

# BMJ Open Validated methods to identify patients with asthma–COPD overlap in healthcare databases: a systematic review protocol

Joseph Emil Amegadzie,<sup>1</sup> Oluwatosin Badejo,<sup>1</sup> John-Michael Gamble,<sup>2</sup> Mark Wright,<sup>3</sup> Jamie Farrell,<sup>1</sup> Brooke Jackson,<sup>3</sup> Kirin Sultana,<sup>3</sup> Maimoona Hashmi,<sup>3</sup> Zhiwei Gao<sup>1</sup>

**To cite:** Amegadzie JE, Badejo O, Gamble J-M, *et al.* Validated methods to identify patients with asthma–COPD overlap in healthcare databases: a systematic review protocol. *BMJ Open* 2019;**9**:e024306. doi:10.1136/bmjopen-2018-024306

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-024306>).

Received 21 May 2018

Revised 26 September 2018

Accepted 31 December 2018



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Department of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada

<sup>2</sup>Department of Pharmacy, University of Waterloo, Kitchener, Ontario, Canada

<sup>3</sup>Clinical Practice Research Datalink (CPRD), Medicines and Healthcare Products Regulatory Agency, London, UK

## Correspondence to

Dr Zhiwei Gao;  
zhiwei.gao@med.mun.ca

## ABSTRACT

**Introduction** Asthma–chronic obstructive pulmonary disease (COPD) overlap (ACO) is characterised by patients presenting symptoms of both asthma and COPD. Many efforts have been made to validate different methods of identifying asthma–COPD overlap cases based on symptoms, spirometry and medical history in epidemiological studies using healthcare databases. There are various coding algorithm strategies that can be used and selection depends on targeted validation. The primary objectives of this systematic review are to identify validated methods (or algorithms) that identify patients with ACO from healthcare databases and summarise the reported validity measures of these methods.

**Methods** MEDLINE, EMBASE databases and the Web of Science will be systematically searched by using appropriate search strategies that are able to identify studies containing validated codes and algorithms for the diagnosis of ACO in healthcare databases published, in English, before October 2018. For each selected study, we require the presence of at least one test measure (eg, sensitivity, specificity etc). We will also include studies, in which the validated algorithm is compared with an external reference standard such as questionnaires completed by patients or physicians, medical charts review, manual review or an independent second database. For all selected studies, a uniform table will be created to summarise the following vital information: name of author, publication year, country, data source, population, clinical outcome, algorithms, reference standard method of validation and characteristics of the test measure used to determine validity.

**PROSPERO registration number** CRD42018087472.

## INTRODUCTION

Asthma and chronic obstructive pulmonary disease (COPD) are the two most common obstructive airway diseases. Recently a new phenotype, referred to as asthma–COPD overlap syndrome or asthma–COPD overlap (ACO), has been identified with its first guidelines for treatment and management in effect since 2015.<sup>1</sup> The Global Initiative

## Strengths and limitations of this study

- To the best of our knowledge, this will be the first study to systematically identify and evaluate methods used to validate asthma–COPD overlap disease in healthcare databases.
- Identification of properly validated algorithms to identify patients with asthma–COPD overlap from healthcare databases will inform more accurate patient selection in future studies.
- Different healthcare databases may validate different codes or algorithms to identify patients with asthma–COPD overlap. This can result in important heterogeneity and therefore limit the generalisability of these algorithms to other settings as they are context-specific depending on the type of database.
- This systematic review will primarily focus on validated methods or algorithms of asthma–COPD overlap recordings in healthcare databases and not on outcome results of studies. This situation may result in publication bias as algorithms without accompanied validity assessment or methods that do not find positive results may be less likely to have been published.

for Asthma (GINA) and Global Initiative for Chronic Obstructive Lung Disease (GOLD) described ACO as ‘persistent airflow limitation with several features usually associated with asthma and several features usually associated with COPD’, and pointed out that ACO includes different clinical phenotypes with several underlying mechanisms.<sup>2</sup> While there have been varied definitions of ACO in the literature, most of the discussions on ACO have primarily focused on reviewing the evidential features of asthma and COPD coexisting at biological,<sup>3</sup> epidemiological levels<sup>4,5</sup> and on its clinical significance.<sup>6,7</sup>

Just as the basic definitions of asthma and COPD are still debatable,<sup>8,9</sup> the primary definition of ACO is not yet clear. The first

guideline for identification of ACO was proposed in the combination of GINA and GOLD guidelines in 2015.<sup>1</sup> The Spanish COPD guideline (GesEPOC) was the first clinical practice guideline to recognise the ACO phenotype, calling it the mixed asthma–COPD phenotype.<sup>10</sup> The GesEPOC and the Spanish Guideline on the Management of Asthma (GEMA) recently came out with a consensus to unify the criteria for the diagnosis of ACO.<sup>11</sup> The GesEPOC/GEMA consensus defined the presence of ACO in a given patient based on three elements: significant smoking exposure, chronic airflow limitation and asthma.

