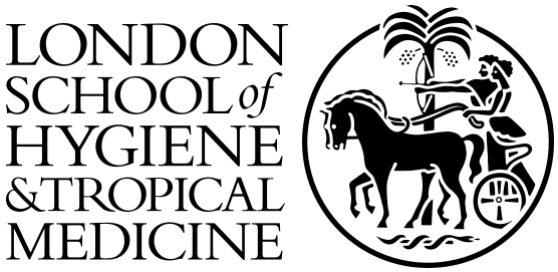


LSHTM Research Online

Nissen, FW; (2019) Asthma in electronic health records: validity and phenotyping. PhD (research paper style) thesis, London School of Hygiene & Tropical Medicine. DOI: https://doi.org/10.17037/PUBS.04653375

Downloaded from: http://researchonline.lshtm.ac.uk/4653375/

DOI: https://doi.org/10.17037/PUBS.04653375

Usage Guidelines:

Please refer to usage guidelines at https://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/

https://researchonline.lshtm.ac.uk

Asthma in electronic health records: validity and phenotyping

Francis Willem Nissen

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy of the University of London

April 2019

Department of Non-communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Scholarship funding by GSK

Affiliated to the LSHTM EHR research group

Advisors: Dr Ian Douglas, Prof Dr Liam Smeeth,

Dr Jennifer Quint and Dr Hana Müllerova

Declaration page

I, Francis Nissen, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been clearly indicated in the thesis.

Francis Nissen

Signature

Date

Use of published work

Three papers have been published based on work undertaken for this thesis, the fourth paper is in press and the fifth paper is under peer review at the time of submission. Work for these papers were carried out as part of the PhD and took place during the period of registration for the PhD. For these papers, Francis Nissen was the lead and corresponding author, and prepared all protocols and drafts of the papers. The contributions of the co-authors were restricted to providing study advice and comments on the drafts prepared by FN.

Abstract

This PhD thesis explores the validation of asthma in electronic health records (EHR) and the characteristics of asthma phenotypes in the UK using CPRD GOLD, HES and ONS data. The absence of a universal case definition, the overlap with other diseases and the incomplete recording of diagnostic markers makes the identification of asthma patients in EHR challenging. Furthermore, asthma phenotypes have previously been established based on cluster analysis in small populations, but their prevalence, treatment and outcomes in the general population have not been investigated.

Firstly, I conducted a systematic review to understand how past epidemiological studies have validated asthma recording in EHR, including a critical appraisal and list of test measure values for the selected studies.

Secondly, I validated algorithms to reliably ascertain the asthma status of patients in CPRD GOLD. This validation study identified multiple algorithms with PPV greater than 80%. The most practical algorithm (presence of a specific asthma diagnostic code) had a PPV of 86.4 (95% CI:77.4-95.4).

Thirdly, I quantified the concomitant occurrence of asthma in COPD patients and vice versa in CPRD GOLD. After detailed case review, concomitant asthma and COPD was concluded in 14.8% of validated asthma patients and in 14.5% of validated COPD patients. However, asthma diagnoses may be unreliable in COPD patients, as over 50% of COPD patients had received an asthma code.

Finally, I examined the prevalence, treatment, outcomes and characteristics of established asthma phenotypes in CPRD GOLD. Only a minority of patients (37.3%) were classified into these phenotypes using stringent inclusion criteria. Exacerbation

rates/1000PY were lowest for benign asthma (106.8 [95% CI:101.2-112.3]), and highest for obese non-eosinophilic asthma (469.0 [95% CI:451.7-486.2]).

In conclusion, this thesis provides information on the validation of asthma diagnoses in EHR and the prevalence, treatment, outcomes of predefined asthma phenotypes in the UK primary care population.

The only true wisdom is in knowing you know nothing.

Socrates (470–399 BC)

Acknowledgements

There are many people that helped me throughout my degree to finish this thesis, and I am indebted to every one of them. In particular:

My main supervisor at the London School of Hygiene and Tropical Medicine, Ian Douglas, for his expertise, kindness and enthusiasm.

My co-supervisor Liam Smeeth for his experience, support and guidance in all areas of epidemiology.

Jennifer Quint for the inspiration she provided for many of the studies included in this thesis, in addition to her advice and direction.

My industry supervisor Hana Müllerova for her timely feedback and guidance.

My past and present office mates and fellow PhD students for their advice, insights and camaraderie, in particular Sam Wilkinson, Masao Iwagami, Kate Reiss, Tessa Roberts, Kieran Rothnie and Eiman Alawadhi.

My co-authors Daniel Morales, Chloe Bloom and Neil Pearce for lending their experience, contributions and feedback.

My upgrading committee Jo Feary and Sinead Langan and my DRDC, Claudia Allemani.

GSK for enabling me to pursue this PhD financially, and CPRD for their support and advice on the handling of the data.

My friends and family for their patience, especially during the writing-up stage of this thesis.

All members of the electronic health records group at the LSHTM.

Finally, and most importantly, to Marjan for her love and unwavering support and encouragement.

Funding

Francis Nissen was supported by GSK to undertake the research degree at the London School of Hygiene and Tropical Medicine. The funder had no role in the choice of the research questions and study designs for the included studies.

Abbreviations

A&E	Accidents and Emergencies
BMI	Body Mass Index
BMJ	British Medical Journal
BTS	British Thoracic Society
CI	Confidence Interval
COPD	Chronic Obstructive Pulmonary Disease
CPP	Cumulative Patient Profile
CPRD	Clinical Practice Research Database
ED	Emergency Department (US term)
EHR	Electronic Health Records
FeNO	Fractured exhaled Nitric oxide
FEV1	Forced Expiratory Volume in 1 second
FVC	Forced Vital Capacity
GWAS	Genome-Wide Association Study
GINA	Global Initiative for Asthma
GP	General Practitioner
GORD	Gastro-Oesophageal Reflux Disease
GPRD	General Practice Research Database (older version of CPRD)
HES	Hospital Episode Statistics
HES APC	Hospital Episode Statistics Admitted Patient Care
HPHC	Harvard Pilgrim Health Care
ICD	International Classification of Disease
ICPI	Integrated Primary Care Information database
ICS	Inhaled CorticoSteroids
IMD	Index of Multiple Deprivation
IQR	Interquartile Range
ISAAC	International Study of Asthma and Allergy in Children
ISAC	Independent Scientific Advisory Committee for MHRA database research
KPNW	Kaiser Permanente North West division

LABA	Long-Acting Beta Agonist
LAMA	Long-Acting Muscarinic Antagonist
LSHTM	London School of Hygiene and Tropical Medicine
MBRN	Medical Birth Registry of Norway
MHRA	Medicines and Healthcare products Regulatory Agency
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
NOS	Not Otherwise Specified
NorPD	Norwegian Prescription Database
NPV	Negative Predictive Value
OCS	Oral Corticosteroids
ONS	Office of National Statistics
OSCAR	Outside Claims database
OXMIS	Oxford Medical Information System
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
QOF	Quality Outcomes Framework
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
Read codes	Standard clinical terminology system used in the CPRD GOLD
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
SABA	Short Acting Beta Agonist
SAGE	Study of Asthma, Genes and the Environment
SIGN	Scottish Intercollegiate Guidelines Network
SES	Socio-Economic Status
TH2 cell	Type 2 Helper Cell
TOPS	The Outpatient Pharmacy System
UTS	Up-To-Standard
WHO	World Health Organisation

Foreword

This thesis is structured as follows: the first chapter describes the thesis background with an overview of asthma, electronic health records and asthma phenotypes, while the second chapter outlines the data sources used and the rationale for using them. The next four chapters are presented as a series of research papers, and the last chapter summarises and discusses the overall results. The research papers are presented as pre-print versions for ease of reading, and all references are numbered and presented at the end of each chapter. Code lists are included in the appendix of this thesis, and other supplementary information specific to a chapter is included at the end of each chapter. All included studies are or will be published using Open Access.

All research papers presented in this thesis are my work (Francis Nissen). I designed the research protocols, obtained the ethical approvals and the data, and performed the data management, analysis and interpretation. I wrote the first draft of the manuscripts, and the final draft after incorporating the comments from the co-authors and advisors. The inclusion and bias assessment of the studies in the systematic review was simultaneously carried out by the author of this thesis and a second PhD student as the second reviewer (Samantha Wilkinson). The assessment of asthma and COPD status based on the questionnaires was carried out by one respiratory physician and one general practitioner familiar with the CPRD GOLD (Jennifer K Quint and Daniel Morales, respectively). The list of codes used in this thesis were developed by the author or, in the case of covariates, by collaborators at the London School of Hygiene and Tropical Medicine and Imperial College. Four papers included in this thesis have been published or are in print, the fifth is under peer review.

Table of contents

Declaration page 3

Abstract 4

Acknowledgements 7

Funding 7

Abbreviations 8

Foreword 10

Table of contents 11

Table of figures 17

Table of tables 18

Chapter 1: Introduction **19**

Summary 19

1.1 Asthma: disease background 20

1.1.1 Definition of asthma 20

1.1.2 Asthma in history 21

1.1.3 Diagnosis of asthma 24

1.1.4 Epidemiology of asthma 25

1.1.5 Pathophysiology of asthma 27

1.1.6 Treatment of asthma 28

1.1.7 Asthma exacerbations 31

1.2 Asthma phenotypes 31

1.2.1 Background 31

1.2.2 Characteristics associated with asthma 33

1.2.3 Phenotypes identified using cluster analysis 41

1.3 Electronic healthcare records 46

1.3.1 Overview 46

1.3.2 Asthma identification in electronic health records 49

1.4 Thesis aim, rationale and objectives 51

1.4.1 Aim 51

1.4.2 Rationale 51

1.4.3 Objectives 51

1.4.4 Thesis organisation 52

Chapter 2: Data sources **53**

Summary 53

2.1 EHR databases and front-end software systems in the UK 54

2.1 Clinical Practice Research Datalink GOLD (CPRD GOLD) 55

2.1.1 Background 55

2.1.2 Data architecture 56

2.1.3 Strengths and limitations 61

2.2 Hospital Episodes Statistics (HES) 65

2.3 Office of National Statistics (ONS) 68

2.4 Asthma questionnaire 69

2.5 Data management 71

Chapter 3: Systematic review: Validation of asthma recording in electronic health records **72**

Summary 72

3.1 Preface 73

3.2 Research paper 1: Protocol for a systematic review 76

Abstract 76

Strengths and limitations of this study 78

Background 78

Research question 80

Methods 81

3.3 Research paper 2: Systematic review 85

Abstract 85

Strengths and limitations of this study 87

Background 87

Methods 89

Results 91

Discussion 107

Conclusion 109

3.4 Appendix 111

Algorithms used for literature review 111

QUADAS-2 112

Chapter 4: Validation of asthma recording in the Clinical Practice Research Datalink (CPRD) **114**

Summary 114

4.1 Preface 115

4.2 Research paper 117

Abstract 117

Article summary 119

Background 119

Methods 120

Results 130

Discussion 133

Conclusion 137

4.3 Appendix 139

Questionnaire for asthma validation study 139

Chapter 5: Quantifying concomitant diagnosis of asthma and COPD in UK primary care 142

Summary 142

5.1 Preface 143

5.2 Research paper 146

Abstract 146

How this fits in 147

Introduction 148

Methods 149

Results 154

Discussion 159

Conclusion 163

Chapter 6: Clinical profile of pre-defined asthma phenotypes in a large cohort of UK primary care patients (CPRD) 165

Summary 165

6.1 Preface 166

6.2 Research paper 170

Abstract 170

Introduction 171

Methods 174

Results 183

Discussion 189

Conclusion 194

6.3 Appendix 196

Research paper cover sheet 198

Chapter 7: Thesis summary and discussion **199**

Summary 199

7.1 Synopsis of findings by research objective 201

Objective 1: Understand how past epidemiological studies have validated asthma in EHR through a systematic review 201

Objective 2: Validate the recording of the diagnosis of asthma in the CPRD GOLD 204

Objective 3: Quantify the concomitant occurrence of asthma in COPD patients and vice versa in CPRD GOLD 205

Objective 4: Identify established asthma phenotypes in CPRD GOLD by studying risk factors and explore the variation of asthma severity (defined by treatment steps) by phenotype 206

Objective 5: Examine the difference in asthma control by asthma phenotype, stratified by treatment step 208

7.2 Overall strengths 209

7.3 Overall limitations 211

Data sources 211

Information bias 212

Confounding 214

Generalisability 214

Sample size 214

Changing or conflicting guidelines 215

7.4 Recommendations for practice 215

1: Asthma may be overdiagnosed in people with COPD 215

2: Asthma patients are identifiable through EHR data 216

3: Established asthma phenotypes are identifiable to a certain extent in primary care 217

7.5 Recommendations for research 217

1: Recommendations for future validation studies of asthma recording in other databases 217

2: Asthma patients are identifiable from the CPRD GOLD 218

3: Identifying patients with concomitant asthma and COPD in EHR 219

4: Asthma phenotyping using primary care EHR 219

7.6 Future work 222

7.7 Overall conclusions 225

REFERENCES 227

APPENDIX 251

Codelist 1: Definite asthma 251

Codelist 2: Possible asthma 254

Codelist 3: Asthma medication 255

Codelist 4: Oral corticosteroid codes 276

Codelist 5: Antihistamines 278

Codelist 6: COPD codes 283

Codelist 7: GORD 283

Codelist 8: Atopy 284

Codelist 9: Anxiety 287

Codelist 10: Depression 288

Table of figures

Chapter 1

Figure 1: Asthma pathophysiology. 21
Figure 2: Summary of asthma management in adults from the BTS/SIGN guidelines. 29
Figure 3: Clinical phenotypes of asthma by Haldar P. 44

Chapter 2

Figure 1: Data flow in CPRD. 58
Figure 2: Strategy for identifying codes for asthma in the CPRD GOLD 61

Chapter 3

Figure 1: PRISMA diagram 92

Chapter 4

Figure 1: Study population 123
Figure 2: PPV as diagnosed by the patient's own GP, and agreement between study physicians 132

Chapter 5

Figure 1: Validated asthma patients 157
Figure 2: Validated COPD patients 158

Chapter 6

Figure 1: Asthma Phenotypes 174
Figure 2: Cohort timeline 176
Figure 3: Flowchart of study eligibility and participation 183
Figure 4: BTS step by phenotype 184
Figure 5: Incidence rate ratios, stratified by treatment step. 187
Figure 6: Time to first exacerbation analysis in years, by phenotype and 95% CI 188
Figure 7: Interaction between age and phenotype 196

Table of tables

Chapter 1

Table 1: Asthma characteristics. 34

Chapter 3

Table 1: Characteristics of studies with validated asthma algorithms 93

Table 2: Test measures of studies with validated asthma algorithms 98

Table 3: Quality assessment using QUADAS-2 106

Table 4: QUADAS-2, adapted from www.quadas.org 112

Chapter 4

Table 1: Characteristics of the 475 patients included in the final study analysis 128

Table 2: The positive predictive value (PPV) and % diagnosed with asthma within each algorithm 132

Chapter 5

Table 1: Baseline characteristics of the validated asthma and validated COPD patients 153

Chapter 6

Table 1: Baseline characteristics of the study population by phenotype 181

Table 2: Exacerbation rates with corresponding rate ratios by phenotype 185

Table 3: Treatment steps of the identifiable phenotypes 196

Chapter 1: Introduction

Summary

This chapter provides the background for the thesis, including an overview of asthma and electronic health records.

The definition, history, diagnosis, epidemiology, treatment and pathophysiology of asthma are described, followed by an overview of asthma characteristics and phenotypes. Asthma is a complex multicomponent syndrome which involves the interactions between the individual patient and their exposure to the environment. The development and symptoms of asthma are affected by a myriad of different risk factors and protective factors, which are described in this chapter. Asthma can therefore not simply be defined or described, despite the considerable investments in asthma research. Asthma phenotyping by grouping patients according to characteristics can offer the opportunity to target specific therapies at patients who are most likely to benefit and develop appropriate therapies for patients who remain poorly controlled.

Electronic healthcare records (EHR) and their use and limitations for asthma research are discussed in detail. The lack of a standard case definition for asthma, the absence of specific symptoms and the overlap with many other diseases render the ascertainment of asthma status in EHR difficult.

The chapter closes with the overall aim and objectives of this PhD project.

1.1 Asthma: disease background

1.1.1 Definition of asthma

The Global Strategy for Asthma Management and Prevention 2018 guidelines from the Global Initiative for Asthma (GINA) define asthma as follows:(1)

"Asthma is a heterogeneous disease, usually characterised by chronic airway inflammation. It is defined by the history of respiratory symptoms such as wheeze, shortness of breath, chest tightness and cough that vary over time and in intensity, together with variable airflow limitation."

Asthma constitutes one of the most prevalent non-communicable diseases in children and adults and affects people of all ages, all ethnicities and both sexes. The core symptoms (cough, wheeze, breathlessness and chest tightness)(2) are non-specific, and asthma is characterised by the pattern of these symptoms and their timings, the response to treatment, asthma triggers, and a variable expiratory airflow limitation which is generally reversible. The disease ranges in severity from milder attacks which can interrupt daily life and work productivity, to more severe and life-threatening attacks,(3) in which case it can greatly hinder the patient's life and ability to perform regular activities and can even cause death. The few signs of asthma are also non-specific: the clinician can look for expiratory wheezing and comorbidities such as obesity, bronchiectasis, eczema and allergic rhinitis to aid with the diagnosis of asthma. Asthma is innately variable, and therefore asthma patients can experience fluctuating symptoms. Figure 1 shows the difference between a normal airway and an airway during asthma symptoms.

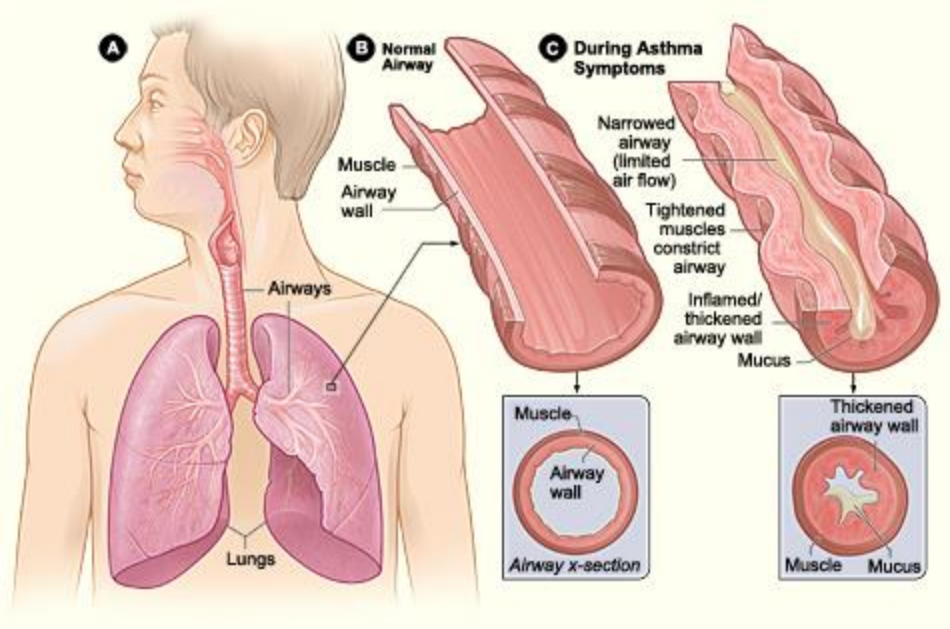


Figure 1: Asthma pathophysiology. A shows the location of the lungs and airways in the body. B shows a cross-section of a normal airway. C shows a cross-section of an airway during asthma symptoms. www.nhlbi.nih.gov/health-topics/asthma. Work of the US Federal government; (public domain free from copyright restriction)

There is widespread consensus that people who suffer from asthma should be offered medication based on a step-wise approach if they require pharmacological treatment and are under regular medical care. Mild asthma might only require few medications on a low dose, while severe disease will require more medications, often at higher doses. The current guidelines of the Global Initiative for Asthma,(3) British Thoracic Society (4) and National Asthma Education and Prevention Program(5) provide direction on the diagnosis and management of asthma.

1.1.2 Asthma in history

Asthma has been recognised before modern times, and the symptoms have been described in multiple ancient civilisations.(6,7) The word asthma itself is derived from the Greek $\alpha\sigma\theta\mu\alpha$ which translates as hard breathing, panting or death rattle.(8) In the

earliest records, asthma was used as a term to describe symptoms rather than a disease entity. Aretaeus of Cappadocia, an ancient Greek clinician living in the fourth century BC, attributed asthma to a "thick and viscid phlegm caused by coldness and humidity of the pneuma".(9) He also noted that women were more susceptible to asthma, men were more likely to die of it, and children had better recovery chances.(10)

Over the centuries, other authors have provided additions to the concept of asthma. Moses Maimonides (1135–1205) stated the need for clean air in asthma patients (11) and Gerolamo Cardano (1501–1576) noted the relationship between a lack of air quality and asthma.(8)

Sir John Floyer (1649–1734) authored the first modern publication on asthma in 1698 and documented the symptoms, triggers, treatment, and prevention of asthma.(12) Henry Hyde Salter (1823–1871) described key features of asthma and gave a more formal definition: "paroxysmal dyspnoea with intervals of healthy respiration between attacks".(13) He also stated that severe asthma can inflict permanent injury to the lungs.(14) However, Salter also viewed asthma as a psychosomatic disorder.(13) Towards the end of the nineteenth and the beginning of the twentieth century, clinicians started to recognise that asthma was associated with allergens and inflammation, and differentiated asthma from hyperventilation.(15–17)

Francis Rackemann was one of the first to recognise the heterogeneity of asthma based on a series of studies conducted at the beginning of the twentieth century. He categorised asthma as either extrinsic (exposure to allergens, younger age) or intrinsic (associated with infection or stress, older age).(18–20) In the 1840s, Hutchinson,(21) used spirometry to show the association between asthma and a variable airflow obstruction in the 1840s while the FEV1/FVC (Forced Expiratory Volume in 1 second over Forced Vital Capacity) ratio as a measure for lung function was introduced by Tiffeneau in the 1940s.(22)

In 1958, Harry Morrow-Brown was the first to notice that systemic corticosteroids had efficacy in patient with clear eosinophils in their sputum smear, but that these corticosteroids were not effective in patients with less eosinophils in their sputum.(23) These finding were largely ignored over the following decades, but have resurfaced with the renewed interest in the heterogeneity of asthma.(24)

In the mid-1960s, the treatment of asthma began to focus on airway hyperresponsiveness through the introduction of selective inhaled β2 agonists.(24) The use of these bronchodilators offered asthma patients more control over their symptoms and improved their quality of life. However, it was also associated with an increase in mortality and acute hospital admissions due to asthma(25,26); this was as a result of the over-reliance on inhaled β2 agonists and lack of anti-inflammatory medication such as corticosteroids.(27)

This led to a treatment shift in the late 1980's towards a more aggressive use of inhaled corticosteroids. Between 1990 and 2005, the increased use of anti-inflammatory medication led to a decrease in fatalities and hospital admissions due to acute asthma, in particular in children.(28) The combination of inhaled long-acting β2 agonists (LABA) and inhaled corticosteroids (ICS) has resulted in better outcomes for many patients.(29,30) However, there is no clear evidence of a correlation between airway hyperresponsiveness and inflammation.(31,32) Recently, new medications have become available, such as monoclonal antibodies.(33) However, there has not been a great deal of progress in key asthma outcomes including preventable deaths since 2002 as the combination of LABAs and ICS is still the basis of most guidelines.(28) The current treatment guidelines are expanded upon in a later section of this chapter.

1.1.3 Diagnosis of asthma

The diagnosis of asthma depends on the identification of a pattern of symptoms and the absence of an alternative explanation for those symptoms. There is no universal case definition for the diagnosis of asthma. Asthma diagnosis is based on probability, symptoms and a variable expiratory airflow limitation. When asthma is suspected, the variable expiratory airflow limitation should be confirmed by use of spirometry and a trial of treatment.(4) In asthma patients, the expiratory airflow typically falls outside the normal range and the FEV1/FVC should be less than 75-80% of the value predicted taking into account the patient's age, sex, height and race in at least one measurement to confirm a diagnosis of asthma in adults.(4) For children, the FEV1/FVC should be less than 90% of the predicted value. A reversibility test is positive if the FEV1 increases by more than 12% or 200mL after administration of a short-acting β2 agonist in adults, or more than 10% in children. The reversibility of airflow obstruction after bronchodilator treatment is the most commonly used test; however, the validity of this test has never been addressed, and it provides no information on the underlying inflammation.(34) Other tests are the bronchial challenge test and exercise challenge test, but they are difficult to do correctly in primary care, and a negative test does not rule out asthma.(35) The use of exhaled nitric oxide (FeNO) to diagnose asthma is controversial. The current British Thoracic Society (BTS) and Global Initiative for Asthma (GINA) guidelines do not recommend using FeNO measurements to diagnose asthma, while the National Institute for Health and Care Excellence (NICE) guidelines endorse it.(3,4,36,37)

FeNO and blood eosinophilia are independent markers of preventable risk in asthma,(38) but these markers cannot predict the severity of the underlying asthma on their own, unlike markers in many other chronic diseases. Some degree of inflammation is present in mild intermittent asthma, and asthma exacerbations can continue even when the inflammation is suppressed (see section 1.1.7 for a more

detailed discussion of exacerbations).(39) Uncontrolled symptoms increase the risk of exacerbation, but several other common non-symptom risk predictors exist.(40) These predictors encompass short-acting β2 agonist (SABA) overuse, a lack of inhaled corticosteroid use, smoking, low lung function, allergies and allergen exposure, viral infections of the upper airways, psychological or socioeconomic troubles, drug abuse, comorbidities including obesity and rhinitis, and high eosinophil counts in blood or sputum.

1.1.4 Epidemiology of asthma

According to the current estimates of the Global Burden of Disease Study 2015, 358 million people worldwide had asthma and 400,000 people died due to asthma in 2015, with a wide variation in prevalence between different countries around the world.(41) However, the prevalence of asthma also depends strongly on the exact asthma definition that was used. The International Study of Asthma and Allergy in Children (ISAAC) and the European Community Respiratory Health Survey have developed questionnaires to assess asthma in the early 1990s.(42–44) These studies showed very large fluctuations in asthma prevalence around the world, with a high prevalence in English-speaking countries and a lower prevalence in developing countries.

Asthma prevalence is lower in girls than in boys, but is 20% higher in women than in men.(45) The high prevalence in boys is thought to be partly due to smaller airways at a young age, genetic and hormonal factors, and differing comorbidities between the sexes.(46) Most commonly, asthma emerges during childhood, but it can also arise during adulthood. Therefore, adult asthma in adults can be either persistent/relapsed childhood disease or true incident adult disease. With the right treatment, symptoms can usually be managed and asthma patients can lead their lives without disruption.(2)

The prevalence of asthma is probably underestimated in developing countries, as patients do not have easy access to healthcare and asthma medications might not be available. Migration studies examining migrants from low-prevalence to high-prevalence countries have been conducted.(47) Asthma prevalence is lower in migrants than in natives of the host country, but it rises with increasing length of residence.

The prevalence of asthma is stable or decreasing in most developed countries with a high socio-demographic index such as those in Western Europe, but increasing in most low-to-middle socio-demographic index countries such as those in Sub-Saharan Africa.(48) Asthma continues to impose a high burden on healthcare systems in both primary and secondary care.(41,48) While the global prevalence of asthma has increased, the global mortality rate of asthma has decreased between 1990 and 2015.(41) Asthma places a significant cost on society through loss of productivity, both because workers are themselves affected by asthma, and because workers might be caring for children suffering from asthma.(49) Frequent exacerbations generally reflect poor asthma control, which is reflected in a lower quality of life and in loss of productivity in the workplace.(50)

Asthma is a major public health issue in the UK, and has a high impact on patients, on healthcare resources and on the wider economy. In the UK, 5.4 million people are currently receiving treatment for asthma of whom 4.3 million are adults,(51) and each year 12.7 million working days are lost due to illness. The direct NHS healthcare expenditure on asthma is more than 1 billion GBP annually,(52) and each day three people die in the UK due to their asthma.(51) The condition accounts for over 65,000 hospital admissions and 1,000 deaths annually.(53)

1.1.5 Pathophysiology of asthma

Asthma results from a chronic inflammation that narrows airways. This narrowing follows a contraction of the smooth muscles around the airways and a build-up of mucus. In some people, asthma attacks happen while they are exposed to a certain trigger, while there is no obvious cause in other people. Common triggers are tobacco smoke, dust mites and pollen.(4) Asthma is often worse at night and during the morning.(3) While most asthma episodes are relatively benign, there is a serious risk when the required oxygen cannot be supplied to the tissues, which results in hypoxaemia.(54) Prevalence of comorbidities (among which gastro-oesophageal reflux, recurrent respiratory infections and psychological disorders) are particularly high in patients with severe asthma, and may be detrimental to asthma control in those individuals.(55,56)

The causes of asthma are a combination of complex and incompletely understood genetic and environmental factors.(57) The endotypes (underlying mechanisms) and phenotypes (observable characteristics) of asthma are complex and result from multiple interactions between host and environment. These interactions occur between genes, cells, tissues and organs at different times. The endotypes describe a subtype of the disease based on intrinsic distinct pathogenic mechanisms, on a cellular and molecular level. Many different genes have been implicated using genome-wide association studies, including polymorphisms for IL33, HLA-DQ, SMAD3 and IL2RB.(58) The endotypes and genetics of asthma fall beyond the scope of this project, whilst the phenotypes are described further in this chapter. The environmental factors include allergens, air pollution and other airborne chemicals.(59) The hygiene hypothesis attempts to explain the increasing rates of asthma as a result of the reduced exposure to bacteria and viruses during childhood due to societal development;(60) however, this hypothesis is controversial and largely superseded.(61)

1.1.6 Treatment of asthma

The main goal of asthma treatment is to attain asthma control: to minimise both the symptom burden and the risk of adverse outcomes (exacerbations, airflow limitation and side-effects of the medication).(62) The treatment of asthma is personalised and includes pharmacological treatment, education, a written asthma management plan,(63) inhaler training,(64) minimisation of risk factors, management of comorbidities and further non-pharmacological treatment.(3) The treatment should be regularly assessed and adjusted based on symptom control, risk factors, comorbidities, side-effects and patient satisfaction. Asthma attack prevention through a healthy diet, exercise, smoking abstinence and trigger avoidance are non-pharmacological interventions that may be enough for patients with mild asthma. There is some evidence for further non-pharmacological treatment of asthma including nocturnal temperature-controlled laminar flow and add-on allergen immunotherapy.(65,66) The decision-making process should be shared with the patient to improve outcomes.(67)

BTS guidelines

The pharmacological treatment recommended by the British Thoracic Society (BTS) and the Global Initiative for Asthma (GINA) is stepwise.(3,4) The exact treatment steps differ slightly by guideline; however, for the purposes of this thesis, the BTS 2016 guidelines for adults are used. The treatment steps in the figure below (Figure 2) are included in the BTS 2016 guidelines and are used for the cohort study included in Chapter 6 of this thesis.

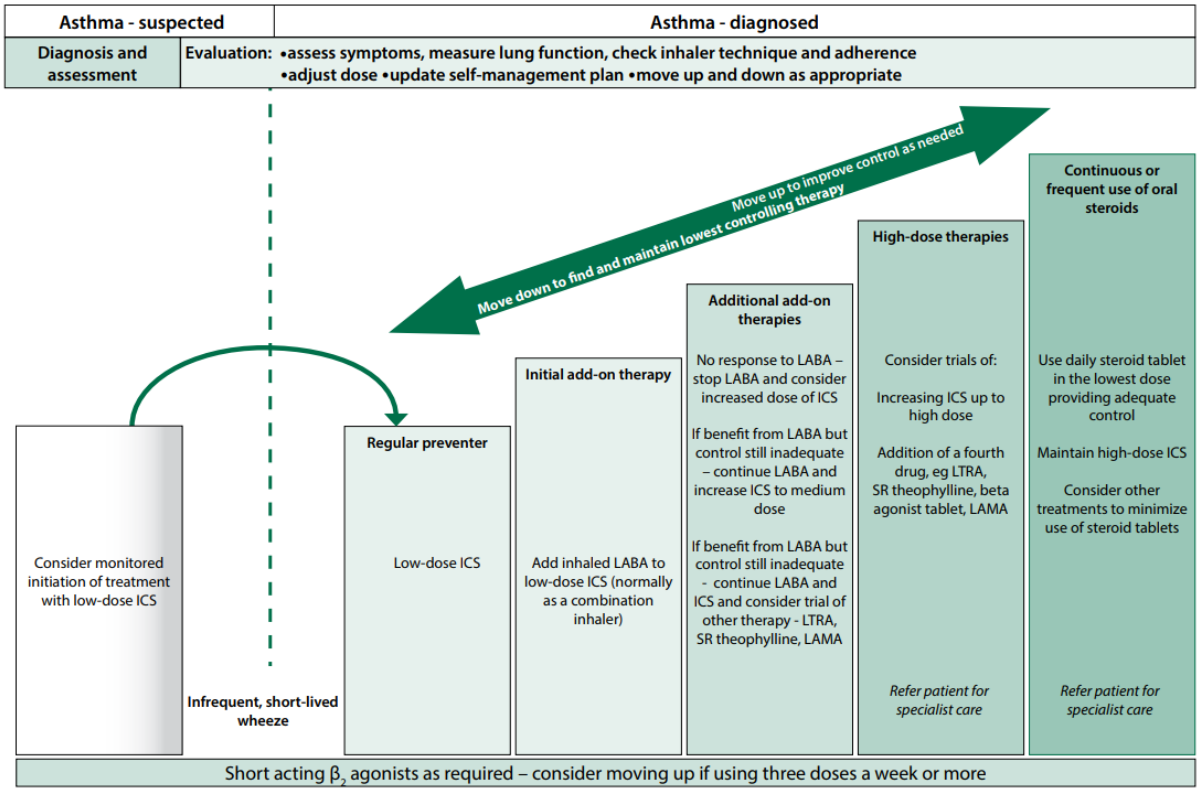


Figure 2: Summary of asthma management in adults. This figure is reproduced from BTS/SIGN British Guideline on the management of asthma by kind permission of the British Thoracic Society. British Thoracic Society (BTS)/Scottish Intercollegiate Guidelines Network (SIGN). British Guideline on the management of asthma. Edinburgh: SIGN; 2016. (QRG 153). [cited 13 08 2018]. Available from URL: http://www.sign.ac.uk

In adults, treatment step 1 is defined by either no maintenance treatment or non-regular treatment with inhaled corticosteroids (ICS). Step 2 includes repeat low-dose ICS, step 3 adds an inhaled long-acting β2 agonist. Step 4 introduces a trial of leukotriene receptor antagonists, theophylline or long-acting muscarinic antagonists (LAMA). Step 5 increases the ICS dose. Step 6 includes the usage of oral corticosteroids, which should be minimised due to their systemic side effects. Additional treatment that can be considered in step 6 are add-on anti-immunoglobulin E (anti-IgE) treatment such as omalizumab, anti-interleukin-5 (anti-IL5) treatment such as mepolizumab/reslizumab, or bronchial thermoplasty.

Asthma medication classes

Inhaled corticosteroids and β2 agonists are the pillar of the pharmacological treatment of asthma. Bronchial inflammation is managed by ICS to prevent exacerbations and breathlessness is relieved by β2 agonists through bronchodilatation. Prescribed β2 agonists are either short-acting (SABA) or long-acting (LABA). SABA use is for quick relief and is not included in the maintenance treatment steps. On the other hand, LABAs tend to act slower and are included in these treatment steps.

Leukotriene receptor antagonists are considered less effective than ICS to treat inflammation, but may be an option for patients unable or unwilling to use ICS.(68) Theophylline in sustained-release medications has only weak efficacy and side effects are common.(69,70) Chromones such as nedocromil sodium and sodium cromoglycate have weak efficacy and are burdensome to use, but have a safer profile compared with theophyllines.(71) LAMA can be used as an add-on therapy in patients at risk of exacerbations, as it modestly improves lung function and increases time to severe exacerbation.(72) Anti-IgE treatments or anti-IL5 treatments are reserved for patients with refractory asthma, partly due to their costs.(32,73–78) Oral corticosteroids may be effective for adults with severe asthma, (36) but often have severe side-effects.(79,80) Other immunosuppressant medications, including ciclosporin or methotrexate, are not recommended.(36) There is only limited evidence on the effectiveness of bronchial thermoplasty.(36) The roles of macrolide antibiotics and antifungal therapy in asthma remain unclear, as there is no conclusive evidence.(36,81,82)

There are some emerging therapies that show promise, including benralizumab (an anti-IL5 antibody) (83) and fevipiprant, a prostaglandin D2 type 2 receptor antagonist.(84) Studies on IL-13 antibodies have been discouraging,(85,86) but a monoclonal antibody targeting both IL-4 and IL-13 (dupilumab) has shown potential in a clinical trial.(87) In general, new medications for refractory patients are considered

the greatest unmet need in asthma. However, from a public health perspective, effective preventative treatment would be more beneficial.(62)

Many controversies in the optimal treatment of asthma persist, such as whether SABA-only treatment should be the initial treatment for asthma, what the criteria for ICS start-ups are and how seasonal asthma should be treated.(62)

1.1.7 Asthma exacerbations

Asthma exacerbations, also known as asthma attacks or flare-ups, are (sub)acute episodes with increased symptoms. Patients with an acute exacerbation show increasing shortness of breath, wheezing, chest tightness, cough, and reduced lung function. The onset time varies and an exacerbation can develop for over more than a week in adults. Asthma exacerbations are expensive to treat, affect quality of life (88) and can be lethal in rare cases.(89) Asthma exacerbations are triggered by multiple different contributing factors, including infections, allergens, atopy, pollution, environment and comorbidities. In about 80% of exacerbations, a respiratory virus infection is one of the causes.(90)

There is some controversy on the use of the word exacerbation, as it allows for multiple different interpretations.(76,77) It has been suggested that the word exacerbation should only be used to describe asthma with poor prognosis that requires immediate attention.(62)

1.2 Asthma phenotypes

1.2.1 Background

Asthma is a heterogeneous disease, this heterogeneity is evident by the existence of observable clusters called asthma phenotypes. A phenotype is defined as the set of observable characteristics of an individual resulting from the interaction of its

genotype with the environment.(93) These phenotypes refer to a pattern of observable characteristics, without regard to the underlying pathophysiology. There are many characteristics that can construct asthma phenotypes including inflammatory profiling based on leucocyte counts (eosinophils, neutrophils and paucigranulocytic), age-of-asthma onset, and airflow measurements.(94–100) Some characteristics may be clinically recognised, for example asthma may be induced by infection, exercise or might be caused by obesity or menstruation. They can also be defined by the frequency of exacerbations, or be more complex and only definable in laboratories or specialised secondary care.

