Chronic Obstructive Pulmonary Diseases:
Journal of the COPD Foundation®

Original Research

Validation of Acute Exacerbation of Chronic Obstructive Pulmonary Disease Recording in Electronic Health Records: A Systematic Review

Elizabeth Moore, MSc[1*] Philip Stone, PhD[1,2*] Ayda Alizadeh, BSc, MBBS[1] Jaspreet Sangha, BSc, MBBS[1]
Saranya Das, BSc, MBBS[1] Shraddha Arshanapalli, BSc, MBBS[1] Jennifer K. Quint, PhD, FRCP[1]

Abstract

Objective: Acute exacerbations of COPD(AECOPD) can have severe impacts on patients with the disease and a heavy burden on health care resources. Electronic health records (EHRs) are a valuable resource for identifying cases of AECOPD and research. Studies have attempted to validate case definitions of AECOPD and this review aims to summarize validated AECOPD definitions in EHRs and to provide guidance on the best algorithms to use to ensure accurate cohorts of AECOPD cases are available for researchers using EHRs.

Methods: MEDLINE and Embase were searched and studies that met the inclusion criteria were reviewed by ≥2 reviewers. Data extracted included the algorithms used to identify AECOPD, the reference standards used to compare against the algorithm, and measures of validity. The risk of bias was assessed using QUADAS-2 adapted for this review.

Results: Out of 2784 studies found by the search strategy, 12 met the inclusion criteria. The clinical terminology used to build algorithms to detect AECOPD included codes from the International Classification of Diseases (ICD) Ninth Revision, Clinical Modification and Tenth Revision (ICD-9-CM and ICD-10), along with the Read codes from United Kingdom general practices. AECOPD can be identified within EHRs using validated definitions, however, the validity of AECOPD definitions varies considerably depending on the algorithm used and the settings to which they are applied.

Conclusion: Although there are validated definitions that can be used to identify AECOPD, there is no clear consensus on which provides the highest validity or the most sensitive and specific definition to use.

1. School of Public Health, Imperial College London, London, United Kingdom
2. University College London, Medical School-Royal Free Campus, London, United Kingdom

* *These authors contributed equally to this work.*

Abbreviations:

AECOPD=acute exacerbation of COPD; **CI**=confidence interval; **COPD**=chronic obstructive pulmonary disease; **CPRD**=Clinical Practice Research Datalink; **DPC**=Diagnosis Procedure Combination; **eCRF**=electronic case report form; **EHR**=electronic health record; **FCE**=functional capacity evaluation; **GOLD**=Global initiative for chronic Obstructive Lung Disease; **GP**=general practitioner; **HES**=Hospital Episode Statistics; **ICD-9-CM**=International Classification of Diseases, Ninth Revision, Clinical Modification; **ICD-10**=International Classification of Diseases, Tenth Revision; **LRTI**=lower respiratory tract infection; **NPV**=negative predictive value; **OCS**=oral corticosteroid; **PPV**=positive predictive value; **SNOMED CT**=Systematized Nomenclature of Medicine Clinical Terms

Funding Support:

Infrastructure support for this research was provided by the National Institute for Health and Care Research's Imperial Biomedical Research Centre.

Citation:

Moore E, Stone P, Alizadeh A et al. Validation of acute exacerbation of chronic obstructive pulmonary disease recording in electronic health records: a systematic review. ***Chronic Obstr Pulm Dis.*** 2025;12(2):190-202. doi: https://doi.org/10.15326/jcopdf.2024.0577

Publication Dates:

Date of Acceptance: January 7, 2025
Published Online Date: February 5, 2025

Address correspondence to:

Elizabeth Moore
School of Public Health
Imperial College London
Sir Michael Uren Hub
86 Wood Lane
London, England W12 0BZ
Email: liz.moore@imperial.ac.uk
Phone: 020 7594 8821

Keywords:

COPD; acute exacerbation of COPD; AECOPD; electronic health records

This article has an online supplement.

Introduction

Chronic obstructive pulmonary disease (COPD) is a disease that is characterized by persistent respiratory symptoms including breathlessness, sputum or cough, and airflow limitation due to damage to the airway and/or alveoli.[1,2] COPD is most commonly caused by cigarette smoke, but pollution and occupational exposures are also risk factors for COPD.[1,2] Patients with COPD can experience episodes of sustained worsening in their symptoms, referred to as an acute exacerbation of COPD (AECOPD), and can be severe enough to require hospitalization.[3] Frequent exacerbations are associated with increased mortality[4] and a decrease in lung function,[5] exercise capacity,[6] and quality of life,[7] and each additional AECOPD increases the risk of a subsequent AECOPD and death.[8] Additionally, hospitalizations for AECOPDs are very costly and can increase the economic burden on health care services.[9-13] In England, the average cost per admission for an AECOPD is estimated to be £1,868,[14] and in the United States for the most severe admissions[11] reportedly as high as an average of $44,909.

Due to the impact of AECOPD admissions on both patients and health care services, there is an impetus[15] to complete research on AECOPDs to discover potential interventions to reduce their frequency. Electronic health records (EHRs) provide a relatively quick and inexpensive[16] source of data to be able to carry out such studies and are increasingly being utilized in research.[17] Diagnoses are recorded in EHRs using a coded clinical terminology set such as International Classification of Diseases (ICD) codes,[18] which are widely used in hospital admission discharge summaries and health care billing databases globally, consisting of 7 characters of letters and numbers to classify diagnoses. In the United Kingdom, in primary care, diagnoses are commonly recorded in databases as Read codes (now increasingly obsolete) or as Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) codes which are coded clinical terms used in general practice primary care databases in the National Health Service.[19]

However, EHRs are not designed with research in mind – their primary focus being to aid physicians in the management of a patient's health care,[20] or for the purpose of insurance claims.[21] For example, the assignment of primary and secondary ICD discharge diagnosis codes for hospitalized patients is often done for reimbursement and, therefore, may be influenced by the anticipated reimbursement for a diagnosis, bringing into question the validity of these data for the identification of patients with a specific condition.[22] Furthermore, different databases (and even different clinicians entering records into those databases) use different coding strategies to classify an AECOPD and there is a lack of consensus over which strategies and definitions to use. To ensure studies utilizing EHRs are examining the condition of interest and are not at risk of misclassification, it is important to use validated definitions of the condition of interest.[23,24] A validated definition will commonly take the form of a list of codes of a particular clinical terminology, along with an algorithm of how to apply those codes. A validation study will then give estimates on the likelihood of a case detected with the algorithm being a true case.[25] Measures of validation include positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity.