In advancing a clearer diagnostic criteria for ACO, Miravittles<sup>12</sup> proposed ‘the five commandments of ACO diagnosis’: (1) a patient with asthma may develop non-fully reversible airflow obstruction, but this is not COPD, not even ACO; it is obstructive asthma. (2) A patient with asthma who smokes may also develop non-fully reversible airflow obstruction, which differs from obstructive asthma and from ‘pure’ COPD, which he categorised as the most frequent type of patient with ACO. (3) Some patients who smoke and develop COPD may have a genetic type 2 immune responses (Th2) background (even in the absence of a previous history of asthma), which can be identified by high eosinophil counts in peripheral blood. These individuals could be included under the umbrella term of ACO. (4) A patient with COPD and a positive bronchodilator test (>200 mL and >12% FEV<sub>1</sub> change) has reversible COPD but is not an asthmatic. Finally, on the fifth commandment, a patient with COPD and a very positive bronchodilator test (>400 mL FEV<sub>1</sub> change) is more likely to have some features of asthma and could also be classified as ACO.

In ACO, combination pharmacotherapy treatment consisting of long-acting  $\beta_2$ -agonists/inhaled corticosteroids may be the first choice of treatment in patients with a history suggestive of the overlap disease.<sup>2</sup> In spite of the uncertainties concerning ACO definition, there is broad agreement that patients with features of both asthma and COPD experience more frequent exacerbations, have poorer quality of life, a more rapid decline in lung function and high mortality and use a disproportionately larger amount of healthcare resources than people with asthma or COPD alone.<sup>1</sup>

There are various kinds of healthcare databases accessible for healthcare research. These databases generally fall into two divisions; administrative (eg, hospital billing data) and electronic health records (EHRs).<sup>13</sup> The increased use of these two categories of databases has added to the popularity of population-based epidemiology and health outcomes research studies. However, the basic functional use of healthcare databases includes but is not limited to hospital billing, administration, provision of care, laboratory procedures, pharmacy dispensing and physician practice.<sup>13</sup> Recently, there has been an increased use of these healthcare databases for epidemiological studies and population outcome studies as researchers have identified these databases as very useful avenues for clinical research.<sup>14–16</sup>

These databases primarily collect longitudinal information in connection with a patient’s demographics, important information regarding healthcare resource utilisation such as hospitalisations, referrals to specialists or secondary care, drug prescription, laboratory tests, imaging and lifestyle.<sup>17 18</sup> Thus, the types of information contained in these databases have become extremely important. The availability of these healthcare databases provide great opportunity and benefits over several major limitations of randomised controlled trials such as lower cost, increased generalisability and increased statistical power due to larger sample size.<sup>13</sup> The applications of these healthcare datasets in observational studies have become desirable as they are well suited in hypothesis generation and in advancing previously tested hypotheses.<sup>13</sup>

Algorithms to identify cases in these structured coded healthcare databases can be developed by a single code, combination of multiple codes or sets of codes. As noted by Nissen *et al*<sup>19</sup> the accuracy of diagnoses recorded in these large databases may be low, which would introduce bias into studies using the data. They developed an algorithm, to increase the ability to identify case definitions for asthma in the Clinical Practice Research Datalink database, using a diagnosis plus spirometry plus specific medication. They found out that extra information on asthma medication prescription (positive predictive value, PPV 83.3%), evidence of reversibility testing (PPV 86.0%) or a combination of all three selection criteria (PPV 86.4%) did not result in a higher PPV.<sup>19</sup> Even though validation of codes or algorithms to correctly identify patients with diseases or medical conditions may be time consuming and labor intensive, unless these algorithms are validated for research, the quality of studies generated from EHRs may be debatable. Identification of properly validated algorithms to identify patients with different health states (diseases and conditions) will inform more accurate patient selection in future studies.