The exact demarcation of potential phenotypes is not well defined in the literature, as there is no universally accepted asthma phenotype categorisation. There have been multiple studies describing asthma phenotypes, involving populations with asthma alone (101–110), or as part of an entity called "obstructive airways disease" together with COPD.(111,112) Classifying asthma into phenotypes can allow one to deconstruct the disease into separate identifiable traits (94) and better understand the disease progression and its response to treatment, which further enables the practice of precision medicine.(113) In particular, phenotyping may be useful in providing long-term prediction of outcomes and determining the effects of specific treatments for selected phenotypes.(113)The classification by eosinophil levels can be particularly meaningful due to a difference in treatment response.(114–116)

Patients with asthma can present with a range of different clinical histories, physiological changes on spirometry and airway inflammation. There are multiple ways to define these phenotypes. For example, it is possible to determine them based on the severity of asthma (103) or to use cluster analysis on a cohort to identify related groups for analysis.(101)

1.2.2 Characteristics associated with asthma

Currently, there is no consensus on the classification of asthma phenotypes. In this thesis, the term phenotype is used to describe a set of characteristics that can be observed in the population. Regardless of how phenotypes are defined, phenotypes consist of a set of characteristics which can be divided in specific clinical, demographic and pathophysiological characteristics. For the purposes of this PhD project, phenotypes are constructed based on the characteristics for asthma in the table below. Some of these characteristics could be defined as phenotypes themselves (like eosinophilia), while others can be regarded as essential components of larger phenotypes. The next section will focus on the specific characteristics which shape the different phenotypes, regardless of the exact delimitations of these phenotypes.

Clinical	Potential relation to asthma	Prevalence(117)	References
Severity	Asthma severity is mostly based on received treatment.		(1)
Atopic eczema	Similar aetiology	13.4%	(118–120)
Rhinitis	Similar aetiology		(121,122)
COPD	Common symptoms, potentially same aetiology	13.4%	(111,123)
GORD	Increased acid reflux, micro-aspirations, reduced sphincter pressure	10.9%	(124–126)
Anxiety	Asthma can give rise to anxiety and vice versa	6.9%	(127,128)
Depression	Asthma can give rise to depression and vice versa	17.3%	(117,129,130)
Allergies	Similar aetiology		(62)
Sleep apnoea	Potentially through obesity		(127,131)
Demographic/Lifestyle			
Age-of-onset	Potentially different aetiology between early and late onset		(96)
Sex-related	More common in boys than girls, but more common in women than men		(132,133)
Obesity	Altered lung dynamics, inflammatory process or common predisposition		(134,135)

Socio-economic status	Different exposures, health status and access to treatment and		(136,137)
Ethnicity	Different genetic predisposition		(138)
Family history	Different genetic predisposition		(139)
Occupational	Different exposures		(140,141)
Smoking	Exposure to toxic substances	24.7%	(142–144)
Pathological			
Early life infections	Influence on the immune system development		(99,145)
Leucocyte levels	Classification based on eosinophil, neutrophil and pauci-granulocytic		(120,146)
Th2 cytokine levels	TH2 helper cells orchestrate immune response		(147,148)

Table 1: Asthma characteristics. GORD=Gastro-Oesophageal Reflux Disease Th2=Type 2 helper cell. Prevalence rates from Weatherburn et al: Comorbidities in adults with asthma: Population-based cross-sectional analysis of 1.4 million adults in Scotland were reported. Estimates from other populations were not included, as comparing would be more difficult.

These characteristics can be aetiological, provoke deterioration of symptoms/exacerbations, or both. In the following paragraphs, the asthma characteristics that form the basis of phenotype categorisations are further described.

Clinical characteristics

An asthma diagnosis requires careful assessment of comorbidities or potential alternative diagnoses. Under-treatment of comorbidities can influence asthma control and quality of life greatly.(127)

Asthma, **atopic eczema** and **chronic rhinitis** often appear together,(149) and asthma and allergic rhinitis are often preceded by atopic eczema. This observation has been seen in multiple longitudinal studies and has been designated the "atopic march".(150,151) Not all atopic patients will develop asthma, however.(152,153) Atopy is the sensitisation followed by the generation of specific IgE antibodies against environmental allergens or can indicate a predisposition to produce increased levels of IgEs after exposure to allergens and develop allergic reactions (type 1). **Atopic eczema** can be assessed though skin prick tests or serum measurements of allergen-specific IgE.(118,154) Fifty to sixty per cent of asthma patients have atopy,(119) but the prevalence is higher in children with severe asthma and adults with early-onset asthma.(120) **Rhinitis,** even in the absence of atopy, is a strong predictor of adult-onset asthma.(121) Inflammation is an important factor in both rhinitis and asthma, which can be caused by exposure of the nose and lung to allergens. Anti-inflammatory strategies targeting both anatomic sites could be beneficial.(122)

Asthma and Chronic Obstructive Pulmonary Disease (**COPD**) share many symptoms.(28)(111,123) It has been proposed that the two conditions are just components of the same airway disease.(59) The idea that asthma and COPD are different components of the same airways disease was first postulated in 1961 as the Dutch hypothesis,(155) and remains controversial.(111,148,156) The term for the overlap syndrome is ACOS

(Asthma-COPD Overlap Syndrome). The viewpoint that asthma and COPD are two distinct disease entities that can co-exist is sometimes called the 'British hypothesis'.(148,157) Asthma also shares inflammatory features with COPD such as neutrophilia, which complicates the differential diagnosis. The airflow limitation is generally less reversible in COPD compared with asthma and there is a reduced elastic recoil and hyperinflation at rest.(158) However, partially reversible or even irreversible airways obstruction has also been described in asthma,(159) and COPD patients may show some degree of reversibility of airways obstruction.(160–162) COPD is more common among older people and smokers. A recent study suggests that asthma can contribute as much as smoking to the development of chronic bronchitis in middle age.(163) The obstructive form of chronic bronchitis is included within the definition of COPD.(164) Outcomes of concomitant asthma and COPD are worse than either disease alone.(165,166)

Most patients with asthma report symptoms related to Gastro-Oesophageal Reflux Disease **(GORD)** or have an abnormal 24h oesophageal pH test.(124,125) There is a strong association between asthma and GORD, but there is not much known on the direction of causality, if indeed any exist.(126) The mechanisms might include increased acid reflux during exacerbations, micro-aspirations, and β2 agonists reducing oesophageal sphincter pressure.

Psychological disorders such as anxiety and depression are more frequent in people with asthma compared with the general population.(127,128,130) Asthma symptoms may be triggered by psychological factors or influence the patient's perception of asthma symptoms. Reverse causality, in which asthma gives rise to anxiety or depression, is also possible and these factors can influence medication adherence.(167). Anxiety symptoms can also mimic asthma exacerbations.

While **allergens** are triggers for exacerbations in those with asthma, their role in development is not crystal clear. How exactly exposure leads to sensitisation, and how sensitisation leads to asthma, is not well understood.

Obstructive sleep apnoea may be associated with asthma, perhaps through obesity as an intermediary variable.(131,168) Fatigue, irritability and decreased concentration are symptoms of obstructive sleep apnoea that are typical in children who also have poorly controlled asthma.(169)

Demographic and lifestyle characteristics

The **age-of-asthma onset** is often used as a determinant of different asthma phenotypes. It has been suggested that early-onset adult asthma, which originates in childhood, is more attributable to atopy and genetic factors, while late-onset adult asthma is more related to environmental risk factors.(96)

Asthma is almost twice as common in boys as in girls,(3) while both **sexes** have the same rates of severe asthma in childhood.(170) In adulthood, the prevalence of asthma is greater in women than men.(132) A possible hormonal influence has been suggested.(133)

People with a **Body Mass Index** higher than 30kg/m² have a higher incidence and prevalence of asthma, particularly in women.(134) The underlying mechanisms are uncertain, but various hypotheses have been proposed. These hypotheses include a common genetic predisposition, altered lung mechanics because of obesity, the presence of a systemic inflammatory process, and an increased prevalence of comorbid conditions such as GORD or sleep apnoea.(135)

Asthma was historically thought to have a higher prevalence in groups with higher **Socio-Economic Status (SES)**, which was an argument in support of the hygiene

hypothesis.(136) However, this might have been due to a difference in diagnosis and reporting of asthma and this view has now been reversed. Asthma prevalence in young adults is higher in individuals living in a low-educational area according to the European Community Respiratory Health Survey.(137)

A **family history** of asthma in first-degree relatives was consistently identified as a risk factor for childhood asthma in a systematic review.(138)

A UK study revealed that there is a **ethnic** difference in the prevalence of asthma, with Caribbean and white populations having a higher prevalence than African or South Asian populations.(139)

Asthma can also be caused by exposure to an agent in the **work environment**. In a large population-based study in young adults in Europe, the highest risk was shown for cleaners, farmers, painters and plastic workers.(140) The proportion of asthma attributable to occupation was 5-10% in this study. This cause of asthma can be missed as GPs may not routinely explore the role of occupation.(141)

Exposure to environmental **tobacco** smoke is associated with an increased reported prevalence of asthma, wheezing and chronic bronchitis.(142) Maternal smoking during pregnancy has been shown to increase the occurrence of physician-diagnosed asthma and wheezing during childhood.(143) Asthma incidence during adulthood has been found to be strongly associated with active cigarette smoking in a UK prospective study using a longitudinal birth cohort.(144) Tobacco smoking could deteriorate the prognosis of people living with asthma.(171)

Pathological characteristics

Exposure to **infections** in early life influences the immune system development. The controversial hygiene hypothesis proposes that this exposure could lead to a reduced

risk of asthma and other allergic diseases.(61,145) Some studies show an increased risk of asthma following infection, while others show a decreased risk.

The heterogeneity of immunology in asthma has been used to better understand the different clinical presentations of asthma. The cytology of sputum or blood can provide evidence of eosinophilic, neutrophilic and pauci-granulocytic inflammation.(146) Around 50% of adult asthma patients have eosinophilic airway inflammation. T helper 2 cells aid the survival and maturation of eosinophils through the production of Interleukin-5. The pathways in non-eosinophilic asthma remain poorly understood.(120) Papi et al. have described the inflammatory pathways in asthma in detail.(62)

The **eosinophilic** phenotype of asthma is defined by the central role that eosinophils play in the pathophysiology of the condition. It is characterised by elevated sputum and/or blood eosinophils and by a significant response to treatments that suppress eosinophilia.(172) The eosinophils are quite rare in serum, but can be common in the airways of asthma patients.(173) They are often considered as effector cells, but also play a role in the regulation of immunity, remodelling and modulating other leukocytes.(174)

Cytokines derived from **T helper type 2 cells** (TH2) play a critical role in orchestrating and amplifying the inflammatory response in asthma.(175) There is a group of asthma patients in which TH2 cytokines predominate and which can be defined by biomarkers and response to therapies targeting this type of immunity.(147)

Protective factors

There are also factors which protect against asthma including certain infections,(176,177) farm and animal exposure (178–180),and vitamin D intake.(181,182)

Many of the characteristics, including allergen exposure in childhood, family history, occupation, early life infections, maternal smoking and TH2 cytokines, cannot be measured reliably in the datasets used for this PhD project, so are not discussed in detail. The implications of this lack of data are further discussed in the last chapter of this thesis.

1.2.3 Phenotypes identified using cluster analysis

Cluster analysis can help reveal hidden arrangements of entities, in this case patients, with similar attributes into groups and differentiate groups of patients with heterogeneous characteristics.(183,184) Patients can be grouped together based on characteristics that make them similar (high intra-class similarity) and separate them from different groups (low inter-class similarity).(185) The patients within a cluster are geometrically grouped together, and the distance between patients in different clusters is greater than the distance between patients within the same cluster. In the context of health data, cluster analysis can be used to identify which patient belongs to which group, and to identify the ideal number of clusters and thus reveal a latent structure within a dataset or group of patients.(186)

There are several different methods of cluster analysis, including k-means, multivariate Gaussian mixture, hierarchical clustering, spectral and nearest neighbour method.(187)(188)

One of the most influential studies using cluster analysis in asthma in order to identify distinct phenotypic groups was conducted by Haldar et al. using cluster analysis of multiple clinical variables.(101) Among 184 patients managed in primary care, three clusters were found: one group with benign asthma, one group with obese non-eosinophilic asthma, and one group with early-onset atopic asthma. Further cluster

analysis of two other asthma populations which were managed in secondary care and were mostly refractory (N=255 total), added an early symptom predominant cluster and an inflammation predominant cluster.

The study by Haldar et al. used the k-clustering algorithm.(101) This algorithm has been used widely and requires the number of groups (k) and a distance metric as inputs.(189) The first step is to associate each data point with one of the k clusters, depending on the distance to the cluster centers (centroids) of each cluster.

The next step is to calculate new centroids and reclassify the data points for the new centroids. This process can then be repeated until there are no more significant changes in centroid position observed at each new step.

One of the main limitations of the k-means algorithm is the a priori setting of the number of clusters, as the final classification of clusters can strongly depend on the choice of number of centroids. The k-means is also not indicated if the clusters have very different sizes,(190,191) and is sensitive to the initial seed selection which determines the initial cluster centres. The advantage of the k-means are the low computational cost (easy to implement and can be faster than alternatives such as hierarchical clustering) and the good results in practical situations such as detection of anomalies within a dataset or grouping patients likely to benefit from a certain intervention through data segmentation(192),

The specific limitations of using clustering analysis on health data is that disease and health is a continuous spectrum, and separating the population into discrete clusters may not be realistic. The study by Haldar et al further mentions that other methods with a more probabilistic approach to cluster grouping could be valuable.(193) In addition, the choice of variables remains subjective as well as the number of clusters chosen for the population.

The authors aimed to choose variables that were measured and could contribute to the clinical evaluation, variables that were considered important in the definition of phenotype definition and avoid variables that would in effect measure the same characteristic twice. The variables were categorised as either symptoms, atopy/allergy, eosinophilic inflammation, psychological status or variable airflow obstruction.

Not all variables were recorded and not all etiologic factors could be explored. The number of clusters in Haldar's study were estimated from the dendrogram plots obtained using Ward's method. Further limitations reported by the study are the question of stability in cluster membership over time and changes in treatment. There was no significant difference in treatments between the clusters. Differences in clusters may have been due to a difference in disease profile and differences in response to treatment.

Other phenotyping studies using similar clinical characteristics found comparable phenotypes.(103,106,194–196) The identified clusters can be found in the figure below. As this categorisation forms the basis for the study included in Chapter 6 of this thesis, these phenotypes are explained in further detail.

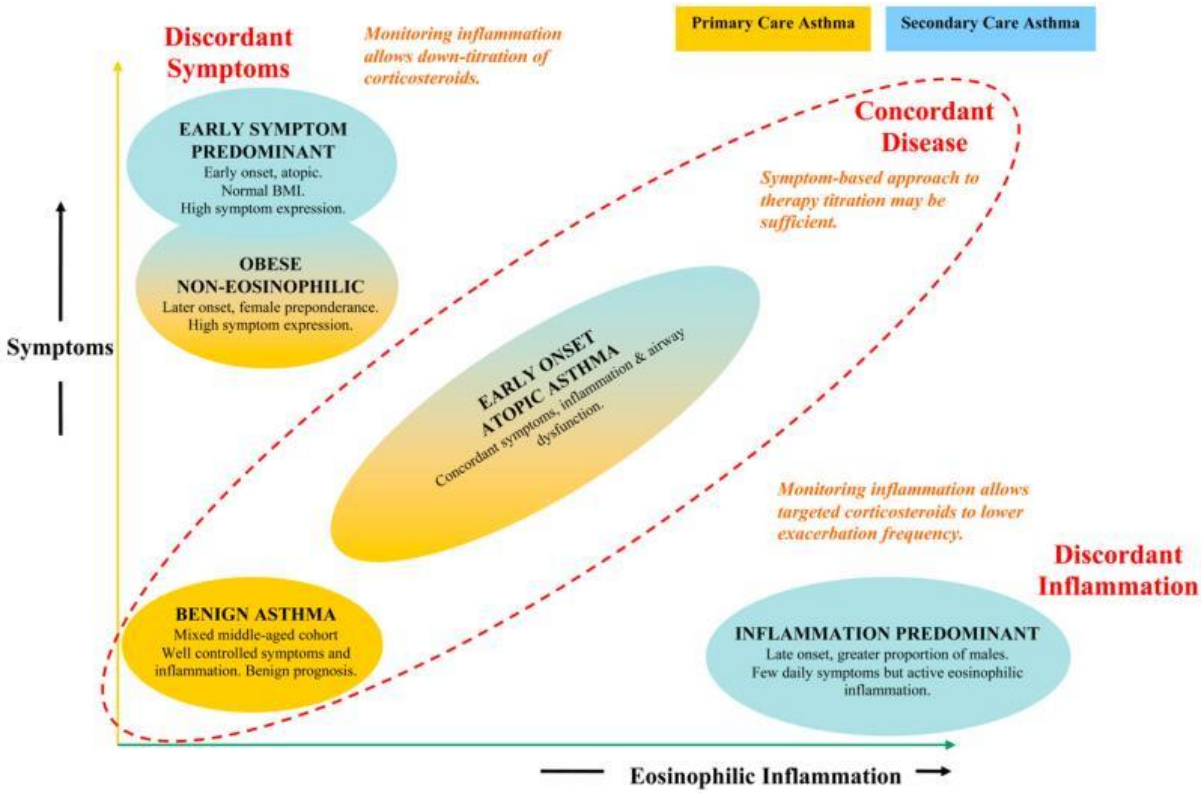


Figure 3: Clinical phenotypes of asthma by Haldar P. Reprinted with permission of the American Thoracic Society. Copyright © 2018 American Thoracic Society. Haldar P et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med. 2008;178(3):218–24. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society

The early-onset atopic phenotype includes primary care patients with airway obstruction reversibility and eosinophilic inflammation and asthma onset in childhood. Obese non-eosinophilic asthma includes mostly female overweight primary care patients with less eosinophilic inflammation. The benign asthma phenotype is mostly composed of primary care patients with good control of symptoms and inflammation, and a favourable prognosis. The early symptom predominant asthma phenotype includes secondary care patients with less inflammation and reversibility, but strong symptom expression. Inflammation predominant asthma is a secondary care phenotype with clear eosinophilic inflammation, but few symptoms.

Another influential study on asthma phenotypes was undertaken by Moore et al. based on the Severe Asthma Research Program (SARP).(94) The defining criteria of these phenotypes were the lung function based on the maximum FEV1 and the age of onset, in which five clusters were found that broadly corresponded to the clusters found in Haldar's study. These clusters were mild atopic asthma, mild to moderate atopic asthma, late-onset non-atopic asthma, severe atopic asthma, and severe asthma with fixed airflow. Moore et al. used Ward's minimum-variance hierarchical clustering method as an unsupervised modelling approach to identify asthma phenotypes within the SARP cohort.(103)

1.3 Electronic healthcare records

1.3.1 Overview

An electronic health record system contains information on the health of an individual and is an electronic version of a patient's medical history. It may include all key administrative and clinical data that are relevant to that person's care, including demographics, medications, past medical history, immunisations, laboratory data, and secondary care reports.(197,198). A clinician or healthcare professional can consult these records for patient care.

The clinical records can be used to access patient information or can be used to improve the efficiency of the clinical practice by generating medication prescriptions or requesting clinical tests.(199) EHR can improve patient safety,(200) but implementation of EHR remains heterogeneous across general practices, healthcare systems and countries. As the pace of implementation differs, this can lead to differences in patient safety outcomes.(201–203) Routinely collected EHR are the predominant setting for pharmacoepidemiological studies, but these EHR are typically not primarily constructed for research purposes. Another kind of health data are the administrative databases; these are used for non-clinical purposes and often exist to facilitate remuneration for care costs.

Strengths of EHR for research

EHR data have an enormous potential for epidemiological and clinical research. Due to their immense size, they can offer high statistical power and can often be representative of a population. Linkage between different EHRs can further improve the completeness of the data.

One of the main benefits of using EHR for research compared with, for example, tailor-made cohort studies is the financial cost, as these databases do not have to be

constructed for the purpose of one study. Other advantages of EHR databases for research include their relatively complete recording of drug prescriptions and diagnoses, minimisation of observer and participant bias, information on potential confounders, and the ability to assess drug effects in the routine clinical care population where the medication is used (often different to the conditions under which randomised trials are conducted). EHR also hold information on the health status of a specific population, which can be used to estimate disease prevalence and incidence or provide additional arguments to support or reject a clinical diagnosis.

Limitations of EHR for research

A major limitation of the use of EHR for research is that the primary goals of EHR are for clinical, administrative or audit purposes. This means that important information can either be missing or wrongly classified. In cohort studies where the data are collected solely for the purpose of epidemiological research, stricter definitions can be used for exposures, covariates and outcomes, and a specific test that is not part of routine clinical practice can be requested or looked into at greater detail than in EHR. In addition, contact with the patient is rarely possible, and contact with the treating physician is only possible in specific cases (such as a validation study). Identification of patients with specific diagnoses or covariates depend on specific algorithms or codes which can be difficult to construct.

The record consists of the results of clinical and administrative appointments between a healthcare provider and a patient during the patient care. As such, the EHR reflects the skill, know-how and job function of the healthcare provider. Due to the nature of EHR, information on potential confounders can also be missing or incomplete. In addition, studies using these databases should be carefully designed to avoid time-related biases such as immortal time bias.(204,205)

Examples of EHR databases

Worldwide, there are several EHR databases that have been proven to be reliable data sources for research purposes. The first two papers, included in Chapter 3 of this thesis, outline different non-claims databases that have been used to identify asthma diagnoses and their validity.

In Europe, the Scandinavian countries are known for the completeness of their databases due to their welfare systems and complete registration. For example, the Danish National Patient Registry contains information on all secondary care for all patients in Denmark,(206) but this database lacks some information on primary care variables. Another well-known database is the Dutch Integrated Primary Care Information (IPCI) database, which was constructed with the primary purpose to conduct pharmacoepidemiologic studies. It contains data on diagnoses in primary care, prescriptions, lifestyle factors and hospitalisation events.

There are several EHR databases containing primary care data in the UK. These databases include THIN (The Health Improvement Network),(207) ResearchOne and CPRD GOLD.(208) They differ in size and availability of linkage to other databases but contain similar data on primary care. Linkage to secondary care of the UK databases varies by UK nation; for example, in England the Hospital Episodes Statistics (HES) contain information on hospitalisations, outpatient and Accident and Emergency attendances.

A separate kind of health data are administrative claims data, whose main purpose is administration of reimbursements to healthcare providers for their services. This contrasts with EHR, which are a digital reflection of the paper medical chart. The quality measures between the two types of data can be markedly different.(207,209)

In general, USA and Canadian databases tend to be more administrative and used for claims and insurance purposes. An issue with this type of data is that the primary purpose of this data is financial and is not always constructed or used by the treating clinician. As a result, these databases might include more diagnoses that maximise profit for the healthcare providers. In addition, they do not contain much information on lifestyle factors such as smoking and BMI, which limits access to confounders. An example are the Kaiser Permanente medical care programmes, a private health insurance scheme covering 8 million people which is based in the US.(210)

Three studies in this thesis use CPRD GOLD data, linked to HES and ONS (Office of National Statistics) data. These databases are described in further detail in the next chapter.

1.3.2 Asthma identification in electronic health records

Identifying patients with asthma in epidemiological studies can be complicated. The absence of a universal case definition for asthma, partly due to the heterogeneity of asthma, remains an issue. The variability of symptom severity, the fact that asthma symptoms are non-specific, and the differential diagnosis with other diseases such as COPD, further convolute the identification of asthma patients. Both bronchodilator reversibility and airway hyperresponsiveness have been used to define asthma in research, but there is no consensus on the cut-off, which can result in noncomparable asthma populations between studies.(95)(211) Asthma medications are commonly prescribed in primary care without assessing lung function, while tests may not even be available. As such, asthma diagnoses in primary care may be inaccurate in up to 30% of cases.(212,213) The evolution of guidelines for the diagnosis and management of asthma further complicates asthma research.(214)

In a review published in 2010, 60 different definitions of childhood asthma were used in 122 epidemiological studies. This variation in case definitions of asthma can lead to

misclassification in research studies. On one hand, if too few inclusion criteria are needed some included patients might not have asthma, while, on the other hand, if all possible inclusion criteria are used to increase the specificity of an asthma diagnosis, people who really have asthma might not be identified as having asthma.(215) Furthermore, if specific criteria or test results (such as sputum eosinophilia) that are not always available are required for the inclusion of a patient in a research study, the study population might not be representative of the general asthma population.

There are several diseases that share symptoms with asthma and can, therefore, be confused with asthma at clinical assessment, including COPD, eosinophilic bronchitis,(216) vocal cord dysfunction,(217) bronchiectasis, anxiety and dysfunctional breathing.(218) Asthma, by its very definition, is variable and patients may present with few symptoms,(219) inflammation (220) or airway hyperresponsiveness at the time of assessment. In addition, inflammation and hyperresponsiveness show only a weak association with each other.(221–223) In clinical and epidemiological research, original asthma diagnoses recorded in patient notes cannot always be verified at a later time due to the inherent variability of asthma, successful treatment, or differing asthma criteria between the clinician and researcher.

These issues are all present when aiming to identify asthma patients from EHR for epidemiological studies. Being able to reliably and transparently identify the asthma status of patients in EHR is vital to conduct asthma research using these EHR. As discussed in this subchapter, there are multiple reasons why this is not straightforward. The ascertainment of the asthma status of patients in EHR is a major part of this PhD thesis, as described in the thesis objectives.

1.4 Thesis aim, rationale and objectives

1.4.1 Aim

The aim of this thesis is to find reliable ways to identify asthma patients from de-identified UK electronic health records and study predefined asthma phenotypes, including patient characteristics and outcomes.

1.4.2 Rationale

Asthma is a common heterogeneous disease that carries a high morbidity and notable mortality worldwide. The identification of asthma patients from electronic health records (EHR) in primary care can be challenging, as there is no universal consensus on what constitutes asthma, asthma shares many symptoms with other diseases such as COPD, and asthma diagnostic tests and markers are not always well recorded. Different asthma phenotypes have previously been established based on cluster analysis in small populations; this categorisation may allow for specific treatment strategies. The prevalence and outcomes of distinct phenotypes are not known.

1.4.3 Objectives

- Understand how past epidemiological studies have identified asthma patients in EHR through a systematic review.
- Validate the recording of the diagnosis of asthma in CPRD GOLD.
- Quantify the concomitant occurrence of asthma in COPD patients and vice versa in CPRD GOLD.
- Identify established asthma phenotypes in CPRD GOLD by studying characteristics and explore the variation of asthma severity (defined by treatment steps) by phenotype.
- Examine the difference in asthma control by asthma phenotype, stratified by treatment step.

1.4.4 Thesis organisation

- Chapter 1 provides the background to the thesis, including an overview of asthma and electronic health records.
- Chapter 2 describes the databases used in this project in detail: the CPRD GOLD, HES, ONS and the questionnaire designed as a reference standard for asthma diagnosis.
- Chapter 3 describes ways in which asthma researchers have identified asthma patients from EHR databases worldwide and test values through a systematic review.
- Chapter 4 provides the results of a validation study of asthma in the CPRD GOLD.
- Chapter 5 quantifies the prevalence of concomitant asthma and COPD in patients with a validated diagnosis of either disease.
- Chapter 6 outlines the results of a study to detect pre-identified asthma phenotypes from CPRD GOLD.
- Chapter 7 summarises and discusses the overall findings of this PhD project.

In conclusion, this thesis provides information on the validation of asthma and the prevalence and control of asthma phenotypes in electronic health records.

Chapter 2: Data sources

Summary

In this chapter, the data sources used in this thesis are described, and their respective advantages and weaknesses are discussed. These data sources include the Clinical Practice Research Datalink GOLD, Hospital Episode Statistics and Office of National Statistics mortality and deprivation data. This chapter also discusses the asthma questionnaire designed as part of this PhD thesis that was applied as reference standard for an asthma diagnosis in the CPRD GOLD. In addition, this chapter reports on the coding system and the data flow from the healthcare provider's practice to the researcher.

2.1 EHR databases and front-end software systems in the UK

Large routine health care databases have been considered as a means of addressing research questions of interest for many years. There are several different primary care EHR databases available in the UK, and multiple front-end software systems that deliver their data into the databases.

The CPRD GOLD is the oldest research database of primary care electronic health records in the UK (224) and has generated the highest number of peer-reviewed publications.(225) Recently, the CPRD Aurum dataset supported by the EMIS front-end software has become available,(226) but this dataset was not available at the start of this PhD programme. There are several different software systems (the front-end systems) available in the UK to record clinical data, for example Vision or EMIS (Egerton Medical Information System), which are used by general practices to record the information that is subsequently uploaded to their respective databases. Published studies in this thesis were included as they were printed and may refer to the CPRD GOLD as "CPRD". This terminology has become more ambiguous with the advent of CPRD Aurum and should be avoided in future studies. Other examples of UK primary care databases include Q-research and THIN (The Health Improvement Network) The data content of these databases is generally comparable with the CPRD GOLD, but their size and linkage availabilities can differ.

Data on secondary care are available in Hospital Episode Statistics (HES), while data on mortality and an area-based socio-economic status are made available by the Office of National Statistics (ONS).

The CPRD GOLD was the main data source for this PhD project, and is supported by the Vision software system. The following subchapter describes how the data are coded, recorded, uploaded, de-identified and made available for research in the CPRD

GOLD database using the Vision front-end software system. The data flow in other EHR databases is comparable, but may have different features (such as the coding system used). The sections thereafter describe HES, ONS and the asthma questionnaire for the study described in Chapter 4.

2.1 Clinical Practice Research Datalink GOLD (CPRD GOLD)

2.1.1 Background

The main database used in this thesis was the UK Clinical Practice Research Datalink GOLD. The Clinical Practice Research Datalink (CPRD) is a governmental, not-for-profit research service for observational and interventional research. It is jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA) and operates as a part of the UK Department of Health.(227) This primary care database is composed of de-identified data on patients from over 650 NHS primary care practices in the UK. In total, there are data on over 11 million patients, of whom 4.4 million were active (alive and currently registered) in 2015.(208)

The database includes data on patient demographics, coded diagnoses using Read codes, prescriptions using Gemscript codes, laboratory test results, and referrals made by general practitioners.(208) The database has been providing de-identified primary care records for public health research since 1987.(208) Over 1,500 articles have been published using CPRD GOLD data, which have led to improvements in drug safety, best practice and clinical guidelines.(227) CPRD GOLD has been used for extensive epidemiological research (208) and is representative of the UK population regarding age and sex.(228) As data are entered for clinical rather than research purpose, data quality can be variable, although the validity of many disease definitions in the CPRD GOLD has proven high.(229) In order to check the data quality for asthma research, the study included in the Chapter 4 of this PhD thesis tests the validity of asthma

recording in the CPRD GOLD. There are two sets of criteria to ensure data quality in the CPRD GOLD. The first criterion is patient acceptability based on registration status, completeness of patient records, age and gender. The second criterion is up-to-standard (UTS) time for practices, which ensures the suitability of the data for research purposes, ensuring or flagging the practice as 'up-to-standard' (UTS).(229) The UTS date is given on practice level and based on the continuity of recording and the number of recorded deaths.(208) The data contain mainly primary care data, but some secondary care data that has been sent to primary care practices may also be recorded in CPRD GOLD through manual entry by clinicians. CPRD was previously called the GPRD (General Practice Research Database).(208)

2.1.2 Data architecture

Coding

Most of the clinical information is registered in Vision using a dictionary known as Read terms with corresponding Read codes. The Read codes are hierarchically structured and are arranged into separate chapters, which broadly correspond to International Classification of Disease (ICD) chapters. The Read terms dictionary has many synonyms and different terms for one specific diagnosis, and it is up to the clinician to choose which one to use. Read codes are also used to record referrals to secondary care and may include specialty, urgency and the distinction between inpatient or outpatient referrals, but this information is not always recorded. Results of tests are also associated with Read codes. These codes can either be uploaded by the pathology department, or manually entered by healthcare providers in primary care.(227) Lifestyle factors and other measured variables such as weight or blood pressure are also coded using Read codes and are directly entered by primary care providers. These Read codes were developed in the early 1980s by Dr James Read and are a standard terminology for describing the care and treatment of patients.(230) The Read coding system has superseded the Oxford Medical Information System (OXMIS), which was developed in the 1970s based on the ICD 8.(231) Read codes are

a predecessor to the newer internationally unified coding system SNOMED CT, which was developed by the International Health Terminology Standards Development Organisation.(232) The Read codes correspond to medcodes in CPRD GOLD, which are ordered by frequency of use to reduce data size (the new codes are added at the end) and are a practical way to identify events in the CPRD GOLD.

Prescriptions of medications are recorded using product codes, based on a Gemscript dictionary in which Gemscript codes are the unique identifiers. The dictionaries can be modified depending on the local prescribing practices. Each Gemscript item includes the product name, administration route, strength, formulation and BNF code. A prescribing clinician can enter the dose, duration and patient advice, or can auto-populate these fields, and print prescriptions directly from the Therapy module. The vast majority of prescribing is electronic and therefore automatically captured. Immunisation records are separate from other therapies.

Data entry

Data entry in the Vision system is performed during clinical practice. Healthcare workers enter clinical information in Vision during consultations. These consultations include different activities which include not only direct patient contacts and attendances, but also telephone calls, administrative duties, repeat prescriptions of medications, information entry from secondary care, or emergency visits.

The data on patients are recorded during routine clinical care by general practitioners or healthcare workers. The clinician can either enter clinical terms as "Read terms" directly into a patient's medical history, or into a structured data area (for example, when the results of a test need to be filled in). The Vision system links Read terms to specific Read codes. Furthermore, the Vision system can be tailored to the preferences of a healthcare professional. One way to do this is to auto-populate preferred terms in the Read term box. Each Read code has a specific date. By default, this is the date of

recording, but can be changed by the clinician to record historical diagnoses or events. In addition to the Read code, a clinician can enter extra information as comments for each Read code, but this free text information was not available for this PhD project.

Figure 1 presents the data flow from the GP practices and hospitals to public health researchers.

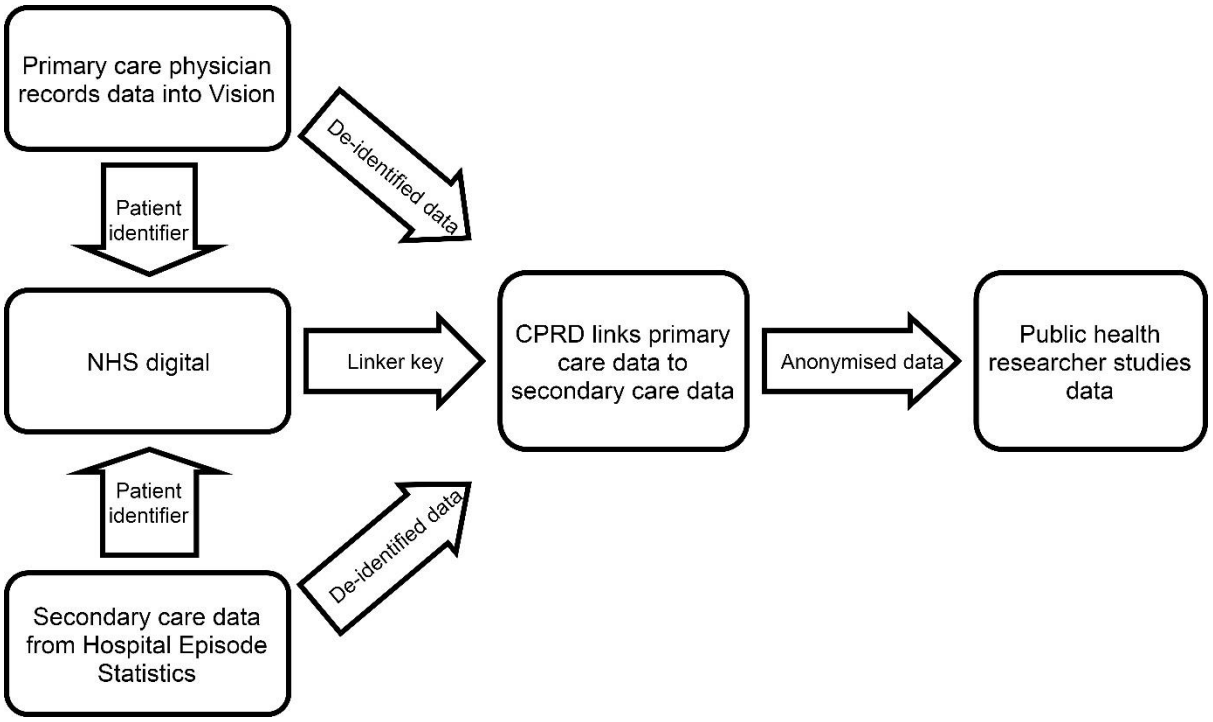


Figure 1: Data flow in CPRD GOLD. Adapted from www.cprd.com/researchpractice/researchgppractice.asp

The data are then uploaded to databases specified by the system in use; in the case of Vision, this is the CPRD GOLD. These de-identified data can then be linked to other data sources such as HES and are made available to researchers as de-identified raw data. There are multiple primary care EHR databases available in the UK with similar data flow schemes.

