USING ROUTINELY COLLECTED DATA TO MAP THE ROLE OF RESPIRATORY PATHOGENS IN ACUTE EXACERBATIONS OF COPD

Shanya Sivakumaran

Submitted to Swansea University in fulfilment of the requirements for the Degree of MSc by Research

Swansea University, 2023

SUMMARY

Linkage of routinely collected microbiology data with other electronic health records (EHRs) could provide important insights into a variety of infection syndromes. In a demonstration of utility, over the course of the thesis, I outline the steps taken in order to retrospectively identify laboratory-confirmed respiratory pathogens associated with hospital admission for acute exacerbations of chronic obstructive pulmonary disease (COPD) in Wales, over a two-year period.

I firstly performed a systematic, scoping review to explore how individuals with COPD were identified in EHRs in the recent literature.

Next, using the Secure Anonymised Information Linkage (SAIL) databank, which contains de-identified health and administrative data covering the entire population of Wales, I created a dataset of individuals admitted to hospitals in Wales with acute exacerbations of COPD over a two-year period, and linked these records to laboratory tests for respiratory pathogens associated with the admission.

Using this dataset, I could then identify what proportion of these emergency admissions were associated with testing for, and detection of, a respiratory pathogen. Additionally, I was able to examine the accuracy of using diagnosis codes (specifically, International Statistical Classification of Diseases and Related Health Problems (ICD) codes) to identify laboratory-confirmed respiratory pathogens associated with COPD exacerbations.

My analysis revealed that respiratory viruses were detected in 46.7% of hospital admissions for COPD exacerbation where testing was undertaken, however diagnostic testing appears to be underutilised (respiratory virus testing carried out in only 4.7% of emergency admissions for COPD). Increasing respiratory viral testing in this population therefore has the potential to enable more effective antimicrobial stewardship.

When comparing ICD codes to microbiology data, the analysis showed that ICD codes have low sensitivity in identifying laboratory-confirmed respiratory pathogens. Large-scale linkage with microbiology data is thus key in order to be able to delineate the burden of specific pathogens with greater accuracy.

DECLARATIONS & STATEMENTS

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed...S Sivakumaran.....

Date.....19/6/23.....

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed...S Sivakumaran.....

Date.....19/6/23.....

I hereby give my consent for my work, if relevant and accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access approved by the University.**

Signed...S Sivakumaran.....

Date.....19/6/23.....

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed...S Sivakumaran.....

Date.....19/6/2.....

Contents

ACKNOWLEDGEMENTS

With thanks to my supervisors – Gwyneth Davies, Ronan Lyons, Jennifer Quint and Mohammad Al-Sallakh – for their time, guidance, and opportunities outside of this work, for which I am very grateful; to the SAIL Databank team and all the data providers who make anonymised data available for research; and to Alistair, for your support in everything I do.

LIST OF TABLES & FIGURES

LIST OF TABLES IN FREE TEXT

LIST OF TABLES IN RESEARCH PAPERS

RESEARCH PAPER I

RESEARCH PAPER II

LIST OF FIGURES IN RESEARCH PAPERS

ABBREVIATIONS

AECOPD	Acute exacerbation(s) of chronic obstructive pulmonary disease
BMI	Body mass index
COPD	Chronic obstructive pulmonary disease
ED	Emergency department
EHR	Electronic health records
FEV1	Forced Expiratory Volume in 1 second
FVC	Forced Vital Capacity
GOLD	Global Initiative for Chronic Obstructive Lung Disease
GP	General Practitioner
HCP	Healthcare practitioner
ICD	International Statistical Classification of Diseases and Related Health Problems
ICS	Inhaled corticosteroid
LABA	Long-acting beta-2 agonist
LAMA	Long-acting muscarinic antagonist
MRC	Medical Research Council
NHS	National Health Service
NIV	Non-invasive ventilation
NICE	National Institute for Health and Care Excellence
PCP	Pneumocystis pneumonia
PEDW	Patient Episode Database for Wales
RCT	Randomised controlled trial
SABA	Short-acting beta-2 agonists
SABD	Short-acting bronchodilator
SAMA	Short-acting muscarinic antagonists
SAIL	Secure Anonymised Information Linkage Databank
UK	United Kingdom
WDS	Welsh Demographic Service
WLIMS	Welsh Laboratory Information Management System
WRRS	Welsh Results Reporting Service

1 INTRODUCTION

1.1 CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Chronic obstructive pulmonary disease (COPD) is a common, chronic condition characterised by persistent respiratory symptoms and irreversible expiratory airflow limitation, usually caused by chronic exposure to inhaled noxious substances.[1] It represents an important cause of morbidity and mortality globally and in those aged over 75, COPD has remained the third leading cause of disability adjusted life years (DALYs) from 1990 to 2019, as reported by the Global Burden of Disease study.[2]

Due to our globally aging population and continued exposures to known risk factors, COPD will likely contribute significantly to the global burden of disease for the foreseeable future.

1.1.1 Epidemiology

Estimating the prevalence of COPD is difficult due to the varying definitions of the condition in use. The Burden of Obstructive Lung Disease (BOLD) study[3] estimates a worldwide prevalence of COPD with GOLD stage 2+ (defined below) of 10.1% (11.8% for men, 8.5% for women), with increasing prevalence by age, and significant heterogeneity across the centres studied. The geographical differences seen globally are likely to reflect differing exposures to risk factors, and prevalence is increasing in low and middle-income countries (LMIC).

Cigarette smoking is recognised as the most important risk factor for COPD in high-income countries. Thus, localities with higher smoking rates have increased prevalence of COPD. In the United Kingdom (UK), the prevalence of cigarette smoking is generally higher among higher deprivation groups according to occupation[4] or annual income.[5] However, not all smokers develop COPD and the individual susceptibility to the condition, through both genetic and environmental factors, is not well understood.

There are also those who develop COPD but have never been smokers, and the proportion of non-smokers with COPD is increased in LMICs, where indoor use of biomass fuels, occupational exposures and tuberculosis play a significant role.[6–9]

1.1.2 Pathophysiology

Environmental insult plus susceptibility to developing COPD results in an inflammatory response in the airways and imbalance of proteases and anti-proteases,[10] leading to proteolysis, loss of elastic tissue in airway walls, disruption of alveolar attachments and airway enlargement and collapse. Inflammation of airway walls, luminal occlusion by exudate and narrowing due to disrupted alveolar attachments leads to obstruction of the small airways. This obstruction is usually slowly progressive over time.

1.1.3 Clinical features

The progressive expiratory airflow limitation and airway inflammation leads to chronic cough, wheeze and gradually worsening dyspnoea, with subsequent limitation in activities of daily living and impact on quality of life. The degree of breathlessness experienced by an individual can be quantified by simple scoring systems, which help to classify the burden of COPD in a standardised manner. The Medical Research Council (MRC) dyspnoea scale[11] is one such tool.

As the condition progresses, people with COPD can develop 'cor pulmonale' – impairment of cardiac function secondary to the lung pathology, which leads to peripheral and pulmonary oedema, worsening the degree of dyspnoea. Further, they can develop chronically raised carbon dioxide levels – type II respiratory failure – which can lead to impairment of higher mental function.

1.1.4 Exacerbations

Definitions: As well as the day-to-day symptoms described above, people with COPD can experience acute "flare-ups" of their symptoms of dyspnoea, wheeze, cough +/- sputum production, termed "exacerbations". There is variability in the definition of acute exacerbations of COPD (AECOPD),[12] and definitions used in research may be symptom-based (relying on worsening of respiratory symptoms) or event-based (relying on hospitalisation, consultation with a General Practitioner (GP) or treatment with antibiotics or systemic steroids). Event-based definitions confer less ambiguity but would miss exacerbations that don't come to the attention of healthcare practitioners (HCPs) due to accessibility or healthcare seeking behaviours. Unlike the troponin level, which is used to detect myocardial infarction as the

"flare-up" of ischaemic heart disease, no biomarker that can detect AECOPD has been found, despite research to this end.[13] Diagnosis, therefore, may involve some subjectivity.

Aetiology: Exacerbations are usually broadly categorised as infective or non-infective in aetiology, but more likely are a result of complex interaction, involving host factors in addition to external factors.[14] Non-infective exacerbations may be triggered in part by air pollutants, and studies have demonstrated association between levels of common air pollutants and frequency of hospitalisations with AECOPD.[15]

Epidemiology: AECOPDs are one of the commonest reasons for emergency hospital admission in the UK. NHS England reports 115,000 emergency admissions for AECOPD per year[16], and the inpatient management of these exacerbations comprise a substantial proportion of the healthcare costs associated with COPD.

In addition to impact on the healthcare service, AECOPDs are associated with a worse quality of life,[17,18] and non-respiratory morbidity such as cardiovascular events.[19,20] Reducing the rate of COPD exacerbations is therefore paramount and would confer both individual and population level benefits.

Some people with COPD experience more frequent exacerbations than others,[21] and are termed 'frequent exacerbators'. The proportion of frequent exacerbators has been seen to rise with each GOLD grade (see 1.1.5 for definitions).[21] However, since COPD with moderate airflow limitation is the more prevalent condition, the frequent exacerbators in this group (22%)[21] result in a greater overall burden than frequent exacerbators with very severe airflow limitation. Understanding the 'frequent exacerbator' phenotype is hugely important, due to the impact of exacerbations on the individual,[17] as well as on healthcare resource utilisation and cost. However, with no single agreed definition of how many exacerbations in any given time period are needed to be deemed a 'frequent exacerbator', comparability between studies is limited.

Management: Treatment of AECOPD varies with the severity of the exacerbation. GOLD employs an event-based classification for the severity of exacerbations, defining mild exacerbations as those treated with short-acting bronchodilators (SABDs) only, moderate as

those treated with SABDs plus antibiotics and/or oral corticosteroids, and severe as those requiring emergency department (ED) attendance or hospital admission.[1] National Institute for Health and Care Excellence (NICE) guidance summarises factors for clinicians to consider when deciding whether patients with AECOPD need management in hospital versus in the community.[22] Management of severe AECOPDs may additionally include nebulised bronchodilators, oxygen therapy if oxygen saturations are low, and non-invasive ventilation (NIV) if acute type II respiratory failure has developed. The small proportion of people who don't respond to these treatments may require invasive ventilation on Intensive Care Units, if that is deemed appropriate.

1.1.5 Diagnosis

Spirometry, a test of lung function that can detect the presence and severity of airflow limitation, is 'key' to diagnosing COPD.[1] A device called a spirometer is used to measure volumes of air moving in and out of the lungs at different times during inspiration and expiration. The Forced Expiratory Volume in 1 second (FEV1) is the volume of air expelled from the lungs in 1 second during a forced expiration, after maximal inspiration. The Forced Vital Capacity (FVC) is the total volume of air that can be expelled from the lungs during a forced manoeuvre after maximal inspiration.

Usually, an individual will be able to expel >70% of their FVC in the first second of a forced expiration. The airflow obstruction or limitation seen in COPD is diagnosed when the FEV1 to FVC ratio is <0.7. The degree of obstruction can be categorised by the FEV1 result, with FEV1 >80% predicted classified as mild (or GOLD grade 1[1]), FEV1 50-79% predicted (GOLD grade 2) as moderate, FEV1 30-49% predicted (GOLD grade 3) as severe, and FEV <30% predicted (GOLD grade 4) as very severe.

However, definitions of COPD in use are heterogeneous and may also include emphysema seen on computerised tomography (CT) scans, symptoms of chronic lung disease with a history of smoking, and a low FEV1 result alone. Some people living with COPD also remain undiagnosed.[23]

1.1.6 Phenotypes

The significant heterogeneity of COPD has been increasingly recognised in recent years and understanding clinically relevant COPD phenotypes is important in order to guide provision of tailored, effective treatment. The most recent treatment guidelines from the Global Initiative for Chronic Obstructive Lung Disease (GOLD)[1] and the UK's National Institute for Health and Care Excellence (NICE)[22] recognise the importance of distinguishing between phenotypes in their algorithms for choice of inhaled therapy. Clinical trials investigating the role of biologic therapies for certain COPD phenotypes are ongoing.[24]

However, there is heterogeneity in approaches to phenotype classification as well as in the COPD population itself. Further, phenotyping has, for the most part, been studied in high-income countries, though is likely to be of greater importance in LMICs where the causative agents are more varied.[6]

Attempts at phenotyping those with COPD began over 60 years ago, when groups of "pink puffers" and "blue bloaters" were described.[25–27] These descriptions continue to persist in medical textbooks,[28,29] though they are likely relevant only to the Caucasian population and are of no clinical utility.

Of more relevance is categorisation which takes place based on pathological features (presence of emphysema on radiological investigation), clinical features (exacerbation frequency, degree of breathlessness, body mass index (BMI), coexisting asthmatic features) and comorbidities (such as the metabolic syndrome and anxiety/depression). Exacerbation frequency, degree of breathlessness and presence/absence of asthmatic features all direct the choice of treatment in current guidelines.[22,30]

1.1.7 Management

The main aims of COPD management are to improve quality of life, reduce exacerbation rate and improve survival. Although there remains unmet need with regard to treatments conferring survival benefit (currently, the only interventions shown to reduce mortality are smoking cessation,[31] oxygen for patients with chronic hypoxaemia, and home NIV for those with type II respiratory failure[32]), there are a number of evidence-based interventions which reduce exacerbation rates and improve symptoms.[33]

Optimal care comprises a bundle of interventions including early smoking cessation, pulmonary rehabilitation, influenza and pneumococcal vaccination and inhaled therapy (which can take the form of long-acting muscarinic antagonists (LAMA), long-acting beta-2 agonists (LABA), LABA and LAMA in combination (LABA/LAMA), LABA in combination with inhaled corticosteroid (LABA/ICS), "triple therapy" (LABA/LAMA/ICS), short-acting muscarinic antagonists (SAMA) and short-acting beta-2 agonists (SABA). Some of the most cost-effective interventions are the non-pharmacological ones, but these have been found to be underused.[34]

National and International Guidelines

The way that common conditions are managed by clinicians is often steered by guidelines produced by national or international bodies, based on available evidence. However, research can be interpreted differently or given different weight by different bodies, which can result in discrepancies in their recommendations. In addition, due to the time between guideline updates and new evidence coming to light in the interim, regional guidelines may be produced which may differ by locality. In reality, clinicians also often prescribe out-with guidelines, perhaps based on their own clinical experience or due to guidelines which lack clarity. There is, therefore, variability in clinical practice.

With respect to COPD, guidelines by both the National Institute for Health and Care Excellence (NICE) and Global Initiative for Chronic Obstructive Lung Disease (GOLD) are widely used for COPD diagnosis and management in the UK.[35] NICE writes their guidance in the context of the UK's National Health Service and the GOLD guidelines are developed by an international committee of experts. Although broadly similar, there have been a number of differences in recommendations over the years. In addition, there remains a significant proportion of patients whose treatment does not conform to either guideline, such as patients with COPD prescribed inhaled corticosteroid as monotherapy,[36,37] despite this not being a licenced or recommended treatment option for many years.

National COPD Audit Programme[34]

A primary care snapshot audit for COPD care in Wales prepared by the Royal College of Physicians, Primary Care Respiratory Society UK and Royal College of General Practitioners

extracted data from January 2014 – March 2015 from the 61% of practices across Wales who agreed to contribute. The results raised questions about accuracy of the diagnosis of COPD (post-bronchodilator spirometry showed results inconsistent with COPD in 26.9% of those who had a diagnosis of COPD), use of resources (overuse of expensive therapies, and under-delivery of more cost-effective measures), and the coding of data.

Recommendations regarding standardisation of coding were made, such as use of standard recording templates, to make it easier for HCPs to record important aspects of care and save time by reducing manual entry.

1.2 ROUTINELY COLLECTED HEALTH DATA

1.2.1 Overview

Research using routinely collected data from electronic health records (EHR) and administrative databases has seen an upsurge in recent years as the abundance of data accumulated as a by-product of routine clinical care has made large, heterogeneous populations accessible to researchers.

The unselected nature of populations accessible via EHR can give researchers access to 'real-world' populations who may be ineligible for inclusion in clinical trials, and longitudinal records enables analysis of outcomes over longer time periods than would otherwise be feasible and affordable.

However, as with any research, there are limitations which may impact on validity of analyses, and some of these are briefly outlined with respect to EHR research below.

1.2.2 Identifying common health conditions in Electronic Health Records (EHR)

To ensure that the results of studies using routinely collected data are valid, the case definitions used must be able to accurately detect individuals with the condition in question. However, given that EHRs and administrative databases are not populated with the purpose of use in research, factors such as missing information, data quality and misclassification can affect the validity of findings.

Studies using EHR often use code sets or algorithms for case identification. Clear reporting of case definition and code sets used is imperative, since changes in the code sets used for case definition can influence important study outcomes.[38] Furthermore, variability in the case definitions used limits comparability between studies.[39–41] Researchers have advocated for the case definitions used by studies to be validated against reference standards,[42] but this does not always happen in practice.

The value of linking primary and secondary care data sources has been highlighted for acute events such as infection[43] and myocardial infarction,[44] with misclassification and underestimation of incidence if only one data source is used.

In Chapter 2, my systematic scoping review summarises the methods used to identify people with COPD in recent EHR-based research.

1.2.3 Limitations and quality assessment in EHR research

As well as the standard methodological issues that affect all research, there are a number of issues specific to EHR-based research. Several groups have produced quality standards to guide such "real world" research and its reporting.[45–47]

Completeness

As EHR data does not originate with the purpose of being used for research, what is or isn't included in EHRs is influenced by a number of factors. Codes in claims databases, such as those in the United States, are likely to reflect reimbursement practices as well as the clinical care provided. In the UK, the data contained in primary care records is affected by financial incentives offered by the Quality and Outcomes Framework.[48,49] These data, sufficient for clinical care, meeting targets, and billing purposes, may not contain the granularity needed to characterise populations, covariates, exposures and outcome measures in order to answer important research questions.

Although diagnoses are usually coded, and thus available in EHR data sources in this format, results of relevant investigations may not be, limiting the ability for researchers to corroborate diagnoses or adjust for confounders such as disease severity. Instead, proxy measures are often used by researchers, but the justification for measures chosen may be unclear.

Patient-reported outcome measures, patient preferences and the thought process of HCPs which may be available in free text are also often not captured by coding, so although inference can be made regarding patient or HCP behaviour, researchers using EHR databases cannot fully understand the nuance involved in decision making.

Although the UK can link healthcare datasets with demographic datasets containing indicators relating to the socioeconomic determinants of health by use of unique identifiers, not all countries have this capability,[50] rendering adjustment for this important confounder in statistical analyses impossible.

Even when important characteristics such as ethnicity and BMI[51,52] are available within a dataset, there can be a high level of missing data, which can be missing at random or there may be increased missingness among certain groups,[52] and this further hinders adjustment for confounding.

It has been suggested that in order to maximise the research potential from EHR data, the paradigm of evidence-based medicine should shift to 'evidence-generating medicine', with creation of a health system that is better able to learn from its own data.[53] In order to realise this, EHR data capture needs to be enhanced, for example by using standardised approaches to data collection of relevant endpoints. This could both aid clinicians and researchers and enhance the quality of patient care.

Chance

With access to a large volume of data, there can be the inclination to investigate associations with a large number of outcomes or perform an array of subgroup analyses. The likelihood of false-positive findings rises concomitantly with the number of tests conducted, and relevant outcomes and subgroup analysis should be pre-defined, have a logical basis and relate to *a priori* hypotheses.

Random error

As well as simple error in the choice of code used, there may be changes in code use and meaning over time, and this needs to be considered when researchers examine historical data.

Misclassification

(i) Case versus control status: We know that with many common health conditions, including COPD,[54] there is an undiagnosed, untreated population. There will also be a proportion of patients that are misdiagnosed as having a certain condition, and whereas in randomised controlled trials (RCTs), the stringent inclusion criteria will mean that this proportion is limited, by using routinely collected data, numbers will likely be higher. Estimates of sensitivity and specificity of case detection algorithms are therefore useful in providing context.

(ii) Exposure misclassification: EHR research presents a number of difficulties to contend with when assessing the effects of exposures. With regard to drug therapeutics, people may obtain prescriptions outwith the data source being examined, leading to incorrect assignment to the 'non-exposed' group. On the contrary, people may be prescribed medication that they do not take, leading to incorrect assignment to the exposed group. With regard to other exposures, such as smoking, people do not always accurately report their behaviours to healthcare professionals.

Non-differential misclassification of these groups can dilute the strength of associations, biasing them towards the null. However, systematic error in classification has the potential to falsely strengthen associations. Classification bias occurs when the likelihood of misclassification is different between exposed or outcome groups. For example, those exposed to a medication will be less likely to be misclassified with regard to diagnoses, as in order to be prescribed a medication, they have interacted with the healthcare system.[55]

Confounding

As patients are not randomised to comparison groups in observational studies as they are in RCTs, there is an increased risk of confounding influencing results. Confounders may be known or unknown, and how well EHR captures details of traditional confounders varies by dataset, as discussed above. Confounding by indication for treatment can hamper studies examining therapeutic effects, if choice of therapy is affected by a factor (e.g. disease severity) which also influences the outcome of interest. Statistical approaches help can help to address these challenges of confounding by indication, as well as measured and unmeasured confounding.[56]

1.3 THESIS AIMS, RESEARCH QUESTIONS, AND OBJECTIVES

1.3.1 Aims

This thesis aims to explore the utility of linking pathology data to more standard EHR datasets in order to answer important research questions, using the COPD population as a case study.

