


BMJ Open Validation of acute exacerbation of chronic obstructive pulmonary disease (COPD) recording in electronic health records: a systematic review protocol

Philip Stone ¹, Nikhil Sood,¹ Johanna Feary,¹ C Michael Roberts,² Jennifer K Quint ¹

To cite: 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. doi:10.1136/bmjopen-2019-032467

► 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-2019-032467>).

Received 19 June 2019
Revised 15 October 2019
Accepted 11 February 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹National Heart and Lung Institute, Imperial College London, London, UK

²UCL Partners, London, UK

Correspondence to

Mr Philip Stone;
p.stone@imperial.ac.uk

ABSTRACT

Introduction Many patients with chronic obstructive pulmonary disease (COPD) experience a sustained worsening in symptoms termed an acute exacerbation (AECOPD). AECOPDs impact on patients' quality of life and lung function, are costly to health services and are an important topic for research. Electronic health records (EHR) are increasingly being used to study AECOPD, requiring accurate detection of AECOPD in EHRs to ensure generalisable results. The aim of this protocol is to provide an overview of studies that validate AECOPD definitions used in EHRs and administrative claims databases.

Methods and analysis Medline and Embase will be searched for terms related to COPD exacerbation, EHRs and validation. All studies published between 1 January 1990 and 30 September 2019 written in English that validate AECOPD in EHRs and administrative claims databases will be considered. Inclusion criteria: EHR data must be routinely collected; the AECOPD detection algorithm must be compared against a reference standard; and a measure of validity must be calculable. Two independent reviewers will screen articles for inclusion, extract study details and assess risk of bias using QUADAS-2. Disagreements will be resolved by consensus or arbitration by a third reviewer. This protocol has been developed in accordance with the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols checklist.

Ethics and dissemination This will be a review of previously published literature therefore no ethical approval is required. Results from this review will be published in a peer-reviewed journal. The results can be used in future research to identify occurrences of AECOPD. **PROSPERO registration number** CRD42019130863.

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is a disease that is characterised by respiratory symptoms such as breathlessness, cough, or sputum production, and airflow limitation due to damage to the airway and/or alveoli.^{1 2} The most common cause of COPD is cigarette smoke, however, pollution and occupational exposures are also risk

Strengths and limitations of this study

- To the best of our knowledge, this will be the first review to evaluate methods of validating acute exacerbation of chronic obstructive pulmonary disease (AECOPD) definitions in electronic health record (EHR) or administrative claims databases.
- On completion this review will give an overview of the validity of current AECOPD definitions in EHR databases and could inform the best choice of AECOPD detection algorithm in future research.
- While many EHR databases use the same clinical coding scheme to denote clinical events, the same code list (algorithm) may have better or worse ability to detect AECOPD in different databases; this possible heterogeneity may affect our ability to synthesise results and reduce the generalisability of results.

factors for developing COPD.^{1 3} Many patients with COPD experience episodes of sustained worsening in their symptoms termed an acute exacerbation of COPD (AECOPD) which require hospitalisation when particularly severe.⁴ Frequent exacerbations are associated with increased mortality⁵ and a decrease in lung function⁶ and quality of life.⁷ AECOPD hospitalisations are very costly to healthcare services^{8–10} costing an estimated average of £1868 per admission in England,¹¹ and as high as an average of \$44909 for the most severe admissions in a US setting.¹⁰

Electronic health records (EHR) are becoming widely used for research purposes, meaning it is essential to ensure that the definition of exposures, outcomes and other covariates used in research using EHRs is accurate.¹² These variables will generally be defined by using 'code lists' of relevant clinical codes from a particular clinical terminology, for example, the 10th Revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10).¹³

Guidelines^{14 15} have been produced to give researchers advice on how best to generate these code lists, with the ultimate aim being to produce an accurate and reusable definition of all variables in a study. Repositories^{16–18} have been created to enable researchers to share their code lists but appear underused.

