M, *et al*. Diagnostic accuracy of blood tests of inflammation in paediatric appendicitis: a systematic review and meta-analysis. *BMJ Open* 2022;12:e056854.

22 Reitsma JB, Glas AS, Rutjes AWS, *et al*. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.

23 Leeflang MMG, Deeks JJ, Rutjes AWS, *et al*. Bivariate meta-analysis of predictive values of diagnostic tests can be an alternative to bivariate meta-analysis of sensitivity and specificity. *J Clin Epidemiol* 2012;65:1088–97.