1.3.2 Research questions

Over the course of the thesis, I will investigate the following research questions:

- How were individuals with COPD identified in EHR in the recent literature? Were methods of case identification validated against reference standards? How were COPD severity and phenotypes defined in EHR-based research? [Chapter 2]
- How many emergency admissions for COPD are with associated testing for, and detection of a respiratory pathogen? How accurately can diagnosis codes used in EHRs identify laboratory-confirmed respiratory pathogens associated with COPD exacerbations? [Chapter 4]

1.3.3 Objectives

In order to answer the above research questions, my objectives were to:

- Conduct a systematic, scoping review examining methods of identifying individuals with COPD in EHR
- Construct a dataset of hospital admissions linked to pathology data, to describe the testing and detection of respiratory pathogens associated with emergency admissions for COPD

1.4 THESIS OUTLINE

The remainder of my thesis is structured as follows:

- Chapter 2 is a systematic, scoping review in which I summarise the methods used to identify and phenotype people with COPD in recent EHR-based research. This is presented as a research paper and has been published in a peer-reviewed journal.

- In Chapter 3, I describe the construction of a dataset in which to examine infections associated with emergency hospital admissions for COPD in the SAIL databank.
- In Chapter 4, I use the dataset described in the previous chapter to estimate the contribution of specific respiratory pathogens to emergency admissions for COPD exacerbations. I also estimate the accuracy of using ICD codes to identify COPD admissions associated with laboratory-confirmed respiratory pathogens. This chapter is presented as a research paper and is currently under review at a peer-reviewed journal.
- Lastly, in Chapter 5, I reflect on the main findings from this thesis, as well as important opportunities for further work that linking routinely collected microbiology data with other electronic health data presents.

2 IDENTIFYING COPD IN ROUTINELY COLLECTED ELECTRONIC HEALTH RECORDS: A SYSTEMATIC SCOPING REVIEW

2.1 RESEARCH PAPER

This paper was originally published in ERJ Open Research, and is available here:

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 Sep 13;7(3):00167-2021. doi: 10.1183/23120541.00167-2021. PMID: 34527726; PMCID: PMC8435805.

Identifying COPD in routinely collected electronic health records: A systematic scoping review

Take-home message: Inconsistency in methods of identifying COPD in electronic health records and the lack of clinically important variables in healthcare databases widely used for research are persisting constraints in harnessing the potential of EHR worldwide.

Authors: Shanya Sivakumaran, Mohammad A. Alsallakh, Ronan A. Lyons, Jennifer K. Quint, Gwyneth A. Davies.

Abstract: Although routinely collected electronic health records (EHR) are widely used to examine outcomes related to chronic obstructive pulmonary disease (COPD), consensus regarding the identification of cases from electronic healthcare databases is lacking. We systematically examine and summarise approaches from the recent literature.

MEDLINE via EBSCOhost was searched for COPD-related studies using EHR published from 1 January 2018 to 30 November 2019. Data were extracted relating to the case definition of COPD and determination of COPD severity and phenotypes.

From 185 eligible studies, we found widespread variation in the definitions used to identify people with COPD in terms of code sets used (with 20 different code sets in use based on the ICD-10 classification alone) and requirement of additional criteria (relating to age (n=139), medication (n=31), multiplicity of events (n=21), spirometry (n=19) and smoking status (n=9)). Only 7 studies used a case definition which had been validated against a reference standard in the same dataset. Various proxies of disease severity were used since spirometry results and patient-reported outcomes were not often available.

To enable the research community to draw reliable insights from electronic health records and aid comparability between studies, clear reporting and greater consistency of the definitions used to identify COPD and related outcome measures is key.

Introduction

Chronic obstructive pulmonary disease (COPD) is a common, chronic condition characterised by persistent respiratory symptoms and irreversible expiratory airflow limitation, usually caused by chronic exposure to inhaled noxious substances. In clinical practice, COPD can be diagnosed in patients suspected to have the condition by use of spirometry, which also aids in the assessment of disease severity. Patients with COPD can also be grouped by their pattern of exacerbations and symptom burden, and this phenotyping guides treatment decisions [1]. Guidelines produced by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) divide patients into 'GOLD groups' based on these characteristics when delineating treatment strategies.

Research using routinely collected data from electronic health records (EHR) and administrative databases to study COPD has seen an upsurge in recent years, as the wealth of data accumulated as a by-product of routine clinical care has made large, diverse populations accessible to researchers. However, this data has not been generated for the purpose of research, and important information such as spirometry results and patient-reported outcome measures are not often accessible in the data sources. Alternative measures are thus frequently used to identify individuals with COPD, as well as determine disease severity and phenotypes, though the extent to which the definitions used have been assessed for validity is unclear.

Although the need to focus on the accuracy of case definitions has been emphasised [2], there is still significant heterogeneity in the definitions used to identify common conditions in routinely collected data [3–5]. In this systematic scoping review, we sought to summarise the range of methods used to identify COPD, its severity and phenotypes in EHR, and determine what proportion of case definitions in use have been validated against reference standards.

Methods

We conducted a systematic scoping review [6] to answer our research questions: 1) how were individuals with COPD identified in EHR in the recent literature, 2) how many methods of case identification had been validated against reference standards, 3) how were COPD severity and phenotypes defined, and 4) what important data are missing from the data sources used in these studies?

Search strategy and eligibility criteria

A broad search strategy was developed to gather studies which used EHR to identify individuals with COPD (supplementary material S1). MEDLINE via EBSCOhost was searched 15th January 2020 for articles published between 1 January 2018 and 30 November 2019. Our search was limited to those written in the English language. There were no limitations as to study design.

EHR included routinely collected, individual level data from administrative databases, disease registries, electronic health records, and any other electronic databases that were generated as a by-product of routine healthcare. Studies using solely survey or trial data were excluded, along with studies not reporting original data. We included studies identifying a study population of individuals with COPD or using COPD as a primary outcome, but not those where COPD was just contained in a list of covariates.

Study selection and data extraction

Articles that did not fit the above eligibility criteria were excluded. Screening was initially conducted using titles and abstracts, and full-texts were accessed when necessary. Information extracted from articles deemed eligible for inclusion related to core study details, definitions of COPD diagnosis, severity and phenotypes, and quality appraisal (supplementary material S2).

Article screening and data extraction were performed independently by two authors (S.S. and M.A.A.) for 20% of studies. S.S. then completed screening and extraction, with discussion with the wider study group when necessary.

Results

Our search strategy identified 1226 articles for screening (supplementary material S1), of which 189 met our eligibility criteria. We were able to access the full text for 185 of these, which are included in our review. Most studies were conducted in North America, Taiwan, and the United Kingdom (UK) (supplementary material S3). We included studies with a range of designs, including retrospective cohort, self-controlled case series, quasi-experimental, nested case control, case crossover and descriptive/exploratory studies.

Identifying COPD

Studies often identified individuals with COPD using clinical codes. The most frequently used coding scheme was the International Statistical Classification of Diseases and Related Health Problems (ICD) [7], either alone or in conjunction with other coding schemes (supplementary material S4). However, studies using the same coding scheme did not always use the same list of codes ('code set') from within the scheme – 57 studies incorporating ICD-10 based code used 20 different code sets to detect COPD (supplementary material S5). Some studies did not report the specific code set used for case identification.

In order to achieve greater accuracy of case definitions, many studies used additional inclusion and exclusion criteria in their definition of COPD. 139 (75%) used age as a criterion, with the lower age limit varying from 18 to 66 (supplementary material S6). Some studies (21, (11%)) required multiple COPD-related event or claim codes. Some (25, (14%)) gave more weight to inpatient care codes, requiring multiple COPD-related codes if arising from primary care or outpatient care, but only one if arising from inpatient care. 19 (10%) mandated presence of a spirometry code, but results of spirometry were not always taken into account. 9 (5%) specified ever-smoking as a criterion (supplementary material S7). A COPD-related medication code was required by 31/171 (18%) studies (not including studies whose aim was to investigate COPD medications, since these would have automatically required presence of the medication). The specific medication requirements and reporting of this varied by study, from requiring 'the prescription of at least one bronchodilator'[8] to mandating a greater frequency of medication use, with 'COPD medication use at least twice per year'[9] (where COPD medications were 'long-acting muscarinic antagonist, long-acting beta-2 agonist (LABA), inhaled corticosteroid (ICS), ICS plus LABA, short-active muscarinic antagonist (SAMA), short-acting beta-2 agonist (SABA), SAMA plus SABA, methylxanthines, systemic

corticosteroids, and systemic beta agonists')[9]. One study stated only that patients were required to have been prescribed a 'respiratory medication' but did not elaborate further[10].

With regard to exclusion criteria, 25 (14%) studies excluded those with a previous asthma diagnosis, some excluded additional comorbid respiratory conditions, and a few excluded individuals using specific medications, such as leukotriene receptor antagonists which are mostly used in people with asthma.

COPD severity

Spirometry was utilised by 25 (14%) studies in their assessment of COPD severity, 8 (4%) as a binary measure, 17 (11%) as an ordinal measure. Most studies did not assess disease severity in any form, specifying that this was because they lacked the clinical data necessary. Proxies of severity were sometimes used, ranging from chronic medication use (n=16 (9%)) and measures relating to exacerbations (n=16 (9%)), to serum bicarbonate levels (n=1 (0.5%)) [11], to algorithms purporting to represent 'complexity' (n=1 (0.5%)) [12].

COPD phenotypes

Coexisting asthma was a phenotype examined by 15 (8%) studies and was generally identified by presence of a previous asthma diagnosis code, but the specific code sets or identification methods were not always reported. 12 (6%) studies compared those with high versus low blood eosinophil counts or concentrations, although the thresholds used to determine high and low differed by study (supplementary material S8) [13–15]. 3 studies performed sensitivity analyses to examine the effect of using differing thresholds [16–18]. 8 (4%) studies categorised individuals by their GOLD groups (i.e. taking into account both exacerbation history and symptom burden), and 18 (10%) examined individuals by exacerbation history. Again, there was variation in the code sets or algorithms used to identify an exacerbation, and the thresholds for high versus low exacerbators.

Validation of case definitions

8 (4%) studies in our review compared their definition of COPD against a reference standard and provided sufficient information that a measure of validity could be calculated, although this may not have been their primary purpose. 2 of these studies identified themselves as 'validation studies' and went on to report measures of validity [19, 20].

Of the remaining 177 studies, a further 7 (4%) used a definition of COPD that had been previously validated against a reference standard in the same database used for their research. Additional studies referred to their case definition being 'based on' validated definitions, but used code sets different to those validated [21–23], or did not report the codes they used [18, 24, 25].

Some studies conducted analyses to justify the validity of their findings in different ways, such as performing sensitivity analyses using different definitions of COPD [26].

Reporting

Only one study referred to the REporting of studies Conducted using Observational Routinely collected Data (RECORD) guidance [27]. 15 (8%) studies stated that they used a particular coding scheme to identify people with COPD but did not report the code set used. 107 (58%) studies did not report whether they could access data related to (one of) smoking or spirometry. 44 (24%) reported that smoking information was not available within their data source (supplementary material S7), and 60 (32%) reported that spirometry events were unavailable.

Discussion

Principal findings

Electronic databases of routinely collected health data are used internationally to advance knowledge about 'real world' COPD by the research community. This systematic scoping review has demonstrated significant variability in the methods used by researchers to identify individuals with COPD and describe disease severity and phenotypes using routine data.

Only a limited number of studies used definitions that had been validated against reference standards in the same database used for their study. Some studies referred to previous validation studies, but as they did not report the code list they had used, it was not clear whether they used the same validated case definition. The RECORD guidance [27] advocates for provision of a 'complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers' in order to enhance research transparency.

Interpretation and implications

Datasets generated as a by-product of routine clinical care are increasingly important and useful in research, given the size, heterogeneity, and unselected nature of the populations they provide access to. However, there are pitfalls to their use, and among them is the use of case definitions with unknown validity. Limitations in the reporting of code sets used by researchers further hinder comparability and reproducibility.

For EHR to provide meaningful insights, the case definitions used must be able to accurately detect individuals with the condition in question. One way of ensuring this is to use definitions validated against a reference standard. However, a definition validated in one database may not be transferable for use in others. One validated definition for COPD in the UK's Clinical Practice Research Datalink [28] specifies being a current or ex-smoker as an inclusion criterion, given that COPD is uncommon in never-smokers in the UK. However, in countries with a higher contribution of alternative risk factors to the development of COPD [29], necessitating being an ever-smoker would likely reduce the sensitivity of this definition, as will happen in all geographies where smoking prevalence is falling. Additionally, different research questions may necessitate different case definitions – if investigators wanted to prioritise specificity over sensitivity, a more restrictive definition would be used, and vice versa – but clarity in the rationale would be useful to readers.

In addition to case definitions, the availability and accuracy of disease severity and phenotyping measures is imperative for studies to be able to adequately adjust for potential confounding. However, many administrative databases do not contain clinical detail at this level, so attempts at adjusting for severity in analyses often use proxy measures, the choice of which may be determined by data availability and not be validated against 'true' disease severity. More often, no attempt at adjusting for disease severity is made, leading to the potential of unmeasured confounding influencing results. Facilitating inclusion of clinically important variables (such as common investigation results and patient-reported outcome measures) into electronic health databases commonly used for research would be a useful and important intervention in improving research outputs. Through the lens of COPD, inclusion of spirometry results and UK Medical Research Council (MRC) dyspnoea scale scores would play a significant role in advancing research in the field. Similarly, inclusion of echocardiogram results and New York Heart Association (NYHA) class would likely be helpful for cardiovascular disease research. However, even when databases do hold such information,

there may be a high rate of missing data (e.g., 65% of patients in one study had no spirometry recorded) [30]. This reflects real world patterns and levels of missingness are likely to vary geographically due to historic differences in clinical practice, or national incentive schemes.

Strengths and limitations

This is the first review, to our knowledge, to systematically examine methods of identifying individuals with COPD in routinely collected electronic health records. This approach has allowed us to objectively demonstrate the variability in research practice in the field. We applied broad inclusion criteria ensuring representation across the research field, but confined our review to recent literature in order to ensure relevancy. We did not include studies where COPD was only relevant due to being contained in a list of covariates (as is often done, for example as part of the Charlson Comorbidity Index) [31] since an accurate case definition for this purpose holds less importance. However, this means that our review does not fully encompass the whole spectrum of the use of 'COPD' in electronic health data research.

Conclusions

Although the interrogation of routinely collected electronic health records is now commonplace in investigating important research questions related to COPD, and provides huge value when used carefully, persistent limitations constrain the quality of this research. The lack of clinically important variables in widely used databases limits researchers' ability to adjust for confounders such as disease severity. Variation in methods to identify COPD and define outcome measures restricts comparability between studies. With the contribution of EHR to COPD research continuing to increase internationally, ensuring greater consistency of case definitions and optimisation of reporting is key to enhancing the reliability of research outputs.

Footnotes

Ethics:

Ethical review was not required since this study was a review of previously published work.

Funding:

This study was funded by Swansea University Medical School with the support of BREATHE – The Health Data Research Hub for Respiratory Health [MC_PC_19004], which is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK.

References

1. Singh D, Agusti A, Anzueto A, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease: the GOLD science committee report 2019. *The European respiratory journal*. 2019;53(5). doi:10.1183/13993003.00164-2019

2. Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. *BMJ*. 2010;341(19). doi:10.1136/bmj.c4226

3. Crossfield SSR, Lai LYH, Kingsbury SR, et al. Variation in methods, results and reporting in electronic health record-based studies evaluating routine care in gout: A systematic review. *PLoS ONE*. 2019;14(10). doi:10.1371/journal.pone.0224272

4. al Sallakh MA, Vasileiou E, Rodgers SE, Lyons RA, Sheikh A, Davies GA. Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. *The European respiratory journal*. 2017;49(6):1700204. doi:10.1183/13993003.00204-2017

5. Rubbo B, Fitzpatrick NK, Denaxas S, et al. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: A systematic review and recommendations. *International Journal of Cardiology*. 2015;187:705-711. doi:10.1016/j.ijcard.2015.03.075

6. Arksey H, O'Malley L. Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology: Theory and Practice*. 2005;8(1):19-32. doi:10.1080/1364557032000119616

7. World Health Organisation. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. 2010 Edition.; 2011. Accessed January 20, 2021. www.who.int

8. Lai CC, Wu CH, Wang YH, Wang CY, Wu VC, Chen L. The association between COPD and outcomes of patients with advanced chronic kidney disease. *International Journal of COPD*. 2018;13:2899-2905. doi:10.2147/COPD.S174215

9. Park SC, Kim YS, Kang YA, et al. Hemoglobin and mortality in patients with COPD: A nationwide population-based cohort study. *International Journal of COPD*. 2018;13:1599-1605. doi:10.2147/COPD.S159249

10. Huang HH, Chen SJ, Chao TF, et al. Influenza vaccination and risk of respiratory failure in patients with chronic obstructive pulmonary disease: A nationwide population-based case-cohort study. *Journal of Microbiology, Immunology and Infection*. 2019;52(1):22-29. doi:10.1016/j.jmii.2017.08.014

11. Vazquez Guillamet R, Ursu O, Iwamoto G, Moseley PL, Oprea T. Chronic obstructive pulmonary disease phenotypes using cluster analysis of electronic medical records. *Health Informatics Journal*. 2018;24(4):394-409. doi:10.1177/1460458216675661

12. Bishwakarma R, Zhang W, Li YL, Kuo YF, Cardenas VJ, Sharma G. Metformin use and health care utilization in patients with coexisting chronic obstructive pulmonary disease and diabetes mellitus. *International Journal of COPD*. 2018;13:793-800. doi:10.2147/COPD.S150047

13. Choi J, Oh JY, Lee YS, et al. The association between blood eosinophil percent and bacterial infection in acute exacerbation of chronic obstructive pulmonary disease. *International Journal of COPD*. 2019;14:953-959. doi:10.2147/COPD.S197361

14. Whittaker HR, Müllerova H, Jarvis D, et al. Inhaled corticosteroids, blood eosinophils, and fev1 decline in patients with COPD in a large UK primary health care setting. *International Journal of COPD*. 2019;14:1063-1073. doi:10.2147/COPD.S200919

15. Hamad GA, Cheung W, Crooks MG, Morice AH. Eosinophils in COPD: How many swallows make a summer? *European Respiratory Journal*. 2018;51(1). doi:10.1183/13993003.02177-2017

16. Li Q, Larivée P, Courteau J, et al. Greater eosinophil counts at first COPD hospitalization are associated with more readmissions and fewer deaths. *International Journal of COPD*. 2019;14:331-341. doi:10.2147/COPD.S187375

17. Müllerová H, Hahn B, Simard EP, Mu G, Hatipoğlu U. Exacerbations and health care resource use among patients with COPD in relation to blood eosinophil counts. *International Journal of COPD*. 2019;14:683-692. doi:10.2147/COPD.S194367

18. Müllerová H, Meeraus WH, Galkin D v., Albers FC, Landis SH. Clinical burden of illness among patients with severe eosinophilic COPD. *International Journal of COPD*. 2019;14:741-755. doi:10.2147/COPD.S194511

19. Ho TW, Ruan SY, Huang CT, Tsai YJ, Lai F, Yu CJ. Validity of ICD9-CM codes to diagnose chronic obstructive pulmonary disease from National Health Insurance claim data in Taiwan. *International Journal of COPD*. 2018;13:3055-3063. doi:10.2147/COPD.S174265

20. Su VYF, Yang KY, Yang YH, et al. Use of ICS/LABA Combinations or LAMA Is Associated with a Lower Risk of Acute Exacerbation in Patients with Coexistent COPD and Asthma. *Journal of Allergy and Clinical Immunology: In Practice*. 2018;6(6):1927-1935.e3. doi:10.1016/j.jaip.2018.01.035

21. Lawson CA, Mamas MA, Jones PW, et al. Association of Medication Intensity and Stages of Airflow Limitation With the Risk of Hospitalization or Death in Patients With Heart Failure and Chronic Obstructive Pulmonary Disease. *JAMA network open*. 2018;1(8):e185489. doi:10.1001/jamanetworkopen.2018.5489

22. To T, Zhu J, Gray N, et al. Asthma and chronic obstructive pulmonary disease overlap in women incidence and risk factors. *Annals of the American Thoracic Society*. 2018;15(11):1304-1310. doi:10.1513/AnnalsATS.201802-078OC

23. Chalmers JD, Poole C, Webster S, Tebboth A, Dickinson S, Gayle A. Assessing the healthcare resource use associated with inappropriate prescribing of inhaled corticosteroids for people with chronic obstructive pulmonary disease (COPD) in GOLD groups A or B: An observational study using the Clinical Practice Research Datalink (CPRD). *Respiratory Research*. 2018;19(1). doi:10.1186/s12931-018-0767-2

24. Oshagbemi OA, Keene SJ, Driessen JHM, et al. Trends in moderate and severe exacerbations among COPD patients in the UK from 2005 to 2013. *Respiratory Medicine*. 2018;144:1-6. doi:10.1016/j.rmed.2018.09.010

25. Landis S, Suruki R, Maskell J, Bonar K, Hilton E, Compton C. Demographic and Clinical Characteristics of COPD Patients at Different Blood Eosinophil Levels in the UK Clinical Practice Research Datalink. *COPD: Journal of Chronic Obstructive Pulmonary Disease*. 2018;15(2):177-184. doi:10.1080/15412555.2018.1441275

26. Mcguire K, Aviña-Zubieta JA, Esdaile JM, et al. Risk of Incident Chronic Obstructive Pulmonary Disease in Rheumatoid Arthritis: A Population-Based Cohort Study. *Arthritis Care and Research*. 2019;71(5):602-610. doi:10.1002/acr.23410

27. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Medicine*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885

28. Quint JK, Müllerova H, DiSantostefano RL, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). *BMJ Open*. 2014;4(7). doi:10.1136/bmjopen-2014-005540

29. Raherison C, Girodet PO. Epidemiology of COPD. *European Respiratory Review*. 2009;18(114):213-221. doi:10.1183/09059180.00003609

30. Ohar JA, Loh CH, Lenoir KM, Wells BJ, Peters SP. A comprehensive care plan that reduces readmissions after acute exacerbations of COPD. *Respiratory Medicine*. 2018;141:20-25. doi:10.1016/j.rmed.2018.06.014

31. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *Journal of Clinical Epidemiology*. 1994;47(11):1245-1251. doi:10.1016/0895-4356(94)90129-5

2.2 SUPPLEMENTARY MATERIAL

Table S1: Search strategy

#	Query	Results
S1	(MH "Medical Records Systems, Computerized+") OR (MH "Medical Record Linkage") OR (MH "Electronic Health Records+") OR (MH "Clinical Coding") OR (MH "International Classification of Diseases")	50,268
S2	(MH "Databases as Topic+")	151,665
S3	medical record linkage or medical records linkage or clinical coding or code or codes or coding or codelist* or codeset* or algorithm* or International Classification of Disease* or ICD* or ICD-10 or ICD-9 or ICD-9-CM or ICD 9 or ICD 10	712,365
S4	emr or electronic medical record* or ehr or electronic health record*	44,324
S5	(data* or record*) and (insurance or claim* or administrative or routine* or electr* or digit* or computer* or linked)	1,667,252
S6	S1 OR S2 OR S3 OR S4 OR S5	2,297,452
S7	(MH "Pulmonary Disease, Chronic Obstructive+")	53,773
S8	copd or chronic obstructive pulmonary disease or coad or chronic obstructive airway disease or chronic obstructive lung disease or emphysema or chronic bronchitis or chronic airflow obstruction* or airflow obstruction, chronic or chronic airway obstruction* or airway obstruction, chronic	114,337
S9	S7 OR S8	114,337
S10	S6 AND S9 Limiters – Date of Publication: 20180101-20191130; English Language; Human	1,226

Table S2: Data extraction table

Variable	Variable value options
Core study details	
PubMed ID	Free text
Article title	Free text
Study location (country)	Free text
Study design	Free text
Data source	Free text
Study aim	Free text
Definition of COPD	
Inclusion criteria notes	Free text
Coding scheme used	ICD-9, ICD-9-CM, ICD-10, ICD-10-CM, ICD-10-CA, Read code, DRG, ATC, Other code, Mix of above codes, No codes reported
Code set used	Free text
Age as criterion?	Y/N
Age - lower limit	Free text
Requirement of multiple claims?	Y/N
More weight to inpatient claims?	Y/N
Requirement of specific treatment?	Y/N
Requirement of spirometry?	Yes spirometry required but results not specified, Yes spirometry required and results specified, No
Requirement of ever-smoker?	Y/N
Exclusion criteria notes	Free text
Exclusion criteria: are patients with asthma excluded?	Y/N

Definitions of COPD severity	
COPD severity notes	Free text
Spirometry- binary	Y/N
Spirometry- ordinal	Y/N
Related to chronic medication use?	Y/N
Related to exacerbations?	Y/N
Other proxies of severity	Free text
Definitions of COPD phenotype	
COPD phenotype notes	Free text
Co-existing asthma	Y/N
Blood eosinophil level	Y/N
Exacerbators	Y/N
GOLD groups	Y/N
Other	Free text
Quality appraisal	
Does the study validate definitions used?	Y/N
Have the definitions used been validated previously in the same dataset used for the study?	Y/N
Other justification of validity?	Free text
Missing data: smoking	Smoking data missing, smoking data not missing, not reported
Missing data: spirometry	Spirometry data missing, spirometry data not missing, not reported
Missing data: other	Free text
Other limitations and measure to minimise limitations	Free text

Reference to RECORD statement	Y/N

Table S3: Geographical distribution

Country	Number of studies
USA	52
Taiwan	29
United Kingdom	20
Canada	15
Spain	11
Korea	9
China	7
Italy	6
Sweden	5
Denmark	4
Netherlands	4
Hong Kong	3
Australia	3
Israel	3
Belgium	2
Poland	2
Austria	2
France	2
Scotland	1
Norway and Germany	1
Ireland	1
Singapore	1
Germany	1
Turkey	1

Table S4: Coding schemes used by studies

Coding scheme	Number of studies
ICD-9-CM	52
ICD-10	35
No codes reported	28
Mix of other named categories	25
ICD-9	23
Read code	15
Other codes*	4
ICD-10-CM	1
ICD-10-CA	1
N/A	1

ICD 10 = the International Statistical Classification of Diseases and Related Health Problems, 10th revision, the replacement of ICD-9. ICD-10-CM = clinical modification of the classification, replacing ICD-9-CM. ICD-10-CA is the Canadian modification.

*'Other codes' includes the Anatomical Therapeutic Chemical Classification System (ATC), Diagnosis-related Group codes (DRG), International Classification of Primary Care (ICPC)

Table S5: ICD-10 code sets used to identify COPD

ICD-10 code sets	Number of studies
J44	14
J41-44	10
J40-44	7
J41, J43-44	5
J43, J44	3
J42-44	2
J43-44, except J43.0	2
J44.1	2
J40, J41.0, J41.1, J41.8, J42, J43.0, J43.1, J43.2, J43.8, J43.9, J44.0, J44.1, J44.9	1
J40, J47	1
J40-42, J44	1
J40-44, J47	1
J41, J43	1
J41.0, J41.1, J41.8, J42, J43.9, J44.0, J44.1, J44.9	1
J41-43	1
J41-44, J47	1
J42, 44	1
J42-44, except J43.0	1

J44.0, J44.1, J44.9	1
J44.1, J44.8, J44.9	1

Includes studies which used ICD-10, ICD-10-CM or ICD-10-CA alone, or in combination with other coding schemes.

Table S6: Lower age limit used in definition of COPD

Lower age limit	Number of studies
40	76
35	13
18	12
20	8
45	6
65	6
66	5
55	5
50	4
30	2
25	1
60	1

Table S7: Smoking data

Is smoking data missing?*	**Number of studies**
Yes	44
No	66
Not reported	75
If study has access to smoking data, are never-smokers included in the analysis?	
Yes	33
No	9
Not reported	24

*Missing here means that information regarding smoking was not available in the database being interrogated

Table S8: Phenotyping by blood eosinophil level – thresholds used by studies

Blood eosinophil threshold	Number of studies
150 cells/μL	4
2% (of total white cell count)	3
300 cells/μL	2
200 cells/μL and/or 2%	1
'Always above', 'fluctuating above and below', and 'never above' cut off points of 100, 150, and 300 cells/μL	1
<2%, 2-4%, >4% and 150, 150–300, 300 cells per μL	1

% = blood eosinophil concentrations as percentage of total white blood cell count; cells/μL = absolute count of blood eosinophils

List of studies included in scoping review

(1) Abad-Arranz M, Moran-Rodríguez A, Mascarós Balaguer E, Quintana Velasco C, Abad Polo L, Núñez Palomo S, et al. Community Assessment of COPD Health Care (COACH) study: A clinical audit on primary care performance variability in COPD care. BMC Medical Research Methodology 2018 /7;18(1).

(2) Aksoy E, Karakurt Z, Gungor S, Ocakli B, Ozmen İ, Yildirim E, et al. Neutrophil to lymphocyte ratio is a better indicator of COPD exacerbation severity in neutrophilic endotypes than eosinophilic endotypes. International Journal of COPD 2018;13:2721-2730.

(3) Amin AN, Bollu V, Stensland MD, Netzer L, Ganapathy V. Treatment patterns for patients hospitalized with chronic obstructive pulmonary disease. American Journal of Health-System Pharmacy 2018 /3;75(6):359-366.

(4) Ando T, Adegbala O, Akintoye E, Ashraf S, Pahuja M, Briasoulis A, et al. Is transcatheter aortic valve replacement better than surgical aortic valve replacement in patients with chronic obstructive pulmonary disease? A nationwide inpatient sample analysis. Journal of the American Heart Association 2018 /4;7(7).

(5) Annavarapu S, Goldfarb S, Gelb M, Moretz C, Renda A, Kaila S. Development and validation of a predictive model to identify patients at risk of severe COPD exacerbations using administrative claims data. International Journal of COPD 2018;13:2121-2130.

(6) Baliatsas C, Smit LAM, Dückers MLA, Van Dijk CE, Heederik D, Yzermans CJ. Patients with overlapping diagnoses of asthma and COPD: Is livestock exposure a risk factor for comorbidity and coexisting symptoms and infections? BMC Pulmonary Medicine 2019 /6;19(1).

(7) Barbosa-Lorenzo R, Ruano-Ravina A, Fernández-Villar A, López-Pardo E, Carballeira-Roca C, Barros-Dios J. COPD prevalence and hospital admissions in Galicia (Spain). An analysis using the potential of new health information systems. Pulmonology 2018 11;24(6):323-329.

(8) Bell CF, Coutinho AD, Farrelly E, Lokhandwala T, Landsman-Blumberg P. Clinical and economic outcomes associated with the use of fluticasone propionate 250 mcg and salmeterol 50 mcg combination versus tiotropium bromide 18 mcg as initial maintenance treatment for chronic obstructive pulmonary disease in managed care. Journal of Medical Economics 2018 /6;21(6):629-638.

(9) Bengtson LGS, DePietro M, McPheeters J, Fox KM. Real-world outcomes in patients with chronic obstructive pulmonary disease initiating long-acting mono bronchodilator therapy. Therapeutic Advances in Respiratory Disease 2018 /5;12.

(10) Bishwakarma R, Zhang W, Li YL, Kuo YF, Cardenas VJ, Sharma G. Metformin use and health care utilization in patients with coexisting chronic obstructive pulmonary disease and diabetes mellitus. International Journal of COPD 2018 /3;13:793-800.

(11) Bloom CI, Elkin SL, Quint JK. Changes in COPD inhaler prescriptions in the united kingdom, 2000 to 2016. International Journal of COPD 2019;14:279-287.

(12) Bloom CI, Slaich B, Morales DR, Smeeth L, Stone P, Quint JK. Low uptake of palliative care for COPD patients within primary care in the UK. European Respiratory Journal 2018;51(2).

(13) Bogart M, Glassberg MB, Reinsch T, Stanford RH. Impact of prompt versus delayed initiation of triple therapy post COPD exacerbation in a US-managed care setting. Respir Med 2018 12;145:138-144.

(14) Bogart M, Stanford RH, Laliberté F, Germain G, Wu JW, Duh MS. Medication adherence and persistence in chronic obstructive pulmonary disease patients receiving triple therapy in a USA commercially insured population. International Journal of COPD 2019;14:343-352.

(15) Bogart M, Stanford RH, Reinsch T, Hull M, Buikema A, Hulbert E. Clinical characteristics and medication patterns in patients with COPD prior to initiation of triple therapy with ICS/LAMA/LABA: A retrospective study. Respir Med 2018 /9;142:73-80.

(16) Broderick J, Mc Grath C, Cullen K, Talbot D, Gilmor J, Baily-Scanlan M, et al. Effects of pulmonary rehabilitation on exercise capacity and disease impact in patients with chronic obstructive pulmonary disease and obesity. Physiotherapy (United Kingdom) 2018 /6;104(2):248-250.

(17) Casas-Mendez F, Abadías MJ, Yuguero O, Bardés I, Barbé F, de Batlle J. Treatment strategies after acute exacerbations of chronic obstructive pulmonary disease: Impact on mortality. PLoS ONE 2018 12;13(12).

(18) Celli BR, Navaie M, Xu Z, Cho-Reyes S, Dembek C, Gilmer TP. Medication management patterns among Medicare beneficiaries with chronic obstructive pulmonary disease who initiate nebulized arformoterol treatment. International journal of chronic obstructive pulmonary disease 2019;14:1019-1031.

(19) Chalmers JD, Poole C, Webster S, Tebboth A, Dickinson S, Gayle A. Assessing the healthcare resource use associated with inappropriate prescribing of inhaled corticosteroids for people with chronic obstructive pulmonary disease (COPD) in GOLD groups A or B: An observational study using the Clinical Practice Research Datalink (CPRD). Respiratory Research 2018 /4;19(1).

(20) Chan HS, Ko FWS, Chan JWM, So LKY, Lam DCL, Chan VL, et al. Comorbidities, mortality, and management of chronic obstructive pulmonary disease patients who required admissions to public hospitals in Hong Kong – Computerized data collection and analysis. International Journal of COPD 2018 /6;13:1913-1925.

(21) Chen CY, Liao KM. The impact of atrial fibrillation in patients with COPD during hospitalization. International Journal of COPD 2018;13:2105-2112.

(22) Chen H, Li Q, Kaufman JS, Wang J, Copes R, Su Y, et al. Effect of air quality alerts on human health: a regression discontinuity analysis in Toronto, Canada. The Lancet Planetary Health 2018 /1;2(1):e2-e3.

(23) Chen HC, Wu CF, Chong IW, Wu MT. Exposure to cooking oil fumes and chronic bronchitis in nonsmoking women aged 40 years and over: A health-care based study. BMC Public Health 2018 /2;18(1).

(24) Chen KY, Wu SM, Liu JC, Lee KY. Effect of annual influenza vaccination on reducing lung cancer in patients with chronic obstructive pulmonary disease from a population-based cohort study. Medicine (United States) 2019 11;98(47).

(25) Chen W, Johnson KM, FitzGerald JM, Sadatsafavi M, Leslie WD. Long-term effects of inhaled corticosteroids on bone mineral density in older women with asthma or COPD: a registry-based cohort study. Archives of Osteoporosis 2018 12;13(1).

(26) Chen W, Sin DD, FitzGerald JM, Sadatsafavi M. The Impact of Care Specialty on Survival-Adjusted Medical Costs of COPD Patients After a Hospitalization: a longitudinal analysis. Journal of General Internal Medicine 2018 /9;33(9):1528-1535.

(27) Choi J, Oh JY, Lee YS, Hur GY, Lee SY, Shim JJ, et al. The association between blood eosinophil percent and bacterial infection in acute exacerbation of chronic obstructive pulmonary disease. International Journal of COPD 2019;14:953-959.

(28) Choi J, Oh JY, Lee YS, Hur GY, Lee SY, Shim JJ, et al. Pseudomonas aeruginosa infection increases the readmission rate of COPD patients. International Journal of COPD 2018;13:3077-3083.

(29) Choi J, Oh JY, Lee YS, Min KH, Hur GY, Lee SY, et al. Harmful impact of air pollution on severe acute exacerbation of chronic obstructive pulmonary disease: Particulate matter is hazardous. International Journal of COPD 2018 /3;13:1053-1059.

(30) Chung WS, Lin CL. Acute respiratory events in patients with bronchiectasis–COPD overlap syndrome: A population-based cohort study. Respir Med 2018 /7;140:6-10.

(31) Contoli M, Campo G, Pavasini R, Marchi I, Pauletti A, Balla C, et al. Inhaled corticosteroid/long-acting bronchodilator treatment mitigates STEMI clinical presentation in COPD patients. Eur J Intern Med 2018;47:82-86.

(32) de Miguel-Díez J, López-de-Andrés A, Esteban-Vasallo M, Hernández-Barrera V, de Miguel-Yanes JM, Méndez-Bailón M, et al. Clostridium difficile infection in hospitalized patients with COPD in Spain (2001–2015). Eur J Intern Med 2018 11;57:76-82.

(33) de Miguel-Díez J, López-De-andres A, Hernández-Barrera V, de Miguel-Yanes JM, Méndez-Bailón M, Jiménez-García R. Trends and hospital outcomes of lung transplantation among patients with and without chronic obstructive pulmonary disease in Spain: A national population-based study (2001–2015). International Journal of COPD 2019;14:729-737.

(34) de Miguel-Díez J, López-de-Andrés A, Hernández-Barrera V, de Miguel-Yanes JM, Méndez-Bailón M, Muñoz-Rivas N, et al. Influence of COPD on outcomes of patients

hospitalized with heart failure: Analysis of the Spanish National Hospital Discharge Database (2001–2015). Int J Cardiol 2018 10;269:213-219.

(35) de Miguel-Díez J, López-de-Andrés A, Hernández-Barrera V, Jiménez-Trujillo I, Méndez-Bailón M, de Miguel-Yanes JM, et al. Postoperative pneumonia among patients with and without COPD in Spain from 2001 to 2015. Eur J Intern Med 2018 /7;53:66-72.

(36) De Schreye R, Smets T, Deliens L, Annemans L, Gielen B, Cohen J. Appropriateness of End-of-Life Care in People Dying From COPD. Applying Quality Indicators on Linked Administrative Databases. J Pain Symptom Manage 2018 10;56(4):541-550.e6.

(37) Desai R, Patel U, Singh S, Bhuva R, Fong HK, Nunna P, et al. The burden and impact of arrhythmia in chronic obstructive pulmonary disease: Insights from the National Inpatient Sample. Int J Cardiol 2019 /4;281:49-55.

(38) Dziankowska-Zaborszczyk E, Bryla M, Ciabiada-Bryla B, Maniecka-Bryla I. Standard expected years of life lost (SEYLL) due to chronic obstructive pulmonary disease (COPD) in Poland from 1999 to 2014. PLoS ONE 2019 /3;14(3).

(39) Epstein D, Nasser R, Mashiach T, Azzam ZS, Berger G. Increased red cell distribution width: A novel predictor of adverse outcome in patients hospitalized due to acute exacerbation of chronic obstructive pulmonary disease. Respir Med 2018 /3;136:1-7.

(40) Faes K, Cohen J, Annemans L. Resource Use During the Last Six Months of Life Among COPD Patients: A Population-Level Study. J Pain Symptom Manage 2018 /9;56(3):318-326.e7.

(41) Flynn RWV, MacDonald TM, Chalmers JD, Schembri S. The effect of changes to GOLD severity stage on long term morbidity and mortality in COPD 11 Medical and Health Sciences 1102 Cardiorespiratory Medicine and Haematology. Respiratory Research 2018 12;19(1).

(42) Fu PK, Tung YC, Wang CY, Hwang SF, Lin SP, Hsu CY, et al. Early and late do-not-resuscitate (DNR) decisions in patients with terminal COPD: A retrospective study in the last year of life. International Journal of COPD 2018;13:2447-2454.

(43) Galvis JN, Vargas MV, Robinson HN, Tyan P, Gu A, Wei C, et al. Impact of chronic obstructive pulmonary disease on laparoscopic hysterectomy outcome. Journal of the Society of Laparoendoscopic Surgeons 2019 /1;23(1).

(44) Gao N, Li C, Ji J, Yang Y, Wang S, Tian X, et al. Short-term effects of ambient air pollution on chronic obstructive pulmonary disease admissions in Beijing, China (2013–2017). International Journal of COPD 2019;14:297-309.

(45) García-Talavera I, Figueira-Gonçalves JM, Gurbani N, Pérez-Méndez L, Pedrero-García A. Clinical characteristics of COPD patients with early-onset desaturation in the 6-minute walk test. Pulmonology 2018 /9;24(5):275-279.

(46) Gayle A, Dickinson S, Morris K, Poole C, Mathioudakis AG, Vestbo J. What is the impact of GOLD 2017 recommendations in primary care? - a descriptive study of patient

classifications, treatment burden and costs. International Journal of COPD 2018;13:3485-3492.

(47) Gedebjerg A, Szépligeti SK, Wackerhausen LMH, Horváth-Puhó E, Dahl R, Hansen JG, et al. Prediction of mortality in patients with chronic obstructive pulmonary disease with the new Global Initiative for Chronic Obstructive Lung Disease 2017 classification: a cohort study. The Lancet Respiratory Medicine 2018 /3;6(3):204-212.

(48) Geraets I, Schermer T, Kocks JWH, Akkermans R, Bischoff E, van den Bemt L. Primary care cohort study in the sequence of diagnosing chronic respiratory diseases and prescribing inhaled corticosteroids. npj Primary Care Respiratory Medicine 2018 12;28(1).

(49) Germini F, Veronese G, Marcucci M, Coen D, Ardemagni D, Montano N, et al. COPD exacerbations in the emergency department: Epidemiology and related costs. A retrospective cohort multicentre study from the Italian Society of Emergency Medicine (SIMEU). Eur J Intern Med 2018 /5;51:74-79.

(50) Gershon AS, Thiruchelvam D, Chapman KR, Aaron SD, Stanbrook MB, Bourbeau J, et al. Health Services Burden of Undiagnosed and Overdiagnosed COPD. Chest 2018 /6;153(6):1336-1346.

(51) Giannico OV, Ambrosino I, Patano F, Germinario C, Quarto M, Moretti AM. Educational level, marital status and sex as social gender discharge determinants in chronic obstructive pulmonary disease exacerbations: A time-to-event analysis. Monaldi Archives for Chest Disease 2019;89(2):7-14.

(52) Gonzalez AV, Coulombe J, Ernst P, Suissa S. Long-term Use of Inhaled Corticosteroids in COPD and the Risk of Fracture. Chest 2018 /2;153(2):321-328.

(53) Goto T, Camargo CA, Faridi MK, Yun BJ, Hasegawa K. Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED. Am J Emerg Med 2018 /9;36(9):1650-1654.

(54) Goto T, Faridi MK, Camargo CA, Hasegawa K. Time-varying readmission diagnoses during 30 days after hospitalization for COPD exacerbation. Med Care 2018 /8;56(8):673-678.

(55) Goto T, Hirayama A, Faridi MK, Camargo CA, Hasegawa K. Obesity and severity of acute exacerbation of chronic obstructive pulmonary disease. Annals of the American Thoracic Society 2018 /2;15(2):184-191.

(56) Goto T, Tsugawa Y, Faridi MK, Camargo CA, Hasegawa K. Reduced Risk of Acute Exacerbation of COPD After Bariatric Surgery: A Self-Controlled Case Series Study. Chest 2018 /3;153(3):611-617.

(57) Granfeldt A, Wissenberg M, Hansen SM, Lippert FK, Torp-Pedersen C, Skaarup SH, et al. Severity of chronic obstructive pulmonary disease and presenting rhythm in patients with out-of-hospital cardiac arrest. Resuscitation 2018 /5;126:111-117.