The development of an algorithm to measure a health outcome from a particular database requires a clear understanding of data provenance and structure. The validity of an algorithm can be assessed against measures based on questionnaires completed by a patient or physician, medical charts review, medical notes, manual review or an independent second database.<sup>19 20</sup> We will conduct a systematic review to evaluate the current body of evidence that have used algorithms or codes based on information in healthcare databases to identify patients with ACO.

## Research question

The primary objectives of this systematic review are to identify validated methods (or algorithms) that identify patients with ACO from healthcare databases and summarise the reported validity measures of these methods.

Specifically, the questions of interest are;

1. What type of healthcare databases have been used to obtain information on the diagnosis of ACO?

2. Which algorithms have been extensively used to define and correctly identify patients with ACO?
3. Against which reference standards were the validity of these algorithms assessed? And what were the diagnostic accuracy estimates?

**METHODS**

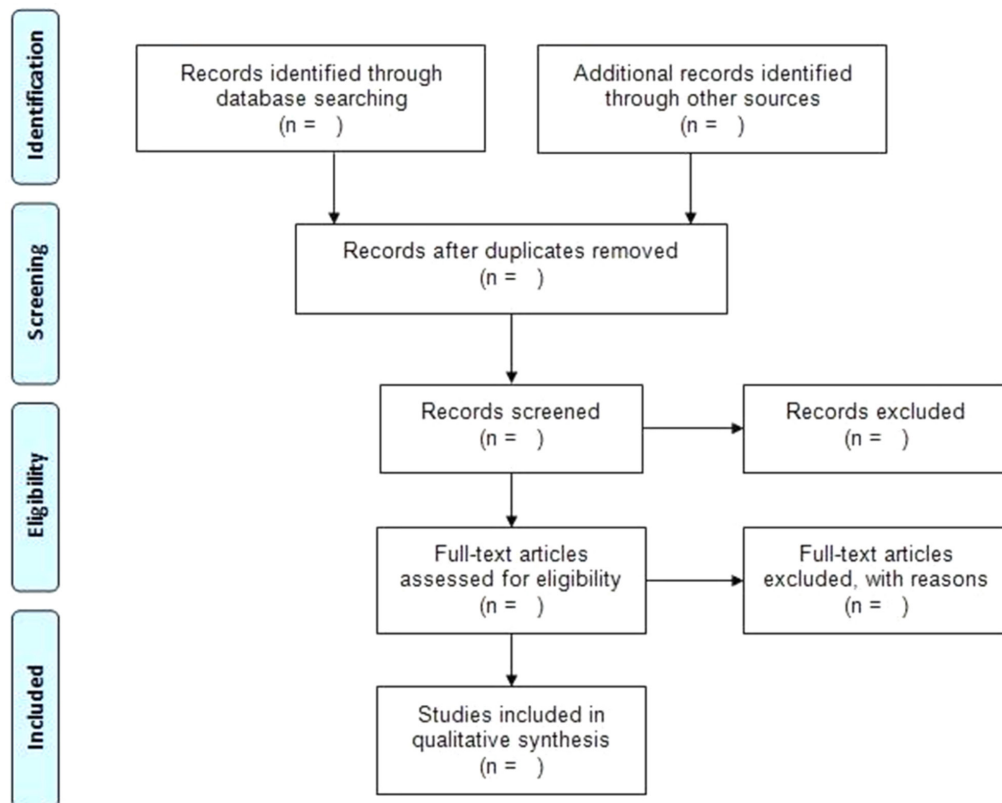
**Literature search**

MEDLINE, EMBASE and the Web of Science will be systematically searched for published peer-reviewed articles. We will use a search strategy based on a combination of: (1) keywords, Medical Subject Headings (MeSH) and title/abstract (tiab) to identify records in association with ‘asthma AND COPD’; (2) terms to identify articles probably containing validity or accuracy measures and (3) a search strategy likely to contain studies on the combination of terms and ACO definitions by Miravittles,<sup>12</sup> Sin<sup>21</sup> and GesEPOC.<sup>11</sup> In addition, reference lists of primary articles will be reviewed to find relevant articles that adopt different standards for asthma–COPD description. An experienced librarian from the Health Science Library of Memorial University along with one of the authors will independently conduct a comprehensive search in MEDLINE, EMBASE and Web of Science to identify potential articles. The MEDLINE, EMBASE and Web of Science searches will be independently reviewed by a more senior librarian and another one of the authors.