Research in practice

The practice data are regularly uploaded to the CPRD servers, after which they are processed, go through quality checks (acceptable patient status and practice-level UTS dates), de-identified and made available for research purposes.

Dictionaries of the Read and drug codes are available and searchable within a code browser. The electronic health records from patients who received either a diagnostic code or drug prescription code can be obtained using the codes and a period of interest.

Study approval for studies using CPRD GOLD data should be sought through a protocol submission to the Independent Scientific Advisory Committee (ISAC) for Medicines & Healthcare products Regulatory Agency (MHRA) Database Research, in addition to institutional ethical committees such as the LSHTM Research Ethics Committee.

After study approval has been obtained, the study data can be downloaded in two stages. The patients are uniquely identified by a patient identifier (patient id or patid). The first stage is the definition of the study population and construction of a list of patids that meet the inclusion criteria. The second stage is the extraction of all records of the patients included in the patid list. All records for the included patients are extracted, even records outside of the study period, to be able to define co-variates. Some variables such as ethnicity, gender, BMI or smoking status might not be recorded in the study period. In some cases, the values of recordings outside the study period can be used if there are no further entries on these variables. In this case, the assumption must be made that the values do not fluctuate greatly. After data extraction, the data from patients are presented in separate data files (including different types of data).

The data are organised in several files as follows:

- Index list: The master list for the specified cohort, which contains one unique patid per patient and date of the clinical diagnosis or medication of interest that led to the selection of the patient.
- Patient file: Demographic data on sex, year of birth, practice id and death date.
- Clinical file: Clinical diagnoses of the patient, including diagnostic Read codes and dates.
- Therapy file: Drug prescriptions of the patient, including therapy code, dates, quantity, dose, indication and formulation.
- Test file: Test records of the patient, including test type, date and result.
- Referral file: Referral records for the patient, including referral date and diagnosis associated with referral.
- Additional file: Records which provide data or measurements on variables such as patient height, weight, BMI and smoking status.

In addition, important medical events occurring prior to registration in a CPRD centre are also recorded.

In order to capture all events of a clinical concept (such as asthma), code lists with all Read codes that correspond to that event should be created. All records of a patient can then be searched using this code list. As such, there are multiple ways of defining a clinical concept in CPRD GOLD and EHR in general. It is useful to reconsider or update a previously used code list for each study as new codes may have been added. The Read code lists for this thesis were either developed by searching for all synonyms of a clinical concept in a code browser and exploring the codes hierarchically above the found codes, or shared with colleagues in the LSHTM and Imperial College for co-variates. All code lists which were used in this thesis are included in the appendix. Figure 2 depicts an example of a strategy to create a code list for asthma.

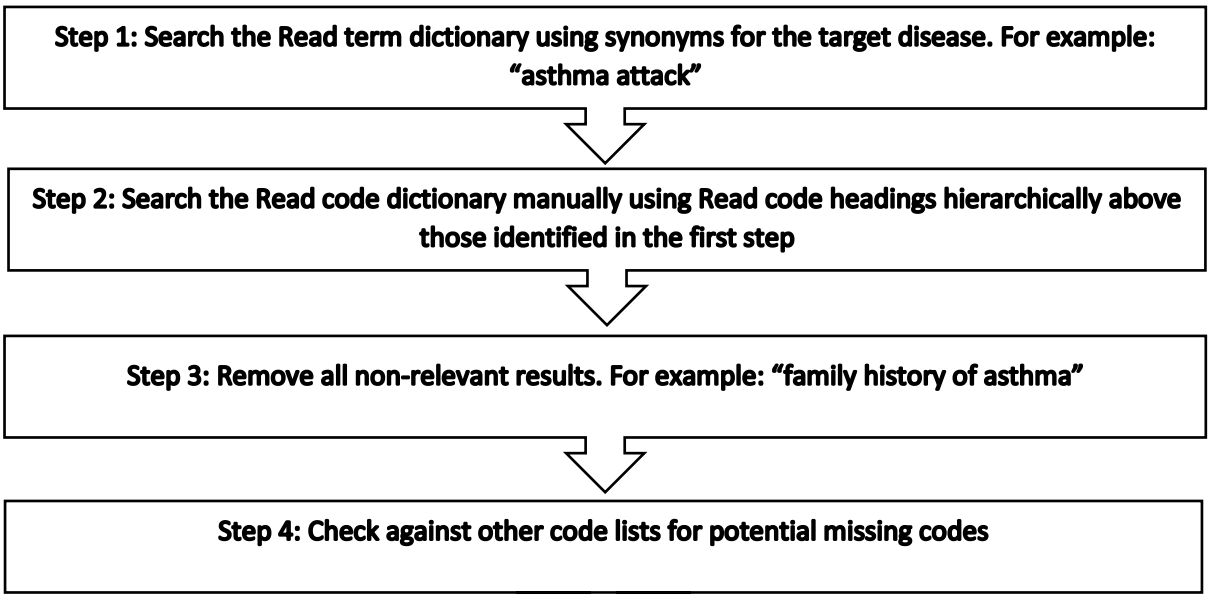


Figure 2: Strategy for identifying codes for asthma in the CPRD GOLD

The CPRD GOLD database can be linked with other data, including Hospital Episodes Statistics (HES) database for inpatient hospitalisations, Office for National Statistics (ONS) data for mortality data and Index of Multiple Deprivation (IMD) data for deprivation indices.

Patients are generally followed up from either the current registration date or date at which the practice was UTS and censored at last collection date or transfer out of practice or death.

2.1.3 Strengths and limitations

Strengths

There are multiple advantages to using the CPRD GOLD for epidemiological research, including the breadth of coverage, the size and long-term follow-up, its representativeness and data quality.(208) The database has data on morbidity and

lifestyle variables, linkage to secondary care via HES and mortality via ONS. It has a median follow-up of 5.1 years, which facilitates long-term epidemiological studies,(129,233) and is broadly representative of the UK population.(208,234,235) Furthermore, validation studies of some diagnoses have shown high positive predictive values (PPVs),(229) and studies on incidence rates have shown similar results to other UK data sources.(236,237) Another advantage of the CPRD GOLD is that the data quality is promoted by the Quality Outcomes Framework (QOF).

Quality Outcomes Framework

The QOF encourages recording of key data by GPs in England through an incentive payment, and therefore influences data quality in the CPRD GOLD.(238) It is an annual reward and incentive programme which gives more information on GP practice achievement results. This programme aims to reward practices for the administration of quality care and helps standardise advancement in the delivery of primary medical services (65). It is a voluntary process for all surgeries in England and was introduced in 2004, so this year was chosen as the index for most work included in this document. The indicators for the QOF are set annually. The QOF awards practices achievement points for the management of common chronic diseases (including asthma since 2006), the management of public health concerns and the implementation of preventative measures. This programme has enhanced aspects of the data in English General Practice.(208) The QOF indicators for asthma include sleeping difficulties, symptoms during the day and interference with daily activities.

Limitations

The primary purpose of the data in CPRD GOLD is to facilitate clinical care rather than research, so the data quality can be variable. The weaknesses of the CPRD GOLD include variability in completeness of data (for example, full blood counts are not conducted for every patient), sparse standard definitions (the need for code lists), missing information from secondary care (these are only available in CPRD GOLD if

the clinician manually records discharge or referral letters) and non-captured data (such as household information or age of disease onset).(208) In detail, if a Read code for a disease is absent in a patient, the disease must be considered as absent in the patient which might not necessarily be the case. There are no standardised definitions of diseases, so Read code lists and algorithms are needed. If secondary care information is not entered manually, this information is not recorded in CPRD GOLD, which might be the case if the information is not directly relevant for patient care. Free text data was not available for the purposes of this PhD project. Finally, some data may be missing, including some lifestyle data, family composition and over-the-counter medication. In addition, the data only provide information on medication prescriptions, not on medication dispensing or adherence to medication.

The CPRD GOLD records prescriptions of medication, which does not guarantee that patients also used their medication. The studies included in this PhD thesis did not directly study at the effects of medication, so the implications of this are limited. The validation study (Chapter 4) defined some of the algorithms using medication use, but the most practical algorithm only used a specific asthma code.

The outcome of the cohort study described in Chapter 6 are asthma exacerbations, and an asthma exacerbation can be defined by the prescription of oral corticosteroids. The medications a patient takes by BTS step were used as measurement of asthma severity. The healthcare practitioner who prescribed the medications assessed the exacerbation or severity of asthma, so the prescription records would be a good proxy for both regardless of whether the patient used the medication.

While the direct adherence cannot be measured in EHR, it is possible to estimate it by studying the percentage of time for which patients at least had medication to cover. For example, if a patient has a prescription every 45 days but the amount prescribed

only covers 30 days, the patient would be covered for 30/45 of the time (the medication possession rate).

Prescription records and patient's self-reported drug exposure were compared in the French PGRx database (Pharmacoepidemiologic General Research eXtension). Self-reported drug exposure itself is not a perfect measure, as it can be affected by memory errors and other biases. The agreement between the two data sources was kappa = 0.83, (95% CI: 0.81-0.85).(239)

Another way to study adherence is estimating the percentage of issued prescriptions that are obtained from the pharmacy, in which case pharmacy-level data are needed. Treatment adherence can also be defined as missing one or more scheduled appointment if recorded, or coding indicating medication non-compliance.(240)

A cohort of asthma patients containing primary and secondary care information can be obtained by using a linked cohort from patients attending practices in England and linking their CPRD GOLD and HES records. Linkage from CPRD GOLD to patient-level datasets, including HES and ONS is only available for consenting English practices. These linkages are present in about 70% of English practices and 55% of all UK practices contributing to the CPRD GOLD.(241,242)

2.2 Hospital Episodes Statistics (HES)

Background

The second data source of this study was the UK Hospital Episodes Statistics database (HES). HES is a data warehouse which holds details of admissions, outpatient appointments and A&E attendances at NHS hospitals in England.(243) This database enables health care providers to be paid for the care that they deliver. The system covers all NHS trusts in England, including primary care and mental health trusts, and emergency care hospitals. HES is an administrative database which is composed of data on patient demographics, clinical diagnoses and procedures performed in the hospital for every NHS hospital admission. Most hospital activity in the UK is funded by the NHS (98-99%).(244)

The Health and Social Care Information Centre prepares and makes the HES databases available for secondary purposes, including service planning, commissioning and academic and pharmaceutical research. The practices contributing to CPRD GOLD located in England have their data linked to HES (not the practices in Wales, Scotland and Northern Ireland). The linkage from CPRD GOLD to HES is available from April 1997 onward and is generally available in bi-annual builds. This affects study designs of epidemiological studies using HES because the data only becomes available twice each year, and the HES data are distributed over several databases. The HES Admitted Patient Care (HES APC) was the main data source on secondary care for this project. Accident and Emergency attendances (HES A&E) and outpatient services (HES outpatient) are held in separate databases.

HES Admitted Patient Care (APC)

The HES Admitted Patient Care (APC) database includes data on hospital admissions including any secondary care-based activity that requires a hospital bed. As such, it includes both emergency and planned admissions, day care and childbirths. The

admission diagnoses in HES APC are coded using a modified ICD-10 (International Classification of Diseases) system. The procedures are coded using standard code for hospital procedures: the OPCS4 codes (Classification of Interventions and Procedures).

The data files in HES APC are first structured by financial year, then by hospitalisations or "spells" (spell is defined as a single stay in the hospital), which can consist of one or more episodes (episode defined as the care under one consultant).(244) Each episode can contain multiple diagnoses. In addition to the data on diagnoses and medical procedures, HES APC also holds data on admission/discharge dates, admission methods, care provider and the individual patient's postcode.(244)

The advantages of using HES APC for epidemiological research are its universal coverage, availability of linkage to other databases, and standardised ICD-10 coding.(244) It is frequently used in health economics, as the information on costs of care are readily available.(245). There are several limitations to the use of HES APC. These include the variation of coding between different hospitals, the sensitivity to admission thresholds (if this differs between hospitals or guidelines) and the patients that opt out of data recording for research purposes (2.3% of episodes).(244) Clinical coders rely on discharge summaries in order to enter data correctly, and as such, data quality can vary between hospitals. In addition, financial incentives exist in order to improve coding.(246) Some conditions have a higher remuneration than others, so hospitals have an incentive to code multiple and specific comorbidities.

HES A&E and HES outpatient databases

The HES A&E and HES outpatient databases add information on patient attendances that do not result in a hospitalisation. However, their data content is lacking compared with HES APC, which limits their usefulness for epidemiological studies.(244) HES

A&E data contain only a limited number of different codes, and HES outpatient data frequently only contain information on the healthcare provider. These databases were not used for the research presented in this thesis.

2.3 Office of National Statistics (ONS)

The third data source of this thesis is the Office of National Statistics database.

The ONS is the executive office of the UK Statistics Authority. Its purpose is to collect and publish statistics related to the economy, population and society of the UK.(247) As the ONS collects death statistics of the UK, the data it provides can be linked to CPRD GOLD to obtain more accurate data on mortality. For areas in Scotland, Northern Ireland and Wales the responsibility for some fields of statistics is transferred to the devolved governments and their ONS data is not readily linkable, so the linked data is only available for England.

A list of the causes of death (as coded on the individual's death certificate) for linked patients can be obtained from the database by providing CPRD headquarters with a list of patient ids. Cause of death is coded according to the WHO ICD-10 standard.

Index of Multiple Deprivation (IMD)

The ONS also provides data on socio-economic status through the IMD. The IMD is a measure of relative deprivation for small English areas, and is available for different time points.(248) I chose the IMD 2015 for the studies in this thesis. The English indices of deprivation measure the relative levels of deprivation in small areas of England, which are called "lower layer super output areas". The indices of deprivation are assigned based on the postal code of residence. Deprivation is described as the decile or quintile of the deprivation index of the patient postcode. The deciles have been calculated by ranking the 32,844 small areas in England from the most deprived to the least deprived area, and subsequently dividing them into 10 equal groups.

The IMD mainly uses seven indicators: income, employment, education, health and disability, barriers to housing and services, living environment and crime.

Deprivation can then be categorised into quintiles or deciles, with 1 being the least deprived and 5 or 10 the most deprived. There is no definitive threshold above which an area can be described as deprived, as the indices form a continuous scale of deprivation.

2.4 Asthma questionnaire

A possible way to ascertain asthma status is by using asthma questionnaires. In large epidemiological studies without EHR, questionnaires on asthma symptoms and history are frequently used.(42,249,250) Questionnaires on symptoms correlate well with a clinical diagnosis of asthma and can provide repeatable results.(251,252)

A two-page questionnaire based on the NICE and BTS guidelines was designed to construct an independent reference standard for the validation of asthma patients in the CPRD GOLD (Chapter 4). The full questionnaire is included in the appendix of Chapter 4. This questionnaire was sent out to the GPs of 684 potential asthma patients and is included in the appendix of Chapter 4. The information by the GP was then reviewed by two study physicians to construct the reference standard.

This questionnaire included several questions in order to ascertain or reject an asthma diagnosis. The first section of the questionnaire requested to confirm whether the patient had asthma, whether this diagnosis was confirmed by a respiratory physician and whether the patient had evidence of reversible airway obstruction. The second section sought information on additional factors that supported the diagnosis, such as history of atopic disorder, wheeze, spirometry results and FeNO measurements and the QOF indicators (including sleeping difficulties, usual asthma symptoms and interference with daily activities). The questionnaire also asked for the patient's smoking status and comorbidities.

A questionnaire was sent to the general practitioners of a random sample of patients who fit in a certain algorithm to obtain information for the gold standard. The questionnaire is based on the "British guideline on the management of asthma" by the British Thoracic Society and Scottish Intercollegiate Guidelines Network 2014 and the asthma QOF indicators.

The main aim of the questionnaire was to allow the study physicians to differentiate individuals with asthma from individuals without asthma. The first question (A) asks the GP's evaluation of the asthma status. If the evaluation was either positive or uncertain, we asked additional questions to check asthma status. These include the confirmation by a respiratory physician (B1), evidence of reversible airway obstruction (B2) and the year of diagnosis (B3).

The BTS 2014 guidelines specify an asthma diagnosis is predominantly based on the recognition of a characteristic pattern of symptoms and signs and the absence of an alternative explanation for those symptoms. Features that increase the probability of asthma include the classic asthma symptoms (wheeze, breathlessness, chest tightness and cough), a family history of asthma and/or atopic disorder, widespread wheeze heard on auscultation of the chest, otherwise unexplained low FEV1 or PEF and otherwise unexplained peripheral blood eosinophilia. We added FeNO measurements due to the scientific interest in these measurements at the time. These additional features were explored in part B4 of the questionnaire.

Question B5 of the questionnaire refers to the QOF indicators. Question B6 asks the smoking status of the patient, as that was important information to be able differentiate asthma from COPD patients who were likely to be picked up with the less stringent algorithms, as the treatment of both diseases overlaps. Question B7 identifies patients with other respiratory conditions that could be identified with the asthma validation algorithms. Question C asks if the patient had a history of asthma,

if they did not have a current asthma diagnosis. This information was useful to assess whether past diagnoses were frequently picked up in the recordings of the last 2 years.

There were some slight changes to the questionnaire on advice from CPRD regarding the remuneration of the GP's after the ISAC protocol. There were also some minor amendments to the questionnaire to clarify the procedure for returning the questionnaire and to insert the patient identifier tables we use. The sentence "To answer this questionnaire, please refrain from using the data recorded in CPRD as the aim of this study is to see how reliable CPRD is." was removed to avoid confusion.

2.5 Data management

The data were stored on a secure server with a backup copy on an encrypted external hard disk and the data required for ongoing research will be kept in line with CPRD and institutional guidance.

Chapter 3: Systematic review: Validation of asthma recording in electronic health records

Summary

- This systematic review found 13 studies with details on their methods for asthma validation and reported test measures.
- Asthma validation studies using EHRs are very varied in their approach to the validation, which seems driven by the nature of the data, the study questions to be answered and the reference standards used.
- There were 3 main reference standard types used for validation: manual validation, comparison with an independent database and comparison with a questionnaire.
- Identifying asthma cases in electronic health records is possible using each of the discussed validation methods with high sensitivity, specificity or positive predictive value, by combining multiple data sources, or by focussing on specific test measures.
- Different case definitions within a single data source have different validity highlighting the importance of testing a range of case definitions.
- Validated case definition algorithms are often specific to the database they were developed in, limiting their generalisability.

3.1 Preface

This chapter reports on a systematic review of validation methods of asthma recording in electronic health records. The primary objectives of this systematic review were to provide an overview of the methods used in the literature for validating asthma diagnosis in EHR and to provide the corresponding estimates of the validation test measures. To do this, I synthesised and appraised the current evidence and test values of strategies to identify asthma patients in electronic health records.

The motivation for this systematic review was to prepare for the subsequent validation study of algorithms to identify asthma patients in the CPRD GOLD (Chapter 4). The first paper describes the protocol of the systematic review and the second paper contains the systematic review itself. The search algorithms for this systematic review are included in the appendix.

Previously to this systematic review, two systematic reviews had been conducted with similar study questions. The first review by Sharifi et al was published in 2013 and contained a review of validation methods to capture acute bronchospasm in administrative or claims data.(253) This study found two validation studies of bronchospasm codes.(254,255) However, the study was limited to administrative and claims databases which originated in either the United States or Canada and only included a symptom (bronchospasm) rather than asthma itself. The second study was published in 2017 by Al Sallakh et al. The authors explored approaches to defining asthma or assessing asthma outcomes using electronic health record-derived data in the literature from 2014 and 2015 and examined the clarity of reporting.(256) This review focussed solely on how asthma was defined and did not include an overview of test measures or validation statistics such as positive predictive values (PPVs), negative predictive values (NPVs), sensitivity or specificity and was published shortly before the systematic review included in this thesis.

In epidemiological studies using electronic health records, the validity of codes and algorithms are quantified using diagnostic accuracy measures: the positive predictive value, negative predictive value, sensitivity and specificity. Several test measures were reported as different study types need to focus on various database measures to identify asthma diagnoses. For example, studies of risk factors for asthma need high sensitivities and PPVs, while studies on asthma prevalence need high sensitivities and Youden indices.(257) PPVs tend to be higher when derived from databases using both diagnosis and prescription data when compared with databases relying only on diagnosis data.(258)

The PPV is the proportion of positive results that are true positive results, while the NPV is the proportion of negative results that are true negative results. Sensitivity measures the proportion of actual positives that are correctly identified as such, and the specificity measures the proportion of actual negatives that are correctly identified as such. The sensitivity and specificity can be combined to form the Youden's index or Youden's J statistic, which is defined as (J= sensitivity +specificity -1).(259) Another measure of validity is face validity, in which researchers compare the prevalence of a disease within a population with the prevalence in the data. This method can only provide rough estimates and is not exact. For example, If the over-and underdiagnosis rates of a disease are similar, the face validity will not be able to measure any of those.The PPV is the most reported test statistic in the CPRD GOLD and EHRs in general,(229) and useful to determine the percentage of patients with asthma codes who actually have asthma. The fourth chapter of this thesis describes a study to find the optimal algorithm to identify asthma patients from the CPRD GOLD using questionnaires to GPs, based on PPVs.

Quality assessment of the studies included in the systematic review was done using the QUADAS-2 tool. The QUADAS-2 tool was constructed to allow for more transparent rating of bias and applicability of primary diagnostic accuracy studies

and is available from the QUADAS website (www.quadas.org). This tool is included in the appendix of this chapter.

In conclusion, the primary objectives of this systematic review were to provide an overview of both the methods with which asthma diagnosis recording has been validated in EHR and the estimates of the validation test measures. Specifically, I listed the EHR databases, algorithms, diagnostic criteria and estimate values of the PPVs, NPVs, sensitivities and specificities.

The protocol of the systematic review was originally published in *BMJ Open*, and is available here:

Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ.: 'Validation of asthma recording in electronic health records: protocol for a systematic review'. BMJ Open. 2017 May 29;7(5), https://www.ncbi.nlm.nih.gov/pubmed/28554919

The systematic review was published in Clinical epidemiology, and is available on:

Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. 'Validation of asthma recording in electronic health records: a systematic review'. Clin Epidemiol. 2017 Dec 1;9:643-656. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5716672/

3.2 Research paper 1: Protocol for a systematic review

Validation of asthma recording in electronic health records: protocol for a systematic review

Authors: Francis Nissen,[1] Jennifer K Quint,[2] Samantha Wilkinson,[1] Hana Mullerova,[3] Liam Smeeth,[1] Ian J Douglas[1]

[1]*Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK*

[2]*National Heart and Lung Institute, Imperial College, London, UK*

[3]*RWD & Epidemiology, GSK R&D, Uxbridge, UK*

ABSTRACT

Background: Asthma is a common, heterogeneous disease with significant morbidity and mortality worldwide. It can be difficult to define in epidemiological studies using electronic health records, as the diagnosis is based on non-specific respiratory symptoms and spirometry, neither of which are routinely registered. Electronic health records can nonetheless be valuable to study the epidemiology, management, health-care utilization and control of asthma. For health databases to be useful sources of information, asthma diagnoses should ideally be validated. The primary objectives are to provide an overview of the methods used to validate asthma diagnoses in electronic health records and summarise the results of the validation studies.

Methods: EMBASE and MEDLINE will be systematically searched for appropriate search terms. The searches will cover all studies in these databases up to October 2016 with no start date and will yield studies that have validated algorithms or codes for the diagnosis of asthma in electronic health records. At least one test validation measure (sensitivity, specificity, positive predictive value, negative predictive value

or other) is necessary for inclusion. In addition, we require the validated algorithms to be compared with an external golden standard, such as a manual review, a questionnaire or an independent second database. We will summarise key data including author, year of publication, country, time period, date, data source, population, case characteristics, clinical events, algorithms, gold standard and validation statistics in a uniform table.

Ethics and dissemination: This study is a synthesis of previously published studies and, therefore, no ethical approval is required. The results will be submitted to a peer-reviewed journal for publication. Results from this systematic review can be used to study outcome research on asthma and can be used to identify case definitions for asthma.

Trial registration number

The protocol has been registered in the PROSPERO database with registration number CRD42016041798.

Keywords

Asthma, Validation, Electronic Health Records, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value

Strengths and limitations of this study

To our knowledge, this is the first systematic review to identify and evaluate methods used to validate a recording of asthma diagnosis in electronic health records.

The review of validation of asthma diagnosis in electronic health records could inform selection of asthma identification algorithms used by future health outcome studies and identify any gaps in quality and scope of validation studies. It will also provide an overview of the algorithms with their PPV, NPV, sensitivity or specificity.

Different databases may validate different algorithms to identify asthma, which might limit the generalisability of these algorithms as they are context-specific.

This review is focused on the methodology of asthma recording validation, and not on all outcome results of studies (except the validation results). Because of this, publication bias might be an issue (methods that do not find positive results may be less likely to have been published).

BACKGROUND

Asthma is a common chronic inflammatory disease of the airways. This condition is characterised by a variable expiratory airflow limitation which is generally reversible. The core symptoms are cough, wheeze, breathlessness and chest tightness.(2) Asthma episodes can range from mild attacks, which interrupt daily life and work productivity, to severe and life-threatening attacks.(3) Asthma is inherently variable, and individuals will experience fluctuating symptoms. Most commonly, asthma emerges in childhood, but it can also arise in adulthood. Therefore, adult asthma consists of both persistent or relapsed childhood disease and true incident adult disease. There is no cure, but with the right treatment, symptoms can usually be managed and asthma patients can lead their lives without disruption.(260)

The widespread adoption of electronic health records (EHR) means that large population-based primary and secondary care databases are available, proving a great opportunity for research on asthma and other diseases. The availability of routinely generated longitudinal records for research has dramatically increased over the last decades.(260) However, the primary function of EHR is to support healthcare clinical decision making, not research purposes. The integrity of the research generated from EHR may be questionable, unless data are thoroughly validated for this purpose. (209,261–263)

EHR are a digital reflection of the paper medical chart, while the main purpose of administrative claims data is administration of reimbursements to healthcare providers for their services. This systematic review will only consider data from EHR, as the quality measures between the two types of data can be markedly different.(264,265)

EHR store information about diagnoses as clinical codes. A single code, or an algorithm consisting of multiple codes, can be used to retrieve records from EHR, and additional restrictions can be applied such as age or exclusion of other diseases. (263,266)Alternatively, several authors have recently used natural language processing and machine learning techniques to automate algorithm generation for the identification of asthma diagnoses from large databases.(255,267,268) The most common method to assess the validity of algorithms is to compare them with a gold standard such as another linkable dataset or request a verification from the treating physician or the patient via a questionnaire.(266) Another approach is active case detection where the databases are constantly screened to identify cases that emerge.(269)

Several limitations apply to the validation of diagnosis recording in EHR. First, individual databases often only cover a single care setting (primary or secondary

care), as such case ascertainment only relies on a partial description of the healthcare pathway [15].(270) Another issue is that the validity of different diseases will not necessarily be the same in a given dataset. For example, mental health disorders such as anxiety or depression might be coded using less specific symptoms, whereas the validity of diagnoses with a very high specificity such as breast cancer is likely to be superior. There have been multiple studies which have measured the validity of specific databases for asthma.(271,272) Sharifi et al. have conducted a systematic review on validated methods to capture acute bronchospasm using administrative or claims data,(253) which yielded two validation studies of bronchospasm codes.(254,255)

This systematic literature review aims to provide an overview of methods used to validate asthma diagnoses, specifically in EHR. Such a study has not yet been published in the medical literature, to the best of our knowledge.

Research question

The primary objectives of this systematic review are to provide an overview of both the methods with which asthma diagnosis recording has been validated in EHR and the estimates of the validation test measures.

The questions of interest for this systematic review are:

- Which EHR that are not only based on claims data have been used to obtain information on the diagnosis of asthma?
- Which algorithms have been used to define an asthma diagnosis (including diagnostic codes, possible spirometry tests and clinical descriptions)?
- How were the diagnostic criteria applied to the data sources and which other approaches have been used to validate a case definition?

- What are the estimates for the PPV, NPV, specificity and sensitivity for a diagnosis of asthma in EHR that are not solely claims-based?

METHODS

MEDLINE and EMBASE will be searched for the terms "asthma", "validation", "electronic databases" and synonyms for each of these terms. In addition, reference lists of review articles and retrieved articles will be reviewed. The PRISMA flow diagram of this protocol, from Moher et al.,(273) can be found in figure 1 and the search strategy can be found in the supplementary file.

Inclusion criteria

Any type of observational study design that used EHR to validate the recording of an asthma diagnosis will be considered. Articles will only be considered if published in English and before October 2016 without any specific start date. Within the databases, we will consider asthma diagnoses based on both structured data (such as laboratory results and prescriptions) and free text data (such as spirometry results). We require the validated algorithms to be compared with an external gold standard, such as a manual review, questionnaires (completed by the patient or their physician) or an independent second database. We will include algorithms formed of single codes, those requiring multiple case characteristics and algorithms generated by natural language processing or machine-learning.

Exclusion criteria

Studies which involve pharmacovigilance databases (signal detection or spontaneous reporting), studies without validation process of asthma recording and conference abstracts will be excluded. Algorithms used in databases originating from only claims data will also be excluded, as a systematic review on the validated methods to capture acute bronchospasm using claims data has been published recently.(253)

Two independent authors will scan the abstracts and titles against the research questions and exclusion criteria and select articles for full-text review. After this full-text article review, eligibility for inclusion in the report will be decided by consensus or arbitration by a third reviewer. A uniform table with information of each included study will be populated after data extraction, which will include information on the author, date of publication, journal, database, algorithms, population, gold standard and test measure(s).

Data synthesis

Studies and study data will be managed using EndNote and Microsoft Excel, respectively. The methods for asthma recording validation will be summarised in a narrative synthesis and tables describing all identified verification processes, and their results. These results will consist of the recorded PPV, NPV, sensitivity and specificity of the included studies. Where possible, these tests will be calculated if they are not reported within the study.

Dissemination and ethics

This study is a synthesis of previously published studies, so no ethical approval is required. The protocol has been registered in the PROSPERO database with registration number CRD42016041798. The results will be submitted for publication and will be disseminated through research conferences and peer reviewed journals.

Funding statement

This work was supported by GlaxoSmithKline (GSK), through a PhD scholarship for FN with grant number EPNCZF5310. The publishing of this study was supported by the Wellcome Trust: grant number 098504/Z/12/Z.

Competing interests

FN and SW are funded by a GSK scholarship during their PhD programs. JQ reports Grants from MRC, BLF, Wellcome Trust, and has received research funds from GSK, AZ, Quintiles IMS and had personal fees from AZ, Chiesi, BI. HM is an employee of GSK R&D and owns shares of GSK Plc. ID is funded by, holds stock in and has consulted for GSK.

Contributors

JQ, ID, LS and HM were responsible for developing the research question and have advised on the data collection and search strategies. FN drafted the manuscript, FN and SW will review the literature and summarise the found papers. ID is responsible for study management and coordination. All authors have read, commented on and approved the final manuscript.

Data sharing statement

Study data will be available on request to FN once the research team has completed pre-planned analyses.

Research paper cover sheet

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Francis Nissen
Principal Supervisor	Ian Douglas
Thesis Title	Asthma in electronic health records: validation & phenotypes.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	BMJ Open. 2017 May 29;7(5)		
When was the work published?	2017 May 29		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have developed the research question and drafted the manuscript for this systematic review protocol.

Student Signature: _____ **Date:** 13/08/2018

Supervisor Signature: _____ **Date:** 14/08/2018

3.3 Research paper 2: Systematic review

Validation of asthma recording in electronic health records: a systematic review

Authors: Francis Nissen,[1] Jennifer K Quint,[2] Samantha Wilkinson,[1] Hana Mullerova,[3] Liam Smeeth,[1] Ian J Douglas[1]

[1]*Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK*

[2]*National Heart and Lung Institute, Imperial College, London, UK*

[3]*RWD & Epidemiology, GSK R&D, Uxbridge, UK*

ABSTRACT

Objective: To describe the methods used to validate asthma diagnoses in electronic health records and summarise the results of the validation studies.

Background: Electronic health records are increasingly being used for research on asthma to inform health services and health policy. Validation of the recording of asthma diagnoses in electronic health records is essential to use these databases for credible epidemiological asthma research.

Methods: We searched EMBASE and MEDLINE databases for studies that validated asthma diagnoses detected in electronic health records up to October 2016. Two reviewers independently assessed the full text against the predetermined inclusion criteria. Key data including author, year, data source, case definitions, reference standard and validation statistics (including sensitivity, specificity, PPV and NPV) were summarised in a uniform table.

Results: Thirteen studies met the inclusion criteria. Most studies demonstrated a high validity using at least one case definition (PPV>80%). Ten studies used a manual

validation as the reference standard; each had at least one case definition with a PPV of at least 63%, up to 100%. We also found 2 studies using a second independent database to validate asthma diagnoses. The PPV's of the best performing case definitions ranged from 46% to 58%. We found one study which used a questionnaire as the reference standard to validate a database case definition; the PPV of the case definition algorithm in this study was 89%.

Conclusions:

Attaining high PPV's (>80%) is possible using each of the discussed validation methods. Identifying asthma cases in electronic health records is possible with high sensitivity, specificity or positive predictive value, by combining multiple data sources, or by focussing on specific test measures. Studies testing a range of case definition show wide variation in the validity of each definition, suggesting this may be important for obtaining asthma definitions with optimal validity.

Strengths and limitations of this study

The review of validation of asthma diagnosis codes in electronic health records informs selection of asthma definitions used by future studies and identify any gaps in quality and scope of validation studies. It also provides an overview of the case definitions and algorithms with their PPV, NPV, sensitivity or specificity.

Validated case definition algorithms are often very specific to the database they were developed in, limiting their generalizability.

Publication bias might be an issue as methods that do not find favorable results may be less likely to have been published.

BACKGROUND

Asthma is one of the most common chronic diseases, and its core symptoms are cough, wheeze, breathlessness and chest tightness.(2) There is no cure, but with the right treatment, symptoms ranging from mild attacks to severe and life-threatening exacerbation[2] can be managed.(3) Despite this, a sizeable percentage of asthma patients are poorly controlled.(274,275)

Electronic health records (EHR) have been widely adopted, which allows for the construction of large population-based patient databases. The availability of these routinely generated longitudinal records for research has greatly increased over the last decades.(260) However, the accuracy of diagnoses recorded in these large databases may be low, which would introduce bias into studies using the data. Unless the data are validated for research, the quality of studies generated from EHR's may be debatable.(209,261–263) Furthermore, the validity of different disease definitions is not always the same in a given dataset. Some diseases (such as asthma) might be coded

using less specific symptoms, whereas the validity of diagnoses with very specific symptoms (such as tension pneumothorax) is likely to be better.

EHRs predominantly store information about diagnoses as clinical codes. A single code, or a case definition consisting of multiple codes (with or without additional information such as tests or prescribing) can be used to retrieve records from EHRs, and additional restrictions can be applied such as age or exclusion of other diseases.(263,266) Validity of coding is generally assessed by comparing a code (or algorithm) with i) the diagnosis as verified by the treating physician either by manual review of the chart notes or in clinic, ii) a reference standard such as another linked dataset or iii) a patient questionnaire.(266) A previous systematic review by Sharifi et al reviewed validation methods to capture acute bronchospasm in administrative or claims data;(253) this review identified two validation studies of bronchospasm codes.(254,255) However, the study was limited to administrative and claims databases, from the United States and Canada. Al Sallakh et al explored approaches to defining asthma or assessing asthma outcomes using electronic health record-derived data in the recent literature (calendar years 2014 and 2015) and examined the clarity of reporting.(256) This systematic review focuses on how asthma was defined and does not include an overview of test measures or validation statistics.

There is currently no consensus on approaches to defining asthma or assessing asthma outcomes using electronic health record-derived data. We explored these approaches in the recent literature and examined the clarity of reporting.

Research objective

The primary objectives of this systematic review are to provide an overview of the methods used in the literature for validating asthma diagnosis in EHR, and the corresponding estimates of the validation test measures.

METHODS

The methods are described in detail in the study protocol.(276) We searched Medline and Embase up to October 2016 for relevant articles. Our search strategy was composed of the following sets of terms: [1] electronic health records or databases AND [2] [validity or validation or case definition or algorithm or sensitivity or specificity or positive predictive value or negative predictive value] AND [3] the medical subject heading terms for asthma. Reference lists of articles of interest were reviewed to add potential additional studies in which a validation of asthma diagnosis was done. The PRISMA flow diagram can be found in figure 1 and the search strategy can be found in the appendix. We considered any type of observational study design that used EHR to validate the recording of a diagnosis of asthma. In addition, we required a clear case definition to define asthma from EHR, including a description of the validation of said case definition through at least one test measure (sensitivity, specificity, Positive Predictive Value (PPV) or Negative Predictive Value (NPV)). Two investigators (FN and SW) separately assessed the abstracts and full text of each potential study against our inclusion criteria; disagreements were resolved through a third investigator or by discussion to reach consensus. The first author independently extracted all relevant data regarding methodologic elements of included studies; author, year of publication, country, time period, date, data source, population, case characteristics, clinical events, algorithms, reference standard and validation statistics. Bias was assesses using QUADAS-2 tailored to this specific review.(277)

The questions of interest for this systematic review are:

- Which EHR databases were used to obtain information on the diagnosis of asthma?