(58) Greiver M, Sullivan F, Kalia S, Aliarzadeh B, Sharma D, Bernard S, et al. Agreement between hospital and primary care on diagnostic labeling for COPD and heart failure in Toronto, Canada: a cross-sectional observational study. npj Primary Care Respiratory Medicine 2018 12;28(1):9-9.

(59) Gulati S, Zouk AN, Kalehoff JP, Wren CS, Davison PN, Kirkpatrick DP, et al. The use of a standardized order set reduces systemic corticosteroid dose and length of stay for individuals hospitalized with acute exacerbations of COPD: A cohort study. International Journal of COPD 2018;13:2147-2156.

(60) Hamad GA, Cheung W, Crooks MG, Morice AH. Eosinophils in COPD: How many swallows make a summer? European Respiratory Journal 2018;51(1).

(61) Heintzman J, Kaufmann J, Ezekiel-Herrera D, Bailey SR, Cornell A, Ukhanova M, et al. Asthma/COPD Disparities in Diagnosis and Basic Care Utilization Among Low-Income Primary Care Patients. Journal of Immigrant and Minority Health 2019 /6;21(3):659-663.

(62) Hirayama A, Goto T, Shimada YJ, Faridi MK, Camargo CA, Hasegawa K. Acute Exacerbation of Chronic Obstructive Pulmonary Disease and Subsequent Risk of Emergency Department Visits and Hospitalizations for Atrial Fibrillation. Circulation.Arrhythmia and electrophysiology 2018 /9;11(9):e006322-e006322.

(63) Ho TW, Ruan SY, Huang CT, Tsai YJ, Lai F, Yu CJ. Validity of ICD9-CM codes to diagnose chronic obstructive pulmonary disease from National Health Insurance claim data in Taiwan. International Journal of COPD 2018;13:3055-3063.

(64) Hong S, Park EC, Kim TH, Kwon JA, Yoo KB, Han KT, et al. Effect of pre existing respiratory conditions on survival of lung cancer patients: A nationwide population-based cohort study. Asia-Pacific Journal of Clinical Oncology 2018 /4;14(2):e71-e80.

(65) Hopke PK, Croft D, Zhang W, Lin S, Masiol M, Squizzato S, et al. Changes in the acute response of respiratory diseases to PM 2.5 in New York State from 2005 to 2016. Sci Total Environ 2019 /8;677:328-339.

(66) Hsu IL, Lu CL, Li CC, Tsai SH, Chen CZ, Hu SC, et al. Population-based cohort study suggesting a significantly increased risk of developing chronic obstructive pulmonary disease in people with type 2 diabetes mellitus. Diabetes Res Clin Pract 2018 /4;138:66-74.

(67) Hu WS, Lin CL. CHA 2 DS 2 -VASc score for ischaemic stroke risk stratification in patients with chronic obstructive pulmonary disease with and without atrial fibrillation: A nationwide cohort study. Europace 2018 /4;20(4):575-581.

(68) Huang HH, Chen SJ, Chao TF, Liu CJ, Chen TJ, Chou P, et al. Influenza vaccination and risk of respiratory failure in patients with chronic obstructive pulmonary disease: A nationwide population-based case-cohort study. Journal of Microbiology, Immunology and Infection 2019 /2;52(1):22-29.

(69) Humenberger M, Horner A, Labek A, Kaiser B, Frechinger R, Brock C, et al. Adherence to inhaled therapy and its impact on chronic obstructive pulmonary disease (COPD) 11

Medical and Health Sciences 1102 Cardiorespiratory Medicine and Haematology. BMC Pulmonary Medicine 2018 10;18(1).

(70) Hurst JR, Dilleen M, Morris K, Hills S, Emir B, Jones R. Factors influencing treatment escalation from long-acting muscarinic antagonist monotherapy to triple therapy in patients with COPD: A retrospective thin-database analysis. International Journal of COPD 2018 /3;13:781-792.

(71) Inabnit LS, Blanchette C, Ruban C. Comorbidities and length of stay in chronic obstructive pulmonary disease patients. COPD: Journal of Chronic Obstructive Pulmonary Disease 2018 /7;15(4):355-360.

(72) Jacobs DM, Noyes K, Zhao J, Gibson W, Murphy TF, Sethi S, et al. Early hospital readmissions after an acute exacerbation of chronic obstructive pulmonary disease in the nationwide readmissions database. Annals of the American Thoracic Society 2018 /7;15(7):837-845.

(73) Janson C, Johansson G, Ställberg B, Lisspers K, Olsson P, Keininger DL, et al. Identifying the associated risks of pneumonia in COPD patients: ARCTIC an observational study. Respiratory research 2018 /9;19(1):172-172.

(74) Januszek R, Siudak Z, Dziewierz A, Rakowski T, Dudek D, Bartus S. Chronic obstructive pulmonary disease affects the angiographic presentation and outcomes of patients with coronary artery disease treated with percutaneous coronary interventions. Polish Archives of Internal Medicine 2018;128(1):915-925.

(75) Jiang X, Xiao H, Segal R, Mobley WC, Park H. Trends in Readmission Rates, Hospital Charges, and Mortality for Patients With Chronic Obstructive Pulmonary Disease (COPD) in Florida From 2009 to 2014. Clin Ther 2018 /4;40(4):613-626.e1.

(76) Jo YS, Lim MN, Han YJ, Kim WJ. Epidemiological study of PM2.5and risk of COPD-related hospital visits in association with particle constituents in Chuncheon, Korea. International Journal of COPD 2018 /1;13:299-307.

(77) Kang HR, Hong SH, Ha SY, Kim TB, Lee EK. Differences in the risk of mood disorders in patients with asthma-COPD overlap and in patients with COPD alone: A nationwide population-based retrospective cohort study in Korea. Respiratory Research 2019 /4;20(1).

(78) Kim JA, Lim MK, Kim K, Park JH, Rhee CK. Adherence to Inhaled Medications and its Effect on Healthcare Utilization and Costs Among High-Grade Chronic Obstructive Pulmonary Disease Patients. Clinical Drug Investigation 2018 /4;38(4):333-340.

(79) Kridin K, Comaneshter D, Batat E, Cohen AD. COPD and lung cancer in patients with pemphigus- a population based study. Respir Med 2018 /3;136:93-97.

(80) Lai CC, Wang YH, Wang CY, Wang HC, Yu CJ, Chen L. Comparative effects of angiotensin-converting enzyme inhibitors and angiotensin ii receptor blockers on the risk of pneumonia and severe exacerbations in patients with COPD. International Journal of COPD 2018 /3;13:867-874.

(81) Lai CC, Wu CH, Wang YH, Wang CY, Wu VC, Chen L. The association between COPD and outcomes of patients with advanced chronic kidney disease. International Journal of COPD 2018;13:2899-2905.

(82) Lai SW, Lin CL. Association between ankylosing spondylitis and chronic obstructive pulmonary disease in Taiwan. Eur J Intern Med 2018 11;57:e28-e29.

(83) Landis S, Suruki R, Maskell J, Bonar K, Hilton E, Compton C. Demographic and Clinical Characteristics of COPD Patients at Different Blood Eosinophil Levels in the UK Clinical Practice Research Datalink. COPD: Journal of Chronic Obstructive Pulmonary Disease 2018 /3;15(2):177-184.

(84) Lane DC, Stemkowski S, Stanford RH, Tao Z. Initiation of triple therapy with multiple inhalers in chronic obstructive pulmonary disease: An analysis of treatment patterns from a U.S. retrospective database study. Journal of Managed Care and Specialty Pharmacy 2018 11;24(11):1165-1172.

(85) Larsson K, Janson C, Ställberg B, Lisspers K, Olsson P, Kostikas K, et al. Impact of COPD diagnosis timing on clinical and economic outcomes: The ARCTIC observational cohort study. International Journal of COPD 2019;14:995-1008.

(86) Lavergne MR, Law MR, Peterson S, Garrison S, Hurley J, Cheng L, et al. Effect of incentive payments on chronic disease management and health services use in British Columbia, Canada: Interrupted time series analysis. Health Policy 2018 /2;122(2):157-164.

(87) Lawson CA, Mamas MA, Jones PW, Teece L, McCann G, Khunti K, et al. Association of Medication Intensity and Stages of Airflow Limitation With the Risk of Hospitalization or Death in Patients With Heart Failure and Chronic Obstructive Pulmonary Disease. JAMA network open 2018 12;1(8):e185489-e185489.

(88) Lee KC, Wu YT, Chen L, Shen C, Chung C, Chien W, et al. Chronic obstructive pulmonary disease combined with vertebral compression fracture increases the risk of temporomandibular disorder: A population-based cohort study. Medicine (Baltimore) 2019;98(37).

(89) Lee R, Lee D, Mamidi IS, Probasco WV, Heyer JH, Pandarinath R. Patients with Chronic Obstructive Pulmonary Disease Are at Higher Risk for Pneumonia, Septic Shock, and Blood Transfusions after Total Shoulder Arthroplasty. Clin Orthop 2019 /2;477(2):416-423.

(90) Lee YM, Kim SJ, Lee JH, Ha E. Inhaled corticosteroids in COPD and the risk of lung cancer. International Journal of Cancer 2018 11;143(9):2311-2318.

(91) Li M, Wang F, Chen R, Liang Z, Zhou Y, Yang Y, et al. Factors contributing to hospitalization costs for patients with COPD in China: A retrospective analysis of medical record data. International Journal of COPD 2018;13:3349-3357.

(92) Li Q, Larivée P, Courteau J, Couillard S, Poder TG, Carrier N, et al. Greater eosinophil counts at first COPD hospitalization are associated with more readmissions and fewer deaths. International Journal of COPD 2019;14:331-341.

(93) Liao KM, Chen YC, Cheng KC, Wang JJ, Ho CH. Trends in intensive care unit admissions of COPD patients from 2003 to 2013 in Taiwan. International Journal of COPD 2018 /6;13:2007-2012.

(94) Liao KM, Tseng CJ, Chen YC, Wang JJ, Ho CH. Outcomes of laparoscopic cholecystectomy in patients with and without COPD. International Journal of COPD 2019;14:1159-1165.

(95) Lim R, Kerr M, Roughead EE. Use of medicines and health services for chronic obstructive pulmonary disease among a cohort of Australians over 50 years. International Journal of COPD 2018;13:3085-3093.

(96) Lin CS, Chen CY, Yeh CC, Chung CL, Chen TL, Liao CC. Defining risk of general surgery in patients with chronic obstructive pulmonary diseases. QJM 2019 /2;112(2):107-113.

(97) Lin HW, Lin LF, Chen HC, Liou TH, Huang SW. Chronic obstructive pulmonary disease with short-acting inhaled pharmacotherapy increases the risk of prostate cancer: A two-stage database approach. PLoS ONE 2018 /9;13(9).

(98) Lin S, Zhang Q, Chen F, Luo L, Chen L, Zhang W. Smooth Bayesian network model for the prediction of future high-cost patients with COPD. Int J Med Inf 2019 /6;126:147-155.

(99) Lin WC, Chen CW, Lu CL, Lai WW, Huang MH, Tsai LM, et al. The association between recent hospitalized COPD exacerbations and adverse outcomes after percutaneous coronary intervention: A nationwide cohort study. International Journal of COPD 2019;14:169-179.

(100) Liou JT, Lin CW, Tsai CL, Wang YH, Lai JH, Hsu YJ, et al. Risk of Severe Cardiovascular Events From Add-On Tiotropium in Chronic Obstructive Pulmonary Disease. Mayo Clin Proc 2018 10;93(10):1462-1473.

(101) Lisspers K, Larsson K, Johansson G, Janson C, Costa-Scharplatz M, Gruenberger JB, et al. Economic burden of COPD in a Swedish cohort: The ARCTIC study. International Journal of COPD 2018 /1;13:275-285.

(102) Lu HY, Liao KM. Risk of empyema in patients with COPD. International Journal of COPD 2018 /1;13:317-324.

(103) Mahta A, Merkler AE, Reznik ME, Burch JE, Yaghi S, Sellke FW, et al. Emphysema: A Potential Risk Factor for Subarachnoid Hemorrhage and Ruptured Aortic Aneurysm. Stroke 2019 /4;50(4):992-994.

(104) Maitre T, Cottenet J, Beltramo G, Georges M, Blot M, Piroth L, et al. Increasing burden of noninfectious lung disease in persons living with HIV: A 7-year study using the French nationwide hospital administrative database. European Respiratory Journal 2018;52(3).

(105) Major JM, Zhou EH, Ding Y, Ly T, Li J, Pinheiro SP, et al. The effect of FDA drug safety communications on patterns of tiotropium dispensing: A U.S. Health Plan Claims Database study. Journal of Managed Care and Specialty Pharmacy 2018 /7;24(7):700-709.

(106) Mcguire K, Aviña-Zubieta JA, Esdaile JM, Sadatsafavi M, Sayre EC, Abrahamowicz M, et al. Risk of Incident Chronic Obstructive Pulmonary Disease in Rheumatoid Arthritis: A Population-Based Cohort Study. Arthritis Care and Research 2019 /5;71(5):602-610.

(107) Meeraus W, Wood R, Jakubanis R, Holbrook T, Bizouard G, Despres J, et al. COPD treatment pathways in france: a retrospective analysis of electronic medical record data from general practitioners. International Journal of COPD 2019;14:51-63.

(108) Morales DR, Flynn R, Zhang J, Trucco E, Quint JK, Zutis K. External validation of ADO, DOSE, COTE and CODEX at predicting death in primary care patients with COPD using standard and machine learning approaches. Respir Med 2018 /5;138:150-155.

(109) Mueller S, Gottschalk F, Groth A, Meeraus W, Driessen M, Kohlmann T, et al. Primary data, claims data, and linked data in observational research: The case of COPD in Germany. Respiratory Research 2018 /8;19(1).

(110) Müllerová H, Hahn B, Simard EP, Mu G, Hatipoğlu U. Exacerbations and health care resource use among patients with COPD in relation to blood eosinophil counts. International Journal of COPD 2019;14:683-692.

(111) Müllerová H, Meeraus WH, Galkin DV, Albers FC, Landis SH. Clinical burden of illness among patients with severe eosinophilic COPD. International Journal of COPD 2019;14:741-755.

(112) Muñoz-Quiles C, López-Lacort M, Díez-Domingo J. Risk and impact of herpes zoster among COPD patients: A population-based study, 2009-2014. BMC Infectious Diseases 2018 /5;18(1).

(113) Nagar S, Patel J, Stanford R. Characteristics and health care resource use of subjects with COPD in the year before initiating LAMA monotherapy or LAMA+LABA combination therapy: A U.S. database study. Manag Care 2018;27(5):40-47.

(114) Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Concomitant diagnosis of asthma and COPD: A quantitative study in UK primary care. British Journal of General Practice 2018 11;68(676):e775-e782.

(115) Niyonsenga T, Coffee NT, Del Fante P, Høj SB, Daniel M. Practical utility of general practice data capture and spatial analysis for understanding COPD and asthma. BMC Health Services Research 2018 11;18(1).

(116) Obi J, Mehari A, Gillum R. Mortality Related to Chronic Obstructive Pulmonary Disease and Co-morbidities in the United States, A Multiple Causes of Death Analysis. COPD: Journal of Chronic Obstructive Pulmonary Disease 2018 /3;15(2):200-205.

(117) Ohar JA, Loh CH, Lenoir KM, Wells BJ, Peters SP. A comprehensive care plan that reduces readmissions after acute exacerbations of COPD. Respir Med 2018 /8;141:20-25.

(118) Ortega H, Llanos JP, Lafeuille MH, Germain G, Duh MS, Bell CF, et al. Burden of disease associated with a COPD eosinophilic phenotype. International Journal of COPD 2018;13:2425-2433.

(119) Oshagbemi OA, Keene SJ, Driessen JHM, Jordan R, Wouters EFM, de Boer A, et al. Trends in moderate and severe exacerbations among COPD patients in the UK from 2005 to 2013. Respir Med 2018 11;144:1-6.

(120) Pacileo G, Tozzi VD, Sotgiu G, Aliberti S, Morando V, Blasi F. Administrative databases and clinical governance: The case of COPD. Int J Health Plann Manage 2019 /1;34(1):177-186.

(121) Park SC, Kim YS, Kang YA, Park EC, Shin CS, Kim DW, et al. Hemoglobin and mortality in patients with COPD: A nationwide population-based cohort study. International Journal of COPD 2018 /5;13:1599-1605.

(122) Parkin L, Khuu W, Stanbrook MB, Tadrous M, Martins D, Gomes T. Trends in the utilisation of COPD therapeutic regimens before and after the introduction of LAMA/LABA combination products: A population-based study. Respir Med 2018 10;143:1-7.

(123) Patel JG, Coutinho AD, Lunacsek OE, Dalal AA. COPD affects worker productivity and health care costs. International Journal of COPD 2018;13:2301-2311.

(124) Paul MC, Dik JWH, Hoekstra T, Van Dijk CE. Admissions for ambulatory care sensitive conditions: A national observational study in the general and COPD population. Eur J Public Health 2019 /4;29(2):213-219.

(125) Pendharkar SR, Ospina MB, Southern DA, Hirani N, Graham J, Faris P, et al. Effectiveness of a standardized electronic admission order set for acute exacerbation of chronic obstructive pulmonary disease. BMC Pulmonary Medicine 2018 /5;18(1).

(126) Petite SE. Characterization of chronic obstructive pulmonary disease prescribing patterns in the United States. Pulmonary Pharmacology and Therapeutics 2018 /4;49:119-122.

(127) Pikoula M, Quint JK, Nissen F, Hemingway H, Smeeth L, Denaxas S. Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. BMC Medical Informatics and Decision Making 2019 /4;19(1).

(128) Qiu H, Tan K, Long F, Wang L, Yu H, Deng R, et al. The burden of COPD morbidity attributable to the interaction between ambient air pollution and temperature in Chengdu, China. International Journal of Environmental Research and Public Health 2018 /3;15(3).

(129) Ran J, Sun S, Yang A, Yang L, Han L, Mason TG, et al. Effects of ambient benzene and toluene on emergency COPD hospitalizations: A time series study in Hong Kong. Sci Total Environ 2019 /3;657:28-35.

(130) Rayner LH, McGovern AP, Sherlock J, Gatenby P, Correa A, Creagh-Brown B, et al. Type 2 diabetes: A protective factor for COPD? Primary Care Diabetes 2018 10;12(5):438-444.

(131) Reid CE, Considine EM, Watson GL, Telesca D, Pfister GG, Jerrett M. Associations between respiratory health and ozone and fine particulate matter during a wildfire event. Environ Int 2019 /8;129:291-298.

(132) Rezaei SS, Rinner C, Ratajczak P, Grossmann W, Gall W, Wolzt M. Use of beta-blocker is associated with lower mortality in patients with coronary artery disease with or without COPD. Clinical Respiratory Journal 2018 12;12(12):2627-2634.

(133) Riisgaard H, Le JV, Søndergaard J, Munch M, Ledderer L, Pedersen LB. Associations between degrees of task delegation and adherence to COPD guidelines on spirometry testing in general practice - A national cross-sectional study. BMC Health Services Research 2019 /7;19(1).

(134) Rothnie KJ, Connell O, Müllerová H, Smeeth L, Pearce N, Douglas I, et al. Myocardial infarction and ischemic stroke after exacerbations of chronic obstructive pulmonary disease. Annals of the American Thoracic Society 2018 /8;15(8):935-946.

(135) Russell REK, Beer S, Pavord ID, Pullinger R, Bafadhel M. The acute wheezy adult with airways disease in the emergency department: A retrospective case-note review of exacerbations of COPD. International Journal of COPD 2019;14:971-977.

(136) Salimi F, Morgan G, Rolfe M, Samoli E, Cowie CT, Hanigan I, et al. Long-term exposure to low concentrations of air pollutants and hospitalisation for respiratory diseases: A prospective cohort study in Australia. Environ Int 2018 12;121:415-420.

(137) Samp JC, Joo MJ, Schumock GT, Calip GS, Pickard AS, Lee TA. Predicting acute exacerbations in chronic obstructive pulmonary disease. Journal of Managed Care and Specialty Pharmacy 2018 /3;24(3):265-279.

(138) Sandelin M, Mindus S, Thuresson M, Lisspers K, Ställberg B, Johansson G, et al. Factors associated with lung cancer in COPD patients. International Journal of COPD 2018 /6;13:1833-1839.

(139) Sansgiry SS, Bhansali A, Serna O, Kamdar M, Fleming M, Abughosh S, et al. Effect of coverage gap on healthcare utilization among Medicare beneficiaries with chronic obstructive pulmonary disorder. Curr Med Res Opin 2019 /2;35(2):321-328.

(140) Scrutinio D, Guida P, Passantino A, Ammirati E, Oliva F, Lagioia R, et al. Acutely decompensated heart failure with chronic obstructive pulmonary disease: Clinical characteristics and long-term survival. Eur J Intern Med 2019 /2;60:31-38.

(141) Serra-Picamal X, Roman R, Escarrabill J, García-Altés A, Argimón JM, Soler N, et al. Hospitalizations due to exacerbations of COPD: A big data perspective. Respir Med 2018 12;145:219-225.

(142) 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. International Journal of COPD 2019;14:693-698.

(143) Shah S, Blanchette CM, Coyle JC, Kowalkowski M, Arthur ST, Howden R. Survival associated with chronic obstructive pulmonary disease among SEER-medicare beneficiaries with non-small-cell lung cancer. International Journal of COPD 2019;14:893-903.

(144) Shah S, Blanchette CM, Coyle JC, Kowalkowski M, Arthur ST, Howden R. Healthcare utilization and costs associated with COPD among SEER-Medicare beneficiaries with NSCLC. Journal of Medical Economics 2018 /9;21(9):861-868.

(145) Shantakumar S, Pwu RF, D'Silva L, Wurst K, Kuo YW, Yang YY, et al. Burden of asthma and COPD overlap (ACO) in Taiwan: A nationwide population-based study. BMC Pulmonary Medicine 2018 /1;18(1).