Rothnie *et al* carried out studies to validate the recording of AECOPD in both primary¹⁹ and secondary²⁰ care (in a UK context). Nissen *et al*²¹ completed a systematic review on the validation of asthma diagnoses in EHRs, and Rimland *et al*²² published a protocol for a systematic review of the validation of COPD in healthcare databases. However, there is no published systematic review of the validation of AECOPD recording in EHRs; this would give researchers confidence in the ability of definitions used in their studies to correctly identify individuals with AECOPD, ensuring results are generalisable, and provide consistency of AECOPD definitions in future studies. It will also benefit disease monitoring using EHR databases (eg, checking prevalence and incidence) as clinicians can be given lists of preferred codes or terms to use when diagnosing AECOPD.

Objective

The primary objective of the systematic review this protocol describes is to provide an overview of the methods and findings of studies that validate AECOPD definitions used in EHRs and administrative claims databases. The target population are people with AECOPD. The intervention measured (index test) is the AECOPD detection algorithm with the comparison group being the reference standard used to confirm AECOPD diagnosis. This means that studies included in this review may use different reference standards—this is to ensure capture of all validation studies. The outcome will be the validity of the AECOPD detection algorithm. These can be studies in any country, using any clinical coding scheme, in any EHR database. In the included studies we will specifically look for:

- ▶ The database and type of EHRs used.
- ▶ The algorithm used to detect the AECOPD.
- ▶ The reference standard used to validate the AECOPD.
- ▶ The estimated validity of the AECOPD detection algorithm.

METHODS AND ANALYSIS

Medline and Embase (via the Ovid interface) will be searched using keywords and Medical Subject Headings terms^{23 24} related to ‘exacerbation of COPD’, ‘electronic health records’ or ‘administrative claims database’, and ‘validation’, including any relevant synonyms. A full draft search strategy can be found in online supplementary file 1. The search strategy used to detect the validation terms will be guided by the strategy developed by Benchimol *et al*²⁵ and strategies used in similar reviews^{21 22 26–28} of validation studies in EHR databases. The reference lists of retrieved articles will also be searched.

Eligibility criteria

All studies written in English published between 1 January 1990 and 30 September 2019 that validate an AECOPD definition in EHRs will be considered. The specific inclusion criteria of the study will be:

- ▶ Data must come from an EHR or administrative claims database where data are routinely collected.
- ▶ The AECOPD detection algorithm must be compared against a reference or gold standard definition (such as a questionnaire completed by a physician to confirm the diagnosis).
- ▶ There must be a measure of validity (positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, c-statistic, and so on) or sufficient information to be able to calculate one.

Studies will be excluded if they only look at COPD diagnosis rather than specifically AECOPD.

Data management and synthesis

Articles identified by the search strategy will be stored in the reference management package EndNote (Clarivate Analytics, Philadelphia, Pennsylvania, USA) and duplicate articles will be removed. Unique article titles and abstracts will then be loaded into Rayyan²⁹ and screened by two independent reviewers. If either reviewer thinks the inclusion criteria are met, then the articles will be included in a full-text review. Articles selected for full-text review will then be independently screened by both reviewers for inclusion in the review with disagreement between reviewers resolved by consensus or arbitration by a third reviewer. Reasons for study exclusion will be recorded. The full-text articles will be read, and both reviewers will independently extract study details and assess risk of bias. These data will be stored in a preformatted Microsoft Excel (Microsoft, Redmond, Washington, USA) form. The data that will be extracted from included studies are:

- ▶ Study details (title, first author, year of publication, DOI).
- ▶ Study aim/research question.
- ▶ EHR database used.
- ▶ Population (location, time period).
- ▶ Type of algorithm(s) used to detect AECOPD (eg, clinical coding scheme).
- ▶ Algorithm(s) used to detect AECOPD (eg, the list of clinical codes used).
- ▶ Reference/gold standard the algorithm(s) was compared against.
- ▶ Measure(s) of validity calculated (eg, PPV).
- ▶ Result(s) of validity measure(s).
- ▶ Prevalence of AECOPD.
- ▶ 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 AECOPD detection algorithm.