- Which case definitions, algorithms or codes were used to define an asthma diagnosis?
- How were the diagnostic criteria applied to the data sources and which other approaches have been used to validate a case definition algorithm?
- What are the estimates for the PPV, NPV, specificity and sensitivity for a diagnosis of asthma in an EHR?

Inclusion criteria

Any type of observational study design which validated the recording of an asthma diagnosis in EHR was considered. Articles were only considered if published in English and published before October 2016 without any specific start date. Within the databases, we considered asthma diagnoses based on both structured data (such as laboratory results and prescriptions) and unstructured data (such as spirometry results). We required the validation case definitions to be compared with an external reference standard, such as a manual review, questionnaires (completed by the patient or their physician) or an independent second database. We included case definitions formed of single codes, those requiring multiple case characteristics and case definitions generated by natural language processing and/or machine-learning.

Exclusion criteria

EHR are a digital reflection of the key facts a healthcare provider needs to record in order to facilitate ongoing and potentially complex clinical care. By contrast, the main purpose of administrative claims data is administration of reimbursements to healthcare providers for their services. This systematic review included only studies from EHR, as the quality measures between the two types of data can be markedly different; studies using administrative claims data were excluded. Studies involving pharmacovigilance databases (signal detection or spontaneous reporting), studies

without validation of asthma recording, and conference abstracts were excluded.(264,265)

Data synthesis

Studies and study data were managed using EndNote and Microsoft Excel, respectively.

The methods for validation of asthma recording in the included studies were outlined in a narrative synthesis. In addition, table 1 summarises the methods and table 2 describes the results, consisting of the recorded PPV, NPV, sensitivity and specificity of the included studies.

RESULTS

In total, 1,346 titles were found in the EMBASE and MEDLINE databases, of which 946 were non-duplicates. Of those, 54 articles were reviewed in full text, we found 13 articles that contained a validation process of asthma diagnosis that met all eligibility criteria. Characteristics of the 13 included studies ordered by year of publication are summarised in table 1, and the study results are displayed in table 2. The asthma prevalence necessary for the interpretation of PPVs and NPVs is presented in table 1, where available.

The reference standard used to validate the asthma diagnosis in the EHR differed between the studies: ten studies used manual validation by a clinician, while two other studies compared the studied records with independent linked databases and one study used patient questionnaires. The test measures also differ between the different papers, encompassing sensitivity, specificity, PPV and NPV. We focus on 13 studies in this review, ordered by reference standard used and by date of publication. Bias assessment results using QUADAS-2 are presented in table 3.

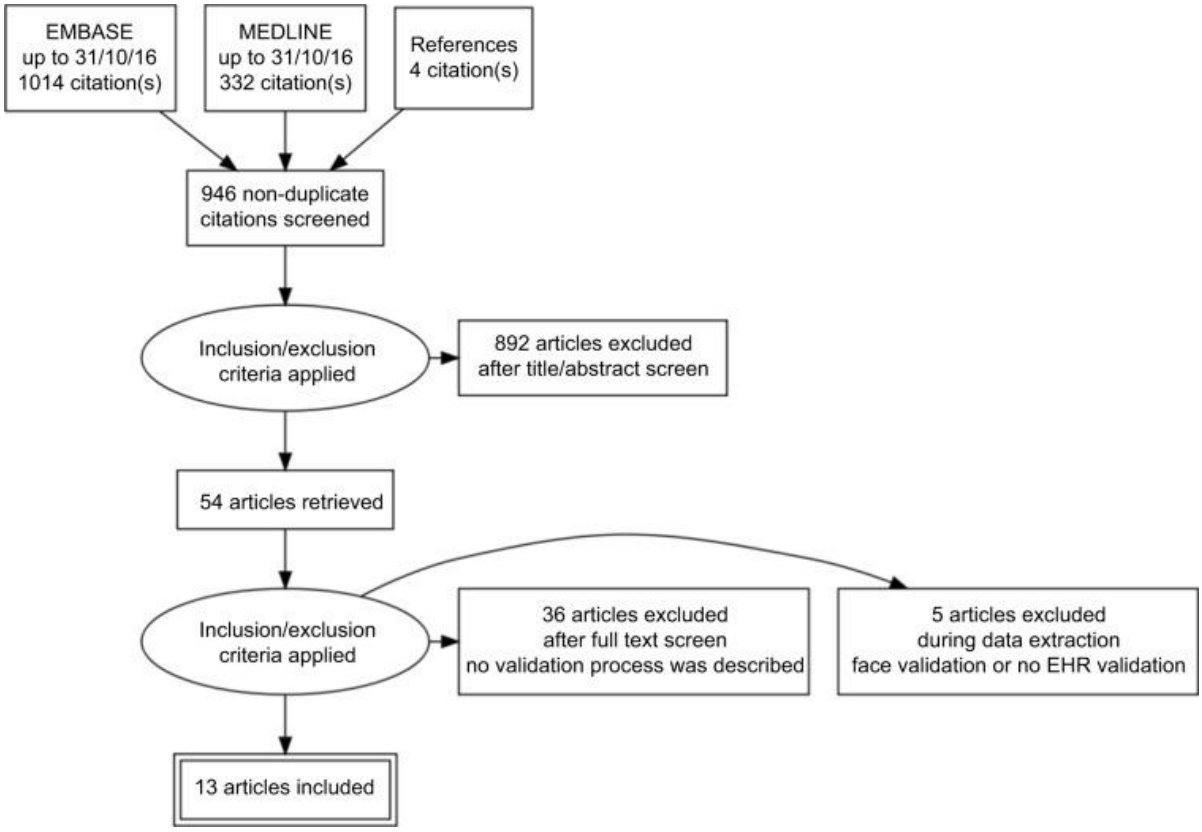


Figure 1: PRISMA diagram

Author, year, country	Data source, population	Sample/case characteristics	Clinical event	Algorithm	Validation
Manual validation					
Xi et al,(271) 2015 Canada	2 large academic primary care clinics Primary care	398 randomly selected patients 16 years and older	Asthma code COPD code Other respiratory condition code Other condition code	Search algorithms: 1. Asthma in disease registry 2. Billing code 3. Asthma in CPP 4. Asthma medications 5. Asthma in chart notes 6. Asthma in CPP OR billing code 493 7. Asthma in CPP OR billing code 493 (exclusion codes 491,492, and 496) 8. (Asthma in chart notes OR asthma medications) AND billing code 493 9. (Billing code 493 OR medications) AND asthma in chart notes 10. Billing diagnostic code 493 AND asthma in chart notes	Manual review
Engelkes et al,(278) 2014 the Netherlands	ICPI: Dutch GP EHR Primary care	63,518 potential cases identified 22,699 cases after automated text validation Children aged 5–18	Definite, probable, and doubtful cases of asthma	Combination of ICPI communication codes, clinician codes, drug names and free text generated by a machine-learning algorithm (RIPPER)	22,699 cases manually validated, 14,303 asthma cases found
Afzal et al,(267) 2013 the Netherlands January 2000–January 2012	ICPI: Dutch GP EHR Primary care	63,618 potential asthma cases identified, children aged 5–18	Definite, probable, and doubtful cases of asthma	Combination of ICPI communication codes, clinician codes, drug names and free text generated by a machine-learning algorithm (RIPPER)	5,032 patients manually validated by clinician
Dexheimer et al,(279) 2013 United States	1 pediatric A&E department	15,163 assessed, 1,100 asthma patients all asthma patients (2–18 years) in a 3-month time window	Asthma code	Bayesian network system previously used on claims data (Sanders)	Paediatric asthma/respiratory distress protocol filled in for identified patients
Wu et al,(280) 2013, 2014 United States	Children enrolled in the Mayo Clinic sick-child daycare program, Secondary care	112 children younger than 4	ICD-9 codes Natural language	Natural language processing (logic) Natural language processing (machine learning)	Manual review by a clinician

Author, year, country	Data source, population	Sample/case characteristics	Clinical event	Algorithm	Validation
Kozyrskyj et al,(281) 2009 Canada	SAGE: birth cohort of 16,320 children born in 1995 in Manitoba, Canada Questionnaire in 2002 had 3,598 responses Manitoba's health care registry records	723 children from the group with completed questionnaires 246 cases, 477 controls	Asthma	Database definitions in health care records	Paediatric allergist diagnosis of asthma
Pacheco et al,(282) 2009 United States	NUgene Project Genome-wide association study	7,970 people with DNA samples, of which 521 had an asthma diagnosis	Asthma diagnosis	**Initial asthma cases algorithm:** Asthma diagnosis and asthma medication prescription on ≥1 visit AND no other chronic lung disease diagnosis on ≥2 visits AND no reported smoking history ≥10 years **Final asthma cases algorithm:** Asthma diagnosis on ≥1 visit AND asthma diagnosis or medication prescription on ≥1 other visit AND no other chronic lung disease diagnosis on ≥2 visits AND no reported smoking history ≥10 years **Initial asthma controls algorithm:** No diagnosis for any respiratory disease or cancer AND no prescription of any astha/COPD/iimmunodepressant medication AND no reported smoking history ≥10 years **Final asthma controls algorithm:** ≥2 visits with any asthma diagnosis or prescriptions AND no diagnosis for any respiratory disease or listed cancer AND no prescription of any asthma/COPD/immunodepressant medication AND no reported smoking history ≥10 years	Manual review of 100 cases for both algorithms

Author, year, country	Data source, population	Sample/case characteristics	Clinical event	Algorithm	Validation
Vollmer et al,(283) 2004 United States July 1998 to January 1999	KPNW, Epic, OSCAR, TOPS ED, secondary care	235,000 patients with continuous health plan eligibility aged 15–55 in January 1999 9,723 asthma patients identified	ICD-9 codes	**Health care utilization profiles used for validation study** 1. Four "high-probable" categories: → Two or more non-urgent care outpatient contacts for asthma → A single non-urgent contact and one or more ED or inpatient contact for asthma → Any Industrial Medicine visit for asthma → Any asthma visit and either of the two medication dispensing criteria 2. Single non-urgent outpatient visit only 3. Four or more β-agonists, with or without a nebulizer treatment order, but no asthma visits of any kind and no ICS dispensing 4. ED or urgent care visit for asthma and nebulizer treatment order, but no other medication criteria met and no other types of asthma visits 5. Hospitalization for asthma, but neither asthma medication criterion met and no outpatient asthma visits of any kind 6. ED or urgent care visit for asthma, but no other types of asthma visits and no asthma medication criteria met 7. Nebulizer treatment but no asthma visits of any kind and no other medication criteria met 8. All other cases	**Criteria used in medical records review Probable asthma** • Two or more asthma health care visits • A single visit for asthma with a chart notation indicating a prior history of asthma • A single health care visit for active symptoms of asthma (wheeze, cough, shortness of breath) • A single visit for an asthma exacerbation that responds to therapy, even if no prior history **Possible asthma** • Patient-reported history of asthma noted in chart, but no evidence of active asthma or treatment for asthma • An uncorroborated ED diagnosis of asthma • Diagnosis of "rule out asthma" with no clear resolution
Donahue et al,(284) 1997 United States	Harvard Pilgrim Health Care (HPHC); Primary, secondary and emergency care	Random sample of 100 patients	Asthma code	Asthma diagnosis and asthma drug dispensing	Manual review by clinicians
Premaratne et al,(285) 1997	Accident and emergency departmentss	All asthma patients January–March 1994	String containing "asth*"	String containing "asth*" in the free text records	**Affirmation of asthma diagnosis:** Final diagnosis of

Author, year, country	Data source, population	Sample/case characteristics	Clinical event	Algorithm	Validation
United Kingdom 1994	of two hospitals	1,185 records, of which 209 did not have enough data			asthma by clinical officer OR symptoms of asthma and (history of asthma or bronchodilators given, with improvement) OR known asthmatic presented with symptoms or for medication **Rejection of asthma diagnosis:** Clear alternative diagnosis Sufficient other information to reject asthma diagnosis
Comparison with an in dependent database					
Engeland et al,(286) 2009 Norway	MBRN: population-based birth registry, all births in Norway since 1967 (more than 2.3 million) NorPD: all dispensed prescriptions from January 2004 in Norway	108,489 pregnancies, of which 4,549 mothers were recorded as having asthma in MBRN	Asthma	Asthma diagnosis in MBRN	NorPD: asthma medication
Coulter et al,(287) 1989 United Kingdom	7 general practices in the Oxford community health project 2,199 patients on medication Primary care	2,443 on digital register Bronchodilators, inhaled CS, prophylactic drugs	Asthma diagnosis	Asthma diagnosis on register	Manual review against the list of patients on long-term medication
Comparison with a questionnaire					
Ward et al,(288) 2004 United Kingdom 1995–2004	GP Practice with 14,830 patients 83 1 controls, 587	833 asthma patients, 659 responses 16–55 years on 1 October 1997	Asthma in GP database	One of the following criteria: 1. Read coded “asthma” diagnosis, H33 2. Attendances recorded on the asthma care screen 3. An intervention for asthma	Questionnaire to determine bronchial hyperreactivity Cases: asthma in database Asthma diagnosis and bronchial

Author, year, country	Data source, population	Sample/case characteristics	Clinical event	Algorithm	Validation
	responses Primary care			recorded 4. A textual entry “asthma” or “wheez” in the medical history 5. Inhaled steroids in the repeat prescriptions 6. Inhaled bronchodilators in the repeat prescriptions 7. Cromolyns in the repeat prescriptions	hyperreactivity: considered positive Asthma diagnosis without bronchial hyperreactivity: further investigated in GP record Controls: bronchial hyperreactivity but no asthma diagnosis

Table 1: Characteristics of studies with validated asthma algorithms

Table 2: Characteristics of studies with validated asthma algorithms

Author, year, country	Algorithm	Sensitivity, 95% CI	Specificity, 95% CI	PPV, 95% CI	NPV, 95% CI	Prev-alence
Manual validation						
Xi et al,(271) 2015 Canada	1. Asthma in disease registry	7% (5–10)	99% (97–100)	67% (38–87)	73% (72–74)	8.1%
	2. Billing code	77% (75–83)	89% (86–92)	74% (67–80)	91% (88–94)	
	3. Asthma in CPP	63% (59–68)	92% (90–95)	76% (68–83)	87% (83–89)	
	4. Asthma medications	79% (75–83)	64% (59–68)	46% (41–50)	88% (84–92)	
	5. Asthma in chart notes	85% (81–88)	76% (72–80)	58% (52–63)	93% (89–95)	
	6. Asthma in CPP OR billing code 493	90% (87–93)	84% (80–88)	69% (63–74)	96% (93–97)	
	7. Asthma in CPP OR billing code 493 (exclusion codes 491, 492, and 496)	87% (83–90)	85% (82–89)	70% (63–76)	94% (91–96)	
	8. (Asthma in chart notes OR asthma medications) AND billing code 493	78% (74–82)	92% (89–95)	79% (72–85)	91% (88–94)	
	9. (Billing code 493 OR medications) AND asthma in chart note	84% (80–88)	84% (80–88)	67% (61–73)	93% (90–95)	
	10. Billing diagnostic code 493 AND asthma in chart notes	74% (70–78)	93% (91–96)	81% (73–87)	90% (87–93)	
Engelkes et al,(278) 2014 Netherlands	Definite, probable and doubtful cases			63%		
Afzal et al,(267) 2013 Netherlands	Definite asthma	98%	95%	66%		6%
	Definite + probable	96%	90%	82%		29%
	Definite, probable and doubtful cases	95%	67%	57%		32%
Dexheimer et al,(279) 2013 United States	Algorithm constructed using a Bayesian network system			64%		7–10%
Wu et al,(280) 2013/	ICD-9 codes	31	93	57	82	4–17%
	Natural language processing: logic	81	95	84	94	

Author, year, country	Algorithm	Sensitivity, 95% CI	Specificity, 95% CI	PPV, 95% CI	NPV, 95% CI	Prev-alence
2014 United States	Natural language processing: machine learning	85	97	88	95	
Kozyrskyj et al,(281) 2009 Canada	At least one asthma hospitalization, or two physician visits, or four prescription medications	47% (35–60)	92% (78–98)	91% (76–98)		11%
	At least one asthma hospitalization, or two physician visits, or two prescription medications	67% (54–78)	92% (78–98)	94% (82–99)		
	At least one asthma hospitalization, or one physician visit, or two prescription medications	77% (65–87)	92% (78–98)	94% (85–99)		
	At least one asthma hospitalization, or one physician visit, or two bronchodilators, or one controller medication	80% (69–89)	89% (74–97)	93% (83–98)		
	At least one asthma hospitalization, or one physician visit, or two bronchodilators, or one bronchodilator and ketotifen or an oral steroid, or one controller medication	80% (69–89)	89% (74–97)	93% (83–98)		
	At least one asthma hospitalization, or one physician visit, or one bronchodilator, or one controller medication	82% (70–90)	83% (67–94)	90% (79–96)		
Pacheco et al, (282) 2009 United States	Initial algorithm	70% (60–78)	100%	100% (90–100)	77% (65–86)	7.2%
	Final algorithm	95% (84–99)	96% (87–99)	95% (84–99)	96% (87–99)	
Vollmer et al,(283) 2004 United States	Algorithm 1: population of 4460			95%		4.1%
	Algorithm 2: population of 2334			90%		
	Algorithm 3: population of 545			70%		
	Algorithm 4: population of 25			100%		
	Algorithm 5: population of 11			50%		
	Algorithm 6: population of 721			80%		
	Algorithm 7: population of 99			27%		
	Algorithm 8: population of 1528			80%		

Author, year, country	Algorithm	Sensitivity, 95% CI	Specificity, 95% CI	PPV, 95% CI	NPV, 95% CI	Prev-alence
Donahue et al,(284) 1997 United States	Asthma code and drug dispensing			86%		3%
Premaratne et al,(285) 1997 United Kingdom	String containing asth* in free text records	80% (75–86)	96% (96–99)	91% (87–94)	94% (93–95)	20.6%
Comparison with an in dependent database						
Engeland et al,(286) 2009 Norway	Asthma in MBRN and NorPD	51% (49–52)	98% (98–98)	46% (45–48)		4.20%
Coulter et al,(287) 1989 United Kingdom	Percentage of people on long term medication and recorded on the register			58%		
Comparison with a questionnaire						
Ward et al,(288) 2004 United Kingdom	Total of all reviewed patients			89%		5.60%
	Cases without bronchial hyperreactivity			73%		
	Controls with bronchial hyperreactivity			78%		

Table 2: Test measures of studies with validated asthma algorithms

Manual validation

We found ten studies that used a manual validation as the reference standard. All studies had at least one case definition algorithm with a PPV of at least 63%. Where other measurements could be calculated, the studies had at least one case definition with a sensitivity of at least 85%, specificity of at least 92% and NPV of at least 94%. Within this group, four studies used case definition algorithms generated by machine learning. Five studies included only children, while two studies included only persons older than 16 years.

Xi and colleagues tested a variety of EHR search algorithms based on two large academic primary care clinics in Hamilton, Canada.(271) The reference standard

consisted of a physician chart review–based diagnosis. The eight case definitions are presented in table 1, and their PPVs in table 2. The algorithm searching for patients who had asthma in their patient profile or had an asthma billing code was the most accurate with a sensitivity of 90% (95% CI (87% to 93%)) and a specificity of 84% (95% CI (80% to 88%)).

Engelkes and colleagues undertook a study to determine the validity of case definitions generated by machine learning to define asthma cases, based on a previous study be Afzal et al.(267,278) Originating from a large Dutch general practitioner database, the authors manually reviewed 22,699 potential asthma cases. Among those, 14,303 asthma cases were found, which resulted in a PPV of 63%.

The study by Afzal et al uses the same dataset and machine-learning algorithm for definite and potential asthma cases as the study by Engelkes.(267,278) Clinicians manually validated 5,032 potential asthma cases identified by a broad search algorithm out of 63,618 patients. This training set was used for the machine-learning algorithm. The test measures are measuring the validity of the machine learning algorithm within the smaller population, not of the broad search algorithm. The PPV, sensitivity and specificity for three case definition algorithms (definite cases; definite and probable cases; definite, probable and doubtful cases) were calculated. The PPV's range from 57% for all definite, probable and doubtful asthma cases to 82% for only the definite asthma cases.

Dexheimer and colleagues evaluated a computerized asthma detection system in an urban, tertiary care paediatric emergency department in a 3-month prospective, randomized controlled trial in 2009.(279) A Bayesian network system screened all emergency department patients for acute asthma. The system identified 1,100 patients with asthma exacerbations, of which 704 were confirmed by a paediatric emergency

care physician within 3 days of the visit. The PPV for the Bayesian network system was 65%.

Wu et al evaluated the accuracy of a computational approach to asthma ascertainment. The authors developed a natural language processing (NLP) system for extracting predetermined asthma from free text in EHR.(280) Manual chart review by a clinician was the reference standard. The patient group consisted of 112 children younger than 4 years. The NLP-generated case definition algorithms had a sensitivity of 85%, specificity of 97%, PPV of 88%, a NPV of 95%. For comparison, the test measures of the ICD-9 asthma codes were calculated (sensitivity 31%, specificity 93%, PPV 57%, NPV 82%).

Kozyrskyj and colleagues described the Study of Asthma, Genes and the Environment (SAGE). The study captures the longitudinal healthcare records of 16,320 children born in 1995 in Manitoba (Canada) and contains detailed information on early-life exposures in relationship to the development of asthma.(281) Within the birth cohort, a nested case-control study with 723 children was partly created to confirm asthma status in children and these data were used to validate healthcare database measures of asthma. These 723 children were chosen by random sampling from the birth cohort; the parents of 288 children with and 435 without a parental report of asthma in the last 12 months agreed to participate. The reference standard for the validation consisted of paediatric allergist-diagnosed asthma, methacholine challenge tests and skin tests. The PPV of asthma definitions varied from 90% to 94%, the sensitivity from 47% to 82% and the specificity from 83% to 92%.

Pacheco and colleagues constructed case definitions to identify asthmatic patients as cases, and healthy patients as controls using data from electronic medical records in the United States. This was done to identify asthma patients for future Genome-Wide Association Studies (GWAS). The case definitions consisted of a combination of

diagnoses, medications, and smoking history.(282) By applying stringent criteria, the study results show a PPV of 95% and a NPV of 96% for identification of asthma cases and controls, using clinician review as the reference standard. Genome-wide association studies require a high specificity, PPV and NPV. A high specificity was achieved but at the loss of 24% of the potential asthma cases.

Vollmer et al used the electronic databases of a large health maintenance organisation to develop a case definition for defining prevalent asthma and to validate it against chart review.(283) The data systems of this organisation, the Kaiser Permanente Northwest Division (KPNW) consist of both EHR (inpatient data, emergency department data, EpicCare) and administrative data: "Outside claims database" (OSCAR) and "The outpatient pharmacy system" (TOPS). Table 2 presents the PPV of the eight different case definition algorithms to define asthma. The fourth case definition based on a combination of an urgent care visit and the order of nebuliser treatment (N=25) had the highest PPV (100%), while the first case definition, based on non-urgent care visits, (N=4460) had a high PPV of 95%.

Donahue and colleagues sought to determine the reliability of identifying asthmatics through automated medical and pharmacy records. All adult members of the Harvard Pilgrim Health Care (HPHC) program who received an asthma diagnosis and at least one asthma drug between April 1988 and 1991 were identified.(284) The authors manually reviewed records of a random sample of 100 patients to validate the asthma diagnosis. The PPV of a coded asthma diagnosis was 86%.

Premaratne and colleagues measured the validity of the string 'asth' in the accident and emergency department (A&E) attendance diagnosis field for identifying patients with asthma-related conditions attending the A&E departments of two hospitals in the UK in 1995.(285) A reception clerk entered the diagnosis field in a database at arrival in the A&E department. The reference standard was a confirmation of the

asthma diagnosis by a clinical officer, or symptoms of asthma plus a history of asthma or bronchodilators given with improvement, or a previously diagnosed asthmatic with symptoms or prescribed asthma medication. An 'attendance diagnosis' of asthma was excluded if there was a clear alternative diagnosis or sufficient other evidence to exclude asthma. The string 'asth' in the attendance diagnosis field had a sensitivity of 80% (75-86%) and a specificity of 97% (96-98%) for a confirmation of asthma.

Linked databases

Our search found 2 studies which used a second independent database to validate asthma diagnoses in the first database. The PPV's ranged from 46% to 58%.

Coulter et al (287) compared repeat prescriptions for asthma, epilepsy and thyroid disease with chronic disease registers stored on general practice computers in the early days of EHR (1989). PPV of an asthma diagnosis on the register was 58% for asthma when using medication prescriptions as the reference standard.

Engeland et al evaluated the reliability of maternal disease registration (diabetes, asthma and epilepsy) in the Medical Birth Registry of Norway (MBRN).(286) The data they examined consisted of the EHR of 108,489 pregnancies between April 2004 and January 2007. The reference standard was the prescriptions in the Norwegian Prescription Database (NorPD). The overall sensitivity of an asthma diagnosis in MBRN was 51% (49-52), but increasing when considering with a higher asthma treatment step in NorPD. The sensitivity was 40% when considering records which only used inhaled selective beta-2-adrenoreceptor agonists (step1), while the sensitivity of asthma diagnosis in records with systemic drugs other than adrenergics for obstructive airway diseases was 73%.

Questionnaires

There was only one study which used a questionnaire as the reference standard for database validation.

Ward and colleagues aimed to determine the degree of under- or over reporting of the diagnosis of asthma for patients aged 16–55 years in one large general practice in the UK.(288) The case definition described in table 1, (based on either codes, text strings or prescriptions) yielded 833 potential asthma cases and 831 age- and sex-matched controls from the GP database. A questionnaire validated for the detection of bronchial hyper-reactivity was sent to all asthma patients and their matched controls. Patients with a diagnosis of asthma and bronchial hyper-reactivity in the questionnaire were considered to have asthma. Evidence of asthma was sought for two groups: patients with asthma and without symptoms of bronchial hyper-reactivity, and controls with symptoms of bronchial hyper-reactivity. The results show an overall PPV of the case definition of 89%.

Table 3: Quality assessment using QUADAS-2

	Patient selection	Index test	Reference standard	Flow and timings
Xi et al, 2015	☺	?	☺	?
Engelkes et al, 2014	☺	☺	☹	☺
Afzal et al, 2013	☹	☺	☺	☺
Dexheimer et al, 2013	☺	☺	☺	☺
Wu et al, 2013,2014	☹	☺	?	☺
Kozyrskyj et al, 2009	☹	☹	☺	☺
Pacheco et al, 2009	☹	☺	☺	☺
Vollmer et al, 2004	☹	☺	☺	☺
Donahue et al, 1997	☺	☹	☹	☺
Premaratne et al, 1997	☺	☹	☺	☺
Engeland et al, 2009	☹	☹	☹	☹
Coulter et al, 1989	☹	☹	☹	?
Ward et al, 2004	☹	☹	☺	☹

Table 3: Quality assessment using QUADAS-2

Note: Happy face: low risk; sad face: high risk; question mark: unclear risk.

DISCUSSION

The main finding of this review is that case definitions and methods of asthma diagnosis validation vary widely across different EHR databases. This is evident in the diversity of databases used by the studies, such as primary care databases, combined EHR and administrative databases, or data from nested case-control studies within larger cohorts. Some databases originate from a single or a few health centres, while others span millions of patients. The source of the EHR databases (primary care, secondary care and urgent care) influences the case definition of asthma and the way the validation is conducted. Patients seeking care for asthma symptoms will present differently in each setting, and the test measures might reflect this.

Case definitions are designed with different purposes in mind, and each of the studied test measures (sensitivity, specificity, PPV and NPV) have different uses. A high sensitivity is needed to identify all asthma patients from a database, but if the aim is to exclude all records which do not have asthma, a high specificity is more important.[32] The PPV reflects the percentage of the records identified with a case definition actually have asthma, while the NPV shows the percentage of records who do not fit the case definition do not have asthma. PPVs and NPVs are directly related to the prevalence of asthma in the population. The PPV will increase with rising prevalence; the NPV will decrease with rising prevalence assuming all other factors remain constant.

Studies whose main aim was not database validation were able to demonstrate a high-test measure to suit their specific needs (PPV, NPV, sensitivity or specificity greater than 80%). If this was not the case, their main study results (not including validation) would not be reliable, and thus potential studies with low validity of asthma diagnosis might not have been conducted or published. In contrast, studies whose main aim was the validation of asthma in databases have a wider range of test measures depending on the case definition. The PPV in these studies range from 46% (286) to 96%.(280)

Manual validation was the most common reference standard in the validation studies included in this systematic review. The computer-generated case definitions studied recently by Engelkes, Afzal, Dexheimer and Wu et al provide ways to create algorithms with high sensitivities and specificities. The PPV's of these methods (whether a person identified as having an asthma diagnosis actually has asthma) might not be sufficient for all purposes (63%-82%). Preselected case definitions were used in five out of ten studies which manually validated the databases. The studies by Xi, Kozyrskyj, Pacheco, Vollmer, Donahue and Premaratne used this approach and all report at least one case definition algorithm with a PPV above 85%. The best results arise when combining diagnostic data and prescription data.

Other studies by Engeland and Coulter used an external data source as reference standard. This approach needs two databases with near complete data, so their test measures are reliable on the quality and completeness of the two databases. It also requires that the validity of the reference standard is already known. However, they are much cheaper to carry out overall. Manual validation requires a considerable amount of time to complete, and questionnaires to hundreds of patients or clinicians can be expensive or unreliable. Coulter et al measured database completeness and integrity by studying different diseases including asthma. Their focus was not on asthma validation, but rather to check whether a digital database can be a valid alternative for analogue registration.

Typical problems of validation studies are the lack of availability of a reliable reference standard and the interdependence of different data sources used for validation. There were four studies, not included in this review, which used face validity to compare the prevalence of asthma using a case definition to the general asthma prevalence. This method was not considered sufficiently exact for inclusion (289–292) and by definition was unable to verify the validity of individual records.

The diagnosis of asthma can represent different conditions in different regions of the world. Thus, several authors used an inclusive strategy and many diagnosis codes in order to maximize sensitivity. Researchers must weigh the benefits of a case-finding algorithm with high sensitivity against the likely lower specificity and PPV, according to the purpose of their research. In future studies using predetermined case definitions, it may be of interest to evaluate the predictive value of a specific set of codes validated by chest physicians or general practitioners working in the health system the database originates from. This group may be more accurate when assigning the diagnosis, and the codes applied may yield a much higher predictive value than when evaluating the same group of codes assigned by all providers. The PPV, NPV, sensitivity and specificity can differ greatly within a single study, as shown in the studies by Xi, Afzal, Kozyrskyj and Vollmer et al. For this reason, the testing of multiple case definitions to obtain the algorithm with the highest test measure needed would be beneficial for future studies.

Conclusion

Asthma validation studies using EHRs are very varied in their approach to the validation. This seems driven by the nature of the data and the reference standards used. The method of sampling records using machine learning in algorithm development allow for measuring all elements of validity. Different case definitions within a single data source have different validity highlighting the importance of testing a range of case definitions.

Dissemination and ethics

This study is a synthesis of previously published studies, so no ethical approval is required. We have registered the protocol in the PROSPERO database with registration number CRD42016041798, and the protocol has been published.(276)

Results from this systematic review can be used to study outcome research on asthma and can be used to identify case definitions for asthma.

Appendices

Search algorithms in the Embase and Medline databases

QUADAS-2

3.4 Appendix

Algorithms used for literature review

MEDLINE

1 (validat* or verif*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
2 (positive predictive value or negative predictive value or likelihood ratio or receiver operating characteristic or kappa).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
3 Validation Studies/ or validation.mp. or Validation Studies as Topic/
4 (electronic* or digital* or computeri?ed or programmed or automated or database or data base).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
5 asthma.mp. or Asthma/ or Asthma, Occupational/ or Asthma, Exercise-Induced/
6 Database Management Systems/
7 1 or 2 or 3
8 4 or 6
9 5 and 7 and 8

EMBASE

1 (validat* or verif*).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
2 validation.mp. or validation study/ or validation process/
3 (sensitivity or specificity or "Sensitivity and Specificity").mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
4 (positive predictive value or negative predictive value or likelihood ratio or receiver operating characteristic or kappa).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
5 (electronic* or digital* or computeri?ed or programmed or automated or database or data base).mp. [mp=title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword]
6 mild persistent asthma/ or nocturnal asthma/ or experimental asthma/ or moderate persistent asthma/ or severe persistent asthma/ or Asthma.mp. or exercise induced asthma/ or occupational asthma/ or intrinsic asthma/ or asthma/ or allergic asthma/ or extrinsic asthma/ or mild intermittent asthma/
7 1 or 2 or 3 or 4
8 5 and 6 and 7

QUADAS-2

QUADAS-2 tool designed to allow for more transparent rating of bias and applicability of primary diagnostic accuracy studies and can be found on the QUADAS website (www.quadas.org).

This tool composes 4 domains: patient selection, index test, reference standard, and flow and timing and is available Each of those four domains are assessed on the risk of bias, and the first three domains are assessed on applicability. Each domain of the QUADAS-2 is applied in four phases.

DOMAIN	***PATIENT SELECTION***	***INDEX TEST***	***REFERENCE STANDARD***	***FLOW AND TIMING***
Description				
Signalling questions(yes/no/unclear)				
Risk of bias: High/low/unclear				
Concerns regarding applicability: High/low/unclear				

Table 4: Appendix QUADAS-2, adapted from www.quadas.org

Research paper cover sheet

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4648
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Francis Nissen
Principal Supervisor	Ian Douglas
Thesis Title	Asthma in electronic health records: validation & phenotypes.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Clin Epidemiol. 2017 Dec 1;9:643-656.		
When was the work published?	2017 Dec 1		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have developed the research question, reviewed the literature with one co-author and summarised the identified papers and drafted the manuscript.

Student Signature: _____ **Date:** 13/08/2018

Supervisor Signature: _____ **Date:** 14/08/2018

Chapter 4: Validation of asthma recording in the Clinical Practice Research Datalink (CPRD)

Summary

- We have estimated the PPV of several different algorithms based on clinical codes for asthma, reversibility testing and asthma medication prescriptions in the CPRD GOLD based on GP questionnaires.
- Diagnoses were confirmed in a high proportion of patients with specific asthma codes, suggesting that epidemiological asthma research conducted using CPRD GOLD data can be conducted with reasonably high validity.
- The PPVs for the algorithms based on specific asthma Read codes and non-specific asthma Read codes in combination with additional evidence were all greater than 0.84.
- A specific asthma code algorithm alone appears to be the most practical approach to identify patients with asthma in CPRD GOLD (PPV=0.86; 95% CI 0.77-0.95).
- The algorithm using non-specific asthma codes, information on reversibility testing, and respiratory medication use scored highest (PPV=90.7%, 95% CI 82.8% to 98.7%), but had a much lower total identifiable population.
- The inclusion of reversibility testing or asthma medications in the algorithm did not clearly improve accuracy.
- In conclusion, people with asthma can be accurately identified from UK primary care records using specific Read codes.

4.1 Preface

The research paper presented in this chapter is a validation study of algorithms to identify people with asthma in the CPRD GOLD. The aim of this study was to test the accuracy of different approaches to identifying asthma in the CPRD GOLD using the positive predictive value (PPV), by comparing the database records with a gold standard constructed from a review by two study physicians based on information provided by asthma patients' general practitioners.

The algorithms consisted of a combination of clinical codes for probable or definite asthma, codes indicating reversibility testing had taken place and codes for asthma medication prescriptions. The validity of these algorithms was tested by randomly selecting patients who qualified for an algorithm and sending extensive questionnaires to the General Practitioners of those patients. These questionnaires were then examined by one chest physician and one study GP to assess how many patients truly had asthma in order to assess the validity of the pre-specified algorithms against a reference standard. Validity of each algorithm was expressed using PPVs.

One reason asthma is difficult to assess in health-care database epidemiological studies is because the diagnostic criteria are based on non-specific respiratory symptoms and variable expiratory airflow limitation. These symptoms and airflow limitation measures are often not recorded in electronic medical records. The other reasons are the overlap with other diseases and the absence of a universal case definition. As the clinical examination necessary for the diagnosis of asthma is time and resource demanding, it can be useful for epidemiological studies to rely on EHR data to obtain accurate records of asthma diagnosis to determine asthma status. As epidemiological research is extremely reliant on data accuracy and misclassification of study variables compromises the validity of study results, validation of algorithms is imperative for valid inference.(293) When using EHR, the usual epidemiological

challenges related to validity of study findings remain in place and may even be amplified.(294,295) Results from validation studies allows researchers to estimate the extent of misclassification and can help coding clinicians to remain motivated to use systematic coding schemes.(296) Validity for algorithms to identify patients with other conditions including COPD were proven high in the CPRD GOLD.(229,297)

There was no validated definition of asthma diagnosis in the CPRD GOLD before the publication of this study. In terms of this thesis, the primary motivation for validating the recording of asthma was to provide a definition to be used in the next studies. The results of this validation study were used to inform patient selection for the studies included in chapter 5 and 6 of this thesis.

The questionnaire described in Chapter 2 of this thesis was used to construct a reference standard for asthma validation. This questionnaire is included in the appendix of this chapter.