A previous systematic scoping review by Sivakumaran et al,[26] aimed to identify how individuals with COPD are identified within EHRs and found widespread variation in the definitions used to identify people with COPD. Of the 185 eligible studies, only 7 used a case definition which had been validated against a reference standard in the same dataset. They argued that the inconsistencies in methods for identifying people with COPD in EHRs are minimizing the potential for harnessing EHRs worldwide. To our knowledge, there has not been another systematic review examining the identification and validation of AECOPDs in EHRs.

Therefore, in this systematic review we aim to summarize all validated definitions of AECOPD for use in EHRs and administrative claims databases, and in cases where multiple similar definitions are available, provide guidance on the best algorithm to use to ensure an accurate cohort of AECOPD cases is available for researchers using EHRs.

Methods

MEDLINE and Embase (via the Ovid interface) were searched using keywords and Medical Subject Headings terms[27,28] related to "exacerbation of COPD," "electronic health records" or "administrative claims database," and "validation," including any relevant synonyms. The full search strategy can be found in Supplementary File 1 in the online supplement. The methodology developed by Benchimol et al,[29] along with search strategies from other similar reviews[30-34] of validation studies in EHR databases, were used to construct the search strategy for this review. To ensure the literature was comprehensively searched, reference lists from studies that were retrieved were also hand searched.

Study Selection Criteria

All studies validating definitions of AECOPD in EHRs were considered for inclusion in this review. Studies had to be written in English and published between 1946 (MEDLINE) or 1947 (Embase) and May 31, 2024. The specific criteria for inclusion were:

1. AECOPD admission data had to come from either an EHR or an administrative claims database that routinely collects health data.
2. The detection algorithms for AECOPD had to be compared against a reference standard or gold standard definition (e.g., chart reviews or questionnaires completed by physicians to confirm and validate the diagnosis).
3. Finally, a measure of validity had to be available (e.g. sensitivity, specificity, PPV, NPV, and c-statistic, etc.) or there had to be a means to calculate one from data within the study.

During the screening process, it became apparent that adding another criterion for inclusion was necessary: potential wider applicability of the algorithm (i.e., the algorithm could be applied to another dataset). As the aim of this review is to recommend algorithms for future research, it was, therefore, decided that studies should be excluded if they could not be easily applied to other datasets. Studies were also excluded if they only validated a diagnosis of COPD, not specifically an AECOPD.

Data Synthesis

The protocol for data management and synthesis is described by Stone et al.[35] Two different reviewers (PS and EM) independently screened the articles selected for full-text review and any disagreement between the reviewers was resolved by consensus or third reviewer (JKQ) arbitration. If studies were excluded the reasons were recorded, and the 2 reviewers extracted study details and assessed risk of bias for the included studies independently. Data were extracted into Microsoft Excel (Microsoft Corporation; Redmond, Washington) and included:

1. Details of the study (including title, first author name, year of publication, doi)
2. The aims of the study or research question
3. Details of the EHR database
4. A description of the studied population (specific groups, location, and time period)
5. A description of the AECOPD detection algorithm(s (e.g., the list of clinical codes used)
6. Details of the reference standard or gold standard that the algorithm(s) were compared against
7. The measure(s) of validity that were used (e.g., PPV, NPV, etc.) along with validity results
8. The prevalence of AECOPD if available

The validity of the AECOPD detection algorithm was the primary outcome measure in this review.

The risk of bias in individual studies was assessed using a quality assessment tool[36] for diagnostic accuracy studies known as the QUADAS-2. The QUADAS-2 was specifically adapted to this review using the reporting checklist developed by Benchimol et al[29] for use in validation studies of health administrative data. A copy of the adapted QUADAS-2 risk of bias assessment used in this study can be found in Supplementary File 2 in the online supplement.

The registered protocol can be found on the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42019130863) and has been published elsewhere.[37]

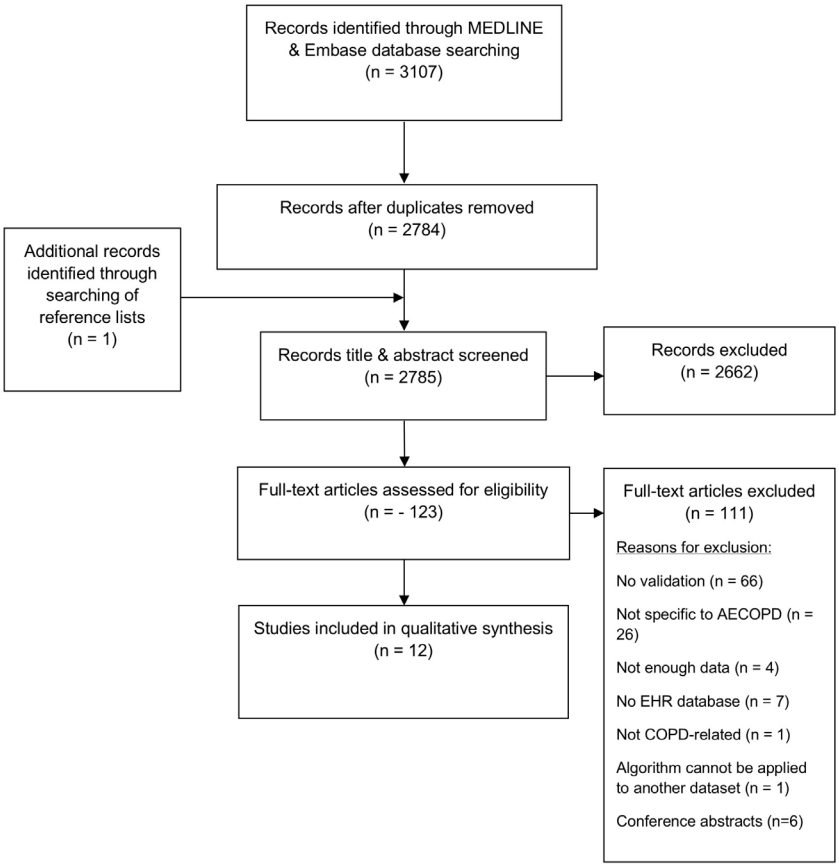
Results

From the 2784 articles found by the search strategy, 12 studies[38-49] were eligible for inclusion and were included in the review (Figure 1). Six of the studies were in databases from the United States, 4 were from English national patient databases, one was from a Japanese database and one came from the Danish National Patient Registry (summarized in Table 1). Full details of each study can be found in Supplementary File 3 in the online supplement). The clinical terminology used to retrieve data on admissions was either ICD-9-Clinicial Modifications (CM) (6 studies), ICD-10 (5 studies), or Read codes (2 studies by Rothnie et al[38,39]). The ages of patients varied between studies with one study using a broad definition of patients aged ≥18 years.[40] whereas another study was more selective, including patients ≥55 years old.[41] There was one conference abstract by Pu et al,[42] that was included and it should be noted that it has not been through peer review, however, sufficient detail was included in the abstract to allow for assessment in this review. For the reference standard, 9 studies used chart review or consensus by physicians and nurses. One study by Rothnie et al used a review of general practitioner (GP) questionnaires[38] and a subsequent study by Rothnie et al[39] utilized hospital discharge summaries. Finally, for their reference standard, Sperrin et al[43] compared the index test with AECOPD events recorded in clinical trial data.