(146) Sharif K, Watad A, Tiosano S, Yavne Y, Blokh Kerpel A, Comaneshter D, et al. The link between COPD and ankylosing spondylitis: A population based study. Eur J Intern Med 2018 /7;53:62-65.

(147) Shi M, Wang J, Zhang L, Yan Y, Miao Yd, Zhang X. Effects of Integrated Case Payment on Medical Expenditure and Readmission of Inpatients with Chronic Obstructive Pulmonary Disease: A Nonrandomized, Comparative Study in Xi County, China. Current Medical Science 2018 /6;38(3):558-566.

(148) Sikjær MG, Løkke A, Hilberg O. The influence of psychiatric disorders on the course of lung cancer, chronic obstructive pulmonary disease and tuberculosis. Respir Med 2018 /2;135:35-41.

(149) Slatore CG, Falvo MJ, Nugent S, Carlson K. Afghanistan and Iraq War Veterans: Mental health diagnoses are associated with respiratory disease diagnoses. Mil Med 2018 /5;183(5-6):e249-e257.

(150) Spila-Alegiani S, Trotta F, Da Cas R, Rossi M, Venegoni M, Traversa G. Comparative Effectiveness of Two Tiotropium Formulations: A Retrospective Cohort Study. COPD: Journal of Chronic Obstructive Pulmonary Disease 2018 /9;15(5):418-423.

(151) Ställberg B, Janson C, Larsson K, Johansson G, Kostikas K, Gruenberger JB, et al. Real-world retrospective cohort study ARCTIC shows burden of comorbidities in Swedish COPD versus non-COPD patients. npj Primary Care Respiratory Medicine 2018 12;28(1).

(152) Stanford RH, Lau MS, Li Y, Stemkowski S. External validation of a COPD risk measure in a commercial and Medicare population: The COPD treatment ratio. Journal of Managed Care and Specialty Pharmacy 2019 /1;25(1):58-65B.

(153) Stanford R, Nag A, Mapel D, Lee T, Rosiello R, Schatz M, et al. Claims-Based Risk Model for First Severe COPD Exacerbation. Am J Manag Care 2018;24(2):e45-e53.

(154) Stephens AR, Wiener RS, Ieong MH. Comparison of Methods to Identify Advance Care Planning in Patients with Severe Chronic Obstructive Pulmonary Disease Exacerbation. J Palliat Med 2018 /3;21(3):284-289.

(155) Su TH, Chang SH, Kuo CF, Liu PH, Chan YL. B-blockers after acute myocardial infarction in patients with chronic obstructive pulmonary disease: A nationwide population-based observational study. PLoS ONE 2019 /3;14(3).

(156) Su VYF, Liao HF, Perng DW, Chou YC, Hsu CC, Chou CL, et al. Proton pump inhibitors use is associated with a lower risk of acute exacerbation and mortality in patients with coexistent COPD and GERD. International Journal of COPD 2018;13:2907-2915.

(157) Su VYF, Yang KY, Yang YH, Tsai YH, Perng DW, Su WJ, et al. Use of ICS/LABA Combinations or LAMA Is Associated with a Lower Risk of Acute Exacerbation in Patients with Coexistent COPD and Asthma. Journal of Allergy and Clinical Immunology: In Practice 2018 11;6(6):1927-1935.e3.

(158) Suissa S, Dell'Aniello S, Ernst P. Comparative effectiveness of LABA-ICS versus LAMA as initial treatment in COPD targeted by blood eosinophils: a population-based cohort study. The Lancet Respiratory Medicine 2018 11;6(11):855-862.

(159) Sujatha-Bhaskar S, Alizadeh RF, Inaba CS, Koh CY, Jafari MD, Mills SD, et al. Respiratory complications after colonic procedures in chronic obstructive pulmonary disease: does laparoscopy offer a benefit? Surg Endosc 2018 /3;32(3):1280-1285.

(160) Sun S, Cao W, Mason TG, Ran J, Qiu H, Li J, et al. Increased susceptibility to heat for respiratory hospitalizations in Hong Kong. Sci Total Environ 2019 /5;666:197-204.

(161) Swanson JO, Vogt V, Sundmacher L, Hagen TP, Moger TA. Continuity of care and its effect on readmissions for COPD patients: A comparative study of Norway and Germany. Health Policy 2018 /7;122(7):737-745.

(162) Szyszkowicz M, Kousha T, Castner J, Dales R. Air pollution and emergency department visits for respiratory diseases: A multi-city case crossover study. Environ Res 2018 /5;163:263-269.

(163) Tian Y, Xiang X, Juan J, Song J, Cao Y, Huang C, et al. Short-term effects of ambient fine particulate matter pollution on hospital visits for chronic obstructive pulmonary disease in Beijing, China. Environmental Health: A Global Access Science Source 2018 /2;17(1).

(164) To T, Zhu J, Gray N, Feldman LY, Villeneuve PJ, Licskai C, et al. Asthma and chronic obstructive pulmonary disease overlap in women incidence and risk factors. Annals of the American Thoracic Society 2018 11;15(11):1304-1310.

(165) Trantham L, Sikirica MV, Candrilli SD, Benson VS, Mohan D, Neil D, et al. Healthcare costs and utilization associated with muscle weakness diagnosis codes in patients with chronic obstructive pulmonary disease: a United States claims analysis. Journal of Medical Economics 2019 /4;22(4):319-327.

(166) Tsai CH, Liao LY, Lin CL, Chung WS. Inhaled corticosteroids and the risks of low-energy fractures in patients with chronic airway diseases: A propensity score matched study. Clinical Respiratory Journal 2018 /5;12(5):1830-1837.

(167) Tseng CH. Metformin and risk of chronic obstructive pulmonary disease in diabetes patients. Diabetes and Metabolism 2019 /4;45(2):184-190.

(168) van Buul AR, Wildschut TS, Bonten TN, Kasteleyn MJ, Slats AM, Chavannes NH, et al. A systematic diagnostic evaluation combined with an internet-based self-management support system for patients with asthma or COPD. International Journal of COPD 2018;13:3297-3306.

(169) Vazquez Guillamet R, Ursu O, Iwamoto G, Moseley PL, Oprea T. Chronic obstructive pulmonary disease phenotypes using cluster analysis of electronic medical records. Health Informatics Journal 2018 12;24(4):394-409.

(170) Vela E, Tényi Á, Cano I, Monterde D, Cleries M, Garcia-Altes A, et al. Population-based analysis of patients with COPD in Catalonia: A cohort study with implications for clinical management. BMJ Open 2018 /3;8(3).

(171) Voorham J, Roche N, Benhaddi H, Van Der Tol M, Carter V, Van Boven JFM, et al. Real-world effectiveness evaluation of budesonide/formoterol Spiromax for the management of asthma and chronic obstructive pulmonary disease in the UK. BMJ Open 2018;8(10).

(172) Vozoris NT, Wang X, Austin PC, O'Donnell DE, Aaron SD, To TM, et al. Incident diuretic drug use and adverse respiratory events among older adults with chronic obstructive pulmonary disease. Br J Clin Pharmacol 2018 /3;84(3):579-589.

(173) Vozoris NT, Wang X, Austin PC, Stephenson AL, O'Donnell DE, Gershon AS, et al. Serotonergic antidepressant use and morbidity and mortality among older adults with COPD. European Respiratory Journal 2018;52(1).

(174) Wallace AE, Shinde MU, Willey VJ, Singer JR, Kaila S, Bayer V, et al. Health care resource utilization and exacerbation rates in patients with COPD stratified by disease severity in a commercially insured population. Journal of Managed Care and Specialty Pharmacy 2019 /2;25(2):205-217.

(175) Association of cardiovascular risk with inhaled long-acting bronchodilators in patients with chronic obstructive pulmonary disease: A nested case-control study. ; /2; : American Medical Association; 2018.

(176) Wang WH, Cheng CC, Mar GY, Wei KC, Huang WC, Liu CP. Improving outcomes in chronic obstructive pulmonary disease by taking beta-blockers after acute myocardial infarction: a nationwide observational study. Heart Vessels 2019 /7;34(7):1158-1167.

(177) Whittaker HR, Müllerova H, Jarvis D, Barnes NC, Jones PW, Compton CH, et al. Inhaled corticosteroids, blood eosinophils, and fev1 decline in patients with COPD in a large UK primary health care setting. International Journal of COPD 2019;14:1063-1073.

(178) Wittbrodt ET, Millette LA, Evans KA, Bonafede M, Tkacz J, Ferguson GT. Differences in health care outcomes between postdischarge COPD patients treated with inhaled corticosteroid/long-acting β 2 -agonist via dry-powder inhalers and pressurized metered-dose inhalers. International Journal of COPD 2019;14:101-114.

(179) Woodcock A, Boucot I, Leather DA, Crawford J, Collier S, Bakerly ND, et al. Effectiveness versus efficacy trials in COPD: How study design influences outcomes and applicability. European Respiratory Journal 2018;51(2).

(180) Wu CX, Hwang CH, Tan WS, Tai KP, Kwek LSL, Chee TG, et al. Effectiveness of a chronic obstructive pulmonary disease integrated care pathway in a regional health system: A propensity score matched cohort study. BMJ Open 2018 /3;8(3).

(181) Yakubek GA, Curtis GL, Khlopas A, Faour M, Klika AK, Mont MA, et al. Chronic Obstructive Pulmonary Disease Is Associated With Short-Term Complications Following Total Knee Arthroplasty. J Arthroplasty 2018 /8;33(8):2623-2626.

(182) Yakubek GA, Curtis GL, Sodhi N, Faour M, Klika AK, Mont MA, et al. Chronic Obstructive Pulmonary Disease Is Associated With Short-Term Complications Following Total Hip Arthroplasty. J Arthroplasty 2018 /6;33(6):1926-1929.

(183) Yen FS, Chen W, Wei JCC, Hsu CC, Hwu CM. Effects of metformin use on total mortality in patients with type 2 diabetes and chronic obstructive pulmonary disease: A matchedsubject design. PLoS ONE 2018 10;13(10).

(184) Zeiger RS, Tran TN, Butler RK, Schatz M, Li Q, Khatry DB, et al. Relationship of Blood Eosinophil Count to Exacerbations in Chronic Obstructive Pulmonary Disease. Journal of Allergy and Clinical Immunology: In Practice 2018 /5;6(3):944-954.e5.

(185) Zhang Z, Chai P, Wang J, Ye Z, Shen P, Lu H, et al. Association of particulate matter air pollution and hospital visits for respiratory diseases: a time-series study from China. Environmental Science and Pollution Research 2019 /4;26(12):12280-12287.

3 CONSTRUCTION OF A DATASET TO EXAMINE INFECTIVE OUTCOMES RELATED TO EMERGENCY ADMISSIONS FOR COPD

3.1 INTRODUCTION

This chapter discusses the creation of a COPD dataset from which I go on to examine the contribution of respiratory pathogens to emergency admissions for COPD.

3.2 METHODS

3.2.1 Ethics and information governance

Ethics approval was not required for this work, since only de-identified data were used. The project was approved by the Secure Anonymised Information Linkage (SAIL) Information Governance Review Panel. [Appendix 1]

3.2.2 The SAIL Databank and data linkage

I accessed my project-specific data views and analysed my data via the SAIL Databank,[57] a remote access system that provides secure data access to approved users. The SAIL Databank contains de-identified health and administrative data from a variety of data sources, covering the entire population of Wales in a Trusted Research Environment.

Linking the different datasets within SAIL is possible using an anonymised person-level identifier, the Anonymised Linkage Field.[58] I used Structured Query Language to extract and amalgamate relevant information from the different datasets to construct my COPD cohort, the content of which is described below.

3.2.3 Data sources

The following datasets were used in the development of my dataset: the Welsh Demographic Service (WDS), Patient Episode Database for Wales (PEDW), and Welsh Results Reporting Service (WRRS). The WDS contains administrative data drawn from GP practices about individuals in Wales that use NHS services, such as address and practice registration history.

PEDW contains administrative and clinical information for all NHS Wales hospital admissions, including spell and episode level data. An episode (or, 'Finished Consultant Episode'), in this context, is the time a patient spends in the continuous care of one consultant within a provider spell. A provider spell is the continuous period of time that an admitted patient spends under the care of one health care provider (Local Health Board in Wales).[59] The data are collected and coded at each hospital. Administrative information is collected from the central Patient Administrative System. For clinical information, after the patient is discharged, the handwritten patient notes are transcribed by clinical coders into medical coding terminology. WRRS is taken (indirectly) from the Welsh Laboratory Information Management System (WLIMS), which is the system used by staff in pathology disciplines.

The SAIL Databank receives data updates at varying frequencies for the different datasets. My project views were created in July 2020 from the most recent extract for each dataset, with a base population of those attaining age ≥35 by the time of the data extraction. Further information regarding coverage for each specific dataset used for the creation of my study population is shown in Table 1, below.

Table 1. Coverage of datasets used to create my AECOPD dataset

Dataset	Coverage	Time period
Welsh Demographic Service (WDS)	Wales	01/01/1990 – 24/04/20
Patient Episode Database for Wales (PEDW)	Wales	01/04/1995 – 28/04/20
Welsh Results Reporting Service (WRRS)	Wales	XX/XX/XX – 10/12/18 (Start date varies by Health Board)

3.2.4 Data items

The data items utilised from each data source for this study are shown in Table 2, below.

Table 2. Data items utilised for the creation of my AECOPD dataset

Data item	Description of content
PEDW	
ALF_PE	Anonymised linking field
ALF_STS_CD	ALF status code (describes quality of linkage)
SPELL_NUM_PE	A unique encrypted number that identifies the provider spell
EPI_NUM	Sequence number for each finished consultant episode in a provider spell
PROV_UNIT_CD	The organisation code of the Healthcare Provider
DIAG_NUM	A number used to identify the position of the associated diagnosis assigned to a patient
DIAG_CD_123	3 characters of the diagnosis (ICD code) for the condition treated or investigated during the relevant episode of healthcare
DIAG_CD_1234	4 characters of the diagnosis (ICD code) for the condition treated or investigated during the relevant episode of healthcare
ADMIS_DT	The date of admission to hospital
ADMIS_YR	The year of admission to hospital
ADMIS_MTHD_CD	Method of admission to a hospital provider spell
WDS	
ALF_PE	Anonymised linking field
WOB	Week of birth
WRRS	
ALF_PE	Anonymised linking field
TESTSETNAME	Description of test
CODE	Description of test
VALUE	Test result
SPCM_COLLECTED_DTTM	Date and time of specimen collection

To create a dataset of individuals with emergency hospital admissions to Wales with COPD over my two year period, I went through the following steps:

- I used the data item 'ADMIS_DT' to identify admissions to hospital from 1/12/16 – 1/12/18
- The hospital admission was required to be in Wales, using a list of values associated with data item 'PROV_UNIT_CD'
- The admission was specified to be an *emergency* admission using a list of values associated with 'ADMIS_MTHD_CD'
- To narrow down to emergency hospital admissions for *COPD,* I used 'DIAG_CD_123' and 'DIAG_NUM' to identify admissions coded with ICD-10 codes J43 or J44 in the first or second position
- To further refine the population to individuals aged ≥35 at admission, 'ADMIS_DT' together with 'WOB' were used to create a new data item, 'age at admission'
- The above required PEDW to be linked to WDS, using the anonymised linkage field, 'ALF_PE', and quality of linkage was specified using ALF_STS_CD
- Linking these admissions to WRRS data again used the anonymised linkage field, 'ALF_PE'
- Use of 'ADMIS_DT' within PEDW and 'SPCM_COLLECTED_DTTM' within WRRS allowed identification of pathology tests collected within a specified window either side of the date of hospital admission
- The pathology tests of interests were specified using 'TESTSETNAME' and 'CODE', and the test result identified using the data item, 'VALUE'

3.2.5 Identifying the COPD population in electronic health records

It is important to remember, as shown in the previous chapter, that the case definition used to identify those with COPD in routinely collected data can take many forms, which are likely to vary in their sensitivity and specificity.

Some individuals are diagnosed with COPD after spirometry testing, and a case definition necessitating presence of a diagnosis code, a spirometry code, and/or the spirometry results in-keeping with COPD would be able to identify these individuals.

However, for some people who have had spirometry testing, this information may not be input into the coded clinical systems available to researchers. Further, some people are diagnosed with COPD based on a clinicians' judgement and would be identified if a case definition required only a diagnosis code to be present but would be missed if the above, more specific case definition were applied. On the other hand, a more inclusive definition using just diagnosis codes runs the risk of including people who have been misdiagnosed or miscoded.

For those receiving treatment for COPD but who may have not received a definite diagnosis, a case definition requiring only medication codes could identify them. This method would also identify those who have been diagnosed with and treated for COPD, but their diagnosis may only be present in free-text format and therefore not in a coded format available in SAIL. However, this definition could also wrongly include individuals who do not have COPD, but use the specified medications for other reasons, such as asthma.

Lastly, there are those individuals who truly have COPD but have been misdiagnosed or have not consulted with HCPs, thus remaining undiagnosed and untreated.

In identifying patients with COPD using the SAIL Databank, I first needed to decide what research question I was attempting to answer, since the case definition used should be tailored according to the aims of a study. In Chapter 4, given that I was looking at those who had had an emergency hospital admission for COPD, my method of identification of individuals differed to the approach I would have taken had I been attempting to answer a research question concerning the whole COPD population in Wales.

4 USING ROUTINE DATA TO ESTIMATE THE CONTRIBUTION OF RESPIRATORY PATHOGENS TO ACUTE EXACERBATIONS OF COPD

This paper was originally published in the Journal of Infection, and is available here:

Sivakumaran S, Alsallakh MA, Lyons RA, Quint JK, Davies GA. Estimating the contribution of respiratory pathogens to acute exacerbations of COPD using routine data. J Infect. 2023 Mar;86(3):233-238. doi: 10.1016/j.jinf.2023.01.012. Epub 2023 Jan 24. PMID: 36706962.

4.1 RESEARCH PAPER

Title: Estimating the contribution of respiratory pathogens to acute exacerbations of COPD using routine data

Running title: Real-world microbiology testing in exacerbations of COPD

Authors: Shanya Sivakumaran, Mohammad A. Alsallakh, Ronan A. Lyons, Jennifer K. Quint, Gwyneth A. Davies.

Highlights

- Linkage of microbiology data to national admissions data over a two year period
- Respiratory viruses are detected in a substantial proportion of COPD exacerbations
- Increased respiratory viral testing could aid antimicrobial stewardship efforts
- ICD codes have low sensitivity in identifying respiratory pathogens in routine data
- Large-scale data linkage delineates the burden of specific pathogens with greater accuracy

Abstract

Objectives: To characterise microbiology testing and results associated with emergency admissions for acute exacerbation of COPD (AECOPD), and determine the accuracy of ICD-10 codes in retrospectively identifying laboratory-confirmed respiratory pathogens in this setting.

Methods: Using person-level data from the Secure Anonymised Information Linkage Databank in Wales, we extracted emergency admissions for COPD from 1/12/2016 to 30/11/2018 and undertook linkage of admissions data to microbiology data to identify laboratory-confirmed infection. We further used these data to assess the accuracy of pathogen-specific ICD-10 codes.

Results: We analysed data from 15,950 people who had 25,715 emergency admissions for COPD over the two-year period. 99.5% of admissions could be linked to a laboratory test within 7 days of admission date. Sputum was collected in 5,013 (19.5%) of admissions, and respiratory virus testing in 1,219 (4.7%). Where respiratory virus testing was undertaken, 46.7% returned any positive result. Influenza was the virus most frequently detected, in 21.5% of admissions where testing was conducted. ICD-10 codes exhibited low sensitivity in detecting laboratory-confirmed respiratory pathogens.

Conclusions: In people admitted to hospital with AECOPD, increased testing for respiratory viruses could enable more effective antibiotic stewardship and isolation of cases. Linkage with microbiology data achieves more accurate and reliable case definitions.

Keywords

‘Respiratory Tract Infections’, ‘Pulmonary Disease, Chronic Obstructive’, ‘Electronic Health Records’

Introduction

Chronic obstructive pulmonary disease (COPD) is characterised by reduced lung function, resulting in progressive dyspnoea and acute exacerbations which necessitate hospital admission if severe. The majority of acute exacerbations of COPD (AECOPDs) are considered to have an infective aetiology,[1] and antibiotic use for this indication is frequent, even though a significant proportion are likely to have a viral trigger.

To develop, implement and evaluate interventions to reduce the burden of respiratory tract infections, pathogen-specific data establishing the frequency and consequences of infection in different populations is imperative. Thus far, the focus of many studies examining the burden of specific respiratory pathogens has been on children.[2–4] Impact on other risk groups, such as those with COPD, requires further study, especially since acute exacerbations of COPD are one of the commonest reasons for emergency hospital admission in the UK,[5] resulting in substantial morbidity, mortality and healthcare cost each year.

International Statistical Classification of Diseases and Related Health Problems (ICD) codes relating to respiratory pathogens are readily available in databases of routinely collected electronic health data and have therefore often been used for research purposes.[6–11] However, the accuracy of diagnostic coding is likely to vary depending on local practices, and not all pathogens map to a specific ICD code.

Large-scale linkage of microbiology data to other routinely collected health data, where possible, is likely to be the more accurate data source in establishing the burden of particular pathogens. Researchers who have had access to microbiology data report that they would have grossly underestimated respiratory pathogen-related hospital admissions had they relied on diagnosis codes alone.[12–15] However, such linkage is currently not possible in many regions, rendering person-level analyses using microbiology data unattainable. Further, even in regions where such linkage is possible, the likelihood of successful linkage may vary depending on age, duration of hospitalisation, ethnic group and rurality.[16]

In this national study, we utilise linkable, routinely collected microbiology and healthcare utilisation data covering all admissions to National Health Service (NHS) hospitals in Wales over a 2 year period.