This article was originally published in the BMJ Open, and is available here:

Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). BMJ Open. 017;7(8). https://www.ncbi.nlm.nih.gov/pubmed/28801439

4.2 Research paper

Validation of asthma recording in the Clinical Practice Research Datalink (CPRD)

Authors: Francis Nissen,[1] Daniel R. Morales,[2] Hana Mullerova,[3] Liam Smeeth,[1] Ian J Douglas,[1] Jennifer K Quint[4]

1. *Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK*
2. *Division of Population Health Sciences, University of Dundee, Dundee, UK*
3. *RWE & Epidemiology, GSK R&D, Uxbridge, UK*
4. *National Heart and Lung Institute, Imperial College, London, UK*

ABSTRACT

Objectives: The optimal method of identifying people with asthma from electronic health records in primary care is not known. The aim of this study is to determine the Positive Predictive Value (PPV) of different algorithms using clinical codes and prescription data to identify people with asthma in the United Kingdom Clinical Practice Research Datalink (CPRD).

Methods: 684 participants registered with a GP practice contributing to CPRD between 1st of December 2013 and 30th of November 2015 were selected according to 1 of 8 pre-defined potential asthma identification algorithms. A questionnaire was sent to the general practitioners to confirm asthma status and provide additional information to support an asthma diagnosis. Two study physicians independently reviewed and adjudicated the questionnaires and additional information to form a

gold standard for asthma diagnosis. The Positive Predictive Value was calculated for each algorithm.

Results: 684 questionnaires were sent, of which 494 (72%) were returned and 475 (69%) were complete and analysed. All 5 algorithms including a specific Read code indicating asthma or non-specific Read code accompanied by additional conditions performed well. The PPV for asthma diagnosis using only a specific asthma code was 86.4% (95% CI 77.4% to 95.4%). Extra information on asthma medication prescription (PPV 83.3%), evidence of reversibility testing (PPV 86.0%) or a combination of all three selection criteria (PPV 86.4%) did not result in a higher PPV. The algorithm using non-specific asthma codes, information on reversibility testing, and respiratory medication use scored highest (PPV 90.7%, 95% CI [82.8% to 98.7%]), but had a much lower identifiable population. Algorithms based on asthma symptom codes had low PPVs (43.1% to 57.8%).

Conclusions: People with asthma can be accurately identified from UK primary care records using specific Read codes. The inclusion of spirometry or asthma medications in the algorithm did not clearly improve accuracy.

Article summary

Strengths:

This study describes algorithms to identify people with asthma from CPRD, a large electronic health records database, and measures the positive predictive value of those algorithms.

Supporting information, including outpatient referral letters, other emergency department discharge letters, airflow measurements and radiography records were used to identify asthma patients and calculate the test measures.

Limitations:

The gold standard to calculate a PPV (GP questionnaire and review by study physicians) is not absolute, even though information from secondary care was used.

GPs of patients with complicated medical histories could be less likely to return the questionnaire, but remuneration makes this less likely.

BACKGROUND

Asthma is one of the most common chronic diseases, with an estimated prevalence of 241 million people worldwide with asthma.(298) The United Kingdom has one of the highest asthma prevalence and mortality rates in Europe.(165,299) The disease is a significant burden to the National Health Service, with 5.4 million people receiving treatment and approximately 65,000 hospital admissions yearly.(300) Cough, wheeze, breathlessness and chest tightness are its core symptoms (301) but it has a wide variety of different presentations.(101)

Electronic health records (EHR) have been adopted worldwide, facilitating the construction of large population-based patient databases that have become available over the last decades for epidemiological research.(260) Validation of diagnoses or outcomes based upon codes recorded in EHRs is required because their accuracy is uncertain, and this may affect the reliability and validity of subsequent observational studies. The quality of studies generated from EHRs may be debatable unless their data are validated for specific research purposes.(261,295,302,303)

The diagnosis of asthma relies on clinical judgement based on a combination of patient history, physical examination and confirmation of the variability or reversibility of airflow obstruction using airflow measurements. This can make it difficult to assess the accuracy of asthma diagnoses in EHR-based epidemiological studies as some symptoms and airflow measurements may not be recorded. In addition, individuals affected by asthma can vary greatly in their presentation and symptoms are sometimes similar to other respiratory diseases such as COPD (Chronic Obstructive Pulmonary Disease).(304,305)

The aim of this study was to test the accuracy of different approaches to identifying asthma in the United Kingdom Clinical Practice Research Datalink (CPRD) using the positive predictive value (PPV), by comparing the database records with a gold standard constructed from a review by 2 study physicians based on information provided by asthma patients' GPs.

METHODS

Dataset

The Clinical Practice Research Datalink (CPRD) is a large UK primary care database containing anonymised data on the people registered with primary care practices from across the UK. CPRD is representative of the UK population with regard to age and sex.(208,306) Within CPRD, diagnostic accuracy has been demonstrated to be high

for many conditions and diseases, including COPD.(229,297,307,308) CPRD contains detailed clinical information on diagnoses, prescriptions, laboratory tests, symptoms and hospital referrals, in addition to basic sociodemographic information recorded by the general practitioners. These general practitioners (GPs) act as primary care providers and gatekeepers for other National Health Service services, and information from other healthcare providers is also transmitted back to the GP. Clinical events and diagnoses are coded as Read codes, a dictionary of clinical terms widely used in the UK National Health Services by both primary and secondary healthcare providers. Validation studies aid to ensure credibility and quality of epidemiological studies done in CPRD.(309)

The random sample of individuals included in the study was constructed from all participants registered in CPRD on or after first of April 2004 who met the inclusion criteria (see below). For the main analysis, a patient was able to contribute to one algorithm only if an asthma medcode was recorded within the 24-month window prior to the end of data collection. It was possible an individual was eligible for more than one algorithm depending on the Read codes used in their medical record. The individuals were randomly selected from the algorithm with the fewest participants first and then removed from the cohort so that they could not be selected for another algorithm. We have chosen this strategy (as opposed to an individual being eligible for a single algorithm only) because we wanted to test strategies to identify asthma patients from a single cohort rather than to test validity of the diagnosis. Further studies could then use a single strategy or their combination to extract an asthma cohort. There was no special measures to ensure less frequent Read codes are used, because we assumed the validity of asthma diagnosis strategy would be not be different between common and less frequent Read codes and the quality of recording would also be comparable. In addition, less frequent Read codes are unlikely to be used in isolation; our experience with validation of COPD recordings had shown that

these infrequent Read codes usually accompanied more commonly used Read codes for the same condition.

Inclusion criteria

The study population consisted of people who had a record for a Read code indicating possible asthma in the two years before the index date (1st of December 2015) and who were registered in a GP practice meeting CPRD quality criteria. The Read code list is included in appendix. The data collection was planned before the index test and reference standard were performed. This timespan was chosen for several reasons: to overcome potential changes in quality of asthma diagnosis and recording over time; to reduce the chance that the database records were out of date; and to ensure the medical records were still available to GPs. People were identified at random based on one of eight pre-defined algorithms exclusively, which means that we populated the algorithm resulting in the smallest population first and subsequently removed these people from the cohort, to prevent them from also being selected for another algorithm. We randomly selected 800 possible asthma cases for validation. Of these, 116 asthma cases were excluded because their GP no longer participated with CPRD at the time questionnaires were sent to the clinicians for validation, as shown in figure 1. Due to changes in CPRD data governance after the start of the study it was not possible to select replacement patients.

- Acceptable user status registered in CPRD.

- Practice was "up to standard" at study start 1/4/2004. From this date onwards, the Quality and Outcomes Framework (QOF) came in effect.

- The patient fulfilled one of the asthma algorithms within the last 24 months

- Patients were still alive and practice was currently active in the CPRD.

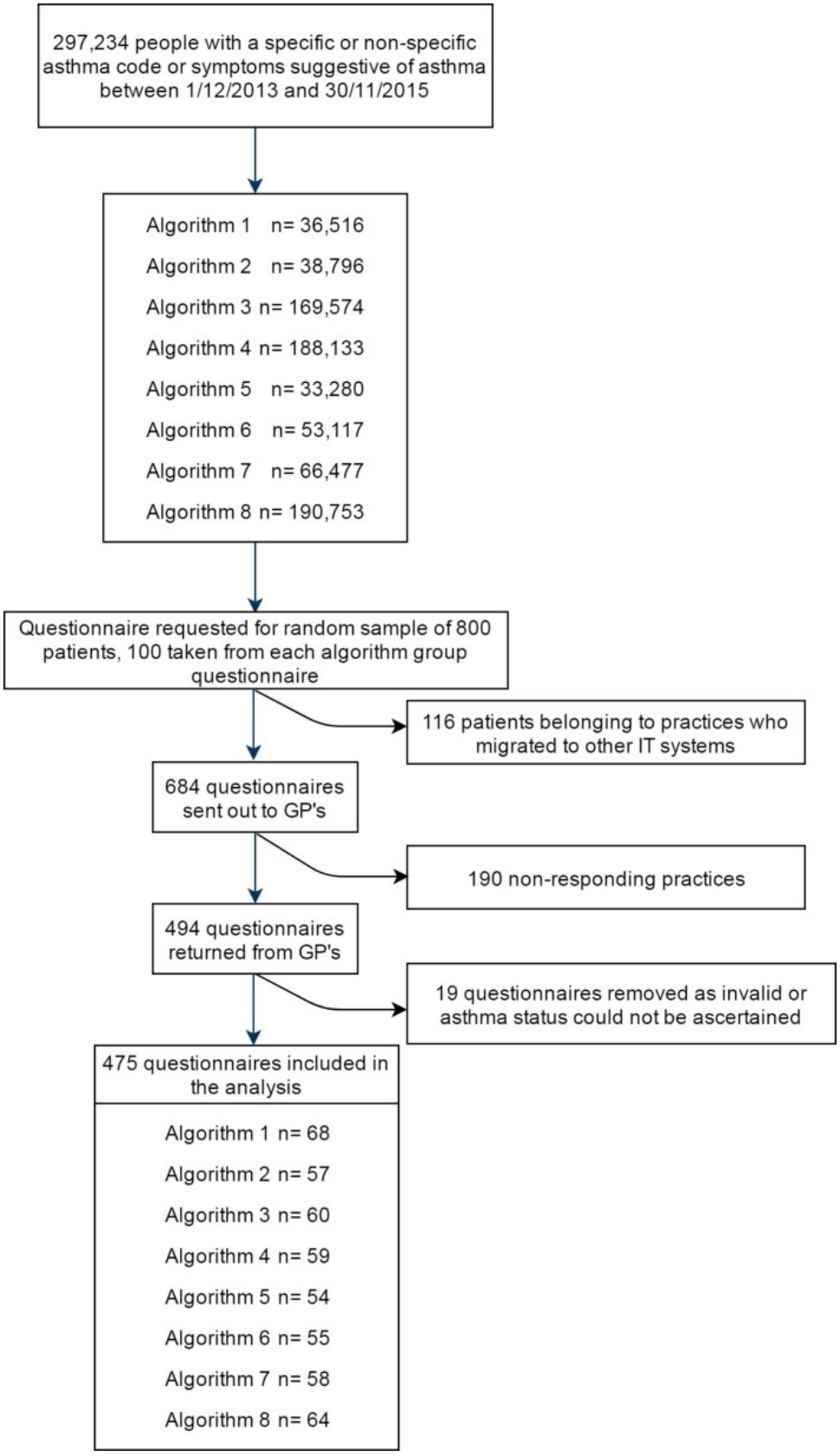


Figure 1: Study population.

GP questionnaire

CPRD mailed a two-page questionnaire to the GPs of the people selected for inclusion as described above, requesting confirmation of current asthma diagnosis and additional information to support this diagnosis. This questionnaire can be found in the appendix. The questionnaire was designed to ascertain the diagnosis of asthma and verify the date of diagnosis. The questions included evidence of reversible airway obstruction, current symptoms, smoking history, respiratory comorbidities and Quality Outcome Framework (QOF) indicators. QOF is a national financial incentive scheme for GPs in the UK encouraging regular disease indicator measurement and recording. Asthma is one of the included diseases, and its indicators including airflow measurements and interference with work and night's rest.(310)

Specific information available from the medical record including spirometry printouts and hospital respiratory outpatient letters were also requested. Data were encrypted twice to ensure anonymity, between practices and CPRD and also from CPRD to researchers. A questionnaire was considered invalid if it was returned blank or every question was answered "unknown".

Code lists and algorithms

Lists of medical codes (Read codes) deemed as specific and non-specific for asthma based on study physicians' opinion were created prior to the start of the study. Read codes are a hierarchical clinical coding system that are used in general practice in the UK and are entered by the GP into a computer programme called Vision. Each Read code is linked to a specific string of text, which refers to a single diagnosis or symptom. These data are then uploaded by CPRD after they have been processed and quality checked. The list of codes used for specific or definite asthma codes and nonspecific or probable asthma codes can be found in the appendix.

Combinations of Read code lists, evidence of reversibility testing and respiratory medication use were used to make up the eight algorithms. A number of different algorithms were constructed with degrees of certainty of asthma using separate indicators. For example, the most stringent algorithm included an asthma code, asthma medication and demonstrated reversibility after trial of treatment. The first four algorithms required a specific asthma diagnosis code, with the first three requiring additional documentation consisting of either respiratory medication use and/or evidence of reversibility testing. The fifth algorithm required a non-specific asthma code and additional documentation of both respiratory medications and reversibility testing; the last three algorithms required respiratory symptom codes indicating asthma symptoms with additional information. The presence of spirometry for inclusion in an algorithm was based on the existence of a specific spirometry Read code in the records rather than an examination of said spirometry, although where spirometry traces were provided as part of the additional information, they were examined. Evidence of reversibility testing only refers to whether airflow measurements or trial of treatment were done and does not reflect the results of these tests. Respiratory medication use was defined as at least two prescriptions of asthma medication for inhaled asthma therapy (Short Acting Beta-Agonists, Long Acting Beta-Agonists and Inhaled Corticosteroids) within 365 days of each other, within the two years before the index date. From the expected most specific to most sensitive, the eight algorithms were constructed as follows:

1. Specific asthma Read code + evidence of reversibility testing (spirometry, variable Peak Expiratory Flow Rate or trial of treatment) + respiratory medications
2. Specific asthma code + evidence of reversibility testing
3. Specific asthma code + respiratory medications
4. Specific asthma code only
5. Non-specific asthma code + evidence of reversibility testing + respiratory medications
6. Asthma Symptoms (wheeze, breathlessness, chest tightness, cough) + evidence of reversibility testing + respiratory medications
7. Asthma Symptoms + evidence of reversibility testing
8. Asthma Symptoms + respiratory medications

Sample size calculations

The reasoning behind the sample size of the validation study was as follows: Assuming an estimated PPV of 0.85 for each algorithm and an accuracy of the PPV (95% CI ± 0.08), a sample size of 77 individuals for each algorithm was needed. A similar study conducted for COPD had a 77.6% response rate and 73.2% of the sent questionnaires were fit to be included in the final analysis.(311) Considering a random sample of fully completed responses of 77 asthma patients for 8 algorithms is needed with 20% extra to account for a potential lower response rate, 800 questionnaires in total were sent. We also assumed that the probability of data being missing was independent of accuracy of the asthma diagnosis. However, we anticipated little missing relevant data in this study based on past research. In addition, the covariates were needed for stratification analysis only, rather than for adjustment.

Missing data

The plan for addressing missing data for the validation study was as follows: We planned to do a complete case analysis, assuming that the probability of data being missing is independent of accuracy of the asthma diagnosis, conditional on covariates. If the amount of missing data was small, any violation of the assumption is unlikely to importantly affect the results.

Algorithm	1. Specific asthma code + reversibility testing + medication	2. Specific asthma code + reversibility testing	3. Specific asthma code + medication	4. Specific asthma code	5. Non-specific asthma code + reversibility testing + medication	6. Symptoms + reversibility testing + medication	7. Symptoms + reversibility testing	8. Symptoms + medication	Total
Individuals, n (%)	68 (100)	57 (100)	60 (100)	59 (100)	54 (100)	55 (100)	58 (100)	64 (100)	475
Asthma diagnosis by own GP	56 (82.4)	49 (86)	48 (80)	51 (86.4)	48 (88.9)	29 (52.7)	23 (39.7)	38 (59.4)	342
Confirmation by respiratory physician before study start	55 (80.9)	29 (50.9)	38 (63.3)	45 (76.3)	34 (63)	23 (41.8)	25 (43.1)	36 (56.3)	285
Evidence of reversible airway obstruction	47 (69.1)	37 (64.9)	32 (53.3)	32 (54.2)	31 (57.4)	26 (47.3)	19 (32.8)	26 (40.6)	250
Mean age	52.3	51.4	47	41.9	45	60.9	61.3	52.1	
Mean age (95% CI)	(47.4-57.2)	(46.2-56.7)	(41.4-52.6)	(36.1-47.6)	(38.7-51.3)	(55.3-66.4)	(57.1-65.5)	(45.4-58.7)	
<18 years old **(%)**	7.35	7.02	15.25	18.64	16.67	7.27	1.72	20.31	11.81
Sex: male	31 (45.6)	17 (29.8)	16 (26.7)	23 (39)	26 (48.1)	28 (50.9)	24 (41.4)	31 (48.4)	196
Current smoker*	11 (16.2)	10 (17.5)	10 (16.7)	5 (8.5)	4 (7.4)	5 (9.1)	8 (13.8)	4 (6.3)	57
Ex-smoker*	16 (23.5)	14 (24.6)	17 (28.3)	16 (27.1)	15 (27.8)	11 (20)	10 (17.2)	12 (18.8)	111
Never smoker*	35 (51.5)	26 (45.6)	25 (41.7)	36 (61.0)	32 (59.3)	18 (32.7)	11 (19.0)	27 (42.2)	210
Individuals with supporting info	23 (33.8)	21 (36.8)	22 (36.7)	14 (23.7)	14 (25.9)	17 (30.9)	14 (24.1)	22 (34.4)	147

*Table 1: Characteristics of the 475 patients included in the final study analysis*As stated by patient's GP on the study questionnaire.*

Primary outcome

The primary outcome was confirmation of a diagnosis of asthma in each of the eight predefined algorithms. The gold standard for the diagnosis of asthma was the adjudicated asthma status agreed by the two study physicians, a respiratory physician and a GP who reviewed all questionnaires and evidence from the patient's GP independently. The reviewers were blinded to the code lists/algorithm. Where opinion differed, the cases were discussed, and agreement was reached by consensus. The reviewing physicians did not know with which algorithm a person was selected.

Statistical analysis

The Positive Predictive Value (PPV) was calculated using the proportion of cases identified by each algorithm that were confirmed as actual cases by the study physicians through a review of the questionnaire and supporting evidence. All analyses were conducted using Stata 14.0. The gold standard consisted of the opinion of 2 medical experts independently reviewing the questionnaires and any additional supporting medical information provided. If there was a disagreement of diagnosis, the case would be discussed by the two experts. If an agreement was not found, a third opinion was sought.

A patient could contribute only to a single algorithm for the main analysis. In the post hoc analysis, individuals could be placed into multiple algorithms where possible to reduce the confidence intervals. The PPV in this analysis was calculated for all individuals who had a specific asthma code compared with those with a specific asthma code and additional information. We also performed a sensitivity analysis to check whether the age and sex for patients whose questionnaire was returned was similar to the age and sex of those patients whose questionnaire was not sent out or were there was no response.

Sample size calculation

As there were 116 patients that could not be evaluated, precision was expected to be slightly lower than in the original sample size calculations. However, a percentage difference in PPV of 0.13 is demonstrable with a sample size of 60 per algorithm (assuming PPV=0.85, alpha=0.05 and power=0.8).

RESULTS

A total of 800 potential asthma cases were selected for validation, of which 116 cases had migrated out of the database at the time the questionnaires were sent. Of the remaining 684 cases, there were 494 returned questionnaires. Nineteen of the returned questionnaires were considered invalid. Thus, 475 valid questionnaires were received, which yielded a response rate of 69.4% (475/684) using the practices that could have answered as denominator, as shown in figure 1. The time interval between the mailing of questionnaires and the review by the study physicians varied, but none of these time intervals was greater than 8 months.

The baseline characteristics of the 475 patients with valid returned questionnaires are shown in table 1. The study populations were mostly middle aged, never smokers and female. There were 97 individuals whose smoking status was not filled in on the questionnaire. Differences in the majority of characteristics were seen among most algorithms.

The positive predictive values of the eight algorithms are displayed in table 2. The PPVs of algorithms containing specific or non-specific asthma codes in algorithms 1-5 (ranging from 83.3% to 90.7%) are markedly higher than the PPVs of the algorithms based on asthma symptoms (ranging from 43.1% to 57.8%). The combination of a specific code and asthma medication prescription and/or evidence of reversibility testing (PPV varies from 83.3% to 86.8%) did not considerably increase the PPV

compared with a specific asthma code alone (PPV 86.4%). The highest PPV was found in the fifth algorithm combining a non-specific asthma code with evidence of reversibility testing and asthma medication use. However, the total number of patients identifiable with this algorithm (n=33,280) was less than one fifth of those identifiable by the fourth algorithm consisting of a specific asthma code alone (n=188,133) in the chosen time period. We have not examined the validity of a non-specific asthma code alone.

A post hoc analysis was performed where individuals were placed in every algorithm they qualified for. In this analysis, we found that the use of additional information on evidence of reversibility testing or medication in an algorithm with a specific asthma code again did not meaningfully increase the PPV. The PPV for all individuals who had a specific asthma code and information on reversibility testing or medication was 86.7% (95% CI 83.3% to 90.1%), and the PPV for individuals with only a specific asthma code was 86.4% (95% CI 83.0% to 89.7%).

When validating the record of possible asthma with a gold standard based on the study physicians' view of extra evidence provided by the GP, the PPV slightly improved across all algorithms. Figure 2 demonstrates the PPV of the different algorithms as diagnosed by the patient's own GP and the study physicians (overall κ=0.81).

There was no considerable difference in age or sex between patients whose questionnaire was returned and patients whose questionnaire was not sent out (age: p=0.74, sex: p=0.73) or were there was no response (age p=0.50, sex p=0.13) using χ^2 tests.

Table 2: The positive predictive value (PPV) and proportion of patients diagnosed with chronic obstructive pulmonary disease within each algorithm

Algorithm	Eligible population	Questionnaires sent out	Valid returned questionnaires (n, %)	Confirmed asthma cases	PPV (95% CI)
Specific asthma code + reversibility testing + medication	36 516	92	68 (60)	61	86.8 (78.5 to 95.0)
Specific asthma code + reversibility testing	38 796	90	57 (63.3)	51	86.0 (76.7 to 95.3)
Specific asthma code + medication	169 574	89	60 (67.4)	51	83.3 (73.6 to 93.0)
Specific asthma code	188 133	84	59 (70.2)	51	86.4 (77.4 to 95.4)
Non-specific asthma code + reversibility testing + medication	33 280	78	54 (69.2)	49	90.7 (82.8 to 98.7)
Symptoms + reversibility testing + medication	53 117	87	55 (63.2)	32	56.4 (42.8 to 69.9)
Symptoms + reversibility testing	66 477	88	58 (65.9)	26	43.1 (30.0 to 56.2)
Symptoms + medication	190 753	78	64 (82.1)	38	57.8 (45.4 to 70.2)

Table 2: The positive predictive value (PPV) and proportion of patients diagnosed with asthma within each algorithm

Medication use was defined as two prescriptions within 365 days. Evidence of reversibility testing does not hold information on the outcome of these tests.

Figure 2: PPV as diagnosed by the patient's own GP, and agreement between the study physicians. GP, general practitioner; PPV, positive predictive value

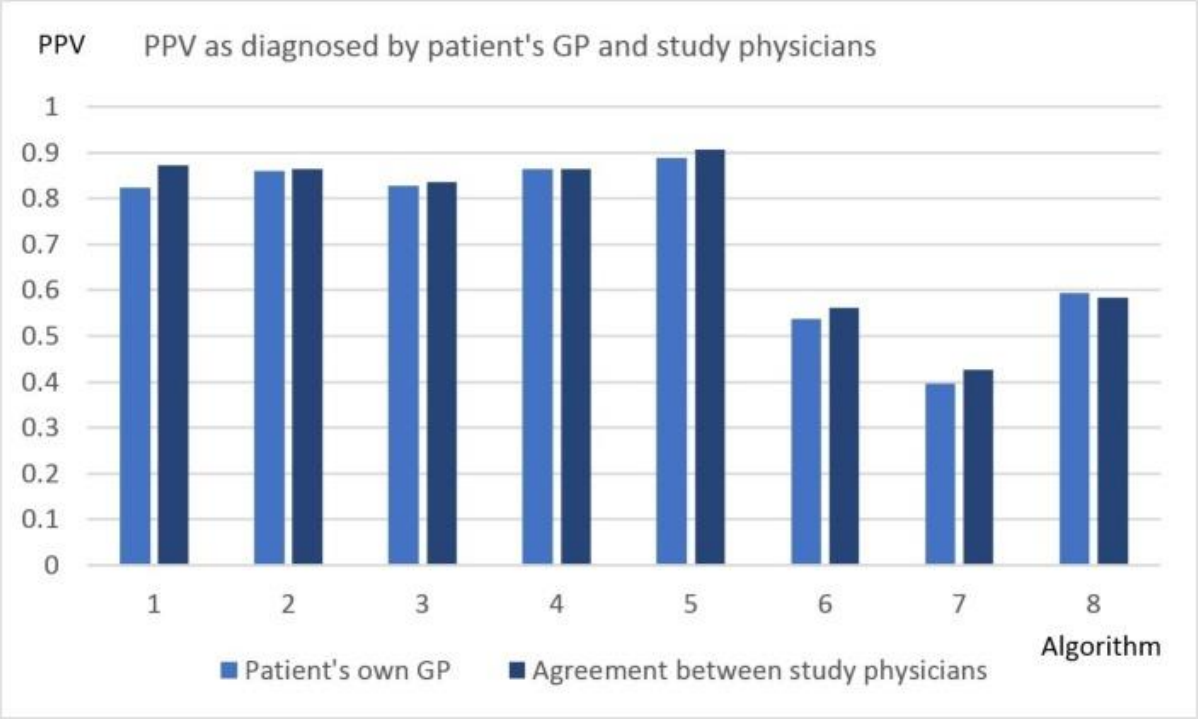


Figure 2: PPV as diagnosed by the patient's own GP, and agreement between the study physicians. GP, general practitioner; PPV, positive predictive value

DISCUSSION

We tested the accuracy of eight algorithms to identify asthma within CPRD using a gold standard constructed using a consensus of the two study physicians. The algorithm with the highest PPV consisted of a combination for nonspecific asthma codes, evidence of reversibility testing and multiple asthma prescriptions within one year (PPV 90.7, 95% CI 82.8 to 0.98.7) followed by a combination for specific asthma codes, evidence of reversibility testing and multiple asthma prescriptions within one year. The confidence interval of this PPV overlaps with the confidence intervals of each of the PPVs of the first four algorithms based on specific asthma codes, so the difference might be due to chance alone. The algorithm with the lowest PPV consisted of asthma symptoms and evidence of reversibility testing (PPV 0.43, 95% CI 0.30-0.55). The results of this validation study suggest that the clinical code-based algorithms that use asthma codes to identify asthma cases have high PPVs (between 0.84 and 0.91). In

this dataset, a specific asthma code algorithm alone appears sufficient to identify current asthma patients from CPRD. As the additional requirements of medication use and evidence of reversibility testing do not appear to significantly increase the PPV, the total number of individuals who can potentially be included in a study increases from 33,280 to 188,133 in the chosen time period (1st of December 2013 to 30th of November 2015). The total identifiable population of people living with asthma is thus much larger when only using a specific asthma code for identification.

Comparison with previous studies

Validity of asthma codes in electronic health records can be assessed by comparison with three different sets of gold standard: comparison with an external database, questionnaire and manual review by a clinician. This validation study uses questionnaires and manual review. Our gold standard consisted of the agreement of the study respiratory physician and study GP, both of whom were experienced with CPRD.

Previous studies which validated asthma in other EHR databases used manual review by clinicians to validate asthma in EHR and all reported at least one algorithm with a PPV above 85%.(271,281–285) In contrast with this study, the best results in previous studies arose when combining diagnostic data and prescription data.

The CPRD has provided anonymised primary care records for public health research since 1987; research was always a focus of interest when it was established. GPs contributing to the CPRD have been trained on how to record data for research use. As a consequence, data quality may be higher than in many other databases, in which research is only a secondary product.

Strengths of this study

This study has several strengths. First, we were able to investigate the accuracy of eight pre-defined different algorithms and how they perform in identification of people with asthma in CPRD, as well as the accuracy of the actual GP diagnosis of asthma using additional information provided. Second, we included supporting information such as outpatient referral letters, other emergency department discharge letters, airflow measurements and radiography records. Finally, we validated asthma diagnoses found in CPRD, which is a primary care database that is extensively used for studying different health outcomes in epidemiological research. This primary care database provides health and medication history of millions of patients. A validated definition in CPRD of asthma allows for informed health-care service planning by increasing the reliability of evidence generated from observational studies.

Limitations of this study

This study has limitations to consider. The gold standard consisting of a GP questionnaire and review by study physicians is not absolute, even if we mitigated this with additional information from secondary care. A GP can look in the electronic health record to see if a specific diagnosis has been recorded for a specific patient when asked. This may lead to an overestimation of the PPV, but there is no suitable practical alternative. Ideally, airflow measurements and reversibility testing on each potential patient would form the optimal gold standard, but this would not be feasible in this setting due to cost. The overall number of questionnaires sent out (n=684) was less than requested (n=800) as some patients and practices were no longer part of CPRD and could not be contacted. However, the precision of PPV estimates was not substantially reduced.

Although practices contributing to CPRD are a sample of all practices in the UK, they are considered representative of the UK population with few patients opting out of contributing data, and is therefore unlikely to bias the results.(208)

GPs of patients with complicated medical histories could be less likely to return the questionnaire. The GPs were remunerated for their participation however, which is likely to have reduced the chance of this happening. Within the returned questionnaires, the amount of missing data was low, which suggests reasonable data quality. In addition, only living patients were assessed, as GPs no longer have access to the patient records after death. This excludes the records of the deceased patients and could result in survival bias. Patients had to be alive to be included, but it is unlikely that coding would differ between living and deceased individuals. If deceased people had died of asthma, the PPV in this study would be underestimated. Our findings are likely to be generalizable to other UK primary care databases using Read coding, but these would ideally still require validation. Databases using other coding systems may need to validate different algorithms to identify asthma, which might limit the generalisability of our findings. Another limitation is that we were not able to assess the Negative Predictive Value (NPV) of asthma diagnoses in CPRD because we evaluated only patients belonging to one of the eight algorithms. We could not calculate the specificity or sensitivity as we had preselected our population of possible asthma cases. We also assumed the validity of asthma diagnoses would not be different between common and less frequent Read codes and the quality of recording would also be comparable for pragmatic reasons. However, the less commonly used codes will by definition identify a smaller proportion of all asthma patients, so the validity we report will apply to the majority of patients. In future practice when identifying patients with asthma, the less commonly used codes will continue to identify a smaller proportion of all asthma patients and so the validity we measure will apply to the majority of patients. Using a GP questionnaire as the source of patient information in order to obtain a gold standard to validate the asthma diagnosis has its limitations as the GP can consult the electronic health record to see if there was an asthma diagnosis. This could lead to an overestimation of the PPV. Incomplete diagnostic information could lead to missing data which we were unaware of which could lead to some inaccuracy in PPV or classification of asthma

probability. Response rate for the questionnaire might have been lower than expected, and the sample size of the completed questionnaires could have been too small. Not all GP practices contribute to CPRD GOLD, and patients might refuse to participate in the CPRD programme. This could have resulted in selection bias.

CONCLUSION

We have successfully estimated the PPV of several different algorithms to identify people with asthma in CPRD. The PPVs for specific asthma Read codes alone and non-specific ones in a combination with additional evidence were all greater than 0.84. A specific asthma code algorithm alone appears to be the most practical approach to identify patients with asthma in CPRD (PPV 0.86; 95% CI 0.77-0.95). Diagnoses were confirmed in a high proportion of patients with specific asthma codes, suggesting that epidemiological asthma research conducted using CPRD data can be conducted with reasonably high validity.

Dissemination and ethics

The protocol for this research was approved by the Independent Scientific Advisory Committee (ISAC) for MHRA Database Research (protocol number15_257) and the approved protocol was made available to the journal and reviewers during peer review. Generic ethical approval for observational research using the CPRD with approval from ISAC has been granted by a Health Research Authority (HRA) Research Ethics Committee (East Midlands – Derby, REC reference number 05/MRE04/87).

The results will be submitted for publication and will be disseminated through research conferences and peer reviewed journals.

Figure legend

Figure 1: Study population

Figure 2: PPV as diagnosed by the patient's own GP, and agreement between the study physicians

Appendices

Appendix 1: General Practitioner questionnaire

4.3 Appendix

Questionnaire for asthma validation study

Study into asthma: questionnaire for £55, further information for £55

The London School of Hygiene and Tropical Medicine is conducting a study to investigate the best way to identify asthma within the Clinical Practice Research Datalink (CPRD). We have developed several methods for identifying asthma in the database, and we would like to obtain some information on the current asthma status of the patient from GPs so that we can decide which method is the most suitable.

We would be very grateful if you could supply us with the following information.

A. Do you agree this patient has a current diagnosis of asthma?

- ☐ Yes: Proceed to question B
- ☐ No: Proceed to question C
- ☐ Uncertain: Proceed to question B

If you answered yes or uncertain to question A:

B1. Has the diagnosis been made or confirmed by a respiratory physician?

- ☐ Yes
- ☐ No

B2. Does this patient have evidence of reversible airway obstruction?

- ☐ Yes
- ☐ No

If yes: **Was this based on;**

- ☐ Spirometry reversibility with a bronchodilator
- ☐ Trial of treatment with oral or inhaled corticosteroids and diurnal variation on a peak flow diary

B3. In what year was the asthma first diagnosed?

B4. Were any other factors taken into consideration in making the diagnosis?

	Yes	No
a. History of atopic disorder	☐	☐
b. Family history of asthma and/or atopic disorder	☐	☐
c. Widespread wheeze heard on auscultation of the chest	☐	☐
d. Otherwise unexplained low FEV (Forced Expiratory Volume) or PEF (Peak Expiratory Flow) on spirometry	☐	☐
e. Otherwise unexplained variability in PEFR (Peak Expiratory Flow Rate) on spirometry	☐	☐

f. Otherwise unexplained peripheral blood eosinophilia	☐	☐
g. FeNO (Fractional exhaled Nitric Oxide) measurement	☐	☐
h. Other (please name)		

B5. Based on the QOF (Quality and Outcomes Framework) indicators:

	Yes	No
a. Does the patient have any difficulty sleeping because of asthma symptoms, including cough	☐	☐
b. Does the patient have the usual asthma symptoms during the day (cough, wheeze, chest tightness of breathlessness)?	☐	☐
c. Does the asthma interfere with the patient's usual activities (housework, work, school, etc.)?	☐	☐

B6. What is the patient's smoking status?

☐ Current smoker
☐ Ex-smoker
☐ Never-smoker

B7. Does the patient have any other respiratory diseases? (Multiple responses possible)

☐ Chronic Obstructive Pulmonary Disease (COPD)
☐ Bronchiectasis
☐ Interstitial Lung Disease
☐ Other, please list:
☐ No

If you answered no to question A:

C. Do you think this patient has a history of asthma?

☐ Yes
☐ No
☐ Uncertain

Please provide anonymised copies of any additional relevant information allowing corroborating asthma diagnosis e.g. medical notes, discharge letters, test values. Payment for further information is £55 per patient.

Please return responses to CPRD in the freepost envelope provided or to our freepost address:
Freepost RSKH-TTAU-UKKX, CPRD, MHRA,
151 Buckingham Palace Rd, London, SW1W 9SZ

Research paper cover sheet

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Francis Nissen
Principal Supervisor	Ian Douglas
Thesis Title	Asthma in electronic health records: validation & phenotypes.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	BMJ Open. 2017 Aug 11;7(8)		
When was the work published?	2017 Aug 11		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I developed the research question, summarised and analysed the questionnaires and drafted the manuscript.

Student Signature: _____ **Date:** 13/08/2018

Supervisor Signature: _____ **Date:** 14/08/2018

Chapter 5: Quantifying concomitant diagnosis of asthma and COPD in UK primary care

Summary

- This study aims to quantify how commonly patients with COPD have a concomitant diagnosis of asthma, and how commonly patients with asthma have a concomitant diagnosis of COPD in UK primary care.
- 400 COPD patients and 351 asthma patients were identified from the Clinical Practice Research Datalink (CPRD) GOLD in separate validation studies and the diseases were confirmed by the review of GP questionnaires by two study physicians.
- We examined the prevalence of concurrent asthma and COPD based on CPRD GOLD coding, GP questionnaires and requested additional information. We also aimed to determine the extent of possible misdiagnosis and missed opportunities for diagnosis.
- A concurrent asthma and COPD diagnosis appears to affect a relative minority of patients with COPD (14.5%, 95% CI 11.2-18.3) or asthma (14.8%, 95% CI 11.3-19.0).
- More than half of the validated COPD patients had ever received an asthma diagnosis Read code, suggesting over diagnosis of asthma in COPD patients commonly occurs, particularly early in the diagnostic process.
- Over diagnosis of COPD in asthma patients and under diagnosis of asthma or COPD in patients with the other disease are less likely.

5.1 Preface

The study included in this chapter quantifies and discusses concomitant diagnosis of asthma and COPD in the CPRD GOLD. In brief, we aimed to quantify the point prevalence of concomitant asthma and COPD in the diagnosed populations of both asthma and COPD patients in the UK using the CPRD GOLD. Validated definitions exist for the identification of both diseases in the CPRD GOLD. In addition, we also examined possible misdiagnosis and missed diagnosis in patients with obstructive lung diseases.

The distinction between the two diseases in electronic health records is not trivial, as they share many symptoms and characteristics. In addition, there was a gap in the current literature on the prevalence of concomitant disease in primary care.