Risk of Bias Assessment

The risk of bias for each study is shown in Table 2. None of the studies had a low risk of bias for all domains assessed. The reference standard was the domain in which studies struggled to score a low risk of bias. Only 2 out of the 12 studies scored a low risk of bias for the reference standard (Mapel et al[44] and Awano et al[40]) and only 3 studies had a low risk of bias under applicability concerns because they used spirometry in the reference standard to confirm a diagnosis of COPD (Thomsen et al,[45] Echevarria et al,[46] and Mapel et al[44]). The reference standard used by Thomsen et al[45] scored high risk of bias because physicians reviewing the charts were not blinded to the diagnosis codes of the

Figure 1. PRISMA Flowchart for Validation of Acute Exacerbation of COPD Definitions in Electronic Health Records: Systematic Review



AECOPD=acute exacerbation of COPD; EHR=electronic health record; COPD=chronic obstructive pulmonary disease

index test and, therefore, this could have influenced the interpretation and classification of the reference standard. This study was also at high risk of bias for flow and timing as it was unclear if all patients were included in the analysis as the busy hospitals (that may have had more severe cases) were unable to return all the details from the patient record. One other study (Sperrin et al[43]) also had an unclear risk of bias for the reference standard as it was unclear if the results were interpreted without knowledge of the index test. In the Stein et al (2012) study,[47] patients who were transferred from another hospital were excluded and, therefore, this study scored unclear risk of bias for patient selection. The patients who were excluded may also have been more severe cases. The Rothnie et al study[38] that validated primary care Read code definitions had a high risk of applicability concerns because they compared the results of Read code definitions against Hospital Episode Statistics (HES) ICD-10 code definitions (as the reference standard) and the results were not validated by physicians (the gold standard). Finally, 2 studies scored a high risk of bias for patient selection (Sperrin et al[43] and Mapel et al[44]) because they used more than one database from which they selected patients and this may have introduced bias as patients were not from one specific setting.

Summary of Results for Studies Validating Use of International Classification of Diseases-Ninth Revision Codes

Studies validating ICD-9-CM codes (Table 3) were all carried out in the United States. All studies validated similar ICD-9-CM codes, and the single AECOPD code of 491.2x provided the best PPV in all studies, ranging between 60% and 100%. Ginde et al[41] demonstrated high PPV (97%) for the detection of AECOPDs using 3 ICD-9-CM codes. However, results from Stein et al[48] (2010) reported lower PPV values, and these varied depending on which algorithm was used (74% for algorithm 1, 62% for algorithm 2, and 60% for algorithm 5), suggesting that the algorithms they used for identifying AECOPDs may identify a substantial number of patients admitted for alternative conditions. A subsequent study by Stein et al[47] in 2012 evaluated the 491.21 ICD-9-CM code in a comparison with other algorithms but found that sensitivity was reduced when using codes for a primary diagnosis of COPD (12.3%) or a secondary diagnosis of COPD with a primary diagnosis of respiratory failure (24.3%). Their results implied that ICD-9-CM codes may undercount hospitalizations for AECOPDs and it is questionable whether researchers should rely on ICD-9-CM codes alone to identify AECOPD admissions. Pu et al[42] also validated the use of ICD-9 -CM code 491.21 and

Table 1. Summary of Studies Included

Author, Year, Country, Period	Population Characteristics	Data Source	Code Type	Reference Standard
Ginde et al[41] 2008 United States July 2005–June 2006	Patients ≥55 years visiting the emergency department	Unspecified EHR database from 2 U.S. hospitals	ICD-9-CM	Chart review by 2 physicians
Stein et al[48] 2010 United States 2000–2006	Patients ≥40 years old with ICD-9-CM code for AECOPD	National Inpatient Sample	ICD-9-CM	Chart abstracted physician diagnosis
Thomsen et al[45] 2011 Denmark January 2008–December 2008	Patients ≥30 years old with a hospital discharge diagnosis of COPD	Danish National Patient Registry discharge codes from 34 Danish hospitals	ICD-10	Physician review of patient medical records.
Stein et al[47] 2012 United States November 2005–October 2006	Patients ≥ 40 years old with a hospital admission	Discharge codes from 2 hospitals in Chicago, United States	ICD-9-CM	Physician chart abstraction
Rothnie et al[38] 2016 United Kingdom January 2004–August 2013	COPD patients ≥35 years old with additional material provided by a GP	CPRD	Read and Product codes	Review of GP questionnaires by 2 physicians
Rothnie et al[39] 2016 United Kingdom January 2004–March 2014	COPD patients ≥35 years old	HES CPRD	ICD-10 Read and Product Codes	Hospital discharge summaries (HES-recorded hospitalization for AECOPD)
Pu et al[42] 2017 United States 2012–2014	Patients discharged with ICD-9 code for AECOPD	Hospital database	ICD-9-CM	Chart review
Sperrin et al[43] 2019 United Kingdom March 2012–October 2014	Patients ≥40 years old who had received a documented diagnosis of COPD from a GP and recorded one or more COPD exacerbations in the previous 3 years	EHRs and electronic case report forms in the Salford Lung Study	Read version 2 or ICD-10 codes.	AECOPD events recorded in clinical trial
Echevarria et al[46] 2020 United Kingdom January 2012–May 2013	Patients admitted to the hospital identified with an AECOPD.	Hospital discharge codes	ICD-10 codes	Consensus of 2 respiratory specialists using GOLD guidelines
Stanford et al[49] 2020 United States January 2009– December 2013	Patients ≥ 40 years old with ICD-9-CM codes for COPD	U.S. health care claims database: Optum Research Database	ICD-9-CM	Review of medical records by physician
Mapel et al[44] 2021 United States January 2010– September 2015	Patients aged ≥ 40 years old with ≥ 1 hospitalization, ≥ 1 emergency department visit, or ≥ 2 outpatient visits with a primary or secondary COPD diagnosis	Two independent EHR systems: Kaiser Permanente Mid-Atlantic States and Reliant Medical Group, Inc.	ICD-9-CM	Chart review by pulmonary nurses using GOLD guidelines
Awano et al[40] 2023 Japan April 2019–March 2021	Patients ≥ 18 years hospitalized in 2 acute-care hospitals in Tokyo.	Diagnosis Procedure Combination database	ICD-10 codes	Physician review of medical records