This systematic review protocol has been prepared according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram from Moher *et al*.<sup>22</sup> can be found in [figure 1](#), and the search strategy can be found in the online supplementary file. The PRISMA flow diagram will allow for more transparent flow of information through the different phases of our systematic review. This protocol has been published in the PROSPERO International Prospective Register of Systematic Reviews with registration number CRD42018087472.

**Inclusion criteria**

Any full-text, peer-reviewed articles published in English before October 2018, that validated the recording of ACO in a healthcare database will be considered for inclusion. We aim to focus on healthcare databases, in which the diagnosis of ACO is primarily based on clinical features, spirometry results, prescription data, radiography and laboratory data. The included studies will be considered if the validated algorithm is compared with an external reference standard such as questionnaires completed by patients or physicians, medical charts review, medical notes, manual review or an independent second database. For each study, we require the presence of at least one study measure such as specificity, sensitivity, positive predictive value or negative predictive value. Also, for



**Figure 1** Study screening process: Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram from Moher *et al*.<sup>22</sup>

our inclusion criteria, we will include algorithms developed from single codes, algorithms formed of multiple case characteristics (eg, disease code plus spirometry code plus prescription code) and algorithms generated by natural language processing or machine-learning (eg, Read code, ICD-9 or ICD-10).

### Exclusion criteria

Studies without validation of ACO recording, conference abstracts, surveys and disease registries will be excluded. In addition, studies involving pharmacovigilance databases (spontaneous reporting, signal detection) will be excluded.

### Selection processes

Two independent reviewers will scan titles and abstracts of identified articles and relevant articles will be retrieved based on our research questions and inclusion/exclusion criteria. Discrepancies in determining whether the study met our inclusion criteria during the full-text review will be resolved by consensus between the reviewers. If a consensus could not be reached, arbitration will be decided by a third reviewer.

### Data extraction

The following information will be extracted from each of the included studies by two reviewers independently.

1. Study characteristics (including title, year, country, journal of publication, date of publication and information on the author).
2. Data source, population
3. Type of healthcare database used (including EHR, hospitalisation discharge data, etc).
4. Sample characteristics
5. Clinical outcome
6. Algorithms; the modality of algorithm development (eg, using logistic regression, Classification and Regression Trees, expert opinion etc).
7. Reference standard of validation.
8. Characteristic of the test measure(s) used to determine validity.

### Risk of bias assessment

Quality assessment of the design and methods on all included primary studies will be assessed using a checklist developed by Benchimol *et al.*<sup>23</sup> Using Standards for Reporting of Diagnostic accuracy<sup>24</sup> criteria as a guide, they created a 40-item checklist of items with which to assess the quality of validation studies of health administrative data and to report studies that validated algorithms or codes for identifying patients with different health states (diseases and conditions).

Two reviewers will independently assess the quality of these studies and report potential bias(es) in a descriptive form. Disagreements will be resolved by discussion or arbitration with a third reviewer. However, no subgroup analysis or publication bias assessment is anticipated.

### Data synthesis

All records will be de-duplicated and screened using Covidence (<https://www.covidence.org>); a web-based software platform that streamlines the production of systematic reviews and EndNote (V. X7, Thomson Reuters) software will be used to manage the study articles and references. An overview for the validation of ACO recording will be summarised in narrative composition and in tables describing the methods and results of the included studies. Possibly, validation statistics will be aggregated and stratified by the kind of healthcare database, the type of EHR coding and the country of origin. However, no formal meta-analysis is planned. These results may include specificity, sensitivity, PPV and NPV of studies that met our inclusion criteria. Where they are not reported, these test results such as 95% CI, PPV and NPV will be calculated if possible.

### Patient and public involvement

No patient will be involved in this review.

### ETHICS AND DESSIMINATION

This protocol was published in the PROSPERO International Prospective Register of Systematic Reviews in February 2018 with registration number CRD42018087472. Findings of this review will be presented at epidemiology and pharmacoepidemiology scientific conferences and disseminated through publication in a peer-reviewed journal.

**Contributors** JEA: was responsible for drafting the protocol and registering it in PROSPERO. JEA, OB, JMG, MW, JF, BJ, KS, MH, ZG: drafted the manuscript and contributed to the development of the research questions, literature search, selection criteria, data extraction criteria, the risk of bias assessment and data synthesis. JEA, OB, JMG, MW, JF, BJ, KS, MH, ZG: have critically read, commented on and approved the final version of the manuscript. ZG: is responsible for the study management and coordination.