The concomitant diagnosis of asthma and COPD has been grounds for controversy within respiratory medicine research. The existence of both diseases in the same patient has been accepted, but the mechanism of the underlying pathology has been cause for discussion. The Dutch hypothesis suggests that both diseases are manifestations of the same disease process, with asthma preceding COPD. The overlap syndrome is then called "Asthma COPD Overlap Syndrome" (ACOS). The other school of thought, sometimes called the British hypothesis, proposes asthma and COPD are distinct disease entities with different causal mechanisms. Asthma and COPD can coexist independently in the same patient according to this hypothesis.(111)

The group of individuals with a concomitant diagnosis merits attention, as patients with both asthma and COPD have more frequent exacerbations, increased morbidity and mortality, faster lung function decline and a poorer quality of life than patients with only asthma or only COPD.(312,313)

Epidemiological studies on concomitant asthma and COPD have been scarce, as the differential diagnosis of both diseases is difficult (single spirometry measurements cannot clearly distinguish between asthma and COPD).(314,315) In addition, many studies have insisted on a separation of both diseases, excluding asthma patients from COPD studies and vice versa to avoid misclassification and these studies are also based on narrow inclusion criteria.(316,317) . The symptoms of asthma and COPD overlap, and the differential diagnosis is not always trivial to make. Information on reversibility testing, the QOF indicators, smoking status, concurrent respiratory diseases and other sources including consultant and hospital discharge letters, lung function tests and radiography results was requested in the questionnaire.

A review of this information by a respiratory consultant and study GP aimed to identify the actual cases of COPD in confirmed asthma patients. This review was used as the gold standard to calculate the PPV, NPV, sensitivity and specificity of recorded GP diagnoses of COPD in the primary care records of asthma patients.

The availability of the data of two validation studies provided the opportunity to look at the prevalence of COPD in validated asthma patients in the CPRD GOLD, and the prevalence of asthma in validated COPD patients in the CPRD GOLD. The data on the validated asthma patients came from the study included in the previous chapter, and the data on the validated COPD patients came from an earlier validation study of COPD recording in the CPRD by Quint JK et al, in which I did not participate.

The validation studies are available here:

- Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). BMJ Open. 017;7(8). https://www.ncbi.nlm.nih.gov/pubmed/28801439
- Quint JK, Mullerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice

Research Datalink (CPRD-GOLD). BMJ Open. 2014;4(7):e005540. https://www.ncbi.nlm.nih.gov/pubmed/25056980

This paper was accepted for publication in the British Journal of General Practice.

- Francis Nissen, Daniel R.Morales, Hana Mullerova, Liam Smeeth, Ian J Douglas, Jennifer K Quint 'Quantifying concomitant diagnosis of asthma and COPD in UK primary care.' BJGP 2018

5.2 Research paper

Quantifying concomitant diagnosis of asthma and COPD in UK primary care

Authors: Francis Nissen,[1] Daniel R. Morales,[2] Hana Mullerova,[3] Liam Smeeth,[1] Ian J Douglas,[1] Jennifer K Quint[4]

1. *Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK*
2. *Division of Population Health Sciences, University of Dundee, Dundee, UK*
3. *RWE & Epidemiology, GSK R&D, Uxbridge, UK*
4. *National Heart and Lung Institute, Imperial College, London, UK*

ABSTRACT

Background: Asthma and COPD share many characteristics and symptoms, and the differential diagnosis between the two diseases can be difficult in primary care. This study explores potential overlap between both diseases in a primary care environment.

Aim: This study aims to quantify how commonly patients with COPD have a concomitant diagnosis of asthma, and how commonly patients with asthma have a concomitant diagnosis of COPD in UK primary care.

Design and Setting: 400 COPD patients and 351 asthma patients were identified from the Clinical Practice Research Datalink (CPRD) in separate validation studies and the diseases were confirmed by review of GP questionnaires.

Method: The prevalence of concurrent asthma and COPD in validated cases of either disease was examined based on CPRD coding, GP questionnaires and requested additional information. We also aimed to determine the extent of possible misdiagnosis and missed opportunities for diagnosis.

Results: More than half (52.5%) of validated COPD patients had ever received a diagnostic asthma Read code. However, when considering additional evidence to support a diagnosis of asthma, concurrent asthma was only likely in 14.5% (95% CI: 11.2%; 18.3%) of validated COPD patients. Of the validated asthma patients, 15.1% have ever received a diagnostic COPD Read code, although COPD was only likely in 14.8% (95% CI: 11.3%; 19.0%) of validated asthma patients.

Conclusion: A concurrent asthma and COPD diagnosis appears to affect a relative minority of patients with COPD (14.5%) or asthma (14.8%). Asthma diagnosis may be over recorded in people with COPD.

How this fits in

The prevalence of concomitant asthma and COPD is likely to be overestimated in studies using only electronic health records as their symptoms are similar. This study reports on this issue by including only validated asthma and COPD patients from two previous validation studies. A concurrent asthma and COPD diagnosis affects a relative minority of patients in primary care with either asthma (14.8%) or COPD (14.5%). Asthma may be over recorded in people with COPD in electronic health records.

INTRODUCTION

Worldwide, 358 million people are estimated to be affected by asthma (299) and 174 million by COPD (Chronic Obstructive Pulmonary Disease).(41) Both diseases can vary greatly in their presentation and imprecision of diagnosis in both diseases remains a problem. (304,305)

Accurate diagnosis of asthma and COPD is essential, as correct treatment of asthma and COPD can reduce the frequency and severity of exacerbations and improve overall quality of life.(41) In addition, information on chronic respiratory disease can help patients to quit smoking.

The differential diagnosis of COPD and asthma rests on differences in clinical presentation, triggering factors, and on demonstration of reversibility of airflow obstruction. This airflow obstruction is not fully reversible in COPD, whereas it is in asthma. However, the differential diagnosis remains difficult and the existence of an overlap syndrome called ACOS (Asthma-COPD Overlap Syndrome) remains controversial,(318,319) as consensus regarding the clinical definition has not yet been reached. Some guidelines classify asthma cases with a persistent airway obstruction as COPD, and the two diseases are often mutually exclusive in studies to obtain unblended populations of asthma and COPD patients. In addition, the prevalence of a concomitant diagnosis of asthma and COPD varies greatly in different studies.

This study aims to quantify the point prevalence of concomitant asthma and COPD in the diagnosed populations of both asthma and COPD patients in the UK using electronic health record databases where validated definitions exist for the identification of both diseases. In addition, we also examine possible misdiagnosis and missed diagnosis in patients with obstructive lung diseases.

METHODS

Study population and validation studies

The study populations consist of people who were included in earlier validation studies.(297,320) and are summarised in figure 1 and 2. Questionnaires were sent out to the GPs of possible asthma and COPD patients with the intent to validate the recording of asthma and COPD in the CPRD. The full selection criteria of both validation studies can be found in their respective articles.(297,320) Patient data for the asthma recording validation study were collected from 1 December 2013 to 30 November 2015, and patient data for the COPD recording validation study was between 1 January 2004 and 31 December 2012. In the asthma validation study, full data was only available for the patients for whom the GP stated a current asthma diagnosis and only current asthma diagnoses were considered. In the COPD validation study, the population was preselected as current or ex-smokers. The two patients populations included in this study have been thoroughly validated in their respective validation studies using these detailed GP questionnaires and requested supporting information including outpatient referral letters, other emergency department discharge letters, airflow measurements and radiography records. In the validation studies, the Positive Predictive Value was 86.5% (77.5-92.3%) for COPD (297) and 86.4% (77.4%-95.4%) for asthma (320) when only using a single diagnostic code for the respective disease.

In the asthma questionnaire, details were requested on evidence of airway obstruction, current symptoms, smoking history, respiratory comorbidities and Quality Outcome Framework (QOF) indicators (QOF is a national financial incentive scheme for GPs in the UK encouraging regular disease indicator measurement and recording). The COPD questionnaire requested information on COPD diagnosis, smoking history, symptoms, spirometry, confirmation by a respiratory physician and respiratory comorbidities. Additional information available from the medical record including spirometry printouts and letters from respiratory physician or hospitals

were also requested. Data were encrypted twice to ensure anonymity. If a questionnaire was returned blank or every question was answered "unknown", it was considered invalid.

Database

The Clinical Practice Research Datalink (CPRD) GOLD is a large anonymised UK primary care database which is representative of the UK population with regard to age and sex.(306) Diagnostic accuracy has been demonstrated to be high in CPRD GOLD for many conditions,(229) including asthma and COPD. This database contains detailed clinical information on diagnoses, prescriptions, laboratory tests, symptoms and hospital referrals of included individuals, in addition to basic sociodemographic information recorded by the general practitioners. In the original validation studies, lists of medical codes (Read codes) deemed as specific for asthma or COPD were used to select algorithms to identify asthma and COPD patients; these codes have a high validity in their respective validation studies. Read codes are a hierarchical clinical coding system that is used in general practices in the UK; each Read code is linked to a specific string of text, which refers to a single diagnosis or symptom.

Primary outcome and measurements

The primary outcome for this study was the proportion of patients with either asthma or COPD who had the other disease in the validated asthma and COPD populations. The presence of a diagnostic asthma Read code and positive reversibility tests supported an asthma diagnosis in the COPD population. The presence of a diagnostic COPD Read code, smoking history and fixed airflow obstruction supported a current COPD diagnosis in the population with validated asthma. Spirometry measurements with at least one airflow measurement with fev1/fvc ≤ 70% were considered as evidence for an obstructive airflow limitation. The quality of the spirometry procedure undertaken in UK primary care to diagnose COPD is high as determined in a previous validation study.(321)

Possible misdiagnosis and/or lacking diagnosis of asthma in validated COPD patients, and vice versa, were examined using spirometry measurements, results of reversibility tests and smoking history. To study the temporality of recorded diagnostic Read codes in patients with concomitantly recorded asthma and COPD, we reported the proportion of patients where the time lapse between the date of validation of one disease and the last known diagnosis of the other disease was greater than two years. This was done as we had learned from the validation studies that a COPD patient would sometimes receive their first asthma diagnosis in the 2 years leading up to the first COPD diagnosis. An asthma code shortly before a first diagnosis of COPD is likely to be a misdiagnosis of asthma. If the asthma code was given multiple years before the COPD diagnosis, asthma before COPD onset is more probable.

Conversely, if the last COPD code was given more than 2 years before the validation of an asthma diagnosis (and we assume the validated asthma diagnosis is true), the COPD might be misdiagnosed as the code was not repeated afterwards.

Asthma and COPD diagnoses are based on symptoms, signs and spirometry, but there is no clear reference test. A panel consisting of two physicians determined whether asthma or COPD were present in the validated patients using all available information, and according to national and international guidelines. Both physicians were blinded to the patient selection algorithm and adjudicated the asthma and COPD statuses independently. Where opinion differed, the cases were discussed, and agreement was reached by consensus.

Statistical analysis

We calculated the proportion of asthma patients with COPD and vice versa with 95% confidence intervals using exact binomial Clopper-Pearson intervals. Cells with less

than 5 entries were merged for presentation. All analyses were conducted using Stata 14.0 in 2017.

Table 1: Baseline characteristics

Data Source	Asthma validation		Total	COPD validation		Total
	COPD Read code	**No COPD Read code**		**Asthma Read code**	**No Asthma Read code**	
Individuals (%)	52 (15%)	299 (85%)	351	210 (52%)	190 (48%)	400
Mean age: (95% CI)	67 (64-70)	45 (42- 47)	48 (46-50)	73 (71-74)	73 (72-75)	73 (72-74)
Sex: male (%)	22 (42%)	114 (38%)	136 (39%)	99 (47%)	104 (55%)	203 (51%)
(Ex-) smoker (%)	43 (82%)	112 (37%)	200 (57%)	*	*	*

**The COPD population was preselected to only include (ex-) smokers*

Table 1: Baseline characteristics of the validated asthma and validated COPD patients

RESULTS

Background characteristics

The baseline characteristics of the 751 patients with confirmed asthma and COPD diagnoses are shown in table 1. Amongst patients with validated asthma, those with a COPD diagnosis were older than those without (67 and 45 years, respectively). There was no noticeable difference in mean age between validated COPD patients with or without an asthma Read code (73 years in both groups). The validated asthma study population was mostly female (61.2%), while the validated COPD population was more evenly split regarding sex (50.7% male). The table is further split into two age categories. In the validated asthma patients, a concomitant COPD diagnosis is more likely when the patient is over 50 years of age. Only a small percentage of validated COPD patients is under 50 years of age.

Validated asthma patients

We studied 351 patients with a validated asthma diagnosis of which 52 (15%) had a recorded COPD Read code. The details are summarised in figure 1. For 6 of the 52 asthma patients with COPD codes, the COPD codes were more than 2 years prior to asthma validation. For the remaining 46, COPD codes were within 2 years of the asthma validation date. Of the 46 with validated asthma and recent COPD codes, 38 were smokers or ex-smokers and 8 were recorded as never-smokers. Out of 299 asthma patients without COPD codes, 112 were (ex-) smokers, while 187 were recorded as never-smokers.

We assumed concomitant asthma and COPD in validated asthma patients in the following cases: if the validated asthma patients had a recent diagnosis of COPD and were (ex-) smokers; or if they showed obstruction on their spirometry and were (ex-) smokers but lacked a COPD code. As such, concomitant asthma and COPD was likely in 52 patients (14.8%, 95% CI: 11.3%-19.0%): 38 of those 52 patients had a recent COPD

diagnosis (within 2 years of their asthma Read code) and were smokers or ex-smokers; the remaining 14 patients had no COPD Read code but showed obstruction on their spirometry and were smokers or ex-smokers.

We assumed solitary asthma (without COPD) in validated asthma patients in three scenarios: either if they did not have a COPD code nor showed obstruction on lung function tests; or if they had a past COPD code more than two years ago (as the coding should have been repeated); or if they had a recent COPD code but no smoking history. As such, a solitary diagnosis of asthma was likely in 299 patients (85.2%; 95% CI 81.0-88.7): 187 never smokers without a COPD Read code, 98 (ex) smokers without obstruction or a COPD Read code, 8 patients with a recent COPD code but no smoking history, and 6 patients whose COPD Read code was more than 2 years since their last asthma code.

Validated COPD patients

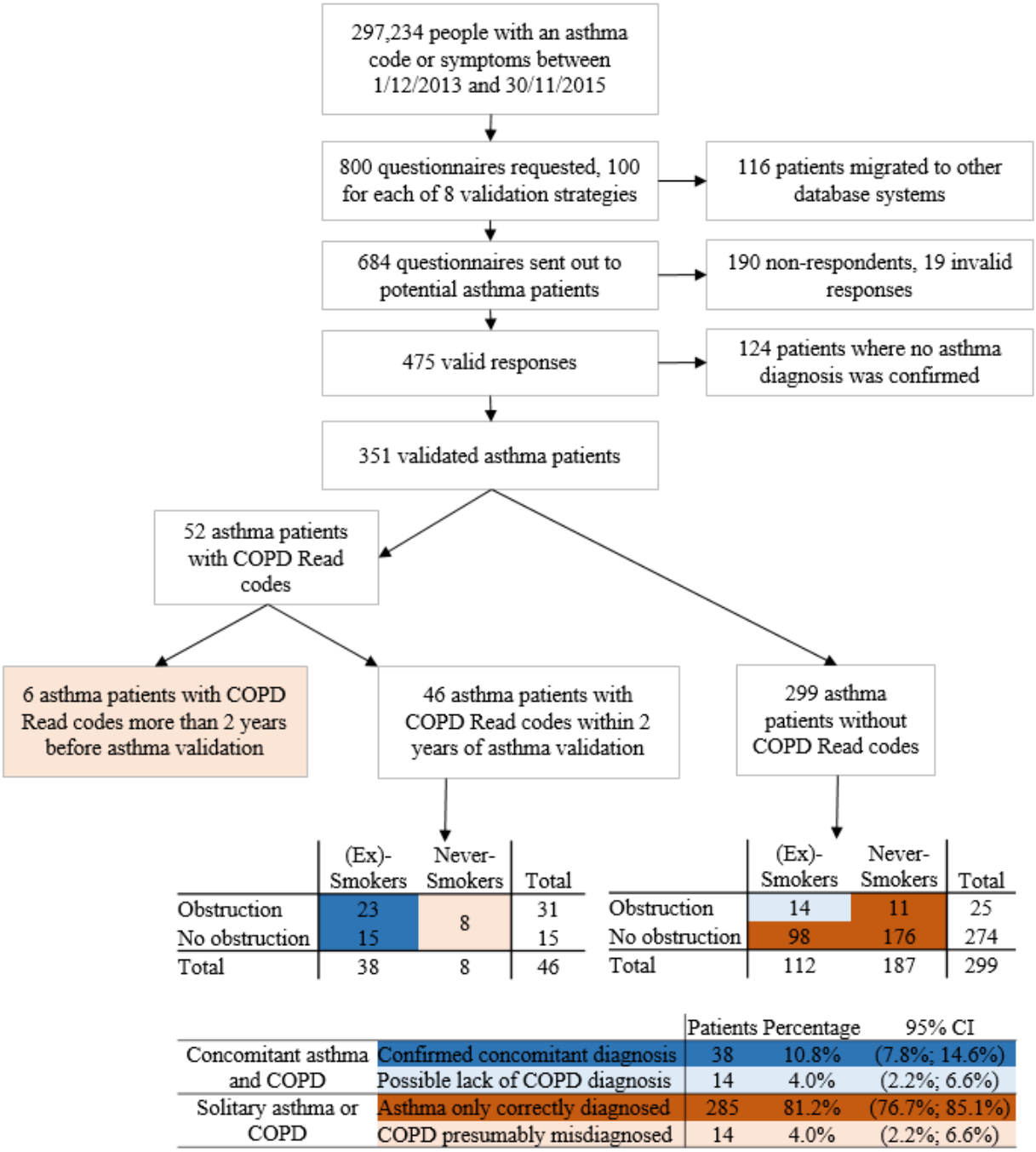
We studied 400 patients with a validated COPD diagnosis, of which 210 (52%) had a recorded asthma Read code. The details are summarised in figure 2. For 82 of the 210 COPD patients with asthma codes, the asthma codes were more than 2 years prior to COPD validation. For the remaining 128, asthma codes were within 2 years of the COPD validation date. Of the 128 with validated COPD and recent asthma codes, 42 had a recording of positive reversibility testing and 86 did not have a recording of positive reversibility testing. Out of 190 COPD patients without asthma codes, 16 had a recording of positive reversibility testing, while 174 did not have lung function tests indicating reversibility of their airflow obstruction.

We assumed concomitant asthma and COPD in validated COPD patients in two scenarios: validated COPD patients with a recently recorded asthma code and a recording of positive reversibility testing; and validated COPD patients without a recent asthma code but with positive reversibility testing recorded. As such,

concomitant asthma and COPD was likely in 58 patients (14.5% (95% CI 11.2%-18.3%)): 42 patients who had a recently recorded asthma diagnosis and positive reversibility testing recorded, in addition to 16 patients without asthma Read codes who had positive reversibility testing recorded.

We assumed solitary COPD (without asthma) in validated COPD patients based on the following criteria: validated COPD patients with no asthma codes nor recording of positive reversibility testing; validated COPD patients where the last asthma code was more than two years before asthma validation (indicating asthma prior to COPD); and validated COPD patients with recent asthma codes but without positive reversibility testing. As such, COPD without clear evidence of current asthma was likely in the remaining 342 patients (85.5% (95% CI: 81.7%-88.8%)): 174 patients with neither asthma Read codes nor positive reversibility testing, 82 patients where the last asthma Read code was more than 2 years before the COPD validation and 86 with recent asthma Read codes (less than 2 years before validation), but without positive reversibility testing.

Figure 1: Validated asthma patients



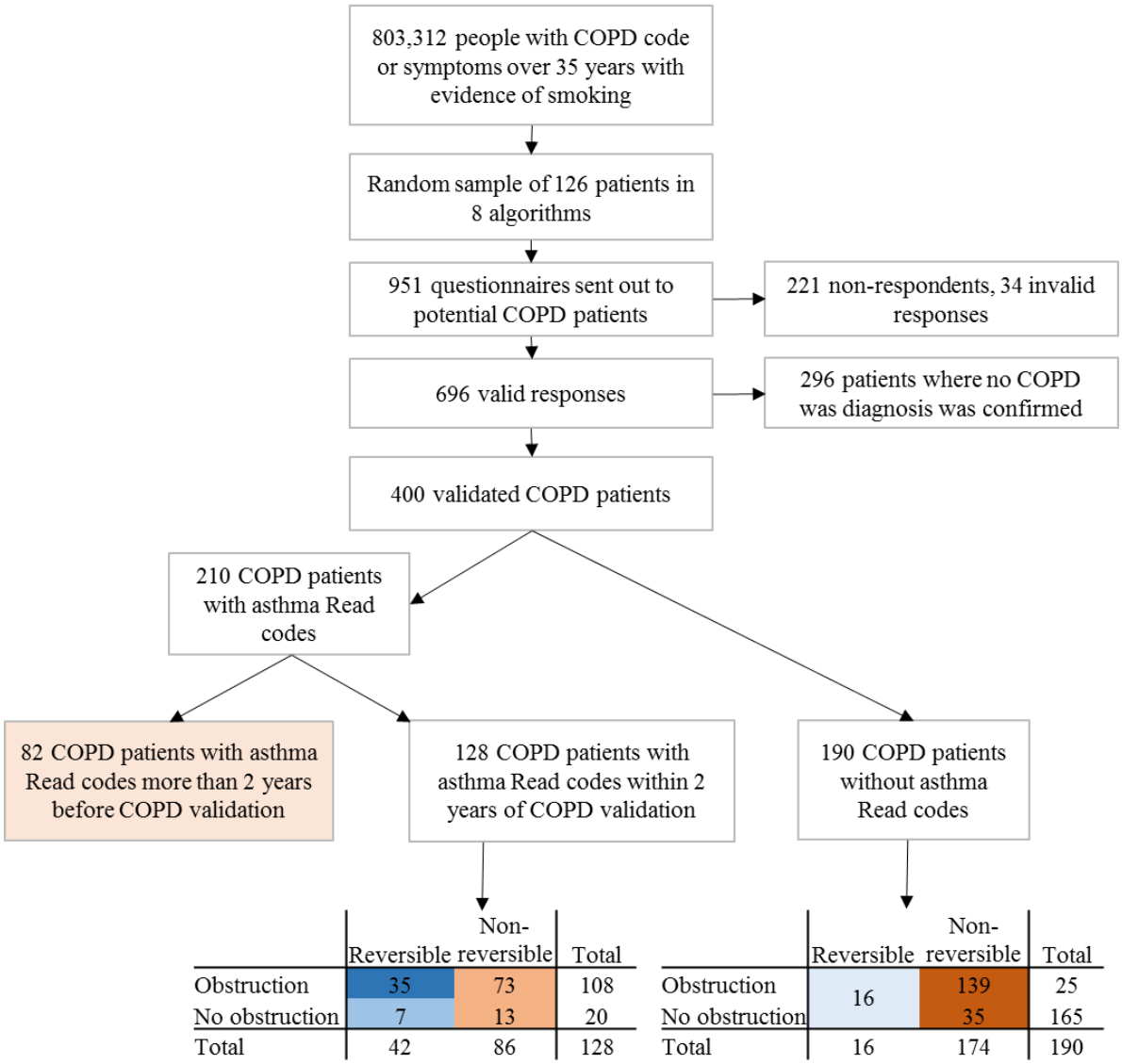
	(Ex)-Smokers	Never-Smokers	Total
Obstruction	23	8	31
No obstruction	15		15
Total	38	8	46

	(Ex)-Smokers	Never-Smokers	Total
Obstruction	14	11	25
No obstruction	98	176	274
Total	112	187	299

		Patients	Percentage	95% CI
Concomitant asthma and COPD	Confirmed concomitant diagnosis	38	10.8%	(7.8%; 14.6%)
	Possible lack of COPD diagnosis	14	4.0%	(2.2%; 6.6%)
Solitary asthma or COPD	Asthma only correctly diagnosed	285	81.2%	(76.7%; 85.1%)
	COPD presumably misdiagnosed	14	4.0%	(2.2%; 6.6%)

Figure 1: Validated asthma patients

Figure 2: Validated COPD patients



	Reversible	Non-reversible	Total
Obstruction	35	73	108
No obstruction	7	13	20
Total	42	86	128

	Reversible	Non-reversible	Total
Obstruction	16	139	25
No obstruction		35	165
Total	16	174	190

		Patients	Percentage	95% CI
Concomitant asthma and COPD	Confirmed concomitant diagnosis	35	8.8%	(6.2%; 12.0%)
	COPD not confirmed	7	1.8%	(0.7%; 3.6%)
	Possible lack of asthma diagnosis	16	4.0%	(2.3%; 6.4%)
Solitary asthma or COPD	COPD only correctly diagnosed	174	43.5%	(38.6%; 48.5%)
	Asthma presumably misdiagnosed	86	21.5%	(17.6%; 25.9%)
	Asthma prior to COPD	82	20.5%	(16.6%; 24.8%)

Figure 2: Validated COPD patients

DISCUSSION

Summary

We were able to investigate the prevalence of COPD in validated asthma patients and vice versa using CPRD GOLD data on smoking, spirometry and reversibility testing; in addition to detailed GP questionnaires and supporting information including outpatient referral letters, other emergency department discharge letters, airflow measurements and radiography records. The main finding of this study is that the 14.8% of validated asthma patients had a concurrent COPD diagnosis, while 14.5% of validated COPD patients had a concurrent asthma diagnosis. Asthma may also be over recorded in CPRD GOLD in COPD patients.

Comparison with existing literature

In primary care, most consultations on respiratory diseases start with a provisional diagnosis made on clinical grounds from the patients' symptoms, in addition to previous specialists' correspondence if available.(322–325) Spirometry is needed to accurately differentiate the diagnosis of asthma and COPD, but is not always used in a primary care setting.(326–329)

The prevalence of asthma in COPD populations is lower compared with many previously published studies, especially those based on electronic health records. However, the prevalence of COPD in asthma populations is similar to a previous cross-sectional study measuring the prevalence fo**r** comorbidities in asthma which reported 13.4% of patients with asthma had a COPD diagnosis compared with 3.4% of the remaining general population. A previous systematic review stated a pooled prevalence of asthma in COPD patients of 27%, with considerable heterogeneity.(330) The definition of asthma and COPD diagnosis tends to differ between studies, which might explain this observation. A study using Norwegian GP data confirmed COPD diagnosis using spirometry in 17.1% of patients with only a previous asthma diagnosis.(331) Among subjects with a spirometry-based study diagnosis of COPD in

GP practices in Scotland and the United States, 51.5% reported a prior diagnosis of asthma without a concurrent chronic bronchitis or emphysema diagnosis.(332) A systematic review on the asthma-COPD overlap syndrome in 2015 found a pooled prevalence of 27% in population-based studies of COPD patients and 28% in hospital-based studies of COPD patients.(330) A recent multicentre study on COPD patients in Japan found an asthma prevalence of 9.2% or 4.2%, depending on the FEV1 cut-off.(333) Other studies report a very wide range of prevalence of concomitant asthma and COPD, as the diagnosis criteria are heterogeneous and a consensus on diagnostic criteria is needed.(305,318)

Strengths

This study has a few strengths. Firstly, we were able to quantify the burden of concomitant diagnosis of asthma and COPD in a large cohort of primary care patients (CPRD GOLD), which is representative in terms of age and sex to the general UK population. Secondly, we used information included in both the CPRD GOLD and in the questionnaires sent out for the original validation studies in order to differentiate between asthma and COPD. Finally, this study adds to the relatively small body of literature on the epidemiology of concomitant asthma and COPD in primary care.

Limitations

This study has potential limitations which need consideration.

First, the results of this study are only applicable to the records in the Clinical Practice Research Datalink, although this database is considered representative of the general UK population.(306)

Second, only asthma and COPD patients for whom their GP responded to verify their diagnosis in the original questionnaire were included in this study. GPs of more complicated cases might be less likely to respond where diagnostic uncertainty may exist. However, this issue is mitigated to an extent as GPs were paid for providing the information for validation, and the baseline characteristics of the individuals for whom a questionnaire was returned were similar to the characteristics for which no questionnaire was returned.(297,320) Data on eligible patients who were not included as there was no returned questionnaire were available from CPRD GOLD but did not contain all the information of a completed questionnaire.

Third, the validation process was mostly based on the GP questionnaires, which are available in the original studies. (297,320) Additional information (discharge letters, spirometry measurements and radiography) were available for 31.5% of asthma patients. This means the strength of evidence for confirmation or rejection of recorded diagnosis in CPRD varied among the participants, and a panel diagnosis for chronic respiratory diseases can be considered as subjective. In the original validation studies, PPV's were calculated separately for people for whom additional information was and was not provided, with similar results in these sensitivity analyses.

Fourth, we are assuming the samples are representative of the asthma and COPD populations, while both sampling methods were based on possible identification strategies. The identification strategies or algorithms used for sampling are described in detail in their respective validation studies. (297,320) Fifth, this study clarifies the burden of concomitant asthma and COPD diagnosis in primary care, but additional information on how these patients are treated is needed in further studies.

Fifth, the COPD population was selected to include only current or ex-smokers. This implies that our findings in the COPD populations are only valid for (ex-) smokers.

Finally, GPs could have more information on the clinical status which was not shared in the questionnaire. This risk is present but diminished as the provision of additional information was remunerated.

Implications for research and/or practice

This study suggests that over diagnosis of asthma in COPD patients is more likely than over diagnosis of COPD in asthma patients. COPD is possibly more conservatively diagnosed as it is considered a more severe disease, while asthma can be more liberally diagnosed. In addition, a COPD patient can be diagnosed with asthma in the years before first COPD diagnosis, after which no further recording of asthma is made. This suggests the asthma diagnosis was likely to be false. In patients with presumed concomitant diagnosis of asthma and COPD, reversibility testing can be used to verify the asthma diagnosis.

The findings from our study have implications on further research into concomitant asthma and COPD. Identifying potential concomitant asthma and COPD using electronic health records should be done cautiously. If only a single code for both diseases is required for the identification algorithm, the prevalence of concomitant diagnosis of asthma and COPD is likely to be overestimated.

In addition, this study also has implications for the management of COPD patients with a past asthma diagnosis, as the previous asthma diagnosis might be either outdated or misdiagnosed. Incorrect management can expose them to adverse effects and incur additional costs for the patient and health system, for example though unnecessary medication regimens such as the usage of montelukast in COPD patients. This study did not go into detail on the current treatment of either validated asthma or validated COPD patients.

CONCLUSION

Concomitant asthma and COPD was likely in 14.8% (95% CI: 11.3%-19.0%) of validated asthma patients, and in 14.5% (95% CI 11.2%-18.3%) of validated COPD patients. However, more than half of the validated COPD patients had ever received an asthma diagnosis Read code, suggesting over diagnosis of asthma in COPD patients commonly occurs, particularly early in the diagnostic process. Over diagnosis of COPD in asthma patients is less likely.

Ethical approval

The protocol for this research was approved by the Independent Scientific Advisory Committee (ISAC) for MHRA Database Research (ISAC protocol 12_065 and 15_257), the approved protocol was made available during peer review. Ethical approval for this this study was also obtained from the London School of Hygiene and Tropical Medicine research ethics committee. Generic ethical approval for observational research using the CPRD with approval from ISAC has been granted by a Health Research Authority Research Ethics Committee (East Midlands – Derby, REC reference number 05/MRE04/87).

Contributors

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content. All authors have read, interpreted the results, commented on and approved the final manuscript.

Research paper cover sheet

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Francis Nissen
Principal Supervisor	Ian Douglas
Thesis Title	Asthma in electronic health records: validation & phenotypes.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	**Yes**	Was the work subject to academic peer review?	**Yes**

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	British Journal of General Practice
Please list the paper's authors in the intended authorship order:	Francis Nissen, Daniel R.Morales, Hana Mullerova, Liam Smeeth, Ian J Douglas, Jennifer K Quint
Stage of publication	**In press**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have drafted the protocol, extracted and analysed the data, and drafted the first and the final manuscript.

Student Signature: _____ **Date:** 13/08/2018

Supervisor Signature: _____ **Date:** 14/08/2018

Chapter 6: Clinical profile of pre-defined asthma phenotypes in a large cohort of UK primary care patients (CPRD)

Summary

- Established asthma phenotypes can be identified in a general asthma population, although many patients did not fit into the specific phenotypes which we studied.
- 3.9% of asthma patients were categorised as benign asthma, 28.6% as atopic asthma, and 4.8% as obese non-eosinophilic asthma.
- 62.7% of patients were included in the asthma NOS (Not Otherwise Specified) group, including asthma NOS without treatment (10.4 %), only on SABA (6.1%) and on maintenance treatment (46.2%).
- Exacerbation rates per 1000 person-years were lowest for benign asthma (106.8 [95% CI:101.2-112.3]), and highest for obese non-eosinophilic asthma (469.0 [451.7-486.2]).
- Asthma incidence rate ratios for all phenotype groups compared with the benign asthma group decreased when stratified by treatment step but remained raised.
- Phenotyping along with knowledge of asthma treatment step helps anticipate future treatment needs and could further aid clinical management.
- However, this is only possible in a minority of primary care patients based on current phenotypes and EHR records so either more complete records or EHR specific phenotypes would be helpful.
- The treatable traits strategy might represent a better conceptual framework towards precision medicine for asthma than phenotyping using primary care EHR at this stage.

6.1 Preface

The research paper presented in this chapter aims to answer the final aim of this thesis by investigating the prevalence, exacerbation risk and patient profile of different asthma phenotypes in a general population. This study builds upon the foundations of the studies included in earlier chapters and on one other study conducted by Chloe Bloom using a similar patient cohort in the CPRD GOLD on which I am second author. (1)

The findings of the research papers included in the previous chapters of this thesis have provided relevant information for the rationale and design of this study. The systematic review indicated possible ways to identify asthma patients from EHR. The validation study identified the optimal algorithm to determine which patients had asthma from the CPRD GOLD, and the study on concomitant asthma and COPD assessed the prevalence and possible misdiagnosis of COPD in asthma patients in the CPRD.

The study by Bloom et al. examined the general asthma population in the UK and their exacerbation risk and patient characteristics by age cohort, as most of the earlier literature has focused on more severe patients or severe exacerbations. The study design was a population-based cohort study using CPRD GOLD and HES, from 2007 to 2015. The population was divided into four age cohorts, under 5, 5-17, 18-54 and older than 55. Poisson regression was used for the regression analysis. The risk factors included gender, socioeconomic status, smoking, body mass index, atopy, rhinitis, gastro-oesophageal reflux, anxiety, depression, COPD and asthma severity defined by asthma treatment according to the British Thoracic Society (BTS) stepwise approach. This study found a total population of 424 326 patients, of whom 60% had mild asthma. The results indicated most UK patients with asthma (60%) had mild asthma (corresponding to BTS steps 1 and 2) and did not have an exacerbation during the

follow-up. Older patients (aged older than 55 years) were more likely to have a higher treatment step and had a higher exacerbation rate. Patients aged between 5 and 18 years were less likely to have a high treatment step and had the lowest exacerbation rates. This study used the asthma algorithm validated in the fourth Chapter of this thesis and similar definition of covariates as the study presented in this chapter.

For the cohort study presented in this chapter, we applied the phenotype categorization identified in primary care by Haldar et al.: benign asthma, early-onset atopic asthma and obese non-eosinophilic asthma.(2) The method section was kept brief to fit the article in a 3000-word limit for publication. The following paragraph expands upon the methods section in the submitted article under section 6.2.

Baseline characteristics were tabulated for each phenotype. Asthma exacerbation incidence rates were calculated for each phenotype using negative binomial regression with lexis expansion for age. We used negative binomial regression over Poisson regression with overdispersion as it provides a better fit to the distribution of data in exacerbation studies.(3,4) A random effects model was used to account for consecutive asthma exacerbations in the same study participant. The minimally adjusted model was adjusted only for age and sex. The fully adjusted model additionally adjusted for confounding by lifestyle factors and comorbidities: smoking, body mass index, socio-economic status through the index of multiple deprivation (IMD), gastro-oesophageal reflux disease (GORD), anxiety, depression and COPD. In addition, we stratified the final treatment step by asthma severity, defined by the received treatment. Sensitivity analyses were performed by including those patients without information on their smoking status and body mass index. We looked for evidence of interaction between phenotypes and age and between phenotypes and sex in the final model using likelihood ratio tests between models with and without interaction terms. Results were displayed using Forest plots and Cumulative Hazard ratios.

This study used negative binomial regression over Poisson regression (a common model in exacerbation studies). The rationale for this was that Poisson regression would not specifically account for interpatient variability, and the distribution of exacerbations would be more dispersed than would be predicted in a Poisson distribution. An overdispersion correction (deviance correction or Pearson correction)(27) could partially resolve these problems, but this would remain only a generic correction and the exacerbation variability in patients would not be an explicit component of the Poisson regression model.(334,335) The data had the potential to be overdispersed (when the conditional variance exceeds the conditional mean. The negative binomial regression model is similar to a Poisson regression model as is has the same mean structure, but it has an extra parameter to model overdispersion. Negative binomial regression could provide a better model for exacerbation data than Poisson regression.(336,337) The assumption of this model is that each individual has their own underlying rate of exacerbations. While the exacerbation count of each individual follows a Poisson distribution, the expected number of exacerbations varies across individuals following a gamma distribution.(334) The confidence intervals using the negative binomial regression would be narrower compared to a Poisson regression model if there is overdispersion of the conditional distribution of the outcome variable. The data was not overly dispersed in retrospect, but nothing was lost by using this negative binomial regression over Poisson regression.(334)

Furthermore, Poisson regression would not account for survivor bias, e.g. if patients who are more likely to exacerbate would also be more likely to be lost to follow-up.(338) Patients who exacerbate often tend to withdraw earlier from follow-up.(338) The survivor bias referenced in the thesis did not refer to the competing risk of death, but rather the time when an exacerbation could not be measured, similar to

immortal time bias.(339,340) In a Poisson model, each unit of time is weighted equally (a patient who is followed up for 12 months is weighted 4 times more than a patient who is followed for 3 months). A negative binomial model includes the modelling of variability between patients in the estimation of rates. If the exacerbation number follows the distribution above, it can account for increased exacerbation rates in patients who withdraw early.(335)

The negative binomial regression analysis presented a model that assumed each participant has their own underlying exacerbation rate.(336,337) Within each participant, the exacerbation number followed a Poisson distribution but the predicted number of exacerbations varied across participants according to a gamma distribution (a variant on the normal distribution).(334) The advantages of negative binomial regression over Poisson regression included the less simplistic assumption about participant variability, and that the negative binomial regression model accounted better for participants with frequent exacerbations withdrawing early.(335,341)

The research paper presented in this chapter has been submitted for peer review to the “Journal of Asthma and Allergy”.