EHR=electronic health record; ICD-9-CM=International Classification of Diseases, Ninth Revision, Clinical Modification; AECOPD=acute exacerbation of COPD; COPD=chronic obstructive pulmonary disease; GP=general practitioner; CPRD=Clinical Practice Research Datalink; HES=Hospital Episode Statistics; GOLD=Global initiative for chronic Obstructive Lung Disease

found that using codes such as this could miss a significant proportion of patients with AECOPDs. In a more recent study, Stanford et al[49] modified the algorithm by Stein et al[47] in 2012 through the addition of further ICD-9-CM codes (493.12, 493.92, 494.1, 466.0) in order to identify exacerbation-related hospital visits and included events for which diagnosis codes may have been a primary or secondary diagnosis. The final algorithm in this study had a high sensitivity of 84.9% and PPV of 67.5%. Finally, a study in 2021 by Mapel et al[44] developed 2 algorithms to identify moderate and severe COPD exacerbations. They used a broader algorithm using 18 different ICD-9-CM codes and required steroid or antibiotic prescriptions to identify moderate exacerbations. For severe exacerbations, the records required an inpatient hospital stay of 2 or more days plus one of 8 different ICD-9-CM codes. For both moderate and severe exacerbations, the PPV was high (98.3% and 96.0% respectively).

Summary of Results for Studies Validating International Classification of Diseases, Tenth Revision Codes

Of the studies using ICD-10 codes to identify AECOPDs (Table 4), 3 were carried out in the United Kingdom[39,43,46] and one was carried out in Japan.[40] All studies validated variations

Table 2. Adapted QUADAS-2 Risk of Bias Results Table for Studies Included

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Ginde et al[41] 2008	Low	Low	?	Low	Low	Low	High
Stein et al[48] 2010	Low	Low	?	Low	Low	Low	High
Thomsen et al[45] 2011	Low	Low	High	High	?	Low	Low
Stein et al[47] 2012	?	Low	Low	Low	Low	Low	High
Rothnie et al[38] 2016 Read codes	Low	Low	?	Low	Low	Low	High
Rothnie et al[39] 2016 HES/ICD-10	Low	Low	?	Low	Low	Low	High
Pu et al[42] 2017	?	Low	?	?	?	Low	High
Sperrin et al[43] 2019	Low	Low	?	Low	High	Low	High
Echevarria et al[46] 2020	Low	Low	?	Low	?	Low	Low
Stanford et al[49] 2020	Low	Low	?	Low	Low	Low	High
Mapel et al[44] 2021	Low	Low	Low	?	High	Low	Low
Awano et al[40] 2023	Low	Low	Low	Low	?	Low	High

Low=Low risk of bias; High=High risk of bias; ?=Unclear risk of bias

HES-Hospital Episode Statistics; ICD-10=International Classification of Diseases, Tenth Revision

of J44 COPD codes, except for Awano et al[40] who validated a broader collection of ICD-10 codes (J410, J411, J42, J43, J44, J449, J841). Specificity and NPV were high in this study (96.1% and 82.9% respectively), however, sensitivity was low (33.7%). In the Danish study,[45] J44 was used as a parent code for primary AECOPD diagnosis resulting in the best PPV (93%), and when testing all 3 algorithms good PPVs were found. In the United Kingdom, in Rothnie et al,[39] the highest sensitivity (87.5%) was found using a COPD code (J44.9) as the primary diagnosis or using codes for AECOPD (J44.0 or J44.1) or lower respiratory tract infections (LRTI) (J22) as either primary or secondary diagnosis codes. The high sensitivity found with this algorithm by Rothnie et al[39] may represent a good compromise between high sensitivity and high PPV because it is similar to the algorithm by Thomsen et al[45] which gave a high PPV. In a more recent U.K. study, Sperrin et al[43] used algorithms for Read codes and ICD-10 codes from both Rothnie et al studies.[38,39] Results were populated from a best-case scenario, using the full algorithm in primary and secondary care, and allowing a maximum gap in the start or end dates of the episodes of up to 15 days. This gave a PPV of 73.6% and a sensitivity of 69.1%. Finally, Echevarria et al,[46] also in the United Kingdom using J44 ICD-10 codes alone, reported a PPV of 63.9%, an NPV of 75.5%, a sensitivity of 70.7%, and specificity of 69.4%.

Summary of Results for Studies Validating Read Codes

Two studies[38,39] validating the use of Read codes were done in the United Kingdom by Rothnie et al in 2016 (Table 5). The first Rothnie et al study validated the use of Read codes in English primary care against a reference standard of GP questionnaires.[38] PPV and sensitivity were used to validate the algorithms and the best compromise was found between the 2 measures when combining their algorithms with a PPV >75%. Using the same definitions as the first, the second study validated the algorithms against HES ICD-10 codes.[39] The combination in the algorithm included antibiotic and oral corticosteroid prescriptions for 5–14 days, a symptom (such as dyspnea, cough, or sputum) in addition to the prescription of antibiotics or oral corticosteroids, an LRTI, or an AECOPD code, and produced a PPV of 85.5% and sensitivity of 62.9%.

A quantitative synthesis was unfortunately not possible because of the limited number of studies in which the same clinical terminology was used, and a lack of data on true and false positives and negatives.

Discussion

This systematic review assessed different methods for validating the recording of acute exacerbations of COPD in EHRs and found that a variety of definitions were used. Studies used ICD-9-CM codes, ICD-10 codes, and different combinations of clinical codes in both primary care (using Read codes) and secondary care settings.

Results from studies validating ICD-9-CM codes suggest that ICD-9-CM codes alone may not accurately identify all patients with AECOPD. The validation measurements varied considerably depending on which codes or algorithms were used. The code 491.21 is used to classify obstructive chronic bronchitis with acute exacerbation and this code had a high PPV of 100% in one study[41] but had low sensitivity in other studies[42,47] suggesting that ICD-9-CM codes alone may underestimate the burden of hospitalizations for COPD. One study[49] modified their algorithms through the addition of further ICD-9-CM codes, for example using those to denote asthma with acute exacerbation, bronchiectasis, and acute bronchitis. Although this improved the sensitivity of the algorithm (84.9%), the ability to detect true positives was not as high (PPV 67.5%). However, using multiple ICD-9-CM codes

Table 3. Summary of International Classification of Disease-9th Revision Validation Studies of Acute Exacerbation of COPD Definitions