**Funding** This study is part of a research project which has been funded with a research grant from Canada Research Respiratory Network (CRRN), Ottawa, Canada, Young Investigator Award, 2017.

**Disclaimer** The study funder was not involved in the study design or the writing of the protocol.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** This review protocol will use previously published studies publicly available without directly involving human participants; hence no ethical approval is required.

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

**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/>.

### REFERENCES

1. ACOS. Diagnosis of diseases of chronic airflow limitation: asthma COPD and asthma-COPD overlap syndrome. *GINA and GOLD* 2015;2015.
2. GINA. Diagnosis of Diseases of Chronic Airflow Limitation: Asthma COPD and asthma-COPD overlap global initiative for asthma. 2017.

3. Piras B, Miravittles M. The overlap phenotype: the (missing) link between asthma and COPD. *Multidiscip Respir Med* 2012;7:8.
4. Shaya FT, Dongyi D, Akazawa MO, *et al.* Burden of concomitant asthma and COPD in a medicaid population. *Chest* 2008;134:14–19.
5. Shaya FT, Maneval MS, Gbarayor CM, *et al.* Burden of COPD, asthma, and concomitant COPD and asthma among adults: racial disparities in a medicaid population. *Chest* 2009;136:405–11.
6. Sin DD, Miravittles M, Mannino DM, *et al.* What is asthma-COPD overlap syndrome? Towards a consensus definition from a round table discussion. *Eur Respir J* 2016;48:664–73.
7. Barrecheguren M, Esquinas C, Miravittles M. How can we identify patients with asthma-COPD overlap syndrome in clinical practice? *Arch Bronconeumol* 2016;52:59–60.
8. Juhn YJ. Risks for infection in patients with asthma (or other atopic conditions): is asthma more than a chronic airway disease? *J Allergy Clin Immunol* 2014;134:247–57.
9. Bhatt SP. Diagnosis of chronic obstructive pulmonary disease: breathing new life into an old debate. *Ann Am Thorac Soc* 2018;15:163–5.
10. Miravittles M, Soler-Cataluña JJ, Calle M, *et al.* [Spanish COPD Guidelines (GesEPOC): pharmacological treatment of stable COPD]. *Aten Primaria* 2012;44:425–37.
11. Plaza V, Álvarez F, Calle M, *et al.* Consensus on the Asthma-COPD Overlap Syndrome (ACOS) Between the Spanish COPD Guidelines (GesEPOC) and the Spanish Guidelines on the management of asthma (GEMA). *Arch Bronconeumol* 2017;53:443–9.
12. Miravittles M. Diagnosis of asthma-COPD overlap: the five commandments. *Eur Respir J* 2017;49:1700506.
13. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
14. Uddin MJ, Groenwold RH, de Boer A, *et al.* Evaluating different physician's prescribing preference based instrumental variables in two primary care databases: a study of inhaled long-acting beta2-agonist use and the risk of myocardial infarction. *Pharmacoepidemiol Drug Saf* 2016;25:132–41.
15. Afonso A, Schmiedl S, Becker C, *et al.* A methodological comparison of two European primary care databases and replication in a US claims database: inhaled long-acting beta-2-agonists and the risk of acute myocardial infarction. *Eur J Clin Pharmacol* 2016;72:1105–16.
16. Yeh JJ, Wei YF, Lin CL, *et al.* Association of asthma-chronic obstructive pulmonary disease overlap syndrome with coronary artery disease, cardiac dysrhythmia and heart failure: a population-based retrospective cohort study. *BMJ Open* 2017;7:e017657.
17. de Jong RG, Gallagher AM, Herrett E, *et al.* Comparability of the age and sex distribution of the UK Clinical practice research datalink and the total dutch population. *Pharmacoepidemiol Drug Saf* 2016;25:1460–4.
18. Herrett E, Gallagher AM, Bhaskaran K, *et al.* Data resource profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;44:827–36.
19. 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.
20. Nissen F, Morales DR, Mullerova H, *et al.* Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). *BMJ Open* 2017;7:e017474.
21. Sin DD. Asthma-COPD Overlap syndrome: what we know and what we don't. *Tuberc Respir Dis* 2017;80:11–20.
22. Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006–12.
23. 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.
24. Bossuyt PM, Reitsma JB, Bruns DE, *et al.* Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Ann Clin Biochem* 2003;40:357–63.