The quality and risk of bias in individual studies will be assessed using QUADAS-2,³⁰ a quality assessment tool for diagnostic accuracy studies. QUADAS-2 will be tailored to

this specific review using a recommended reporting checklist developed by Benchimol *et al*²⁵ for use in validation studies of health administrative data. A draft version of our tailored QUADAS-2 risk of bias assessment can be found in online supplementary file 2. Where multiple validations are reported in a study, quality of reporting and risk of bias will be assessed for each validation. Results from the review will be presented in a narrative synthesis with information presented in the text and in tables to summarise study details, the algorithms used to validate AECOPD in EHRs, the reference standard used to validate the algorithm, the validity of the algorithms and the risk of bias in studies.

Where studies have validated algorithms in similar databases that use the same clinical terminology (eg, ICD-10), the methods and results of the validations will be compared to assess the best algorithm to use when using that particular clinical terminology. Where studies are sufficiently homogeneous and have been carried out in similar populations using similar reference standards, we will use bivariate random effects regression to calculate summary measures of sensitivity and specificity³¹ or PPV and NPV³² (where no sensitivity and specificity values are provided).

Limitations

One potential issue with studies of validity is publication bias where a detection algorithm found to have an undesirable validity may be less likely to be published. Validity may also be calculated in a population with a higher prevalence of the condition than would be found in the general population to produce a greater PPV. Publication bias can be difficult to assess but studies that provide information on prevalence can be checked to ensure it matches that of the general population. There may also be an issue with reuse of algorithms in different EHR databases. While many databases use the same clinical terminology and could therefore share detection algorithms, it is possible that a detection algorithm for one database may not have the same level of validity in another database. This will be particularly true for databases with data quality improvement programmes where coding will be much more accurate compared with those without such programmes. Another limitation is that some AECOPDs may be managed at home by patients using a rescue pack of antibiotics and oral corticosteroids; this may be those with less severe symptoms. These exacerbations will not be detected by the EHRs as the patient will not visit a doctor in either primary or secondary care.

Patient and public involvement

This protocol was designed without patient involvement and no plans exist for patient involvement in the review.

ETHICS AND DISSEMINATION

This review will collate publicly available information and therefore no ethical approval is required. This protocol has been registered on PROSPERO: International Prospective Register of Systematic Reviews.³³ Findings of the review will be disseminated via presentation at

relevant scientific conferences and publication in a peer-reviewed journal.

The Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols checklist³⁴ was used to aid production of this protocol.

Contributors JKQ and PS developed the research question. PS developed the search strategy with input from NS, JF and JKQ. PS drafted the manuscript. PS, JKQ, JF and CMR contributed to production of the final manuscript. All authors read, commented on and approved the protocol and final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Patient consent for publication Not 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/>.