Our aims were: (i) to illustrate the utility of a national microbiology dataset that can be linked to other routinely collected electronic health data; (ii) to quantify the proportion of emergency admissions for COPD undergoing testing for respiratory pathogens, and the proportion tested where viral or bacterial pathogens were identified; and (iii) to determine the accuracy of using ICD-10 diagnosis codes to identify laboratory-confirmed respiratory pathogens associated with COPD exacerbations.

Methods

We accessed complete coverage, person-level, anonymised datasets from the Secure Anonymised Information Linkage (SAIL) Databank,[17] which receives routinely collected data from all NHS hospitals in Wales, to examine all emergency hospital admissions due to COPD over a two-year period (1/12/2016 to 30/11/2018). This report of our findings was prepared according to the REporting of studies Conducted using Observational Routinely-collected Data (RECORD) statement.[18] The study was approved by the SAIL Information Governance Review Panel (project number 0996).

Data sources, population, and case definitions

Hospital admissions data were extracted from the Patient Episodes Database for Wales (PEDW), a dataset which includes diagnosis codes for all NHS hospital admissions in Wales. Emergency admissions for COPD were defined as those with an ICD-10 COPD code (J43, J44) recorded in the first or second position in any Finished Consultant Episode (FCE) in the population aged ≥35 at date of admission.

We defined two methods of classifying emergency COPD admissions as being associated with specific respiratory pathogens: (1) using pathogen-specific ICD-10

diagnosis codes (the full code set used is available in **supplementary material S1**), and (2) using microbiology test results. Microbiology data were extracted from the Welsh Results Reporting Service (WRRS), which receives results from all NHS hospital laboratories in Wales. Person-level linkage with hospital admissions data was possible using the Anonymised Linkage Field, which has been described previously.[19] We included individuals with high quality linkage (where there was either exact matching of NHS number or matching probability ≥ 90% (using name, sex, date of birth, postcode)). Laboratory diagnosis of respiratory viruses was all based on reverse transcription polymerase chain reaction (RT-PCR) and PCR testing, but it could not be established whether nasal or throat swabs were collected. Bacteria were detected by culture of sputum samples.

Pathogens of interest were influenza (A and B), respiratory syncytial virus (RSV), parainfluenza, rhinovirus, human metapneumovirus (hMPV), adenovirus, enterovirus, *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Moraxella catarrhalis* and *Pseudomonas aeruginosa*. The choice of respiratory viruses was dictated by those included on multiplex panels over the time period studied.

We used laboratory data to determine the proportion of COPD admissions where testing for bacterial and/or viral respiratory pathogens was conducted and estimated the proportion of COPD admissions with laboratory-confirmed infection (with the total number tested as our denominator). We defined admissions associated with laboratory-confirmed infection as those where the collection date of the specimen with a positive result was within (+/-) 7 days of hospital admission date. We assessed the seasonality of virus-associated admissions using methods adapted from Li et al,[20] where seasonality exists if at least 75% of the annual total of positive cases occur in ≤ 5 consecutive months in a year.

Accuracy of ICD-10 coding

Using COPD admissions associated with laboratory-confirmed infection as our reference standard, we assessed the accuracy of ICD-10 diagnosis codes in identifying

respiratory pathogens associated with COPD admissions. Measures of accuracy were positive predictive value (PPV), negative predictive value (NPV), sensitivity, and specificity, each with a 95% confidence interval. The sampling frame for these analyses was all COPD admissions where testing for the pathogen in question was conducted within (+/-) 7 days of admission. PPV was calculated as the proportion of COPD admissions with a pathogen-specific ICD-10 code that had a laboratory-confirmed respiratory pathogen on testing. NPV was calculated as the proportion of COPD admissions without a pathogen-specific ICD-10 code when there was a negative respiratory pathogen test. Sensitivity was calculated as the proportion of COPD admissions with a laboratory-confirmed respiratory pathogen that had a pathogen-specific ICD-10 code. Specificity was calculated as the proportion of COPD admissions with a negative respiratory pathogen test and without a pathogen-specific ICD-10 code.

Results

We analysed data from 15,950 people who had 25,715 emergency admissions for COPD over the two year period 1/12/2016 to 30/11/2018. Median age at admission was 73 (interquartile range 66 – 80). Emergency admissions for COPD were fewer over the summer months (**figure 1**), and peaked at a mean of 1,480 admissions in the month of January.

99.5% (25,596/25,715) of emergency COPD admissions could be linked to a laboratory record with a specimen collection date within (+/-) 7 days of the admission date. The proportion of admissions successfully linked to laboratory data was >99% in the majority of NHS Providers in Wales (**supplementary material S2**).

Respiratory pathogens associated with emergency COPD admissions: laboratory data

A specimen for sputum culture was collected in 5,013 (19.5%) out of 25,715 COPD admissions. Of those admissions where a sputum sample was collected, 1,232 (24.6%)

were associated with growth of one of the aforementioned bacteria. The most frequently observed bacteria was *H.influenzae,* which was cultured in 16.7% of admissions where a sputum sample was collected. The other bacteria under investigation all occurred at an overall percent positivity of <5% **(table 1)**.

A specimen for respiratory virus testing was collected in 1,219 (4.7%) of COPD admissions, and of those admissions associated with any respiratory virus testing, 569 (46.7%) returned any positive result. The most frequently detected virus was influenza, identified in 21.5% of admissions where testing was carried out, followed by rhinovirus (12.4%), RSV (6.8%) and hMPV (5.2%).

Of the 491 admissions where both sputum collection and viral testing was carried out, 41 (8%) admissions were associated with both bacterial growth in sputum and a positive respiratory virus test.

Testing frequency and percent positive rate varied by pathogen (**table 1**), with sputum collection more frequent than viral testing, and the highest percent positive levels seen for influenza, rhinovirus and *H.influenzae.* There was also marked variation in testing frequency and percent positivity by month of admission for the viral pathogens, with influenza percent positivity ranging from 0% in the UK summer months, to > 30% in January and February (**table 2**). Over 75% of the total influenza-associated COPD admissions occurred over the UK winter months (December, January, February). RSV and hMPV-associated COPD admissions also showed seasonality (>75% of total admissions occurring over November to January for RSV, and December to March for hMPV), in contrast to admissions associated with rhinovirus, parainfluenza, and the bacterial pathogens. Counts of positive adenovirus and enterovirus tests were small, and thus not analysed by month.

Accuracy of ICD-10 diagnosis codes in detecting respiratory pathogens associated with emergency COPD admissions

The ICD-10 diagnosis codes used to identify influenza-associated COPD admissions performed best in our data, compared to those used for other pathogens, with a

sensitivity of 59.2% (95% confidence interval 52.9 – 65.1%), PPV of 86.1% (80.6 – 90.2%), NPV of 89.7% (88.3 – 91.0%) and specificity of 97.4% (96.2 – 98.3%) (**table 3**).

Sensitivities of the ICD-10 codes related to the other respiratory pathogens were substantially lower, although those of the bacteria were generally higher than the non-influenza viruses (**supplementary material S3**). *P.aeruginosa* ICD-10 codes had the highest sensitivity of all the non-influenza pathogens (42.5% (95% CI 35.6 – 49.7%), but also had a lower PPV (48.0% (41.6 – 54.5%)), compared to a PPV >80% for all other pathogens where PPV can be reported; **table 4**). Specificities were high (>95%) for all pathogens where this can be reported.

Due to low counts in the contingency tables for parainfluenza and hMPV, all accuracy measures cannot be reported for these pathogens, but sensitivities were <8.9% (parainfluenza) and <7.9% (hMPV). *M.catarrhalis* is not associated with any pathogen-specific ICD-10 codes, and so could not be included in these analyses.

Discussion

Our national level analyses of microbiology testing related to emergency COPD admissions in Wales reveals widespread under-utilisation of diagnostic microbiology for severe AECOPD, and a high percent positivity for respiratory viruses, with 46.7% of admissions in which testing was carried out returning a positive result. Given the frequency with which AECOPD admissions occur, the prevalent use of antibiotics for this presentation, and the growing problem of antimicrobial resistance, our data suggests that more widespread use of viral testing could be beneficial in advancing antibiotic stewardship. In addition, increased viral testing would enable timely isolation of cases in healthcare settings in order to reduce transmission to other, often vulnerable, hospital inpatients.

We demonstrate how using ICD-10 diagnosis codes to detect respiratory pathogens associated with COPD admissions in our study would have considerably underestimated the burden of these infections. Researchers using ICD-10 codes as substitutes for raw microbiology data should be aware that coding deficiencies may lead to a substantial loss in accuracy.

Furthermore, our study illustrates the feasibility and utility of linking national microbiology data with other routinely collected electronic health data, which has innumerable potential applications in future research.

The strengths of our study include the complete population coverage, with a base dataset covering all admissions to hospitals in Wales, and the high level (99.5%) of successful linkage between hospital admissions and microbiology data. This was possible due to the recent development of a national system in Wales, the Wales Results Reporting System, which has brought together the results of all laboratory investigations from all providers, whether instituted in community or hospital settings. Whilst there are linkages in individual hospital based systems in the UK, there is no other system that links laboratory results to other routinely collected health data at a national population level. This system has allowed us to gain insight into real-world patterns of microbiology testing in emergency admissions with COPD, estimate the contribution of bacterial and viral pathogens to severe AECOPD, and report the accuracy of pathogen-specific ICD-10 diagnosis codes.

Our study has some limitations. People undergoing microbiology tests may be those whom clinicians consider most likely to have an infective aetiology of their AECOPD, therefore extrapolating the proportion of positive tests to the total number of AECOPD hospitalisations would likely be an overestimate. Nevertheless, the majority of AECOPD are thought to be infective in nature,[1] the proportion of our admissions associated with respiratory viruses was consistent with previous studies,[22,23] and the high percent positivity rate suggests a large number who were not tested would also test positive. Patterns of respiratory virus transmission change year to year, and the winter of 2017/2018 saw a high level of influenza in the UK. We were confined to analysing a 2 year period due to data availability, so how representative our data is of other years

is uncertain, though general patterns are likely to remain consistent. Co-infection between respiratory viruses and bacteria may be underestimated in our data, since some pathogens were not tested for or not included in our study, such as the seasonal coronaviruses. We have not reported subtypes of the respiratory viruses such as influenza or parainfluenza due to paucity of data in this regard. Lastly, the role of microorganisms in the aetiology of AECOPD has previously been unclear due to potential contamination of sputum samples by the upper airway, chronic bacterial colonisation that can occur in the lower airways of those with COPD, a lack of definitive improvement in AECOPDs treated with antibiotics in randomised controlled trials,[24] and studies reporting asymptomatic detection of rhinovirus.[25] However, studies demonstrating the development of specific immune responses to bacteria after exacerbations support the causative role of bacteria in the exacerbation,[26,27] trials may have been weakened due to inclusion of AECOPD with non-bacterial aetiology, and respiratory viruses aside from rhinovirus are not commonly identified in stable disease.[25] Nevertheless, we are unable to infer from our data that any microorganisms detected directly contributed to development of the COPD exacerbation.

In the context of other research, prospective studies[28–34] have provided useful estimates of pathogen distribution in AECOPD due to avoidance of selective testing strategies. However, due to the expense involved with such study designs, they are mostly limited to small, selected populations or regions. Our results broadly reflect those previously published. For example, in Brendish et al's prospective study,[35] 42% of patients presenting to hospital with an exacerbation of their airways disease tested positive for respiratory viruses, compared to 46.7% in our data. Others have also demonstrated increased percent positivity of respiratory viruses in COPD admissions during winter months.[29]

Data from a randomised controlled trial has shown that point-of-care testing (POCT) for influenza can allow timely treatment with antivirals and swift isolation in healthcare settings, minimising nosocomial spread to frail populations.[36] Another study demonstrates increased early discontinuation of antibiotics in those with underlying airways disease testing positive for a respiratory virus on POCT.[35]

Evidence of the importance of the non-influenza respiratory viruses on the adult population has grown in recent years. RSV has been shown to be associated with 11% of COPD hospitalisations in a prospective US-based study,[37] and >900,000 inappropriate antibiotic prescriptions each season in a UK-based modelling study.[38] With progress in the development of vaccination against RSV,[39] further data regarding the burden of RSV-associated disease by age, geography, and comorbidity is key to guide future vaccination strategy,[40] and our data contributes to this. With regard to previous research examining routinely collected microbiology data, other studies have also identified underutilisation of diagnostic microbiology testing[41] and low sensitivity of ICD codes to identify laboratory-confirmed non-influenza respiratory infection.[42,43]

To summarise, our findings of a high percent positivity of respiratory viruses amongst those hospitalised with AECOPD, low utilisation of microbiology testing, and limitations of ICD-10 codes in identifying the burden of respiratory infections generate several key targets for further work. These include increasing the use of diagnostic testing in AECOPD, along with translation of this into reduced antimicrobial prescribing and timely infection control measures, and increasing the ability to perform linkage of existing electronic health databases to routinely collected microbiology data, which would have almost countless future applications.

Footnotes

Acknowledgements:

This study made use of anonymised data held in the SAIL Databank. We would like to acknowledge all the data providers who make anonymised data available for research.

Funding:

This study was funded by Swansea University Medical School with the support of BREATHE – The Health Data Research Hub for Respiratory Health [MC_PC_19004], which is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK.

The funding source had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Ethics:

We were granted permissions from SAIL's independent Information Governance Review Panel to conduct this study. Ethical review was not required as only anonymised data were used.

Conflict of Interest statement:

The authors declare no conflict of interests.

Data availability:

The anonymised person-level data used in this study are held by the SAIL Databank and not publicly available. All proposals to use the SAIL Databank are carefully

reviewed by an independent Information Governance Review Panel to ensure proper and appropriate use of data (https://www.saildatabank.com/application-process). When approved, access is then provided through the SAIL Gateway, a privacy-protecting safe haven and a secure remote access system.

Contributors:

SS conceived the study in collaboration with all authors, conducted the analysis with advice from MA, and drafted the manuscript. All authors critically reviewed and approved the final version of the manuscript.

References

1. Sapey E, Stockley RA. COPD exacerbations·2: Aetiology. *Thorax*. 2006;61(3):250-258. doi:10.1136/thx.2005.041822

2. Wahl B, O'Brien KL, Greenbaum A, et al. Burden of Streptococcus pneumoniae and Haemophilus influenzae type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *Lancet Glob Health*. 2018;6(7):e744-e757. doi:10.1016/S2214-109X(18)30247-X

3. Nair H, Nokes DJ, Gessner BD, et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *The Lancet*. 2010;375(9725):1545-1555. doi:10.1016/S0140-6736(10)60206-1

4. Troeger C, Blacker B, Khalil IA, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis*. 2018;18(11):1191-1210. doi:10.1016/S1473-3099(18)30310-4

5. National Institute for Health and Care Excellence. Quality standards and indicators briefing paper: Chronic obstructive pulmonary disease. Published 2015. Accessed March 22, 2021. https://www.nice.org.uk/guidance/qs10/documents/briefing-paper

6. Hauge SH, Bakken IJ, de Blasio BF, Håberg SE. Risk conditions in children hospitalized with influenza in Norway, 2017–2019. *BMC Infect Dis*. 2020;20(1). doi:10.1186/s12879-020-05486-6

7. Hungerford D, Ibarz-Pavon A, Cleary P, French N. Influenza-associated hospitalisation, vaccine uptake and socioeconomic deprivation in an English city region: An ecological study. *BMJ Open*. 2018;8(12):23275. doi:10.1136/bmjopen-2018-023275

8. Moss JWE, Davidson C, Mattock R, Gibbons I, Mealing S, Carroll S. Quantifying the direct secondary health care cost of seasonal influenza in England. *BMC Public Health*. 2020;20(1). doi:10.1186/s12889-020-09553-0

9. San-Román-Montero JM, Gil Prieto R, Gallardo Pino C, Hinojosa Mena J, Zapatero Gaviria A, Gil De Miguel A. Inpatient hospital fatality related to coding (ICD-9-CM) of the influenza diagnosis in Spain (2009-2015). *BMC Infect Dis*. 2019;19(1). doi:10.1186/s12879-019-4308-5

10. Tong S, Amand C, Kieffer A, Kyaw MH. Incidence of Respiratory Syncytial Virus Related Health Care Utilization in The United States. *J Glob Health*. 2020;10(2):1-10. doi:10.7189/jogh.10.020422

11. Huh K, Kim YE, Ji W, et al. Decrease in hospital admissions for respiratory diseases during the COVID-19 pandemic: a nationwide claims study. *Thorax*. Published online March 29, 2021:thoraxjnl-2020-216526. doi:10.1136/thoraxjnl-2020-216526

12. Hardelid P, Verfuerden M, McMenamin J, Gilbert R. Risk factors for admission to hospital with laboratory-confirmed influenza in young children: birth cohort study. *Eur Respir J*. 2017;50(3):1700489. doi:10.1183/13993003.00489-2017

13. Nicholson KG, McNally T, Silverman M, Simons P, Stockton JD, Zambon MC. Rates of hospitalisation for influenza, respiratory syncytial virus and human metapneumovirus among infants and young children. *Vaccine*. 2006;24(1):102-108. doi:10.1016/j.vaccine.2005.02.004

14. Lim FJ, Blyth CC, Fathima P, de Klerk N, Moore HC. Record linkage study of the pathogen-specific burden of respiratory viruses in children. *Influenza Other Respir Viruses*. 2017;11(6):502-510. doi:10.1111/irv.12508

15. Alchikh M, Conrad T, Hoppe C, et al. Are we missing respiratory viral infections in infants and children? Comparison of a hospital-based quality management system with standard of care. *Clinical Microbiology and Infection*. 2019;25(3):380.e9-380.e16. doi:10.1016/j.cmi.2018.05.023

16. Moore HC, de Klerk N, Keil AD, et al. Use of data linkage to investigate the aetiology of acute lower respiratory infection hospitalisations in children. *J Paediatr Child Health*. 2012;48(6):520-528. doi:10.1111/j.1440-1754.2011.02229.x

17. Jones KH, Ford D v., Thompson S, Lyons RA. A profile of the SAIL databank on the UK secure research platform. *Int J Popul Data Sci*. 2019;4(2). doi:10.23889/ijpds.v4i2.1134

18. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885

19. Lyons RA, Jones KH, John G, et al. The SAIL databank: Linking multiple health and social care datasets. *BMC Med Inform Decis Mak*. 2009;9(1):3-3. doi:10.1186/1472-6947-9-3

20. Li Y, Hodgson D, Wang X, Atkins KE, Feikin DR, Nair H. Respiratory syncytial virus seasonality and prevention strategy planning for passive immunisation of infants in low-income and middle-income countries: a modelling study. *Lancet Infect Dis*. 2021;0(0). doi:10.1016/s1473-3099(20)30703-9

21. Hurst JR, Cumella A, Niklewicz CN, Philip KEJ, Singh V, Hopkinson NS. Long-Term Acceptability of Hygiene, Face Covering, and Social Distancing Interventions to Prevent Exacerbations in people living with Airways Diseases. *medRxiv*. Published online January 1, 2021:2021.04.09.21255189. doi:10.1101/2021.04.09.21255189

22. Clark TW, Medina M jo, Batham S, Curran MD, Parmar S, Nicholson KG. Adults hospitalised with acute respiratory illness rarely have detectable bacteria in the absence of COPD or pneumonia; viral infection predominates in a large prospective UK sample. *Journal of Infection*. 2014;69(5):507-515. doi:10.1016/j.jinf.2014.07.023

23. Mohan A, Chandra S, Agarwal D, et al. Prevalence of viral infection detected by PCR and RT-PCR in patients with acute exacerbation of COPD: A systematic review. *Respirology*. 2010;15(3):536-542. doi:10.1111/j.1440-1843.2010.01722.x

24. Vollenweider DJ, Frei A, Steurer-Stey CA, Garcia-Aymerich J, Puhan MA. Antibiotics for exacerbations of chronic obstructive pulmonary disease. *Cochrane Database of Systematic Reviews*. 2018;2018(10). doi:10.1002/14651858.CD010257.pub2

25. Falsey AR, Walsh EE, Esser MT, Shoemaker K, Yu L, Griffin MP. Respiratory syncytial virus–associated illness in adults with advanced chronic obstructive pulmonary disease and/or congestive heart failure. *J Med Virol*. 2019;91(1):65-71. doi:10.1002/jmv.25285

26. Murphy TF, Brauer AL, Grant BJB, Sethi S. Moraxella catarrhalis in chronic obstructive pulmonary disease: Burden of disease and immune response. *Am J Respir Crit Care Med*. 2005;172(2):195-199. doi:10.1164/rccm.200412-1747OC

27. Sethi S, Wrona C, Grant BJB, Murphy TF. Strain-specific Immune Response to Haemophilus influenzae in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2004;169(4):448-453. doi:10.1164/rccm.200308-1181oc

28. Wilkinson TMA, Aris E, Bourne S, et al. A prospective, observational cohort study of the seasonal dynamics of airway pathogens in the aetiology of exacerbations in COPD. *Thorax*. 2017;72(10):919-927. doi:10.1136/thoraxjnl-2016-209023

29. Greenberg SB, Allen M, Wilson J, Atmar RL. Respiratory viral infections in adults with and without chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2000;162(1):167-173. doi:10.1164/ajrccm.162.1.9911019

30. Rohde G, Wiethege A, Borg I, et al. Respiratory viruses in exacerbations of chronic obstructive pulmonary disease requiring hospitalisation: A case-control study. *Thorax*. 2003;58(1):37-42. doi:10.1136/thorax.58.1.37

31. Seemungal T, Harper-Owen R, Bhowmik A, et al. Respiratory viruses, symptoms, and inflammatory markers in acute exacerbations and stable chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2001;164(9):1618-1623. doi:10.1164/ajrccm.164.9.2105011

32. Beckham JD, Cadena A, Lin J, et al. Respiratory viral infections in patients with chronic, obstructive pulmonary disease. *Journal of Infection*. 2005;50(4):322-330. doi:10.1016/j.jinf.2004.07.011