Study	Algorithm (Codes)	Gold Standard Reference	N	PPV/ Derived PPV (95% CI)	NPV/ Derived NPV (95% CI)	Sensitivity/ Derived Sensitivity (95% CI)	Specificity/ Derived Sensitivity (95% CI)
Ginde et al[41] 2008	491.2x	Consensus by 2 emergency physicians from abstracted chart data	181	100% (98–100)	-	-	-
	491.2x, 492.8, or 496	(same as above)	200	97% (93–99)	-	-	-
Stein et al[48] 2010	Algorithm 1: 491.21 primary diagnosis	Primary diagnosis recorded in physician notes	Sample of 200	74%	-	-	-
	Algorithm 2: 491.x, 492.x, or 496	(same as above)		62%	-	-	-
	Algorithm 5: 491.0, 491.1, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 493.22, or 496 primary diagnosis OR 518.81, 518.82, or 518.84 primary diagnosis AND 491.0, 491.1, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 493.22, or 496 secondary diagnosis	(same as above)		60%	-	-	-
Stein et al[47] 2012	Primary diagnosis of COPD (491.0, 491.1, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 493.22, 496) OR primary diagnosis of respiratory failure (518.81, 518.82, 518.84) AND secondary diagnosis of COPD (defined using same codes as primary diagnosis) (age ≥40)	Physician chart abstraction: physician diagnosis of COPD; presence of cough, dyspnea, or sputum production on presentation; and hospitalization for one of these respiratory symptoms	46	85.4%	93.9%	24.3%	99.7%
	Primary diagnosis of AECOPD: 491.21 (age≥40)	(same as above)	20	97.2%	93%	12.3%	100%
Pu et al[42] 2017	491.21 (AECOPD)	Chart review	620	91%	31% (27–35)	57% (54–61)	76% (70–81)
Stanford et al[49] 2020	Claims-based algorithm (modified from the Stein 2012 algorithm through the addition of further ICD-9 codes - 493.12, 493.92, 494.1, 466.0)	Review of exacerbation history in medical records by patient's physician.	402	(88–93) 67.5%	-	84.90%	-
Mapel et al[44] 2021		Chart review by trained pulmonary nurses using GOLD COPD 2017 definition.	298	98.3% (96.1–99.5)	75.0% (65.3, 83.1)	-	-
	Severe exacerbations: At least 1 inpatient hospital stay or 2 or more days with any of the following ICD-9-CM codes as primary diagnosis: 491, 492, 493.20, 493.22, 496, 518.81, 518.82, 518.84.	(same as above)	225	96.0% (92.5–98.2)	95.0% (88.7, 98.4)		

PPV=positive predictive value; CI=confidence interval; NPV=negative predictive value; COPD=chronic obstructive pulmonary disease; AECOPD=acute exacerbation of COPD; ICD-9-CM=International Classification of Diseases, Ninth Revision, Clinical Modification; GOLD=Global initiative for chronic Obstructive Lung Disease

alongside additional information on treatment from care records such as the prescription of steroids or antibiotics, gave high PPVs for moderate (98.3%) and severe (96.0%) AECOPD in another study.[44]

Our review also found that, as with ICD-9-CM codes, using ICD-10 codes alone in the algorithms may not effectively identify admissions for AECOPD in EHRs. In the United Kingdom, Echevarria et al[46] found that using ICD-10 codes alone missed almost a third of patients admitted with AECOPD in their study. By contrast, the Danish study[45] found that using a J44 parent code as primary diagnosis gave a high PPV (93%). However, in this study, the reviewers were not blinded to the diagnosis codes and, therefore, knowledge of this could have influenced the results of the physicians' assessment. The recent study in the Japanese database Diagnosis Procedure Combination (DPC)[40] combined multiple ICD-10 codes in addition to J44, including those for bronchitis (J40, J411, J42), emphysema (J43), and acute interstitial pneumonitis (J841). Although the specificity and NPV were high (96% and 83% respectively), sensitivity was low (34%). The authors presumed that diagnoses for chronic diseases such as COPD had not been recorded in the DPC database as specific tests or treatments were not required during hospitalization. This suggests the use of other clinical data in addition to ICD-10 codes would improve the identification of hospitalizations for AECOPD. Rothnie et al completed 2 studies[38,39] in 2016 validating the recording of AECOPD cases within U.K. health records. In the first study,[38] the data collected was

Table 4. Summary of International Classification of Diseases-10th Revision Validation Studies of Acute Exacerbation of COPD Definitions

Study	Algorithm (Codes)	Gold Standard Reference	N	PPV/ Derived PPV (95% CI)	NPV/ Derived NPV (95% CI)	Sensitivity/ Derived Sensitivity (95% CI)	Specificity/ Derived Sensitivity (95% CI)
Thomsen et al[45] 2011	J44 (COPD) as primary diagnosis	Physician review of patient medical records	1223	93% (92–95)	-	-	-
	Pneumonia (J13-J18) without J44		1432	-	82% (80–84)	-	-
Rothnie et al[39] 2016 (HES/ICD-10)	Specific AECOPD code (J44.0 or J44.1) or LRTI code (J22) in any position or COPD code (J44.9) in the first position in any FCE during spell	Hospital discharge summary	40	-	-	87.5% (72.4–94.9)	-
Sperrin et al[43] 2019	Algorithms for both Read codes from Rothnie et al[38] 2016 *AND* ICD-10 codes from Rothnie et al[39] Results from a "*best-case scenario*," using the full algorithm in primary and secondary care, and allowing a maximum gap in the start or end dates of the episodes of up to 15 days	Moderate and severe AECOPD episodes reported in the eCRF for a clinical trial	3042	73.6%	-	69.1%	-
Echevarria et al[46] 2020	COPD codes J44	Consensus of 2 respiratory specialists using GOLD guidelines	1014	63.9%	75.5%	70.7%	69.4%
Awano et al[40] 2023	COPD codes J410, J411, J42, J43, J44, J449, J841	Physician review of patient medical records	92	72.1%	82.9%	33.7%	96.1%

PPV=positive predictive value; CI=confidence interval; NPV=negative predictive value; COPD=chronic obstructive pulmonary disease; HES=Hospital Episode Statistics; ICD-10=International Classification of Diseases, Tenth Revision; AECOPD=acute exacerbation of COPD; LRTI=lower respiratory tract infection; FCE=functional capacity evaluation; eCRF=electronic case report form; GOLD=Global initiative for chronic Obstructive Lung Disease