6.2 Research paper

Clinical profile of pre-defined asthma phenotypes in a large cohort of UK primary care patients (CPRD).

Short Title: Pre-defined asthma phenotypes in CPRD

Authors: Francis Nissen,[1] Ian J Douglas,[1] Hana Mullerova,[2] Neil Pearce,[1] Chloe Bloom,[3] Liam Smeeth,[1] Jennifer K Quint[3]

1. *Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK*
2. *RWE & Epidemiology, GSK R&D, Uxbridge, UK*
3. *National Heart and Lung Institute, Imperial College, London, UK*

ABSTRACT

Background:

Distinct asthma phenotypes have previously been suggested, including benign asthma, atopic asthma and obese non-eosinophilic asthma. This study aims to establish if these phenotypes can be identified using data recorded in primary care clinical records and reports on patient characteristics and exacerbation frequency.

Methods: A population-based cohort study identified 193,999 asthma patients in UK primary care from 2007 to 2017. We used linked primary and secondary care data from the Clinical Practice Research Datalink, Hospital Episode Statistics and Office of National Statistics. Patients were classified into predefined phenotypes or included in an asthma "not otherwise specified" (NOS) group. We used negative binomial

regression to calculate the exacerbation rates and adjusted rate ratios. Rate ratios were further stratified by asthma treatment step.

Results:

In our cohort, 3.9% of patients were categorized as benign asthma, 28.6% atopic asthma, and 4.8% obese non-eosinophilic asthma. 62.7% of patients were asthma NOS, including asthma NOS without treatment (10.4 %), only on SABA (6.1%) and on maintenance treatment (46.2%). Crude severe exacerbation rates per 1000 person-years were lowest for benign asthma (106.8 [95% CI:101.2-112.3]), and highest for obese non-eosinophilic asthma (469.0[451.7-486.2]). Incidence rate ratios for all phenotype groups decreased when stratified by treatment step but remained raised compared with benign asthma.

Conclusion:

Established phenotypes can be identified in a general asthma population, although many patients did not fit into the specific phenotypes which we studied. Phenotyping patients and knowledge of asthma treatment step could help anticipate clinical course and therefore could aid clinical management but is only possible in a minority of primary care patients based on current phenotypes and EHR records.

INTRODUCTION

Asthma is a heterogeneous disease with recognisable clusters called asthma phenotypes.(62,113,342,343) These phenotypes are defined as the set of observable characteristics of an individual resulting from the interaction of its genotype with the environment.(93) Classifying asthma into phenotypes allows to deconstruct the disease into separate identifiable and treatable traits (8) and better understand disease

progression and response to treatment, further enabling practice of precision medicine.(113)

There have been multiple studies describing asthma phenotypes, (101–110,344) involving populations with asthma alone or as part of an entity called "obstructive airways disease" together with COPD.(111,112) Criteria to distinguish asthma phenotypes include inflammatory profiling based on leucocyte counts (eosinophils, neutrophils and paucigranulocytic), symptom expression, age-of-asthma onset, and airflow measurements.(94–100) Classification by eosinophil counts has been found to be particularly important due to treatment response.(114–116)

One of the most impactful studies on clinical asthma phenotypes was conducted by Haldar et al using cluster analysis of multiple clinical variables.(101) Among 184 patients managed in primary care, three clusters were found: one group with benign asthma, one group with obese non-eosinophilic asthma and one group with early-onset atopic asthma (figure 1). Cluster analysis of two further mostly refractory asthma populations managed in secondary care (N=255 total) added an early symptom predominant cluster and an inflammation predominant cluster. Other phenotyping studies using comparable clinical variables found similar phenotypes.(103,106,194–196)

In this study, we examined if it is possible to identify asthma patients with one of three phenotypes identified in primary care by Haldar et al from electronic health records and report their characteristics and medication use. To accurately classify patients into a phenotype, strict criteria were applied. The exact criteria are described in the methods section. Blood eosinophil tests were used as they are well recorded in CPRD, unlike sputum eosinophils.(345)

We used the CPRD (Clinical Practice Research Datalink) GOLD database to identify asthma patients. To define asthma exacerbations, we also linked to the Hospital Episodes Statistics (HES) and Office of National Statistics data.

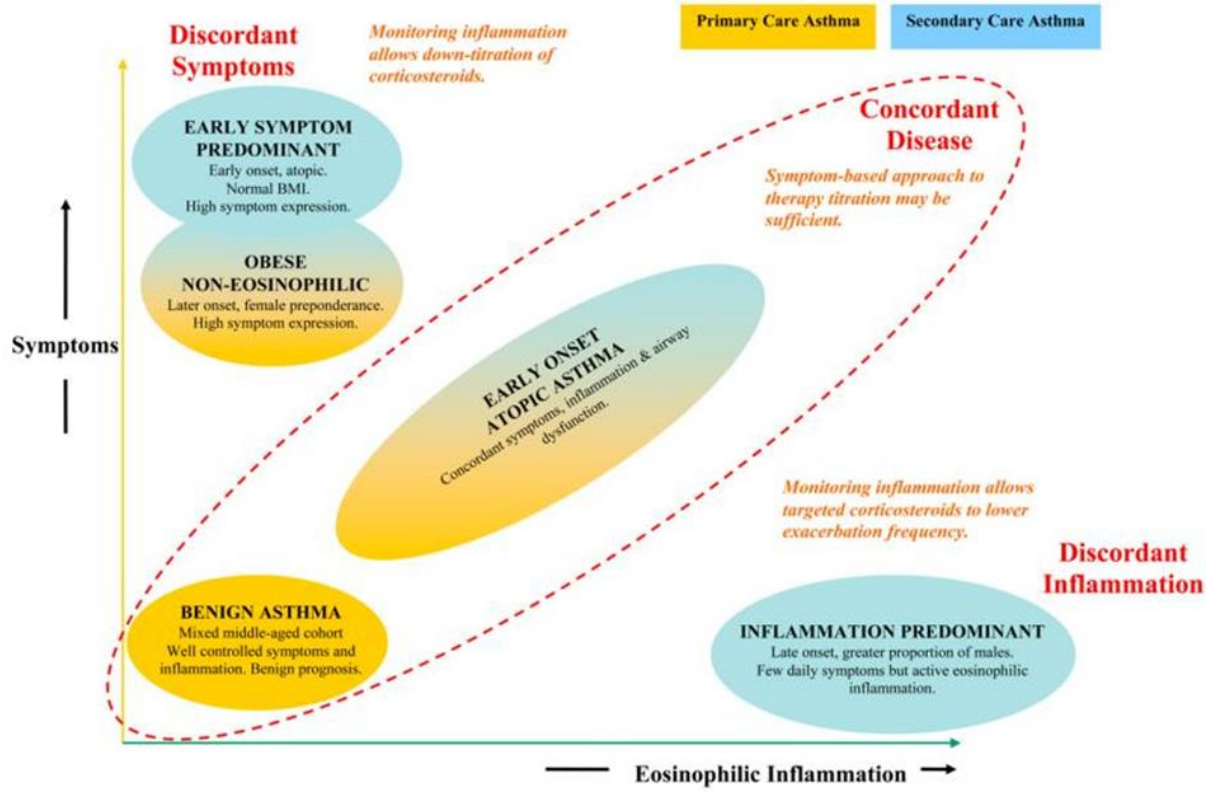


Figure 1: Asthma Phenotypes, based on cluster analysis (quote from Haldar et al., reproduced with permission. Reprinted with permission of the American Thoracic Society. Copyright © 2018 American Thoracic Society. Haldar P et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med. 2008;178(3):218–24. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.)

Aim and Objectives:

To evaluate the extent to which three previously suggested asthma phenotypes (benign asthma, atopic asthma, obese non-eosinophilic asthma) could be identified using data included in routinely collected electronic health records, and to assess the exacerbation frequency, clinical profile and medication use by phenotype.

METHODS

Data sources

We used the July 2017 dataset of the Clinical Practice Research Datalink (CPRD), a large UK primary care database containing anonymized data of people registered

with primary care practices from across the UK. CPRD is representative of the UK population with regard to age and sex.(208,306) Diagnostic accuracy is high in CPRD, including for asthma and COPD,(229,297,308,320) and CPRD can be used to identify individuals at risk of recurrent asthma attacks.(346,347) Only patients with linkage to Hospital Episodes Statistics (HES) and Office for National Statistics (ONS) were considered for inclusion. Linked data are available for patients registered at consenting English practices. This study used only data on patients who were linked to Hospital Episodes Statistics (HES) for all hospital inpatient admissions and emergency visits and Office for National Statistics for deaths and socio-economic status through the index of multiple deprivation (IMD).

Study population and follow-up

Adult patients in CPRD (18 years of age or older) with linkage to HES and ONS and a validated asthma Read code between April 2007 and July 2017 in addition to a valid blood eosinophil count, BMI and determinable smoking status were eligible for inclusion.(320) Patients entered the cohort at the latest date of: 1 year of follow-up from the practice up-to-standard date (contributing research quality data to CPRD); reaching 18 years of age; available linkage and from April 2007 at earliest. Once all these criteria were fulfilled, participants were included in one of the phenotypes below after one year of continuous follow-up in which their SABA and inhaler use is measured and remained in their respective phenotype(s) during further follow up. The index date was the time point after the one year of follow-up, when the participants are included in a phenotype. The follow-up ended when the patient was transferred out, death or end of study period. During this year, their reliever medication (short-acting beta agonist, SABA) and maintenance treatment prescriptions (see covariate section) were measured. The time point after one year when patients were assigned to a phenotype and after which exacerbations were measured was designated as the index date (figure 2). We used SABA prescription count as proxy for symptom expression,(348,349) as asthma symptoms are non-

specific and often not recorded in CPRD. Blood eosinophil counts, routinely recorded in primary care, were used as proxy for sputum eosinophil counts.(345,347,350) Patients were assigned their phenotype group on index date, after which severe asthma exacerbations were ascertained and counted. Patients remained in their respective phenotype group and followed-up until the earliest date of transfer-out of CPRD practice, last collection date, death or end of study period (01/07/2017).

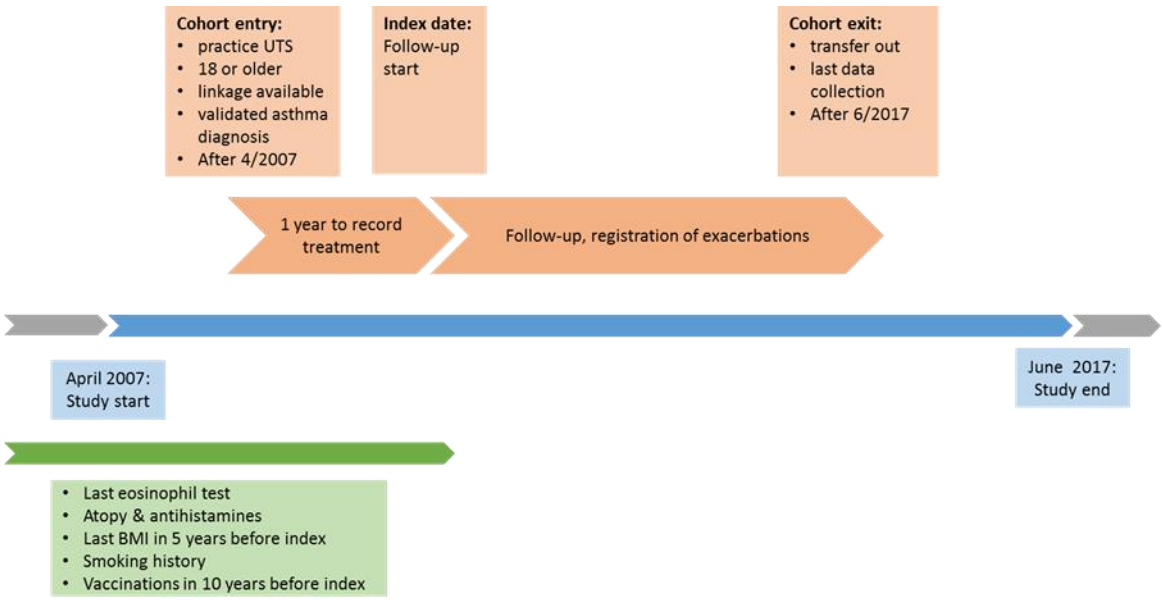


Figure 2: Cohort timeline

Definition of the phenotypes

Each patient was assigned to a single phenotype group on index date based on previously recorded information. Stringent inclusion criteria were used to keep the phenotype groups specific. All code lists for covariates and comorbidities are included in attachment and on Data compass. Patients were only allowed to be classified into a single phenotype. Obese non-eosinophilic asthma held priority over atopic asthma as the former phenotype was deemed more specific. Phenotype groups were defined as follows:

1) Benign asthma: low eosinophil counts on the latest blood test from April 2007 onward (< 300 cells/μL and <4% of leucocytes), absence of SABA prescriptions and of severe exacerbations of asthma in the year before index date, and aged 40-60 at study entry.

2) Atopic asthma: occasional SABA prescriptions (2-4 in year before index date, excluding 23% of patients without SABA prescriptions and 26% with 5+ prescriptions) and ≥1 atopy or ≥2 antihistamine codes ever recorded.

3) Obese non-eosinophilic asthma: low eosinophil counts on the last blood count (ie blood eosinophil levels less than 300 cells/μL (36) and less than 4% of blood leucocytes), female, frequent SABA prescriptions (≥3 prescriptions in year before index date, which corresponds to 42% of patients) and at least one record for BMI >30 in last five years before index date.

4) Asthma NOS (Not Otherwise Specified): patients who did not fit in previous phenotypes were split into three sub-groups: (1) patients without any asthma medication prescriptions, (2) patients only on SABA and (3) patients with at least one maintenance treatment prescription during the one year before index date. We described these groups to determine whether they fit other phenotypes described in the literature.

Definition of severe asthma exacerbations

An exacerbation was defined as any of the following: prescription of ≤300mg oral corticosteroids (OCS) outside an annual asthma review (4) or an A&E visit, acute hospital visit of <1 day duration, overnight hospitalisation or death due to asthma. This corticosteroid dose cut-off was chosen to eliminate chronic oral corticosteroid use for other conditions than asthma. Exacerbations within fourteen days of a previous exacerbation were excluded.

Definition of covariates

Age was defined in 10-year age bands, socio-economic status was assigned at patient level using the ONS Index of Multiple Deprivation. Smoking status was categorized as either current smoker, ex-smoker or never-smoker. Co-morbid conditions were determined by Read codes: COPD, atopic dermatitis, GORD (Gastro-oesophageal Reflux Disease), atopy (eczema or rhinitis), anxiety and depression. Influenza and pneumococcal vaccinations during last 10 years were included as covariates. The final model was stratified by disease severity based on the stepwise approach in the 2016 British Thoracic Society Asthma Management Guidelines (BTS/SIGN) which includes inhaled ICS thresholds.(4) Step 1 was defined by absence of maintenance asthma treatment. Step 2 by regular prescription of low-dose ICS. Step 3 added long acting beta agonists (LABA). Step 4 by medium-dose ICS with or without additional therapies (LABA, theophyllines, leukotriene receptor antagonists or long-acting antimuscarinics). Step 5 was defined by high-dose ICS and step 6 by continuous/frequent use of oral corticosteroids.

Data analysis

Baseline characteristics were tabulated for each phenotype. Asthma exacerbation incidence rates and rate ratios were calculated using negative binomial regression with a random effects model and lexis expansion for age. We used negative binomial regression over Poisson regression with overdispersion as it provides a better fit to the distribution of the data.(334,335) The minimally adjusted model included age and sex only. The fully adjusted model additionally controlled for smoking status, body mass index, socio-economic status, GORD, pneumococcal and influenza vaccinations, anxiety, depression and COPD. In addition, we stratified incidence rate ratios (IRRs) by severity, defined by prescribed treatment. Stata 15.0 was used for data analysis. Results were displayed using Forrest plots and a Kaplan-Meier survival plot to display time to first exacerbation.

Sample size calculations

The reasoning behind the sample size of the cohort study was as follows: We would need 3786 asthma patients in each of the different phenotypes to detect an exacerbation rate ratio of 0.9 (alpha 0.05 and power 0.9) considering an equal amount of asthma patients in 2 phenotypes. A 95% confidence interval for baseline characteristics and rate ratios quantified the random error associated with our estimate.

Missing data

The plan for addressing missing data for the cohort study was as follows:Two phenotypes require an absence of high eosinophilia counts, but tests for eosinophils have not been recorded for all asthma patients. Feasibility counts on the number of patients with asthma from 2004-2015 show that 71% of records have a valid eosinophil count. We anticipated a small degree of missingness for the BMI and smoking covariates based on previous studies. Decisions regarding how to deal with missing values were based on the proportion of missing data, and assumptions regarding whether data was missing at random (MAR) or not. Where appropriate we would undertake a complete case analysis.

If data was thought to be MAR we would consider using multiple imputation, however this MAR assumption did not seem likely. MAR means that there might be systematic differences between the missing and observed variables, but these can be entirely explained by other observed variables. We would not be able to predict variables such as eosinophil values based on other observed variables. Where data was not missing at random, for example with BMI or smoking data, but where we expect the data to be ~80% complete (based on previous studies), we would use a complete case analysis but will discuss biases that may occur as a result of adopting that approach. BMI data is unlikely to be MAR as patients with overweight would be more likely to have their BMI recorded as it would be more clinically relevant. If multiple imputation was not appropriate and large quantities of data would be

missing, we would consider using those covariates only as part of a secondary analysis and discuss any biases and limitations that would occur as a result of that.

	Benign Asthma	Atopic Asthma	Obese non-eosinophilic	NOS no med	NOS with only SABA	NOS with maintenance	Total
Total	7,495	55,455	9,372	20,204	11,926	89,547	193,999 (100.0%)
Percentage of total	3.9%	28.6%	4.8%	10.4%	6.1%	46.2%	100.0%
Follow-up median & IQR	4.28 (2.05;6.36)	4.34 (2.10;6.46)	4.54 (2.21;6.84)	3.46 (1.41;5.26)	4.11 (1.93;6.11)	4.35 (2.10;6.46)	4.24 (1.99;6.34)
Gender							
Female	4,727 (63.1%)	38,060 (68.6%)	9,372 (100.0%)	13,627 (67.4%)	7,142 (59.9%)	53,828 (60.1%)	126,756 (65.3%)
Age category							
18-30 y	0	10,693 (19.3%)	813 (8.7%)	6,536 (32.4%)	1,966 (16.5%)	9,600 (10.7%)	29,608 (15.3%)
31-50 y	4,201 (56.1%)	19,762 (35.6%)	3,127 (33.4%)	6,810 (33.7%)	4,736 (39.7%)	26,585 (29.7%)	65,221 (33.6%)
51-70 y	3,294 (43.9%)	17,005 (30.7%)	3,888 (41.5%)	4,365 (21.6%)	3,451 (28.9%)	32,381 (36.2%)	64,384 (33.2%)
≥71 y	0	7,995 (14.4%)	1,544 (16.5%)	2,493 (12.3%)	1,773 (14.9%)	20,981 (23.4%)	34,786 (17.9%)
Socio-economic status							
1: least deprived	2,092 (27.9%)	13,647 (24.6%)	1,260 (13.4%)	4,696 (23.2%)	2,465 (20.7%)	17,971 (20.1%)	42,131 (21.7%)
2	1,724 (23.0%)	12,153 (21.9%)	1,644 (17.5%)	4,306 (21.3%)	2,405 (20.2%)	18,813 (21.0%)	41,045 (21.2%)
3	1,513 (20.2%)	11,090 (20.0%)	1,801 (19.2%)	4,154 (20.6%)	2,450 (20.5%)	18,352 (20.5%)	39,360 (20.3%)
4	1,219 (16.3%)	10,189 (18.4%)	2,149 (22.9%)	3,844 (19.0%)	2,423 (20.3%)	17,655 (19.7%)	37,479 (19.3%)
5: most deprived	943 (12.6%)	8,350 (15.1%)	2,513 (26.8%)	3,192 (15.8%)	2,180 (18.3%)	16,702 (18.7%)	33,880 (17.5%)
Smoking status							
Current smoker	1,473 (19.7%)	10,523 (19.0%)	2,485 (26.5%)	4,637 (23.0%)	3,541 (29.7%)	22,165 (24.8%)	44,824 (23.1%)
Ex-smoker	2,884 (38.5%)	22,930 (41.3%)	3,813 (40.7%)	6,679 (33.1%)	4,090 (34.3%)	38,000 (42.4%)	78,396 (40.4%)
Never smoker	3,138 (41.9%)	22,002 (39.7%)	3,074 (32.8%)	8,888 (44.0%)	4,295 (36.0%)	29,382 (32.8%)	70,779 (36.5%)
BMI							
<20	278 (3.7%)	3,257 (5.9%)	0	1,477 (7.3%)	756 (6.3%)	6,018 (6.7%)	11,786 (6.1%)
20-25	2,041 (27.2%)	16,819 (30.3%)	0	6,596 (32.6%)	3,488 (29.2%)	25,468 (28.4%)	54,412 (28.0%)
25-30	2,703 (36.1%)	18,814 (33.9%)	0	6,267 (31.0%)	4,080 (34.2%)	31,068 (34.7%)	62,935 (32.4%)
>30	2,473 (33.0%)	16,565 (29.9%)	9,369 (100.0%)	5,864 (29.0%)	3,602 (30.2%)	26,993 (30.1%)	64,866 (33.4%)

Table 1: Baseline characteristics of the study population by phenotype

	Benign Asthma	Atopic Asthma	Obese non-eos	NOS no med	NOS with only SABA	NOS with maintenance	Total
BTS step							
BTS1	4,401 (58.7%)	11,612 (20.9%)	754 (8.0%)	20,204 (100.0%)	11,926 (100.0%)	0	48,897 (25.2%)
BTS2	1,186 (15.8%)	18,898 (34.1%)	2,526 (27.0%)	0	0	29,809 (33.3%)	52,419 (27.0%)
BTS3	520 (6.9%)	6,220 (11.2%)	1,164 (12.4%)	0	0	12,309 (13.7%)	20,213 (10.4%)
BTS4	978 (13.0%)	11,710 (21.1%)	2,625 (28.0%)	0	0	26,173 (29.2%)	41,486 (21.4%)
BTS5	263 (3.5%)	5,346 (9.6%)	1,831 (19.5%)	0	0	15,825 (17.7%)	23,265 (12.0%)
BTS6	3 (.0%)	243 (.4%)	129 (1.4%)	0	0	1,457 (1.6%)	1,832 (.9%)
Non BTS	144 (1.9%)	1,426 (2.6%)	343 (3.7%)	0	0	3,974 (4.4%)	5,887 (3.0%)
Comorbid conditions							
Atopy	3,069 (40.9%)	45,632 (82.3%)	3,965 (42.3%)	9,283 (45.9%)	1,388 (11.6%)	26,577 (29.7%)	89,914 (46.3%)
GORD	1,291 (17.2%)	11,331 (20.4%)	2,374 (25.3%)	2,965 (14.7%)	1,859 (15.6%)	18,242 (20.4%)	38,062 (19.6%)
Anxiety	2,161 (28.8%)	17,841 (32.2%)	3,584 (38.2%)	5,835 (28.9%)	3,397 (28.5%)	26,182 (29.2%)	59,000 (30.4%)
Depression	2,874 (38.3%)	21,826 (39.4%)	5,125 (54.7%)	7,240 (35.8%)	4,394 (36.8%)	34,056 (38.0%)	75,515 (38.9%)
COPD	351 (4.7%)	5,676 (10.2%)	1,836 (19.6%)	1,005 (5.0%)	1,313 (11.0%)	20,512 (22.9%)	30,693 (15.8%)
Eosinophils							
Eosinophils <300/μL & <4%	7,495 (100.0%)	24,967 (45.0%)	9,372 (100.0%)	9,800 (48.5%)	6,439 (54.0%)	38,539 (43.0%)	96,612 (49.8%)
Eosinophils >300/μL / >4%	0	30,488 (55.0%)	0	10,404 (51.5%)	5,487 (46.0%)	51,008 (57.0%)	97,387 (50.2%)
SABA							
0 prescriptions	7,495 (100.0%)	0	0	20,204 (100.0%)	0	16,306 (18.2%)	44,005 (22.7%)
1-2 prescriptions	0	0	(.0%)	0	7,430 (62.3%)	21,671 (24.2%)	68,381 (35.2%)
3-9 prescriptions	0	16,175 (29.2%)	7,519 (80.2%)	0	3,585 (30.1%)	36,703 (41.0%)	63,982 (33.0%)
10+ prescriptions	0	0	1,853 (19.8%)	0	911 (7.6%)	14,867 (16.6%)	17,631 (9.1%)

Table 1: Baseline characteristics of the study population by phenotype

RESULTS

Background characteristics

Of 323,862 asthma patients with complete linkages and eligible for inclusion, 193,999 (59.9%) had at least one year of follow-up and an eosinophil count, BMI value and smoking variables and formed the analysis population (figure 3). Study participants were followed up for a median of 4.24 years (IQR:1.99-6.34); the median age at study entry was 51 years (IQR 37-66). 65.3% were female and 63.5% were smokers or ex-smokers.

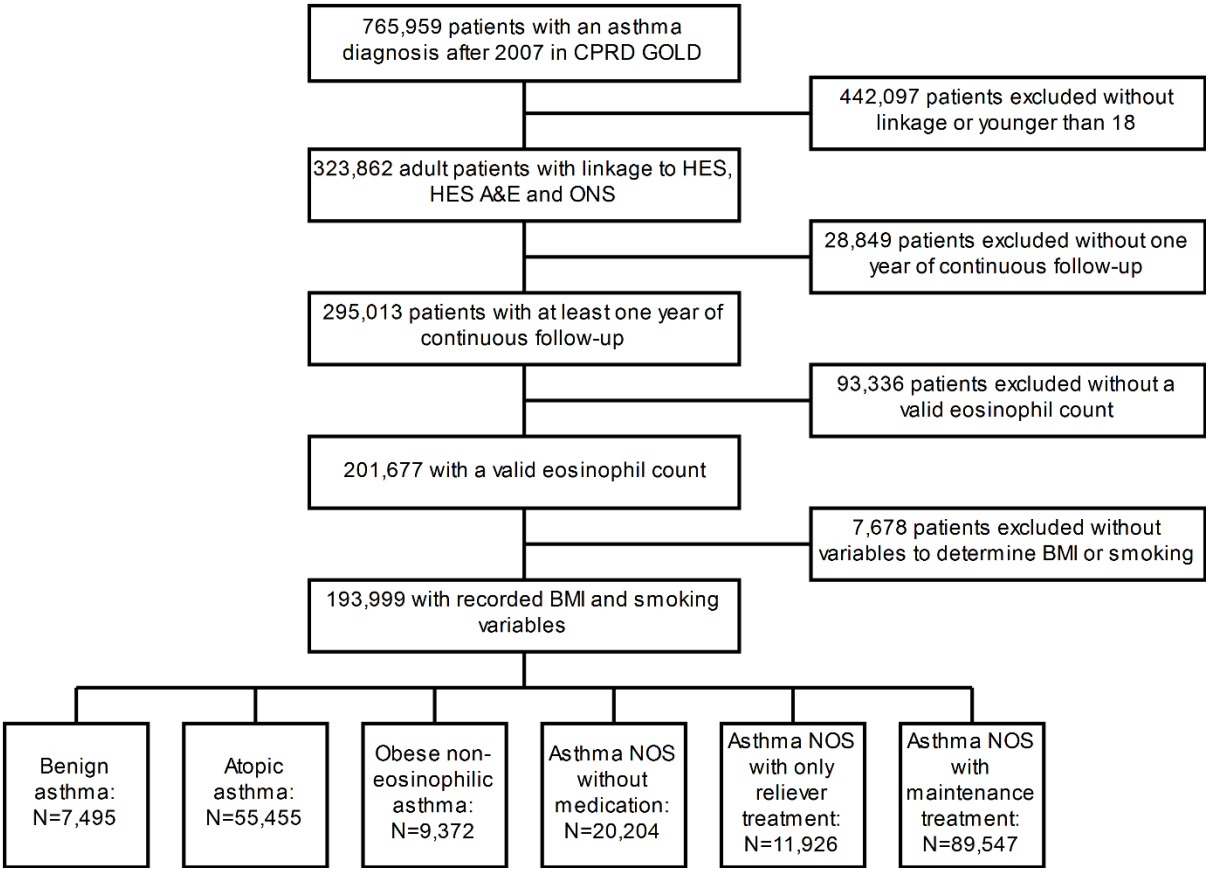


Figure 3: Flowchart of study eligibility and participation

In this primary care asthma population, 7,495 (3.9%) were classified into the benign asthma group, 55,455 (28.6%) as atopic asthma, and 9,372 (4.8%) as obese non-eosinophilic asthma (Table 1). Of the remaining patients classified as asthma NOS, 20,204 (10.4%) did not receive any asthma medication, 11,926 (6.1%) had only SABA

prescription codes and 89,547 patients (46.2%) had maintenance treatment in the year before index date.

The patient characteristics and total follow-up duration varied between phenotypes. The asthma NOS group with maintenance treatment had the highest mean age on study entry (55 years, SD 18 years). Average BTS step was highest in the same group (mean 3.38), followed by obese non-eosinophilic and atopic asthma (Figure 4). GORD and anxiety were most common in the obese non-eosinophilic group (25.3% and 38.2%, resp.), followed by atopic asthma and asthma NOS with maintenance treatment. Comorbid COPD was most common in the asthma NOS group with maintenance asthma treatment (22.9%). The last available eosinophil count was elevated in 50.2% of all patients and 65.8% were overweight or obese.

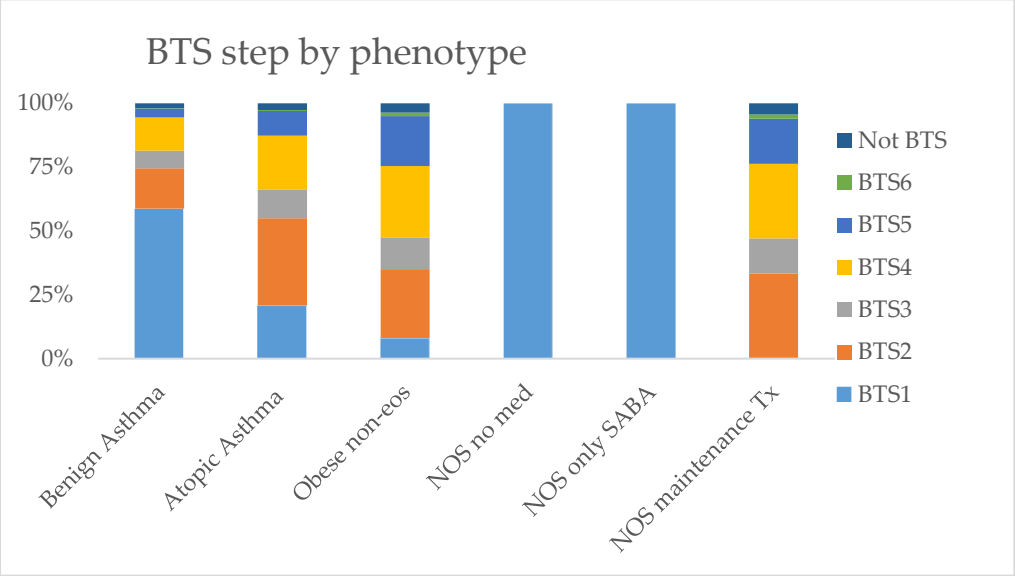


Figure 4: BTS step by phenotype

Severe Exacerbation Rates

The study participants were followed for a total of 819,619 years and 258,388 exacerbations were recorded (Table 2). Exacerbation rates (per 1000 person-years) were highest in the obese non-eosinophilic group and lowest in the benign asthma group. Minimally adjusted exacerbation rates per phenotype were as follows: 116.2

for benign asthma, 286.9 for atopic asthma, 454.9 for obese non-eosinophilic asthma, 148.1 for asthma NOS without medication, 208.6 for asthma with only SABA prescriptions and 389.4 for asthma NOS with maintenance medication.

Fully adjusted exacerbation rates controlling for lifestyle factors and comorbidities, show a similar relation between asthma phenotypes with event rates of: 143.2 for benign asthma, 322.1 for atopic asthma, 439.3 for obese non-eosinophilic asthma, 174.6 for asthma NOS without medication, 240.0 for asthma NOS with only SABA prescriptions and 414.0 for asthma NOS with maintenance treatment.

Table 2: Exacerbation rates with corresponding rate ratios by phenotype

Phenotype	No. of Events	Time at Risk (1000 person-y)	Crude Rate/1,000 Person-y (95% CI)	Minimally Adjusted	Adjusted rates
Benign Asthma	3,431	30.867	106.8 (101.2-112.3)	116.2 (110.1-122.3)	143.2 (135.6-150.8)
Atopic Asthma	68,143	239.664	283.2 (278.6-287.7)	286.9 (282.3-291.5)	322.1 (316.6-327.5)
Obese non-eos	19,263	42.471	469.0 (451.7-486.2)	454.9 (438.0-471.7)	439.3 (422.5-456.2)
NOS no med	10,495	70.978	143.8 (139.3-148.3)	148.1 (143.4-152.7)	174.6 (169.0-180.2)
NOS with reliever Tx	9,529	48.238	200.7 (193.2-208.1)	208.6 (200.9-216.3)	240.0 (231.1-249.0)
NOS with maint Tx	147,527	387.403	388.2 (383.5-392.9)	389.4 (384.6-394.1)	414.0 (408.6-419.3)

Minimally adjusted: adjusted for age and sex
Adjusted rates: adjusted for age, sex, smoking, bmi, imd, anxiety, depression, COPD, GORD
Note: Unless otherwise indicated, values are given as rates (95% CI).

Table 2: Exacerbation rates with corresponding rate ratios by phenotype

Rate ratios

Benign asthma had the lowest rate of exacerbations and was used as reference group for the calculation of incidence rate ratios (Figure 5). IRRs, (fully adjusted models) for asthma exacerbation were 2.28 (95% CI 2.16-2.41) for those with atopy; 3.11 (95% CI 2.91-3.32) for obese non-eosinophilic asthma, 1.23 (95% CI 1.16-1.31) for asthma NOS without medication, 1.69 (95% CI 1.58-1.80) for asthma NOS with SABA and 2.92 (95% CI 2.77-3.08) for asthma NOS with maintenance treatment. When stratified by BTS treatment step, the IRRs of all phenotypes compared with benign asthma decreased across all steps, but difference in incidence rates between the groups and benign asthma was still notable. Time to first exacerbation analysis showed a pattern comparable with rates derived by negative binomial regression (Figure 6). Shortest median time to exacerbation was observed in the obese non-eosinophilic asthma group, and longest in the benign asthma group. No clinically important interaction between phenotype and age or phenotype and gender was observed. Sensitivity analyses including patients with missing BMI or smoking status found similar results to the main analysis.