33. Glezen WP, Greenberg SB, Atmar RL, Piedra PA, Couch RB. Impact of respiratory virus infections on persons with chronic underlying conditions. *J Am Med Assoc*. 2000;283(4):499-505. doi:10.1001/jama.283.4.499

34. Sethi S, Murphy TF. Bacterial infection in chronic obstructive pulmonary disease in 2000: A state-of-the-art review. *Clin Microbiol Rev*. 2001;14(2):336-363. doi:10.1128/CMR.14.2.336-363.2001

35. Brendish NJ, Mills S, Ewings S, Clark TW. Impact of point-of-care testing for respiratory viruses on antibiotic use in adults with exacerbation of airways disease. *Journal of Infection*. 2019;79(4):357-362. doi:10.1016/j.jinf.2019.06.010

36. Clark TW, Beard KR, Brendish NJ, et al. Clinical impact of a routine, molecular, point-of-care, test-and-treat strategy for influenza in adults admitted to hospital (FluPOC): a multicentre, open-label, randomised controlled trial. *Lancet Respir Med*. Published online 2020. doi:10.1016/S2213-2600(20)30469-0

37. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE. Respiratory Syncytial Virus Infection in Elderly and High-Risk Adults. *New England Journal of Medicine*. 2005;352(17):1749-1759. doi:10.1056/NEJMoa043951

38. Fleming DM, Taylor RJ, Lustig RL, et al. Modelling estimates of the burden of Respiratory Syncytial virus infection in adults and the elderly in the United Kingdom. *BMC Infect Dis*. 2015;15(1):443. doi:10.1186/s12879-015-1218-z

39. Sadoff J, de Paepe E, DeVincenzo J, et al. Prevention of Respiratory Syncytial Virus Infection in Healthy Adults by a Single Immunization of Ad26.RSV.preF in a Human Challenge Study. *J Infect Dis*. Published online January 5, 2021. doi:10.1093/infdis/jiab003

40. Campbell H, Bont L, Nair H. Respiratory syncytial virus (RSV) disease - new data needed to guide future policy. *J Glob Health*. 2015;5(2). doi:10.7189/jogh.05.020101

41. Bathoorn E, Groenhof F, Hendrix R, et al. Real-life data on antibiotic prescription and sputum culture diagnostics in acute exacerbations of COPD in primary care. *International Journal of COPD*. 2017;12:285-290. doi:10.2147/COPD.S120510

42. Higgins TL, Deshpande A, Zilberberg MD, et al. Assessment of the Accuracy of Using ICD-9 Diagnosis Codes to Identify Pneumonia Etiology in Patients Hospitalized With Pneumonia. *JAMA Netw Open*. 2020;3(7):e207750. doi:10.1001/jamanetworkopen.2020.7750

43. Moore HC, Lehmann D, Klerk N de, et al. How accurate are international classification of diseases-10 diagnosis codes in detecting influenza and pertussis hospitalizations in Children? *J Pediatric Infect Dis Soc*. 2014;3(3):255-260. doi:10.1093/jpids/pit036

Tables

Table 1: Respiratory pathogen testing and results for specimens associated with emergency admission for COPD

Respiratory pathogen	Count of COPD admissions tested	Count of COPD admissions positive (percent positive*)
H.Influenzae	5,013	837 (16.7%)
S.Pneumoniae	5,013	154 (3.1%)
M.Catarrhalis	5,013	131 (2.6%)
P.Aeruginosa	5,013	200 (4.0%)
Influenza	1,219	262 (21.5%)
Rhinovirus	1,208	150 (12.4%)
RSV	1,214	83 (6.8%)
Parainfluenza	1,208	56 (4.6%)
hMPV	1,208	63 (5.2%)
Adenovirus	1,007	10 (1.0%)
Enterovirus	1,007	10 (1.0%)

*Percent positive = detected on laboratory test within 7 days of admission as a percentage of the total tested

Table 2: Respiratory pathogen testing, results and percent positive rate for specimens associated with emergency admission for COPD, by month of admission (1/12/2016 to 30/11/2018 data combined)

Month of admission with COPD	Influenza (percent positive*)	Rhinovirus	RSV	Parainfluenza	hMPV	HINF	SPNE	MCAT	PAER
January	109/338 (32.2%)	30/333 (9.0%)	29/333 (8.7%)	6/333 (1.8%)	21/333 (6.3%)	119/578 (20.6%)	19/578 (3.3%)	16/578 (2.8%)	16/578 (2.8%)
February	70/202 (34.7%)	14/202 (6.9%)	5/202 (2.5%)	<5/202 (0.5 – 1.0%)	8/202 (3.9%)	82/491 (16.7%)	15/491 (3.1%)	7/491 (1.4%)	15/491 (3.1%)
March	32/122 (26.2%)	11/122 (9.0%)	<5/122 (0.8 – 3.3%)	<5/122 (0.8 - 3.3%)	5/122 (4.1%)	77/440 (17.5%)	15/440 (3.4%)	11/440 (2.5%)	15/440 (3.4%)
April	11/72 (15.2%)	14/72 (19.4%)	<5/72 (1.4 – 5.6%)	14/72 (19.4%)	<5/72 (1.4 - 5.6%)	73/405 (18.0%)	10/405 (2.5%)	13/405 (3.2%)	13/405 (3.2%)
May	<5/49 (2.0 – 8.2%)	5/49 (10.2%)	0/49	<5/49 (2.0 – 8.1%)	<5/49 (2.0 – 8.1%)	61/376 (16.2%)	10/376 (2.7%)	11/376 (2.9%)	13/376 (3.5%)
June	0/42	7/42 (16.7%)	0/42	5/42 (11.9%)	<5/42 (2.4 – 9.5%)	50/358 (14.0%)	10/358 (2.8%)	5/358 (1.4%)	20/358 (5.6%)
July	0/29	<5/29 (3.4 – 13.8%)	0/29	<5/29 (3.4 – 13.8%)	<5/29 (3.4 – 13.8%)	66/334 (19.8%)	5/334 (1.5%)	7/334 (2.1%)	14/334 (4.2%)
August	0/23	5/23 (21.7%)	0/23	<5/23 (4.3 – 17.4%)	<5/23 (4.3 – 17.4%)	50/307 (16.3%)	6/307 (2.0%)	7/307 (1.0%)	19/307 (5.8%)
September	0/31	10/31	<5/31	<5/31	<5/31	53/344	8/344	6/344	22/344

		(32.3%)	(3.2 – 12.9%)	(3.2 – 12.9%)	(3.2 – 12.9%)	(15.4%)	(2.3%)	(1.7%)	(6.4%)
October	<5/71 (1.4 – 5.6%)	22/69 (31.9%)	<5/71 (1.4 – 5.6%)	5/69 (7.2%)	<5/69 (1.4 – 5.8%)	66/394 (16.8%)	8/394 (2.0%)	9/394 (2.3%)	17/394 (4.3%)
November	<5/86 (1.2 – 4.7%)	8/82 (9.8%)	9/86 (10.5%)	<5/82 (1.2 – 4.9%)	<5/82 (1.2 – 4.9%)	63/471 (13.4%)	21/471 (4.5%)	19/471 (4.0%)	22/471 (4.7%)
December	30/154 (19.5%)	20/154 (13.0%)	32/154 (20.8%)	7/154 (4.5%)	14/154 (9.1%)	87/515 (16.9%)	27/515 (5.2%)	20/515 (3.9%)	14/515 (2.7%)

*Percent positive = detected on laboratory test within 7 days of admission as a percentage of the total tested

Small counts have been masked as <5.

HINF = *H.Influenzae*, SPNE = *S.Pneumoniae*, MCAT = *M.Catarrhalis*, PAER = *P.Aeruginosa*

Table 3: Contingency table showing accuracy of influenza ICD-10 codes associated with emergency admission for COPD, as compared to laboratory test result

	Laboratory test		
Diagnosis	**Positive**	**Negative**	**Total**
ICD-10 code present	155	25	180
ICD-10 code absent	107	932	1039
Total	262	957	1,219

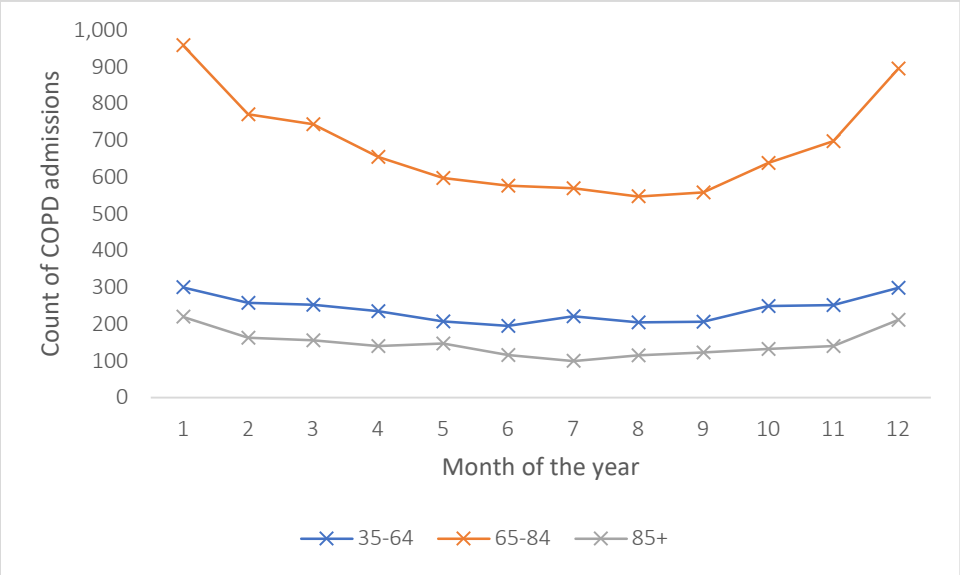
Sensitivity = 59.2% (52.9 – 65.1%); Specificity = 97.4% (96.2 – 98.3%); PPV = 86.1% (80.6 – 90.2%); NPV = 89.7% (88.3 – 91.0%)

Table 4: Contingency table showing accuracy of *P.aeruginosa* ICD-10 codes associated with emergency admission for COPD, as compared to laboratory test result

	Laboratory test		
Diagnosis	**Positive**	**Negative**	**Total**
ICD-10 code present	85	92	177
ICD-10 code absent	115	4,721	4836
Total	200	4,813	5,013

Sensitivity = 42.5% (35.6 – 49.7%); Specificity = 98.1% (97.7 – 98.5%); PPV = 48.0% (41.6 – 54.5%); NPV = 97.6% (97.3 – 97.9%)

Figure 1: Mean count of monthly emergency admissions for COPD in Wales, by age group



4.2 SUPPLEMENTARY MATERIAL

S1: ICD-10 codes for specific respiratory pathogens

Influenza - J09, J10, J11; Respiratory syncytial virus - J12.1, J20.5, J21.0, B97.4; Parainfluenza - J12.2, J20.4; Rhinovirus - J20.6; Human metapneumovirus - J12.3, J21.1; Adenovirus - J12.0, B34.0, B97.0; Enterovirus - B34.1, B97.1; *Haemophilus influenzae* - J14, J20.1, A49.2, B96.3, A41.3, *Streptococcus pneumoniae* - J13, B95.3, A40.3, and *Pseudomonas aeruginosa* - J15.1, B96.5. *Moraxella catarrhalis* is not associated with any pathogen-specific ICD-10 codes.

S2: Emergency admission for COPD and laboratory data linkage by NHS provider

NHS Provider	**COPD admissions linked to a laboratory record within 7 days of admission (%)**
Aneurin Bevan Local Health Board	6,341/6,369 (99.6%)
Abertawe Bro Morgannwg University Local Health Board	5,636/5,664 (99.5%)
Betsi Cadwaladr University Local Health Board	4,267/4,287 (99.5%)
Hywel Dda Local Health Board	3,503/3,515 (99.7%)
Cardiff and Vale University Local Health Board	3,051/3,065 (99.5%)
Cwm Taf Local Health Board	2,760/2,761 (99.9%)
Powys Teaching Local Health Board	24/40 (60.0%)
Velindre NHS Trust	14/14 (100%)

S3: Contingency tables showing accuracy of pathogen-specific ICD-10 codes associated with emergency admission for COPD, as compared to laboratory test result

Rhinovirus:

	Laboratory testing		
Diagnosis	**Positive**	**Negative**	**Total**
ICD-10 code present	0	0	0
ICD-10 code absent	150	1058	1208
Total	150	1058	1,208

Sensitivity = 0.0% (95% confidence interval 0.0 – 2.4%); Specificity = 100% (99.7 – 100%); PPV = n/a; NPV = 87.6% (85.6 – 89.4%)

Respiratory syncytial virus:

	Laboratory testing		
Diagnosis	**Positive**	**Negative**	**Total**
ICD-10 code present	10	0	10
ICD-10 code absent	73	1,131	1,204
Total	83	1,131	1,214

Sensitivity = 12.1% (95% CI 5.9 – 21.0%); Specificity = 100% (99.7 – 100%); PPV = 100%; NPV = 93.9% (93.5 – 94.4%)

H.influenzae:

	Laboratory test		
Diagnosis	**Positive**	**Negative**	**Total**
ICD-10 code present	176	16	192
ICD-10 code absent	661	4,160	4,821
Total	837	4,176	5,013

Sensitivity = 21.0% (95% CI 18.3 – 24.0%); Specificity = 99.6% (99.4 – 99.8%); PPV = 91.7% (86.9 – 94.8%); NPV = 86.3% (85.9 – 86.7%)

***S.pneumoniae*:**

	Laboratory test		
Diagnosis	**Positive**	**Negative**	**Total**
ICD-10 code present	31	13	44
ICD-10 code absent	123	4,846	4,969
Total	154	4,859	5,013

Sensitivity = 20.1% (95% CI 14.1 – 27.3%); Specificity = 99.7% (99.5 – 99.9%); PPV = 70.5% (56.0 – 81.7%); NPV = 97.5% (97.3 – 97.7%)

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.**

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	2	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.	2
				RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract.	2
				RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	2
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported	3,4		
Objectives	3	State specific objectives, including any prespecified hypotheses	4		
Methods					

Study Design	4	Present key elements of study design early in the paper	4, 5		
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	4		
Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case	NA	RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided.	5,6
				RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided.	NA
				RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers.	5,6	RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect	5

		Give diagnostic criteria, if applicable.		modifiers should be provided. If these cannot be reported, an explanation should be provided.	
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5,6		
Bias	9	Describe any efforts to address potential sources of bias	NA		
Study size	10	Explain how the study size was arrived at	NA		
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why	NA		
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	6		

Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	5
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	4, 5
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram	NA	RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	5
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)	NA		

Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures	NA		
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	19-24		
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses	NA		
Discussion					
Key results	18	Summarise key results with reference to study objectives	9		
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10	RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data,	10, 11

				and changing eligibility over time, as they pertain to the study being reported.	
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9,10,11		
Generalisability	21	Discuss the generalisability (external validity) of the study results	9,10,11		
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	13		
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	13

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

5 DISCUSSION

5.1 SUMMARY

It is undisputed that respiratory tract infections result in huge morbidity and mortality each year, impacting people across age groups and geographical boundaries,[60–64] with the World Health Organisation ranking LRTIs 4th amongst the leading causes of death globally in 2019.[65] However, estimates of impact vary by research group, and are – necessarily – extrapolated from the limited data available.[66] To develop, implement and evaluate interventions to reduce the burden of respiratory tract infections, high quality pathogen-specific data establishing the rate and severity of infection in different regions and patient groups is important.

ICD codes which relate to respiratory pathogens are often readily available in linkable electronic health record databases and have therefore been used for these purposes previously.[67–74] However, as described in my systematic scoping review, variation in case definitions can limit the research community's ability to draw reliable insights from EHR-based research. Furthermore, for diagnoses based on microbiology results, I show in Chapter 4 how linkage with laboratory data enables greater accuracy in identifying cases, as compared to use of clinical coding systems. Our data shows that ICD codes have a low sensitivity to identify respiratory pathogens (ranging from 0-59.2%, depending on the pathogen in question), and so relying on these codes for retrospective case ascertainment would significantly underestimate the burden of disease conferred by the pathogen under study. Other groups who had access to microbiology data also report that they would have grossly underestimated respiratory pathogen-related hospital admissions had they relied on diagnosis codes alone.[75–78]

Our data also shows that there is currently an under-utilisation of diagnostic microbiology for AECOPD, but when tested for, respiratory viruses were commonly detected in people admitted to hospital with AECOPD. Increased testing and thus, identification, of people with a viral rather than bacterial trigger for their exacerbation could help with timely isolation in healthcare settings, treatment with antivirals when appropriate, and the targeting of antibiotic therapy (especially useful in the face of growing antimicrobial resistance). The increased respiratory virus testing implemented during the COVID-19 pandemic may facilitate more widespread use in routine care going forward.

5.2 STRENGTHS AND LIMITATIONS

The strengths of this work include near-complete population coverage, with a base dataset covering all admissions to hospitals in Wales, and the high level (99.5%) of successful linkage between hospital admissions and microbiology data. This has enabled description of real-world microbiology testing in emergency admissions for COPD, an estimation of the contribution of bacterial and viral pathogens to severe AECOPD, and assessment of the accuracy of pathogen-specific ICD-10 diagnosis codes in this context.

There are some limitations to our work. The pattern of respiratory infections varies by year and by geographical region, and so the data presented here do not necessarily characterise patterns in other countries or current times. However, our estimates of the proportion of hospitalised AECOPD testing positive for respiratory viruses are similar to previous estimates.[79–81] Further, the large decline in emergency admissions for COPD during the COVID-19 pandemic (both in the UK and elsewhere),[82,83] when circulation of other respiratory viruses had plummeted,[84,85] adds weight to the impression that respiratory viruses trigger a substantial proportion of AECOPD.

The definition of COPD used was based on the diagnosis codes ascribed to hospital admissions, along with the added criterion of needing to be age ≥35 at the time of admission. However, it is unlikely that all individuals included have spirometry-confirmed COPD and/or are on maintenance treatment for COPD. I used my working definition in order to try to identify all admissions where the individual was treated as having an AECOPD. I felt that stipulating additional inclusion criteria would not be beneficial to the balance of sensitivity and specificity, especially since not all GP practices release data to SAIL, and so data regarding investigations and treatment would be missing for a proportion of people.

5.3 FUTURE WORK

This study has shown just one example of the utility of linking microbiology data to other routinely collected healthcare data. Further applications are almost limitless, and given our national dataset, could include estimating the incidence of a number of important infections in defined populations, for which data currently remain elusive.

For example, in the sphere of COPD, our dataset could be used to estimate the incidence of rarer infection outcomes, such as pulmonary aspergillosis, pneumocystis pneumonia (PCP), and

mycobacterial infection. By linking to primary care data, it would also be possible to investigate whether inhaled corticosteroid use in COPD was a risk factor for development of these infections, especially important given that a sizeable number of individuals are prescribed inhaled corticosteroid without a clear indication.[36] It would also be possible to provide long term follow up data regarding individuals with COPD receiving long-term antibiotics for exacerbation prevention, including a comparison of the effectiveness of different prophylactic antibiotic regimens, one of the areas highlighted for future research by NICE.[86]

Another group in which long-term, large-scale linked microbiology and healthcare data is invaluable is the immunosuppressed population. Growing numbers of individuals are receiving immunomodulatory therapy, but to date, estimating the risk of specific infections conferred by each therapy is limited by trials of limited duration and relatively small numbers, together with the fact that often pathogen-specific detail is not recorded or reported. This means that ambiguity and inconsistency exist regarding the populations that particular preventative strategies should be applied to. For example, use of antimicrobial prophylaxis for pneumocystis pneumonia has been advocated where risk of infection exceeds the risk of severe adverse events (around 3%),[87] but the risk of PCP conferred by many immunomodulatory therapies and disease states currently remains unknown.

Finally, with antimicrobial resistance a growing problem, linking resistance data to other routinely-collected healthcare data could help to highlight which patient groups are experiencing the biggest problem in this regard, and were most in need of antimicrobial stewardship interventions. Supporting the availability and access of linked data to enable analysis of antimicrobial resistance has been highlighted as one of the UK's commitments in tackling antimicrobial resistance in a national 5-year action plan 2019-2024.[88]

Several hurdles need to be overcome to achieve full potential from our data. These include cleaning of the available data without losing granularity, standardisation of the naming of tests and the units of measurement reported across the same test in the raw data (or careful clinical review to establish code sets for given diagnoses).[89]

5.4 Conclusions

To conclude, accurate and reproducible case definitions are imperative to draw reliable insights from EHR-based research. For diagnoses based on microbiology tests, linkage with

microbiology datasets enables greater accuracy in case ascertainment. Although currently, large-scale linkage of anonymised microbiology data to other routinely collected healthcare data for retrospective study is not possible in many regions, advances seen in health data research during the COVID-19 pandemic may prompt wider accessibility and linkage to microbiology data going forward. Linking microbiology data with other routinely collected healthcare data presents an array of future applications.

APPENDICES

Appendix 1: Information Governance Review Panel approval

SAIL

Shanya Sivakumaran

27/03/2020

Dear Shanya,

Re: Effects of withdrawal of inhaled corticosteroids in patients with COPD: a national population cohort study in Wales

Your proposal to use the SAIL databank has been assessed by the SAIL Collaboration Review System (CRS). The CRS consists of the SAIL Management Team and the Information Governance Review Panel (IGRP). The membership of the IGRP is comprised of senior representatives from:

- British Medical Association (BMA)
- National Research Ethics Service (NRES)
- Public Health Wales
- NHS Wales Informatics Service (NWIS)
- Swansea Bay University Health Board (SBUHB)
- Consumer Panel for Data Linkage Research

After careful consideration the proposal has been given approval to commence with analysis.

The project has been given a SAIL project number of [redacted] Please quote this number in all correspondence regarding this project.

Creation of project specific data view
Work will now commence on the creation of the project specific data view. The lead analyst contact for this will be Caroline Brooks and they will be in contact with you to confirm your data specification.