Table 5. Summary of Read Code Validation Studies of Acute Exacerbation of COPD

Study	Algorithm (Codes)	Gold Standard Reference	N	PPV/ Derived PPV (95% CI)	NPV/ Derived NPV (95% CI)	Sensitivity/ Derived Sensitivity (95% CI)	Specificity/ Derived Sensitivity (95% CI)
Rothnie et al[38] 2016 (subset with additional patient data)	OCS prescription	Review of GP questionnaires and other relevant material from patient notes by 2 respiratory physicians (with additional information provided by GPs)	367	72.2% (66.5–77.9)	-	22.7% (16.1–29.2)	-
	Antibiotic prescription		2245	61.3% (58.3–64.3)	-	63.4% (55.4–71.4)	-
	LTRI code and OCS (on the same day)		621	84.5% (80.6–88.5)	-	20.6% (15.2–26.0)	-
	AECOPD code		350	98.3% (96.9–99.6)	-	26.8% (19.7–33.9)	-
Rothnie et al[38] 2016 (subset with additional patient data-combined algorithms)	Prescription of antibiotics and OCS for 5–14 days; or Symptom definition with prescription of antibiotic or OCS; or LRTI code; or AECOPD code	Review of GP questionnaires and other relevant material from patient notes by 2 respiratory physicians (with additional information provided by GPs)		85.5% (82.7–88.3)	-	62.9% (55.4–70.4)	-
	All algorithms combined			63.8% (61.0–66.6)	-	88.1% (82.9–93.4)	-
Rothnie et al[39] 2016 (CPRD/Read)	AECOPD hospitalization code	HES: Specific AECOPD code (J44.0 or J44.1) or LRTI code (J22) in any position or COPD code (J44.9) in the first position in any FCE during spell		50.2% (48.5–51.8)	-	4.1% (3.9–4.3)	-
	AECOPD identified using validated algorithm and hospitalization code			43.3% (42.3–44.2)	-	5.4% (5.1–5.7)	-

PPV=positive predictive value; CI=confidence interval; NPV=negative predictive value; OCS=oral corticosteroid; LRTI=lower respiratory tract infection; AECOPD=acute exacerbation of COPD; GP=general practitioner; CPRD=Clinical Practice Research Datalink; HES=Hospital Episode Statistics; COPD=chronic obstructive pulmonary disease; FCE=functional capacity evaluation

purely from primary health care via the Clinical Practice Research Datalink (CPRD) database using Read codes and product codes. It was suggested that using multiple codes increased the validity, in this case, AECOPD, LRTI codes, antibiotics, and oral corticosteroid codes were utilized. This combination of codes led to a PPV of 85.5% but a lower sensitivity of 62.9%, suggesting that although the strategy was valid it would underestimate the number of events. The

second study[39] then aimed to identify hospitalizations for AECOPD in CPRD using secondary care data linked to HES and found a sensitivity of 87.5%. However, when using a code suggesting hospitalization for an AECOPD in primary care data alone without HES linkage, a much lower PPV of 50.2% and a sensitivity of 4.1% were found. This implies that primary care data alone does not accurately identify hospitalizations for AECOPD, and researchers should use primary care data that are linked to data from secondary care.

As the screening process was undertaken, it became clear that one study by Shah et al[50] stood out for using very different algorithms from the other studies. For the index test, they compared 6 models with different combinations of clinical and administrative data detailing care steps for patients admitted to the hospital including COPD "power plans," bronchodilator protocol use, billing diagnosis, and treatments administered such as steroid use and oxygen management. Unlike other studies in which ICD-10 codes were used as the index test, this study used the final billing ICD-10 diagnosis for AECOPD as the reference standard for comparing model performance. Since the aim of this review is to provide guidance on which algorithms provide the most accurate cohort of AECOPDs, it became apparent that researchers should be able to apply the recommended algorithms to other datasets, and, therefore, our additional exclusion criteria ruled out this study from our review.

None of the studies had a low risk of bias for all domains assessed meaning that the validity of all the studies may be overestimated. Most studies scored a high risk of bias for the applicability of the reference standard because they did not use spirometry to confirm the COPD diagnosis. Spirometry is a key component of a COPD diagnosis and, therefore, not including it to confirm COPD in the reference standard could increase the risk of bias. Stein et al[47](2012) explained the reason why they deliberately did not confirm COPD diagnosis in their reference standard with spirometry as *"it would have led to a narrowly selected (and potentially biased) sample with which to evaluate the validity of ICD-9-CM algorithms*." In this case, the authors were aiming for sensitivity over specificity. However, their definition should still be considered at risk of bias because it makes it more likely that the reference standard could include non-COPD cases.

A similar systematic review was conducted looking at the validation of codes for asthma within EHRs.[30] They conducted a search and found 13 studies that fit their inclusion criteria, particularly choosing to focus on the databases and codes used, along with any sensitivity or specificity measures. As in our review in which the validity of definitions of AECOPD varied across different databases and settings, they found that case definitions and methods of asthma diagnosis validation also varied widely across different EHR databases. The authors suggested that the source of the EHR databases (primary care, secondary care, and urgent care) could influence the case definition of asthma and the way the validation is conducted. For example, patients seeking care for asthma symptoms might present differently in each setting, and the test measures, therefore, might reflect this.

In this study, we have found that using single codes to search for case definitions of AECOPD in EHRs may not effectively identify admissions for AECOPD. Some of the research has shown that modifying algorithms with additional codes may improve sensitivity but at the expense of accurately identifying true positives. This review and others[26,30] have shown that different research questions may necessitate different case definitions. For example, if researchers want to prioritize specificity over sensitivity, a more restrictive definition of AECOPD would be used, and vice versa. The Stein et al[48] (2010) findings suggested that the selection of an algorithm should depend on its intended purpose. For example, if the intent is to identify patients for quality measurement, an algorithm with the highest PPV would be desirable (e.g., their first algorithm using ICD-9-CM code 491.21). However, if the intent is to estimate the overall burden of disease, then the authors suggested using a more inclusive approach. We propose that a Delphi study would be useful to obtain the consensus of expert clinicians and researchers to decide which algorithms would be recommended in different research scenarios.

There are some strengths and limitations to our study. To our knowledge, this is the first review to systematically review studies that validated definitions of AECOPD in EHRs. We used broad search criteria which meant that we could review a variety of different codes and algorithms used in different databases globally. However, we found that in many studies, the clinical codes utilized were not well reported or were difficult to obtain. Our risk of bias assessment, the adapted QUADAS-2, may have unfairly scored studies that did not use spirometry in the reference standard with a high risk of bias because spirometry was unavailable to confirm the diagnosis of COPD. However, this highlights the importance and need for spirometry data in EHRs. Finally, we were unable to carry out a quantitative analysis because of the limited number of studies included in our review.

COPD and acute exacerbations are underdiagnosed in the general population[51] and this is related to the underuse of spirometry as we found in many of the studies. Furthermore, recordings of AECOPDs in EHRs tend to capture events that lead to health care utilization, such as moderate and severe exacerbations, therefore, limiting the capture of mild exacerbations. These are important points for researchers to consider in the future when devising methods to identify AECOPDs in EHRs, and to find ways of balancing sensitivity versus specificity.