ORCID iDs

Philip Stone <http://orcid.org/0000-0002-2326-4987>

Jennifer K Quint <http://orcid.org/0000-0003-0149-4869>

REFERENCES

- 1 Global Initiative for Chronic Obstructive Lung Disease (GOLD). Pocket Guide to COPD Diagnosis, Management, and Prevention: A Guide for Health Care Professionals - 2019 Edition. Available: <https://goldcopd.org/wp-content/uploads/2018/11/GOLD-2019-POCKET-GUIDE-DRAFT-v1.7-14Nov2018-WMS.pdf> [Accessed 9th May 2019].
- 2 British Lung Foundation. What is COPD?. [Online] British Lung Foundation. Available: <https://www.blf.org.uk/support-for-you/copd/what-is-it> [Accessed 29th Sep 2017].
- 3 British Lung Foundation. Causes of COPD. [Online] British Lung Foundation. Available: <https://www.blf.org.uk/support-for-you/copd/causes> [Accessed 29th May 2018].
- 4 National Institute for Health and Care Excellence. Chronic obstructive pulmonary disease in over 16s: diagnosis and management. recommendations. NICE guideline (NG115). Available: <https://www.nice.org.uk/guidance/ng115/chapter/Recommendations> [Accessed 28th Feb 2019].
- 5 Soler-Cataluña JJ, Martínez-García MA, Román Sánchez P, *et al*. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax* 2005;60:925–31.
- 6 Donaldson GC, Seemungal TAR, Bhowmik A, *et al*. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 2002;57:847–52.
- 7 SEEMUNGAL TAR, Donaldson GC, PAUL EA, *et al*. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157:1418–22.
- 8 Geitona M, Hatzikou M, Steiropoulos P, *et al*. The cost of COPD exacerbations: a university hospital – based study in Greece. *Respir Med* 2011;105:402–9.
- 9 Mittmann N, Kuramoto L, Seung SJ, *et al*. The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. *Respir Med* 2008;102:413–21.
- 10 Dalal AA, Shah M, D'Souza AO, *et al*. Costs of COPD exacerbations in the emergency department and inpatient setting. *Respir Med* 2011;105:454–60.
- 11 National Institute for Health and Care Excellence. Resource impact report: chronic obstructive pulmonary disease in over 16s: diagnosis and management (update) (NG115). Available: <https://www.nice.org.uk/guidance/ng115/resources/resource-impact-report-pdf-6602803741> [Accessed 8th May 2019].
- 12 Herrett E, Thomas SL, Schoonen WM, *et al*. Validation and validity of diagnoses in the general practice research database: a systematic review. *Br J Clin Pharmacol* 2010;69:4–14.
- 13 World Health Organization. International statistical classification of diseases and related health problems – 10th revision. [Online]. 5th ed, 2016. Available: https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2016.pdf



- 14 Williams R, Kontopantelis E, Buchan I, *et al.* Clinical code set engineering for reusing EHR data for research: a review. *J Biomed Inform* 2017;70:1–13.
- 15 Watson J, Nicholson BD, Hamilton W, *et al.* Identifying clinical features in primary care electronic health record studies: methods for codelist development. *BMJ Open* 2017;7:e019637.
- 16 Springate DA, Kontopantelis E, Ashcroft DM, *et al.* ClinicalCodes: an online clinical codes Repository to improve the validity and reproducibility of research using electronic medical records. *PLoS One* 2014;9:e99825.
- 17 Farr Institute of Health Informatics Research. CALIBER. [Online]. Available: <https://www.caliberresearch.org/portal> [Accessed 31st May 2018].
- 18 University of Cambridge. CPRD @ Cambridge - Code Lists. [Online] Primary Care Unit. Available: http://www.phpc.cam.ac.uk/pcu/cprd_cam/codelists/ [Accessed 31st May 2018].
- 19 Rothnie KJ, Müllerová H, Hurst JR, *et al.* Validation of the recording of acute exacerbations of COPD in UK primary care electronic healthcare records. *PLoS One* 2016;11:e0151357.
- 20 Rothnie KJ, Müllerová H, Thomas SL, *et al.* Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. *Clin Epidemiol* 2016;8:771–82.
- 21 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.
- 22 Rimland JM, Abraha I, Luchetta ML, *et al.* Validation of chronic obstructive pulmonary disease (COPD) diagnoses in healthcare databases: a systematic review protocol. *BMJ Open* 2016;6:e011777.
- 23 National Center for Biotechnology Information. PubMed Help. [Online] National Center for Biotechnology Information (US), 2018. Available: <https://www.ncbi.nlm.nih.gov/books/NBK3827/> [Accessed 1st Jun 2018].
- 24 Baumann N. How to use the medical subject headings (mesh). *Int J Clin Pract* 2016;70:171–4.
- 25 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.
- 26 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.
- 27 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.
- 28 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.
- 29 Ouzzani M, Hammady H, Fedorowicz Z, *et al.* Rayyan—a web and mobile APP for systematic reviews. *Syst Rev* 2016;5:210.
- 30 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.
- 31 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.
- 32 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.
- 33 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. [Online] PROSPERO, 2019. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42019130863 [Accessed 8th May 2019].
- 34 Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.