User access
Once the project specific data view has been created you will be allocated a SAIL Gateway user account if you require direct access to the data. Please contact [redacted] for details of how to apply for an account.

Publication statement
All publications must acknowledge the use of SAIL data.

Yours sincerely

[redacted]

SAIL DATABANK

[redacted]

REFERENCES

1. Global Initiative for Chronic Obstructive Lung Disease. *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease 2020 Report*.; 2020. Accessed January 7, 2021. www.goldcopd.org

2. Abbafati C, Machado DB, Cislaghi B, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-1222. doi:10.1016/S0140-6736(20)30925-9

3. Buist AS, McBurnie MA, Vollmer WM, et al. International variation in the prevalence of COPD (The BOLD Study): a population-based prevalence study. *Lancet*. 2007;370(9589):741-750. doi:10.1016/S0140-6736(07)61377-4

4. Public Health England. Smoking prevalence in adults - current smokers aged 18 to 64 by socio - economic group - GOV.UK. Published January 28, 2020. Accessed December 15, 2020. https://www.gov.uk/government/publications/smoking-prevalence-in-adults-current-smokers-aged-18-to-64-by-socio-economic-group

5. Public Health England. Wider Impacts of COVID-19. Published December 10, 2020. Accessed December 15, 2020. https://analytics.phe.gov.uk/apps/covid-19-indirect-effects/

6. Siddharthan T, Gupte A, Barnes PJ. Chronic obstructive pulmonary disease endotypes in low-and middle-income country settings: Precision medicine for all. *Am J Respir Crit Care Med*. 2020;202(2):171-172. doi:10.1164/rccm.202001-0165ED

7. Smith KR. National burden of disease in India from indoor air pollution. *Proc Natl Acad Sci U S A*. 2000;97(24):13286-13293. doi:10.1073/pnas.97.24.13286

8. Chan-Yeung M, Aït-Khaled N, White N, Ip MS, Tan WC. The burden and impact of COPD in Asia and Africa. *The International Journal of Tuberculosis and Lung Disease*. 2004;8(1):2-14.

9. Wang C, Xu J, Yang L, et al. Prevalence and risk factors of chronic obstructive pulmonary disease in China (the China Pulmonary Health [CPH] study): a national cross-sectional study. *The Lancet*. 2018;391(10131):1706-1717. doi:10.1016/S0140-6736(18)30841-9

10. Barnes PJ, Shapiro SD, Pauwels RA. Chronic obstructive pulmonary disease: Molecular and cellular mechanisms. *European Respiratory Journal*. 2003;22(4):672-688. doi:10.1183/09031936.03.00040703

11. Bestall JC, Paul EA, Garrod R, Garnham R, Jones PW, Wedzicha JA. Usefulness of the Medical Research Council (MRC) dyspnoea scale as a measure of disability in patients with chronic obstructive pulmonary disease. *Thorax*. 1999;54(7):581-586. doi:10.1136/thx.54.7.581

12. Pauwels R, Calverly P, Buist AS, et al. COPD exacerbations: The importance of a standard definition. *Respir Med*. 2004;98(2):99-107. doi:10.1016/j.rmed.2003.09.001

13. Mathioudakis AG, Janssens W, Sivapalan P, et al. Acute exacerbations of chronic obstructive pulmonary disease: In search of diagnostic biomarkers and treatable traits. *Thorax*. 2020;75(6):520-527. doi:10.1136/thoraxjnl-2019-214484

14. Wedzicha JA, Seemungal TA. COPD exacerbations: defining their cause and prevention. *Lancet*. 2007;370(9589):786-796. doi:10.1016/S0140-6736(07)61382-8

15. Ko FWS, Tam W, Tze WW, et al. Temporal relationship between air pollutants and hospital admissions for chronic obstructive pulmonary disease in Hong Kong. *Thorax*. 2007;62(9):779-784. doi:10.1136/thx.2006.076166

16. NHS England. *Overview of Potential to Reduce Lives Lost from Chronic Obstructive Pulmonary Disease (COPD)*.; 2014. Accessed January 7, 2021. https://www.england.nhs.uk/wp-content/uploads/2014/02/rm-fs-6.pdf

17. 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 PART I):1418-1422. doi:10.1164/ajrccm.157.5.9709032

18. Miravitlles M, Ferrer M, Pont À, et al. Effect of exacerbations on quality of life in patients with chronic obstructive pulmonary disease: A 2 year follow up study. *Thorax*. 2004;59(5):387-395. doi:10.1136/thx.2003.008730

19. Rothnie KJ, Connell O, Müllerová H, et al. Myocardial infarction and ischemic stroke after exacerbations of chronic obstructive pulmonary disease. *Ann Am Thorac Soc*. 2018;15(8):935-946. doi:10.1513/AnnalsATS.201710-815OC

20. Kunisaki KM, Dransfield MT, Anderson JA, et al. Exacerbations of chronic obstructive pulmonary disease and cardiac events a post hoc cohort analysis from the SUMMIT randomized clinical trial. *Am J Respir Crit Care Med*. 2018;198(1):51-57. doi:10.1164/rccm.201711-2239OC

21. Hurst JR, Vestbo J, Anzueto A, et al. Susceptibility to Exacerbation in Chronic Obstructive Pulmonary Disease. *New England Journal of Medicine*. 2010;363(12):1128-1138. doi:10.1056/nejmoa0909883

22. National Institute for Health and Care Excellence. *Chronic Obstructive Pulmonary Disease in over 16s: Diagnosis and Management NICE Guideline*.; 2018. Accessed December 15, 2020. www.nice.org.uk/guidance/ng115

23. Nacul L, Soljak M, Samarasundera E, et al. COPD in England: a comparison of expected, model-based prevalence and observed prevalence from general practice data†. *J Public Health (Oxf)*. 2011;33(1):108-116. doi:10.1093/pubmed/fdq031

24. National Institutes of Health. Pivotal Study to Assess the Efficacy, Safety and Tolerability of Dupilumab in Patients With Moderate-to-severe COPD With Type 2 Inflammation. Accessed December 15, 2020. https://clinicaltrials.gov/ct2/show/NCT03930732

25. Dornhorst AC. RESPIRATORY INSUFFICIENCY. *The Lancet*. 1955;265(6876):1185-1187. doi:10.1016/S0140-6736(55)90689-9

26. Robin ED, O'Neill RP. The fighter versus the nonfighter. *Arch Environ Health*. 1963;7(2):125-129. doi:10.1080/00039896.1963.10663506

27. Johnson MA, Woodcock AA, Rehahn M, Geddes DM. Are "pink puffers" more breathless than "blue bloaters"? *Br Med J*. 1983;286(6360):179-182. doi:10.1136/bmj.286.6360.179

28. Colledge NR, Walker BR, Ralston S, Davidson S. *Davidson's Principles and Practice of Medicine 21st Edition*. 21st ed. (Colledge NR, Walker BR, Ralston S, eds.). Edinburgh ; New York : Churchill Livingstone/Elsevier, 2010.; 2010.

29. Goldman L, Schafer A. *Goldman's Cecil Medicine*. Vol 1. 24th ed. (Goldman L, Schafer A, eds.). Elsevier Saunders; 2012.

30. Global Initiative for Chronic Obstructive Lung Disease. *Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease (2021 Report)*.; 2020. Accessed December 15, 2020. https://goldcopd.org/wp-content/uploads/2020/11/GOLD-REPORT-2021-v1.1-25Nov20_WMV.pdf

31. Anthonisen NR, Skeans MA, Wise RA, Manfreda J, Kanner RE, Connett JE. The effects of a smoking cessation intervention on 14.5-year mortality: A randomized clinical trial. *Ann Intern Med*. 2005;142(4):233-239. doi:10.7326/0003-4819-142-4-200502150-00005

32. Wilson ME, Dobler CC, Morrow AS, et al. Association of Home Noninvasive Positive Pressure Ventilation with Clinical Outcomes in Chronic Obstructive Pulmonary Disease: A Systematic Review and Meta-analysis. *JAMA - Journal of the American Medical Association*. 2020;323(5):455-465. doi:10.1001/jama.2019.22343

33. Viniol C, Vogelmeier CF. Exacerbations of COPD. *European Respiratory Review*. 2018;27(147). doi:10.1183/16000617.0103-2017

34. Royal College of Physicians. Time to Take a Breath. National Chronic Obstructive Pulmonary (COPD) Audit Programme: Clinical Audit of COPD in Primary Care in Wales 2014–15.; 2016. www.rcplondon.ac.uk/COPD

35. Mak V. Treatment guidelines for COPD - Going for GOLD? *Primary Care Respiratory Update*. 2017;4(2):19-24. Accessed December 15, 2020. https://www.pcrs-uk.org/sites/pcrs-uk.org/files/GoingforGOLD_PCRUSept2017.pdf

36. Price D, Miravitlles M, Pavord I, et al. First maintenance therapy for COPD in the UK between 2009 and 2012: A retrospective database analysis. *NPJ Prim Care Respir Med*. 2016;26:16061. doi:10.1038/npjpcrm.2016.61

37. Gruffydd-Jones K, Brusselle G, Jones R, et al. Changes in initial COPD treatment choice over time and factors influencing prescribing decisions in UK primary care: In UK primary care: A real-world, retrospective, observational. *NPJ Prim Care Respir Med*. 2016;26(1):1-7. doi:10.1038/npjpcrm.2016.2

38. Crooks CJ, Card TR, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. *BMC Health Serv Res*. 2012;12(1):392. doi:10.1186/1472-6963-12-392

39. al Sallakh MA, Vasileiou E, Rodgers SE, Lyons RA, Sheikh A, Davies GA. Defining asthma and assessing asthma outcomes using electronic health record data: a systematic scoping review. *Eur Respir J*. 2017;49(6):1700204. doi:10.1183/13993003.00204-2017

40. Crossfield SSR, Lai LYH, Kingsbury SR, et al. Variation in methods, results and reporting in electronic health record-based studies evaluating routine care in gout: A systematic review. Nurmohamed M, ed. *PLoS One*. 2019;14(10):e0224272. doi:10.1371/journal.pone.0224272

41. Rubbo B, Fitzpatrick NK, Denaxas S, et al. Use of electronic health records to ascertain, validate and phenotype acute myocardial infarction: A systematic review and recommendations. *Int J Cardiol*. 2015;187:705-711. doi:10.1016/j.ijcard.2015.03.075

42. Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. *BMJ*. 2010;341(19). doi:10.1136/bmj.c4226

43. Millett ERC, Quint JK, de Stavola BL, Smeeth L, Thomas SL. Improved incidence estimates from linked vs. stand-alone electronic health records. In: *Journal of Clinical Epidemiology*. Vol 75. Elsevier USA; 2016:66-69. doi:10.1016/j.jclinepi.2016.01.005

44. Herrett E, Shah AD, Boggon R, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: Cohort study. *BMJ (Online)*. 2013;346(7909). doi:10.1136/bmj.f2350

45. Roche N, Reddel H, Martin R, et al. Quality standards for real-world research: Focus on observational database studies of comparative effectiveness. In: *Annals of the American Thoracic Society*. Vol 11. American Thoracic Society; 2014:99-104. doi:10.1513/AnnalsATS.201309-300RM

46. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):e1001885. doi:10.1371/journal.pmed.1001885

47. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *Journal of the American Medical Informatics Association*. 2013;20(1):144-151. doi:10.1136/amiajnl-2011-000681

48. Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. Quality of clinical primary care and targeted incentive payments: An observational study. *British Journal of General Practice*. 2007;57(539):449-454. Accessed May 20, 2021. /pmc/articles/PMC2078183/

49. Minchin M, Roland M, Richardson J, Rowark S, Guthrie B. Quality of Care in the United Kingdom after Removal of Financial Incentives. *New England Journal of Medicine*. 2018;379(10):948-957. doi:10.1056/nejmsa1801495

50. Raman SR, Curtis LH, Temple R, et al. Leveraging electronic health records for clinical research. *Am Heart J*. 2018;202:13-19. doi:10.1016/j.ahj.2018.04.015

51. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *Journal of Public Health (Oxf)*. 2014;36(4):684-692. doi:10.1093/pubmed/fdt116

52. Petersen I, Welch CA, Nazareth I, et al. Health indicator recording in UK primary care electronic health records: Key implications for handling missing data. *Clin Epidemiol*. 2019;11:157-167. doi:10.2147/CLEP.S191437

53. Embi PJ, Payne PRO. Evidence generating medicine: Redefining the research-practice relationship to complete the evidence cycle. *Med Care*. 2013;51(8 SUPPL.3). doi:10.1097/MLR.0b013e31829b1d66

54. Diab N, Gershon AS, Sin DD, et al. Underdiagnosis and overdiagnosis of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2018;198(9):1130-1139. doi:10.1164/rccm.201804-0621CI

55. Cox E, Martin BC, van Staa T, Garbe E, Siebert U, Johnson ML. Good research practices for comparative effectiveness research: Approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: The international society for pharmacoeconomics and outcomes research good research practices for retrospective database analysis task force report - Part II. *Value in Health*. 2009;12(8):1053-1061. doi:10.1111/j.1524-4733.2009.00601.x

56. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. Good research practices for comparative effectiveness research: Analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: The ISPOR good research practices for retrospective database analysis task force report - Part III. *Value in Health*. 2009;12(8):1062-1073. doi:10.1111/j.1524-4733.2009.00602.x

57. Jones KH, Ford D v., Jones C, et al. A case study of the secure anonymous information linkage (SAIL) gateway: A privacy-protecting remote access system for health-related research and evaluation. *J Biomed Inform*. 2014;50(100):196-204. doi:10.1016/j.jbi.2014.01.003

58. Lyons RA, Jones KH, John G, et al. The SAIL databank: Linking multiple health and social care datasets. *BMC Med Inform Decis Mak*. 2009;9(1):3-3. doi:10.1186/1472-6947-9-3

59. Digital Health and Care Wales. NHS Wales Data Dictionary. Published September 15, 2022. Accessed December 15, 2022. http://www.datadictionary.wales.nhs.uk/

60. Lafond KE, Porter RM, Whaley MJ, et al. Global burden of influenza-associated lower respiratory tract infections and hospitalizations among adults: A systematic review and meta-analysis. Kretzschmar MEE, ed. *PLoS Med*. 2021;18(3):e1003550. doi:10.1371/journal.pmed.1003550

61. Li Y, Johnson EK, Shi T, et al. National burden estimates of hospitalisations for acute lower respiratory infections due to respiratory syncytial virus in young children in 2019 among 58 countries: a modelling study. *Lancet Respir Med*. 2021;9(2):175-185. doi:10.1016/S2213-2600(20)30322-2

62. Wang X, Li Y, Deloria-Knoll M, et al. Global burden of acute lower respiratory infection associated with human metapneumovirus in children under 5 years in 2018: a systematic review and modelling study. *Lancet Glob Health*. 2021;9(1):e33-e43. doi:10.1016/S2214-109X(20)30393-4

63. Shi T, Denouel A, Tietjen AK, et al. Global Disease Burden Estimates of Respiratory Syncytial Virus-Associated Acute Respiratory Infection in Older Adults in 2015: A Systematic Review and Meta-Analysis. *J Infect Dis*. 2020;222(7):S577-S583. doi:10.1093/infdis/jiz059

64. Troeger C, Blacker B, Khalil IA, et al. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis*. 2018;18(11):1191-1210. doi:10.1016/S1473-3099(18)30310-4

65. World Health Organisation. The top 10 causes of death. Published December 9, 2020. Accessed December 15, 2022. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death

66. Murdoch DR, Howie SRC. The global burden of lower respiratory infections: making progress, but we need to do better. *Lancet Infect Dis*. 2018;18(11):1162-1163. doi:10.1016/S1473-3099(18)30407-9

67. Hauge SH, Bakken IJ, de Blasio BF, Håberg SE. Risk conditions in children hospitalized with influenza in Norway, 2017–2019. *BMC Infect Dis*. 2020;20(1). doi:10.1186/s12879-020-05486-6

68. Hungerford D, Ibarz-Pavon A, Cleary P, French N. Influenza-associated hospitalisation, vaccine uptake and socioeconomic deprivation in an English city region: An ecological study. *BMJ Open*. 2018;8(12):23275. doi:10.1136/bmjopen-2018-023275

69. Moss JWE, Davidson C, Mattock R, Gibbons I, Mealing S, Carroll S. Quantifying the direct secondary health care cost of seasonal influenza in England. *BMC Public Health*. 2020;20(1). doi:10.1186/s12889-020-09553-0

70. Lau K, Hauck K, Miraldo M. Excess influenza hospital admissions and costs due to the 2009 H1N1 pandemic in England. *Health Economics (United Kingdom)*. 2019;28(2):175-188. doi:10.1002/hec.3834

71. Spagnuolo PJ, Zhang M, Xu Y, et al. Effects of antiviral treatment on influenza-related complications over four influenza seasons: 2006–2010. *Curr Med Res Opin*. 2016;32(8):1399-1407.

72. San-Román-Montero JM, Gil Prieto R, Gallardo Pino C, Hinojosa Mena J, Zapatero Gaviria A, Gil De Miguel A. Inpatient hospital fatality related to coding (ICD-9-CM) of the influenza diagnosis in Spain (2009-2015). *BMC Infect Dis*. 2019;19(1). doi:10.1186/s12879-019-4308-5

73. Tong S, Amand C, Kieffer A, Kyaw MH. Incidence of Respiratory Syncytial Virus Related Health Care Utilization in The United States. *J Glob Health*. 2020;10(2):1-10. doi:10.7189/jogh.10.020422

74. Huh K, Kim YE, Ji W, et al. Decrease in hospital admissions for respiratory diseases during the COVID-19 pandemic: a nationwide claims study. *Thorax*. Published online March 29, 2021:thoraxjnl-2020-216526. doi:10.1136/thoraxjnl-2020-216526

75. Lim FJ, Blyth CC, Fathima P, de Klerk N, Moore HC. Record linkage study of the pathogen-specific burden of respiratory viruses in children. *Influenza Other Respir Viruses*. 2017;11(6):502-510. doi:10.1111/irv.12508

76. Hardelid P, Verfuerden M, McMenamin J, Gilbert R. Risk factors for admission to hospital with laboratory-confirmed influenza in young children: birth cohort study. *Eur Respir J*. 2017;50(3):1700489. doi:10.1183/13993003.00489-2017

77. Nicholson KG, McNally T, Silverman M, Simons P, Stockton JD, Zambon MC. Rates of hospitalisation for influenza, respiratory syncytial virus and human metapneumovirus among infants and young children. *Vaccine*. 2006;24(1):102-108. doi:10.1016/j.vaccine.2005.02.004

78. Alchikh M, Conrad T, Hoppe C, et al. Are we missing respiratory viral infections in infants and children? Comparison of a hospital-based quality management system with standard of care. *Clinical Microbiology and Infection*. 2019;25(3):380.e9-380.e16. doi:10.1016/j.cmi.2018.05.023

79. Clark TW, Medina M jo, Batham S, Curran MD, Parmar S, Nicholson KG. Adults hospitalised with acute respiratory illness rarely have detectable bacteria in the absence of COPD or pneumonia; viral infection predominates in a large prospective UK sample. *Journal of Infection*. 2014;69(5):507-515. doi:10.1016/j.jinf.2014.07.023

80. Mohan A, Chandra S, Agarwal D, et al. Prevalence of viral infection detected by PCR and RT-PCR in patients with acute exacerbation of COPD: A systematic review. *Respirology*. 2010;15(3):536-542. doi:10.1111/j.1440-1843.2010.01722.x

81. Brendish NJ, Mills S, Ewings S, Clark TW. Impact of point-of-care testing for respiratory viruses on antibiotic use in adults with exacerbation of airways disease. *Journal of Infection*. 2019;79(4):357-362. doi:10.1016/j.jinf.2019.06.010

82. Cookson W, Moffatt M, Rapeport G, Quint J. A Pandemic Lesson for Global Lung Diseases: Exacerbations Are Preventable. *Am J Respir Crit Care Med*. 2022;205(11):1271-1280. doi:10.1164/rccm.202110-2389CI

83. Alsallakh MA, Sivakumaran S, Kennedy S, et al. Impact of COVID-19 lockdown on the incidence and mortality of acute exacerbations of chronic obstructive pulmonary disease: national interrupted time series analyses for Scotland and Wales. *BMC Med*. 2021;19(1):124. doi:10.1186/s12916-021-02000-w

84. Huang QS, Wood T, Jelley L, et al. Impact of the COVID-19 nonpharmaceutical interventions on influenza and other respiratory viral infections in New Zealand. *Nat Commun*. 2021;12(1):1001. doi:10.1038/s41467-021-21157-9

85. Groves HE, Papenburg J, Mehta K, et al. The effect of the COVID-19 pandemic on influenza-related hospitalization, intensive care admission and mortality in children in Canada: A population-based study. *The Lancet Regional Health – Americas*. 2022;7. doi:10.1016/j.lana.2021.100132

86. National Institute for Health and Care Excellence. Chronic obstructive pulmonary disease in over 16s: diagnosis and management. Published July 26, 2019. Accessed December 15, 2022. https://www.nice.org.uk/guidance/ng115/resources/research-recommendations-from-an-individual-piece-of-guidance

87. Green H, Paul M, Vidal L, Leibovici L. Prophylaxis for Pneumocystis pneumonia (PCP) in non-HIV immunocompromised patients. In: Green H, ed. *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd; 2007. doi:10.1002/14651858.CD005590.pub2

88. Department of Health & Social Care. Tackling antimicrobial resistance 2019 to 2024: addendum to the UK's 5-year national action plan.

89. Davies G, Akbari A, Abbasizanjani H, et al. The Welsh Results Reports Service (WRRS) Data. Published April 2022. Accessed December 15, 2022. https://saildatabank.com/wp-content/uploads/2022/04/Data-Explained-The-Welsh-Results-Reports-Service-WRRS-Data.pdf