Conclusion

The methods used for validating definitions of AECOPDs in electronic health care vary, with different algorithms and case definitions used in different databases globally and in different settings such as primary and secondary care. Using single codes to identify COPD exacerbations (for example ICD-9-CM code 491.21 or ICD-10 code J44) was found to have a high PPV in some studies but low sensitivity in others. This means that the algorithms used can positively identify cases of AECOPD within datasets but may not accurately identify all cases. At present, there is no clear consensus on which definition provides the highest validity or the most sensitive and specific results when searching EHRs for AECOPD cases. The variation between studies in defining COPD exacerbations restricts the ability of researchers to reliably compare findings and provide robust evidence. Consensus from experts is required to guide researchers on which definitions to use in different research scenarios. Researchers should endeavor to make all their disease definitions easily accessible so that others can validate and replicate them.

Acknowledgments

Author contributions: JKQ, EM, and PS were responsible for the conception and design of the work. All authors were responsible for the acquisition of data, data analysis, and interpretation. EM and PS were equally responsible for writing the manuscript. In addition, all authors have read and approved of the manuscript.

The authors would like to thank Dr. Nikhil Sood for his research contribution towards this review.

Declaration of Interests

EM, PS, AA, JS, SD, and SA have nothing to declare. JKQ has been supported by institutional research grants from the Medical Research Council, the National Institute for Health and Care Research, Health Data Research, GSK, Boehringer Ingelheim, AstraZeneca, Insmed, and Sanofi and received personal fees for advisory board participation, and consultancy or speaking fees from GSK, Chiesi, and AstraZeneca.

References

1. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Pocket guide to COPD diagnosis, management, and prevention: a guide for health care professionals, 2022 Report. Published 2021. Accessed October 2024. https://goldcopd.org/wp-content/uploads/2021/12/GOLD-POCKET-GUIDE-2022-v1.1-22Nov2021_WMV.pdf

2. Asthma and Lung UK. What is chronic obstructive pulmonary disease (COPD)? Published June 2022. Accessed October 2024. https://www.blf.org.uk/support-for-you/copd/what-is-copd

3. National Institute for Health and Care Excellence. Chronic obstructive pulmonary disease in over 16s: diagnosis and management. Updated July 2019. Accessed October 2024. https://www.nice.org.uk/guidance/ng115/chapter/Recommendations.

4. Soler-Cataluna JJ, Martinez-Garcia MA, Roman Sanchez P, Salcedo E, Navarro M, Ochando R. Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease. *Thorax*. 2005;60(11):925-31. https://doi.org/10.1136/thx.2005.040527

5. Donaldson GC, Seemungal TAR, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax*. 2002;57(10):847-52. https://doi.org/10.1136/thorax.57.10.847

6. Crook S, Busching G, Keusch S, et al. The association between daily exacerbation symptoms and physical activity in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis*. 2018;13:2199-206. https://doi.org/10.2147/COPD.S156986

7. Seemungal TAR, Donaldson GC, Paul EA, Bestall JC, Jeffries DJ, Wedzicha JA. Effect of exacerbation on quality of life in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 1998;157(5):1418-1422. https://doi.org/10.1164/ajrccm.157.5.9709032

8. Anzueto A. Impact of exacerbations on COPD. *Eur Respir Rev*. 2010;19(116):113-118. https://doi.org/10.1183/09059180.00002610

9. Geitona M, Hatzikou M, Steiropoulos P, Alexopoulos EC, Bouros D. The cost of COPD exacerbations: a university hospital-based study in Greece. *Respir Med*. 2011;105(3):402-409. https://doi.org/10.1016/j.rmed.2010.09.020

10. Mittmann N, Kuramoto L, Seung SJ, Haddon JM, Bradley-Kennedy C, Fitzgerald JM. The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. *Respir Med*. 2008;102(3):413-421. https://doi.org/10.1016/j.rmed.2007.10.010

11. Dalal AA, Shah M, D'Souza AO, Rane P. Costs of COPD exacerbations in the emergency department and inpatient setting. *Respir Med*. 2011;105(3):454-60. https://doi.org/10.1016/j.rmed.2010.09.003

12. Friedman M, Hilleman DE. Economic burden of chronic obstructive pulmonary disease. Impact of new treatment options. *Pharmacoeconomics*. 2001;19(3):245-54. https://doi.org/10.2165/00019053-200119030-00003

13. Miravitlles M, Murio C, Guerrero T, Gisbert R. Pharmacoeconomic evaluation of acute exacerbations of chronic bronchitis and COPD. *Chest*. 2002;121(5):1449-55. https://doi.org/10.1378/chest.121.5.1449

14. National Institute for Health and Care Excellence (NICE). Chronic obstructive pulmonary disease in over 16s: diagnosis and management. NICE website. Updated July 2019. Accessed October 2024. https://www.nice.org.uk/guidance/ng115

15. Pavord ID, Jones PW, Burgel PR, Rabe KF. Exacerbations of COPD. *Int J Chron Obstruct Pulmon Dis*. 2016;11 Spec Iss(Spec Iss):21-30. https://doi.org/10.2147/COPD.S85978

16. Ambinder E. Electronic health records. *J Oncol Pract*. 2005;1(2):57-63. https://doi.org/10.1200/jop.2005.1.2.57

17. Kim E, Rubinstein SM, Nead KT, Wojcieszynski AP, Gabriel PE, Warner JL. The evolving use of electronic health records (EHR) for research. *Semin Radiat Oncol*. 2019;29(4):354-361. https://doi.org/10.1016/j.semradonc.2019.05.010

18. Centers for Disease Control and Prevention (CDC), National Center for Health Statistics. International Classification of Diseases, Ninth Revision (ICD-9). CDC website. Archived May 2024. Accessed October 2024. https://archive.cdc.gov/#/details?url=https://www.cdc.gov/nchs/icd/icd9.htm

19. National Health Service (NHS). Read codes, 2023. NHS website. Updated December 7, 2023. Accessed October 2024. https://digital.nhs.uk/services/terminology-and-classifications/read-codes

20. Souri S, Symonds NE, Rouhi A, et al. Identification of validated case definitions for chronic disease using electronic medical records: a systematic review protocol. *Syst Rev*. 2017;6(1):38. https://doi.org/10.1186/s13643-017-0431-9

21. Ellili N, Nobanee H, Alsaiari L, et al. The applications of big data in the insurance industry: a bibliometric and systematic review of relevant literature. *J Finance Data Sci*. 2023;9(100102). https://doi.org/10.1016/j.jfds.2023.100102

22. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health Serv Res*. 2005;40(5 Pt 2):1620-1639. https://doi.org/10.1111/j.1475-6773.2005.00444.x

23. 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(11):e0292876. https://doi.org/10.1371/journal.pone.0292876

24. Ehrenstein V, Petersen I, Smeeth L, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. *Clin Epidemiol*. 2016;2016(8):49-51. https://doi.org/10.2147/CLEP.S104448

25. Fox MP, Lash TL, Bodnar LM. Common misconceptions about validation studies. *Int J Epidemiol*. 2020;49(4):1392-1396. https://doi.org/10.1093/ije/dyaa090

26. Sivakumaran S, Alsallakh MA, Lyons RA, Quint JK, Davies GA. Identifying COPD in routinely collected electronic health records: a systematic scoping review. *ERJ Open Res.* 2021;7(3):00167-2021. https://doi.org/10.1183/23120541.00167-2021

27. National Center for Biotechnology Information (NCBI). PubMed Help. NCBI website. Published June 2018. Accessed October 2024. https://www.ncbi.nlm.nih.gov/books/NBK3827/

28. Baumann N. How to use the medical subject headings (MeSH). *Int J Clin Pract.* 2016;70(2):171-174. https://doi.org/10.1111/ijcp.12767

29. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol.* 2011;64(8):821-829. https://doi.org/10.1016/j.jclinepi.2010.10.006

30. Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. Validation of asthma recording in electronic health records: a systematic review. *Clin Epidemiol.* 2017;2017(9):643-656. https://doi.org/10.2147/CLEP.S143718

31. 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(6):e011777. https://doi.org/10.1136/bmjopen-2016-011777

32. Jayatunga W, Stone P, Aldridge RW, Quint JK, George J. Code sets for respiratory symptoms in electronic health records research: a systematic review protocol. *BMJ Open.* 2019;9(3):e025965. https://doi.org/10.1136/bmjopen-2018-025965

33. Leong A, Dasgupta K, Bernatsky S, Lacaille D, Avina-Zubieta A, Rahme E. Systematic review and meta-analysis of validation studies on a diabetes case definition from health administrative records. *PLoS One.* 2013;8(10):e75256. https://doi.org/10.1371/journal.pone.0075256

34. Shiff NJ, Jama S, Boden C, Lix LM. Validation of administrative health data for the pediatric population: a scoping review. *BMC Health Serv Res.* 2014;14:236. https://doi.org/10.1186/1472-6963-14-236

35. Stone P, Sood N, Feary J, Roberts CM, Quint JK. Validation of acute exacerbation of chronic obstructive pulmonary disease (COPD) recording in electronic health records: a systematic review protocol. *BMJ Open.* 2020;10(2):e032467. https://doi.org/10.1136/bmjopen-2019-032467

36. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

37. Stone P. Sood N, Feary J, Roberts CM, Quint JK. Validation of acute exacerbation of chronic obstructive pulmonary disease (COPD) recording in electronic health records: a systematic review. *BMJ Open.* 2021;10:e032467. https://doi.org/10.1136/bmjopen-2019-032467

38. Rothnie KJ, Mullerova 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(3):e0151357. https://doi.org/10.1371/journal.pone.0151357

39. Rothnie KJ, Mullerova H, Thomas SL, et al. Recording of hospitalizations for acute exacerbations of COPD in UK electronic health care records. *Clin Epidemiol.* 2016;2016(8):771-782. https://doi.org/10.2147/CLEP.S117867

40. Awano N, Urushiyama H, Yamana H, et al. Validity of diagnoses of respiratory diseases recorded in a Japanese administrative database. *Respir Investig.* 2023;61(3):314-320. https://doi.org/10.1016/j.resinv.2023.01.009

41. Ginde AA, Tsai CL, Blanc PG, Camargo CA, Jr. Positive predictive value of ICD-9-CM codes to detect acute exacerbation of COPD in the emergency department. *Jt Comm J Qual Patient Saf.* 2008;34(11):678-680. https://doi.org/10.1016/S1553-7250(08)34086-0

42. Pu C, Nunez Lopez R, Aryal K, Quesada N, Tulaimat A. QA project: accuracy of case definition used for COPD epidemiologic studies. B49. COPD: symptom and clinical assessment. *Am J Respir Crit Care Med.* 2017;195:A3664.

43. Sperrin M, Webb DJ, Patel P, et al. Chronic obstructive pulmonary disease exacerbation episodes derived from electronic health record data validated using clinical trial data. *Pharmacoepidemiol Drug Saf.* 2019;28(10):1369-1376. https://doi.org/10.1002/pds.4883

44. Mapel DW, Roberts MH, Sama S, et al. Development and validation of a healthcare utilization-based algorithm to identify acute exacerbations of chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2021;16:1687-1698. https://doi.org/10.2147/COPD.S302241

45. Thomsen RW, Lange P, Hellquist B, et al. Validity and underrecording of diagnosis of COPD in the Danish National Patient Registry. *Respir Med.* 2011;105(7):1063-1068. https://doi.org/10.1016/j.rmed.2011.01.012

46. Echevarria C, Steer J, Bourke SC. Coding of COPD exacerbations and the implications on clinical practice, audit and research. *COPD.* 2020;17(6):706-710. https://doi.org/10.1080/15412555.2020.1841745

47. Stein BD, Bautista A, Schumock GT, et al. The validity of international classification of diseases, ninth revision, clinical modification diagnosis codes for identifying patients hospitalized for COPD exacerbations. *Chest.* 2012;141(1):87-93. https://doi.org/10.1378/chest.11-0024

48. Stein BD, Charbeneau JT, Lee TA, et al. Hospitalizations for acute exacerbations of chronic obstructive pulmonary disease: how you count matters. *COPD.* 2010;7(3):164-171. https://doi.org/10.3109/15412555.2010.481696

49. Stanford RH, Engel-Nitz NM, Bancroft T, Essoi B. The identification and cost of acute chronic obstructive pulmonary disease exacerbations in a United States population healthcare claims database. *COPD.* 2020;17(5):499-508. https://doi.org/10.1080/15412555.2020.1817357

50. Shah P, McWilliams A, Howard D, Roberge J. A comparison of methodologies for the real-time identification of hospitalized patients with acute exacerbations of COPD. *Int J Chron Obstruct Pulmon Dis*. 2019;14:693-698. https://doi.org/10.2147/COPD.S175296

51. Almagro P, Soriano JB. Underdiagnosis in COPD: a battle worth fighting. *Lancet Respir Med*. 2017;5(5):367-368. https://doi.org/10.1016/s2213-2600(17)30133-9