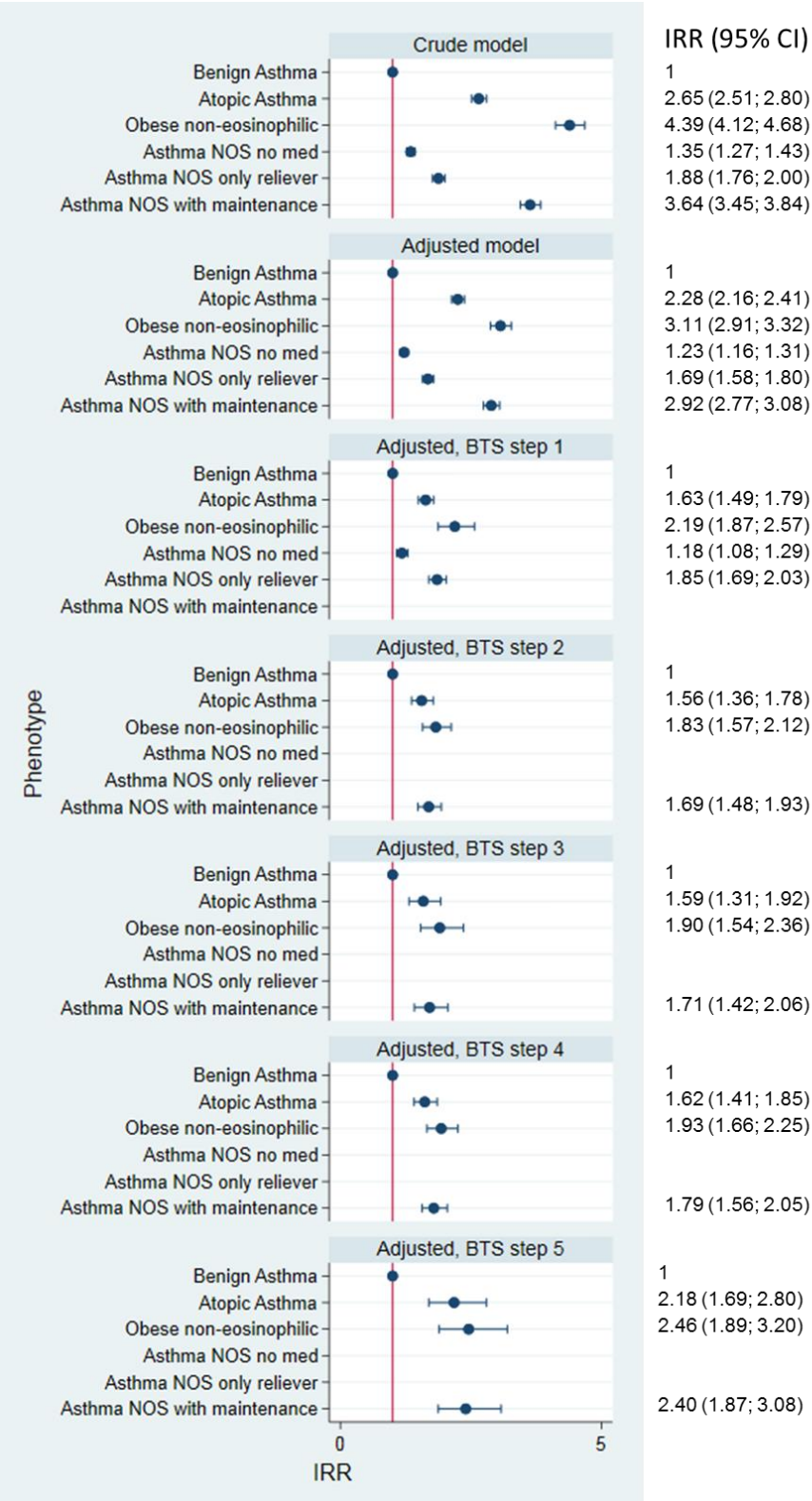


Figure 5: Incidence rate ratios, stratified by treatment step. Adjustment for step 6 resulted in very wide confidence intervals due to low sample size. This made the Forrest plot unreadable, so this adjustment is not displayed.

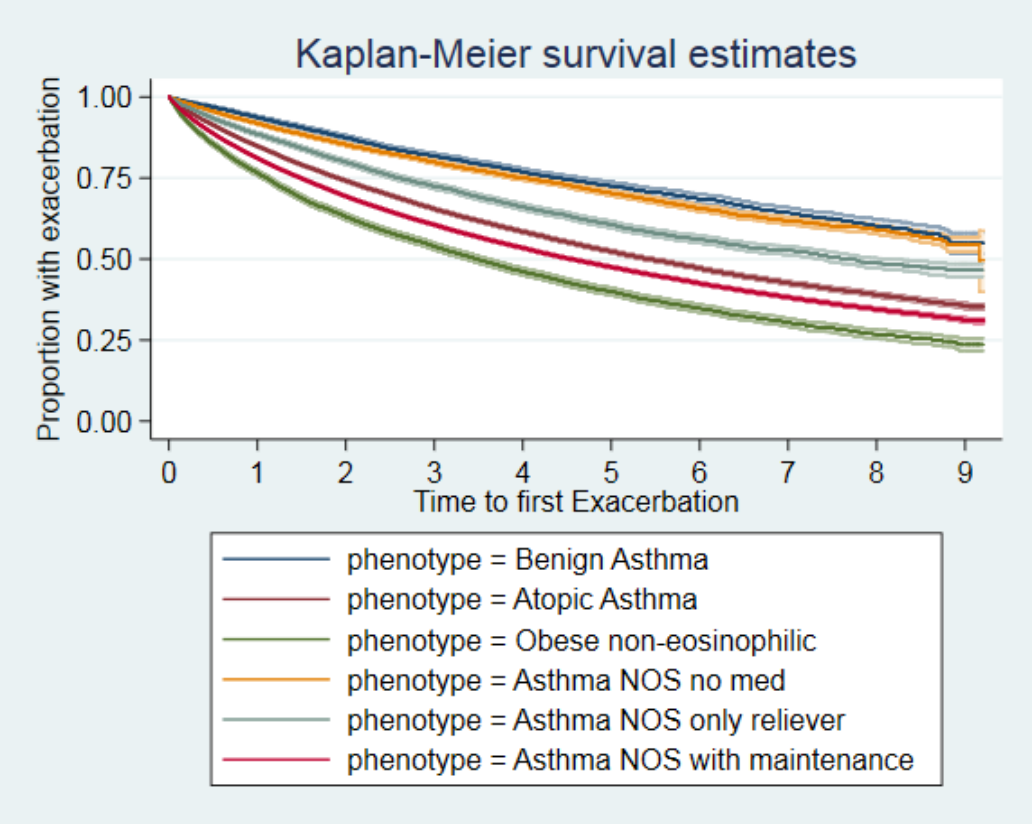


Figure 6: Time to first exacerbation analysis in years, by phenotype and 95% CI

DISCUSSION

Summary

In this study on asthma phenotypes in a large general asthma population, we were able to identify patients who fitted three previously suggested phenotypes: benign, early-onset atopic and obese non-eosinophilic asthma. Due to strict criteria used to define the three main phenotypes, most patients with asthma (62.7%) were not included in any of three predefined primary care phenotypes and were categorized as asthma NOS. Patients in the asthma NOS groups partly reflected some of the established phenotypes, for example inflammation predominant or early symptom predominant phenotypes. For example, in the asthma NOS with maintenance treatment group, 57.0% of patients had high eosinophil counts; possibly indicating some patients with an inflammation predominant phenotype were included in this group. However, these patients also had more SABA prescriptions (and presumably more symptoms) than the total cohort. This group may include undiagnosed COPD patients based on their higher average age and treatment step. There was a higher exacerbation burden in those with obese non-eosinophilic atopic asthma, and a lower exacerbation rate in those with benign asthma compared with the asthma NOS group (with and without medication) in the crude model. These rate differences persisted after adjustment for lifestyle factors and comorbidities. When stratifying the patients by treatment step, differences in incidence rates between phenotypes remained but were decreased. Phenotyping a greater proportion of asthma patients based on their primary care health records could be possible by either constructing different phenotypes or by creating more complete records (for example more full blood counts). However, this might not be the most efficient way to offer precision medicine to asthma patients. The recently proposed treatable traits strategy (24,351) might represent a better conceptual framework towards precision medicine for asthma than phenotyping using primary care EHR at this stage.(352,353) This strategy focuses

asthma management on single traits that are identifiable and treatable, such as airway inflammation (measured using eosinophil counts) or airflow limitation.(24)

There are multiple strengths to the current study. The study is population-based and representative of the population of England (30) which allows estimation of asthma phenotype prevalence, and the median length of follow-up is considerable (4.1 years). Further strengths include the detailed methods to define the dependent (exacerbation rate) and independent (phenotype) variable and the inclusion of exacerbations in primary, secondary and emergency care, in addition to asthma deaths. The asthma codes in the CPRD have been validated in a previous validation study using GP questionnaires.(40)

Comparison with existing literature

In the past, phenotype categorisation was mostly based on variables such as age of onset, severity, reaction to treatment or comorbidities. More recently, cluster analysis of clinical variables including airflow measurements has been used to describe phenotypes. These cluster analyses have all been limited in terms of sample size,(101,103,108,194,195) or were preselected such as severe asthma populations.(354) To the best of our knowledge, this is the first longitudinal study on a general population-wide asthma cohort. As such, it is difficult to draw direct comparisons between this population-based asthma study and previous phenotyping studies.

In the categorisation by Haldar et al, the three phenotypes we focused on here were described in a primary care cohort of 184 patients. In this cohort, 96 (52%) patients had benign asthma, 61 (33%) had early-onset atopic asthma and 27 (15%) had obese non-eosinophilic asthma. In addition, two more phenotypes were identified from two separate populations (including a secondary care and a longitudinal study of mostly refractory patients). The early onset symptom predominant phenotype has high symptom expression and a tendency towards overtreatment, while the inflammation-

predominant asthma phenotype has a lower symptom expression, but active eosinophilic inflammation. Another well-known categorisation of asthma phenotypes was undertaken by Moore et al (103) using cluster analysis in the U.S. Severe Asthma Research Program based on respiratory function and age-of-onset. While this analysis was heavily based on the latter (unfortunately routinely collected electronic records often lack information on age-of-onset), the obese non-eosinophilic asthma and early-onset atopic asthma were also identified. A recent cluster analysis on two populations of severe asthma patients (n= 238 total) defined and validated four severe asthma phenotypes mainly differentiated by lung function and level of eosinophilic inflammation.(14) The exacerbation frequency is similar to those of previous studies on asthma exacerbation rates in the UK.(346,347) Comparison of exacerbation rates between countries remains difficult without consensus on the definition of asthma.(3,24)

Limitations

The main limitation of this study is due to the nature of routinely collected data. For example, the CPRD does not hold information on the age of onset, which is one of the defining traits of the early-onset atopic asthma phenotype. Our inability to identify phenotype for a sizable proportion of the population highlights the need for developing phenotypes that can be more readily identified from routine care records, as well as the need for improving routine care records so that important phenotypes can be identified.

Residual confounding remains possible, despite the adjustment for several potential confounders. Misclassification of asthma is possible, but Read codes for asthma have a high PPV (86%) in CPRD.(320) The exacerbation cut-off of ≤300mg oral corticosteroids might have misclassified some patients. The blood eosinophil cut-off at <300 cells//μL for "eosinophilic asthma" is not absolute, and multiple different eosinophil levels are used in the literature.(76,355–357) We included only people with

full linkage, however exclusions were at practice level so unlikely to bias estimates. Asthma phenotypes might change over follow-up, but this would remain true even in a cohort study identified in real time. Similarly, we assume eosinophil levels do not change greatly over time. This assumption may not always hold as eosinophil levels are fluid and depend on the level of steroid treatment and inflammation (such as hay fever or recent viral infections). CPRD contains information on only prescriptions of treatments, without information on adherence to those treatments. In the case of SABA prescriptions, not all reliever treatment that was prescribed is necessarily used. BTS guidelines evolve over the years, so the treatment step given might not correspond exactly to the step at time of prescription. Nonetheless, the BTS 2016 guidelines were used for consistency. SABA prescriptions are an imperfect measure for asthma symptoms, as some practices may prescribe SABA as part of a patient's repeat prescription, and some symptomatic patients may only use maintenance inhalers. SABA or inhaler use is not a perfect measure for asthma symptoms, as it would only count patients who visit their GP, obtain a SABA/inhaler prescription and have their prescription recorded. Specific Read codes have been chosen by a respiratory physician to maximize the sensitivity of diagnosing asthma. Ultimately, however, we are limited by the acumen of the clinician recording the diagnosis.

Exacerbations were captured using CPRD, HES and ONS based on OCS prescriptions and hospitalisations. An exacerbation was defined as ≤300mg oral corticosteroids (OCS) (not prescribed during an annual asthma review), or an A&E visit, or an acute hospital visit of <1 day duration, an overnight hospitalisation or an asthma-related death. Exacerbations starting 14 days after the index one will be considered as part of the same exacerbation. We anticipated that this was unlikely to bias estimates of the rate ratio however, assuming missed outcomes are equally likely in each phenotype. We did not have any information on the age of onset of asthma, as this information is not available in CPRD. However, this is one of the criteria by which the asthma phenotypes were defined in the cluster analysis (6).

Problems with dichotomising eosinophil countsThere have been multiple studies that used blood eosinophil measurements as a proxy for eosinophilic inflammation, with a plethora of different thresholds in blood eosinophil counts used. In patients with eosinophilic asthma, inhaled corticosteroids form the pillar of treatment but are often insufficient in patients with severe asthma. Biological treatments have been developed in recent years, (358) but the treatment responses are heterogeneous and eosinophil counts are used to identify patients with the highest expected benefits. However, the eosinophil thresholds vary considerably between studies. The choice of cut-off seems largely driven by the positioning of drug manufacturers. The thresholds used to define asthma have varied considerably, and there is no clear consensus as to which would be the most appropriate.(350) Blood eosinophil cut-offs that have been used before are 150/µl, (359) 260/µl, (359,360) 300/µl (76,83,87,361–364), 400/µl (77,347,365–367) and 500/µl (368).

Adjusting the eosinophil threshold in the cohort study in Chapter 6 to a different count (instead of <300 cells/µL and <4% of leucocytes) would change the amount of people that could be classified into one of the predefined phenotypes. Any alteration in diagnosis of eosinophilic asthma and practice could be associated with changes in clinical outcomes of patients, as blood eosinophilia is a risk factor for future asthma exacerbations.(369,370) The association between blood eosinophil counts and exacerbation risk is likely to be continuous rather than dichotomised in asthma, as it is in COPD.(371)

Changing the diagnostic threshold of eosinophilia could also change the pharmacological management of patients if the eosinophil counts are used to inform treatment. These treatment changes could then further impact exacerbation rates. Lowering the thresholds of eosinophilia may qualify more patients for biologic treatment, in specific anti-IL5s (including mepolizumab, benralizumab and

reslizumab). ICS treatment is a crucial component of severe asthma management and not dependent on eosinophil counts according to the GINA 2018 guidelines.(1)

Progressively lowering diagnostic thresholds may result in misdiagnosis, initiation of non-cost-effective treatments and overall poorer clinical outcomes for patients. Lowering the diagnostic thresholds would also potentially increase the number of patients referred to secondary care and increase the use of secondary prevention. Potential future clinical trials' success would depend on the selection and determination of the population which define the normal reference range. We assumed eosinophil counts remains stable over time, but this assumption might not always hold although blood eosinophilia has been proven reasonably stable in the CPRD GOLD in COPD patients.(372) For any patient with an eosinophil count close to the threshold of the dichotomised count, a step-wise change does not exist between eosinophil counts with risk, as would be the case with a true dichotomised variable. Cut-offs used with biomarkers that behave as continuous variables remain arbitrary, but this does not necessarily undermine their utility though.(373)

CONCLUSION

Primary care asthma phenotypes can be identified from large electronic healthcare databases, although a large proportion could not be classified. Exacerbation frequencies are lowest in the benign phenotype and highest for the obese non-eosinophilic phenotype. Phenotyping along with knowledge of asthma treatment step could help anticipate future treatment needs but is only possible in a minority of asthma patients based on current phenotypes and primary care records.

Ethical Approval

The protocol for this research was approved by the Independent Scientific Advisory Committee (ISAC) for MHRA Database Research (ISAC protocol 17_152A), the approved protocol was made available during peer review. Generic ethical approval for observational research using the CPRD with approval from ISAC has been granted by a Health Research Authority Research Ethics Committee (East Midlands – Derby, REC reference number 05/MRE04/87). Ethical approval for this study was also obtained from the London School of Hygiene and Tropical Medicine research ethics committee.

All code lists for covariates and comorbidities are included in the appendix and on Data compass. (*http://datacompass.lshtm.ac.uk/)*

6.3 Appendix

Appendix 1: Treatment steps of the identifiable phenotypes

	Benign Asthma	Atopic Asthma	Obese non-eosinophilic	NOS all	Total
BTS step					
BTS1	4,401 (58.7%)	11,612 (20.9%)	754 (8.0%)	32,130 (26.4%)	48,897 (25.2%)
BTS2	1,186 (15.8%)	18,898 (34.1%)	2,526 (27.0%)	29,809 (24.5%)	52,419 (27.0%)
BTS3	520 (6.9%)	6,220 (11.2%)	1,164 (12.4%)	12,309 (10.1%)	20,213 (10.4%)
BTS4	978 (13.0%)	11,710 (21.1%)	2,625 (28.0%)	26,173 (21.5%)	41,486 (21.4%)
BTS5	263 (3.5%)	5,346 (9.6%)	1,831 (19.5%)	15,825 (13.0%)	23,265 (12.0%)
BTS6	3 (0.0%)	243 (0.4%)	129 (1.4%)	1,457 (1.2%)	1,832 (0.9%)
Non BTS	144 (1.9%)	1,426 (2.6%)	343 (3.7%)	3,974 (3.3%)	5,887 (3.0%)
Total	7,495 (3.9%)	55,455 (28.6%)	9,372 (4.8%)	121,1677 (62.7%)	193,999 (100%)

Table 3: Treatment steps of the identifiable phenotypes

Appendix 2: Interaction between age and phenotype

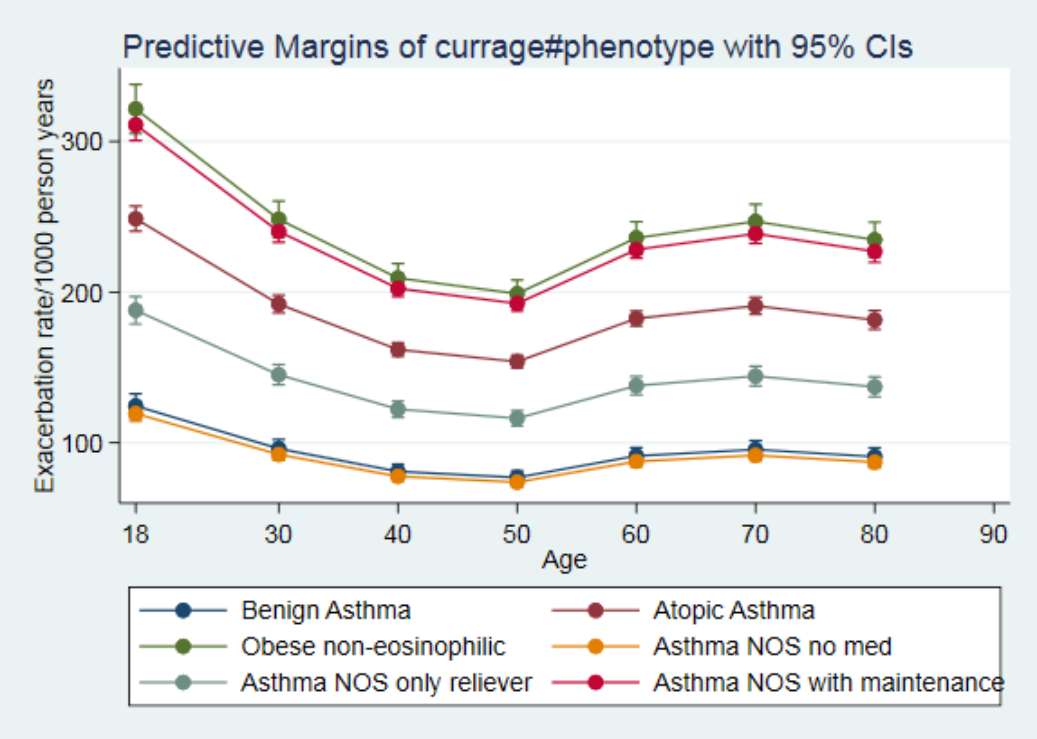


Figure 7: Interaction between age and phenotype

FIGURE legends

- Figure 1: Cluster analysis and clinical asthma phenotypes Reprinted with permission of the American Thoracic Society. Copyright © 2018 American Thoracic Society. Haldar P et al. Am J Respir Crit Care Med. 2008 Aug 1;178(3):218-224. The American Journal of Respiratory and Critical Care Medicine is an official journal of the American Thoracic Society.
- Figure 2: Cohort timeline
- Figure 3: Flowchart of study eligibility and participation
- Figure 4: BTS step by phenotype
- Figure 5: Incidence rate ratios, stratified by treatment step. Adjustment for step 6 resulted in very wide confidence intervals due to low sample size. This made the Forrest plot unreadable and is not displayed.
- Figure 6: Time to first exacerbation analysis in years, by phenotype and 95% CI
- Figure 7: Interaction between age and phenotype

TABLE legends

- Table 1: Baseline characteristics of the study population by phenotype
- Table 2: Exacerbation rates by phenotype
- Table 3: Treatment steps of the identifiable phenotypes

Research paper cover sheet

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Francis Nissen
Principal Supervisor	Ian Douglas
Thesis Title	Asthma in electronic health records: validation & phenotypes.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

**If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.*

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	Journal of asthma and allergy
Please list the paper's authors in the intended authorship order:	Francis Nissen, Ian J Douglas, Hana Mullerova, Neil Pearce, Chloe Bloom, Liam Smeeth, Jennifer K Quint
Stage of publication	**Submitted**

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I have drafted the protocol, extracted and analysed the data, and have drafted the first and the final manuscript.

Student Signature: _____ **Date:** 13/08/2018

Supervisor Signature: _____ **Date:** 14/08/2018

Chapter 7: Thesis summary and discussion

Summary

This chapter presents an overall synopsis of the results for each of the original objectives of the thesis determined at the time of the upgrading seminar. In addition, this chapter lists the major strengths and weaknesses of the thesis as a single research body. The specific strengths and limitations of each study have been discussed in their respective chapters. This chapter also provides recommendations for clinical practice and future research that follow from the results of the studies included in this thesis. The section on recommendations for future research includes an assessment on the usefulness of asthma phenotyping based on primary care electronic health records. This chapter closes with an overall conclusion to the thesis.

The focus of this PhD thesis is the validation of asthma in de-identified EHR and the study of the treatment, severity and outcomes of different asthma phenotypes in the UK. Information on asthma risk factors, treatment and outcomes were retrieved from the CPRD GOLD, Hospital Episodes Statistics and Office of National Statistics data. I undertook this project as there is no universal consensus on what constitutes asthma and asthma overlaps with many other diseases, which makes the identification of asthma patients from electronic health records (EHR) in primary care difficult. Asthma continues to carry a high morbidity and notable mortality worldwide. Distinct asthma phenotypes have previously been established based on cluster analyses in small populations, but how asthma phenotypes are related to disease outcomes was not known.

This thesis is based on a series of studies on asthma in electronic health records. The systematic review (Chapter 3) and validation study (Chapter 4) show reliable ways to identify asthma patients from EHR using positive predictive values. The systematic review examined previous studies which encompassed a validation process of asthma in different electronic healthcare databases worldwide, and the validation study developed and validated algorithms to identify asthma patients in the CPRD GOLD database. Furthermore, this thesis demonstrates the prevalence and characteristics of a concomitant diagnosis of asthma and COPD in the UK primary care population through a study based on validated asthma and COPD patients in the CPRD GOLD, as presented in Chapter 5. In Chapter 6, I examined the presence, treatment and outcomes of different asthma phenotypes in primary care using a large cohort study of asthma patients in the CPRD GOLD. This study also investigated patient characteristics, comorbidities and exacerbation rates of people with these phenotypes. This study adds to the relatively small body of research on the epidemiology of asthma phenotypes in the general population.

7.1 Synopsis of findings by research objective

This subchapter discusses and summarises the findings of each of the original research objectives of this thesis.

In order, the original objectives of this PhD thesis were:

1. Understand how past epidemiological studies have identified asthma patients in EHR through a systematic review.
2. Validate the recording of the diagnosis of asthma in the CPRD GOLD.
3. Quantify the concomitant occurrence of asthma in COPD patients and vice versa in the CPRD GOLD.
4. Identify established asthma phenotypes in the CPRD GOLD by studying risk factors and explore the variation of asthma severity (defined by treatment steps) by phenotype.
5. Examine the difference in asthma control by asthma phenotype, stratified by treatment step.

Objective 1: Understand how past epidemiological studies have validated asthma in EHR through a systematic review

The validity of asthma diagnoses in electronic health records presents a problem for asthma researchers. Most asthma symptoms are non-specific and there is no consensus on the exact clinical definition of asthma and its key outcomes, including disease severity, asthma control and exacerbations.(62,304,374,375) Both the overdiagnosis and underdiagnosis of asthma have been reported and have been cause for concern.(213,376,377) In addition, tests that suggest an asthma diagnosis such as airflow measurements or trial of treatment are often poorly recorded in electronic health records. The approaches to defining asthma in electronic health records and the

validity of these definitions are diverse and complicate the critical appraisal and comparison of results from EHR-based studies.

To understand how asthma researchers have dealt with the issue of defining and validating asthma diagnoses in electronic health records in the past, I have conducted a systematic review on this topic. The systematic review which is presented in Chapter 3, provided an overview of the methods used in the literature for validating asthma diagnosis in EHR and presented a summary of the corresponding estimates of the validation test measures. In addition, this systematic review informed potential algorithms for the validation study of asthma recording in the CPRD GOLD covered in Chapter 4.

The exploration of methods used in the literature for validating asthma diagnosis in EHR found a wide variety of approaches to validation. The variety was mostly based on the nature of the data, on the potential requirements for further studies in terms of test measures and on the availability of potential reference standards. Four types of validation methods were seen, with three using different types of reference standard. Ten studies included in the review used a manual validation as the reference standard;(267,271,278–285) in this case a clinician or researcher manually verified the electronic records with the patients' physical charts or discharge notes. Two studies used a second independent database to validate asthma diagnoses.(286,287) This reference standard depends heavily on the availability and reliability of the second database. A veritable independent database that includes data on asthma status would not be available for most primary care databases, which makes this reference standard unfeasible for validation studies in most databases. One study retrieved by the systematic review used a questionnaire as the reference standard(288) to validate a case definition of asthma. There was a fourth method of validation, face validity, which consists of checking the prevalence of asthma in the database against the known prevalence in the population. This method was not considered as exact enough, as it

would not be able to detect a difference if the net overdiagnosis and net underdiagnosis of asthma are comparable. Furthermore, this systematic review suggested that there are method of sampling records using machine learning to develop algorithms that can measure all frequently used test measures: the PPV, NPV, sensitivity and specificity.

The summary of corresponding estimates of validation test measures demonstrated that the results of the test measures are strongly dependent on the underlying study question, case definition and data source. This is apparent in the diversity of the databases of the retrieved studies. For example, the records could have included either primary care, secondary care or emergency care records; they also greatly varied in total size from data on a single health centre to data on millions of patients. All included studies were able to validate a case definition for asthma. In the ten studies using manual validation as the reference standard, each study included at least one case definition with a PPV of at least 63%, up to 100%. In the two studies using a second independent database as reference standard, the PPV's of the best performing case definitions ranged from 46% to 58%. In the last study using a questionnaire as the reference standard, the PPV of the case definition algorithm was 89%. Differing case definitions for asthma within a single data source greatly impacted the validity of a specific algorithm. As such, testing a range of case definitions when studying the validity of an asthma recording would be essential.

In the light of these results, we opted to combine the reference standard of GP questionnaires with manual validation of the questionnaires and patient data in the CPRD GOLD records to validate the recording of asthma in the CPRD GOLD.

Objective 2: Validate the recording of the diagnosis of asthma in the CPRD GOLD

The aim of the study presented in Chapter 4 was to validate the recording of asthma in the CPRD GOLD database. To do this, eight algorithms to identify asthma patients in the CPRD GOLD were constructed. These algorithms consisted of a combination of specific and non-specific asthma Read codes, evidence of reversibility testing and recorded airflow measurements. Subsequently, we requested CPRD to send questionnaires to the GPs of 880 patients which qualified for one of the algorithms (110 for each algorithm). A total of 684 questionnaires were sent out (the GPs of the remaining 196 asthma patients could not be contacted as they had recently migrated to a different system); 494 questionnaires were subsequently returned, and 475 were valid and analysed. The reference standard consisted of the review of GP questionnaires and additional materials by a respiratory physician and a study GP to test the eight algorithms. Out of the eight tested algorithms, five algorithms reported a PPV higher than 80%. The 95% confidence intervals for the PPVs overlapped, which means the difference in PPVs between these five algorithms was consistent with random chance. The algorithm with the highest PPV consisted of a combination of nonspecific asthma codes, evidence of reversibility testing and multiple asthma prescriptions within one year (PPV 90.7, 95% CI 82.8 to 98.7). The most practical algorithm, however, was the algorithm which consisted of only a specific asthma code (PPV 86.4, 95% CI 77.4 to 95.4). The additional requirements of medication prescription codes and evidence of reversibility testing did not appear to significantly increase the PPV of the algorithms. The total number of individuals who potentially could be included in a study on asthma increased almost six-fold when the algorithm did not include these requirements. As a result, the total identifiable population of people living with asthma is much larger when only using a specific asthma code for identification. In conclusion, a specific asthma Read code had a reasonably high PPV (86.1%) and was used to identify asthma patients from the CPRD GOLD in this thesis

and more widely, can be adopted by others carrying out asthma research in the CPRD GOLD.

Objective 3: Quantify the concomitant occurrence of asthma in COPD patients and vice versa in CPRD GOLD

The study included in Chapter 5 investigated the prevalence of COPD in validated asthma patients, as well as the prevalence of asthma in validated COPD patients using data from the GP questionnaires of two validation studies and patient data recorded in the CPRD GOLD. The data on the validated asthma patients were available from the validation study of asthma recording included in Chapter 4, and the data on the validated COPD patients were available from a similar validation of COPD recording in the CPRD GOLD.(297) The data included in the analysis encompassed smoking history, spirometry records and reversibility testing results, detailed GP questionnaires and supporting information including outpatient referral letters, emergency department discharge notes and radiography records. Based on this information, and assuming the validated status of each of the asthma and COPD patients identified by their validation studies held true, I was able to review whether these patients had a record of concomitant diagnosis of asthma and COPD in the CPRD GOLD and whether this was likely to be a legitimate diagnosis.

The main finding of this study indicated that more than half (52.5%) of validated COPD patients had received a diagnostic asthma Read code. When additional evidence that could support the diagnosis of asthma in these COPD patients was considered, concurrent asthma was only likely in 14.5% (95% CI: 11.2%; 18.3%) of the validated COPD patients as many had either no indication of airflow reversibility or the last asthma code was more than two years before the COPD code. The same pattern was not observed in the validated asthma patients. Only 15.1% of validated asthma patients had ever received a diagnostic COPD Read code, and a COPD

diagnosis was likely in 14.8% (95% CI: 11.3%; 19.0%) of those validated asthma patients. In conclusion, a concurrent asthma and COPD diagnosis was only likely in a minority of patients with validated COPD (14.5%) or validated asthma (14.8%), and asthma diagnoses appear to be over-recorded in people with COPD.

Objective 4: Identify established asthma phenotypes in CPRD GOLD by studying risk factors and explore the variation of asthma severity (defined by treatment steps) by phenotype

The fourth and fifth objective of this PhD thesis were addressed in the final study included in this thesis, which is presented in Chapter 6. This study evaluated the extent to which three previously suggested asthma phenotypes could be identified using data included in routinely collected electronic health records. In addition, this study also reported on the severity of asthma (defined by medication use) by phenotype, among other considerations.

The categorisation into phenotypes was based on research by Haldar et al., which used cluster analysis of multiple clinical variables to identify potential asthma phenotypes.(101) This study identified five clusters, of which three were identified in primary care. Only the three phenotypes identified in primary care (benign asthma, obese non-eosinophilic asthma and early-onset atopic asthma) were studied in Chapter 6, as primary care EHR data was the main data source for this thesis.

In the primary care asthma population, 7,495 (3.9%) were classified into the benign asthma group, 55,455 (28.6%) as atopic asthma, and 9,372 (4.8%) as obese non-eosinophilic asthma. The remaining 121,167 (62.7%) patients were included in the asthma Not Otherwise Specified (NOS) group. In the original article, the asthma NOS group was further split by presence of treatment: one group without any asthma

treatment, one group with SABA prescriptions, and one group with asthma maintenance treatment.

The variation of asthma severity was examined using medication prescriptions, based on the treatment steps described in the BTS/SIGN 2016 guidelines.(4) The classification into treatment steps expanded on earlier collaborative work.(346) The first treatment step was defined by the absence of maintenance asthma treatment (maintenance treatment does not include SABA use). The second treatment step was defined by the regular prescription of low-dose ICS, and the third treatment step included long-acting beta agonists prescriptions. The fourth treatment step was defined by a higher dose of ICS with or without additional therapies such as LABAs, theophyllines, leukotriene receptor antagonists or long-acting antimuscarinics. The fifth treatment step was defined by high-dose ICS, and the sixth treatment step by continuous or frequent use of oral corticosteroids.

For easier comparison of treatment steps between the phenotype groups, the asthma NOS group has not been subdivided by treatment in the third table of Chapter 6 (included in the appendix of Chapter 6), which presents a summary of the BTS treatment step of the patients with different phenotypes. The benign asthma phenotype had the lowest average BTS step, followed by atopic asthma and asthma NOS with the obese non-eosinophilic asthma group on the highest average BTS step, and presumably the most severe asthma.

In conclusion, we were only able to classify a minority (37.3%) into one of the predefined asthma phenotypes using stringent inclusion criteria. The asthma severity defined by BTS treatment step varied markedly among the different asthma phenotypes.

Objective 5: Examine the difference in asthma control by asthma phenotype, stratified by treatment step

The final objective, addressed in the study presented in Chapter 6, was to investigate asthma control by phenotype, and further stratify by treatment step. The study reported the following adjusted exacerbation rates per 1000 person-years: 143.2 for benign asthma, 322.1 for atopic asthma, 439.3 for obese non-eosinophilic asthma, 174.6 for asthma NOS without medication, 240.0 for asthma NOS with only SABA prescriptions, and 414.0 for asthma NOS with maintenance treatment. The exacerbation rates were adjusted for age, sex, smoking, BMI, socio-economic status anxiety, depression, COPD and GORD. Benign asthma was used as the reference group for the calculation of incidence rate ratios, as it had the lowest exacerbation rate. Incidence rate ratios (IRRs) in the fully adjusted models for asthma exacerbation were 2.28 (95% CI 2.16-2.41) for those with atopic asthma; 3.11 (95% CI 2.91-3.32) for obese non-eosinophilic asthma, 1.23 (95% CI 1.16-1.31) for asthma NOS without medication, 1.69 (95% CI 1.58-1.80) for asthma NOS with SABA, and 2.92 (95% CI 2.77-3.08) for asthma NOS with maintenance treatment. When stratified by BTS treatment step, the IRRs of all phenotypes compared with benign asthma decreased across all steps although they remained elevated.

This study expands upon the findings of previous studies on asthma exacerbations using the CPRD GOLD, including one study exploring the age variation of the general asthma population in the UK by Bloom et al., to which I contributed. (346,347,378) The findings of this study are briefly explained as the methodology of classifying asthma patients by asthma severity (defined by BTS treatment step) were shared between the two studies. This study by Bloom et al. examined the general asthma population in the UK and their exacerbation risk and characteristics by age cohort, as most of the earlier literature has focused on patients with either more severe asthma or more severe exacerbations. It was a population-based cohort study using CPRD GOLD,

ONS and HES, from 2007 to 2015 using a similar patient cohort as the study included in Chapter 6. The study population was divided into four age cohorts and their exacerbation rates were calculated using Poisson regression. The study found a total population of 424,326 patients, of whom 60% had mild asthma. Older patients over 55 years were more likely to have more severe asthma and had a higher exacerbation rate compared with the general cohort. The patients aged between 5 and 18 years were less likely to have a high treatment step and had the lowest exacerbation rates of the whole study population.

In conclusion, exacerbation frequencies were lowest in the benign phenotype and highest for the obese non-eosinophilic phenotype. Stratifying by treatment step decreased the exacerbation rate ratios of each phenotype compared with benign asthma, but remained raised. Phenotyping along with knowledge of asthma treatment step could help anticipate future treatment need but remains limited as only a minority of patients could be classified into one of the phenotypes.

7.2 Overall strengths

The specific strengths of the studies included in this thesis are discussed in their respective chapters. However, there are several strengths to the integral thesis.

Firstly, the systematic literature review informed the design of the algorithms to identify asthma cases in the CPRD GOLD. Multiple methods to determine the asthma status of potential patients from the CPRD GOLD database were considered and evaluated.

Secondly, the breadth of the data used for the conduct of this thesis is a strength of the research presented here. The CPRD GOLD includes not only information on disease diagnoses, medication prescriptions and clinical tests such as airflow measurements, but also on important life-style factors such as BMI and smoking. In particular, the

availability of information on the smoking status of potential asthma patients was vital for the conduct of the last two studies included in this thesis. In addition, multiple additional linked data sources were used throughout the thesis: the HES and ONS databases for the cohort study presented in Chapter 6, and GP questionnaires, discharge letters, radiography records and airflow measurements that were not recorded in the CPRD GOLD in the studies presented in Chapters 4 and 5.

Thirdly, the categorisation of asthma treatment in steps based on the BTS/SIGN guidelines allowed me to stratify the analysis of the final cohort study by severity, defined by treatment step.(4) As there are many different drug classes that are used in the treatment of asthma, controlling or stratifying for each of them separately would be unfeasible. Furthermore, the exposures, covariates and outcomes of the studies included in this thesis were clearly defined using code lists. The validity of many of these covariates in CPRD GOLD have been found to be high, e.g. BMI and smoking.(234,379) The code lists of the covariates and medications are included in the appendix of this thesis. In addition, the cut-offs and standards for measured continuous variables such as Body Mass Indices, eosinophil levels and airflow measurements were clearly defined and stated in the studies where appropriate.

Fourthly, as the CPRD GOLD is representative of the UK population with respect to age and sex, the findings of the cohort and validation study can be generalised to the general UK asthma population. The practices contributing to CPRD GOLD are a sample of all UK practices, but are considered representative of the UK population and there are only few patients opting out.(208,229) In addition, the relatively large sample size allows for enough power to precisely estimate asthma exacerbation rates of patients with different asthma phenotypes. The consequential power also allowed the stratified analysis by treatment step to be carried out.

Finally, the sample size of the CPRD GOLD allowed for the improvement of the precision of the estimates of the studies in this thesis. However, the accuracy of these estimates could still be affected by biases and systematic errors (which is a different issue altogether).

7.3 Overall limitations

The limitations which are specific to a single study were discussed in their respective chapters. Overall limitations are discussed below.

Data sources

(a) CPRD (Clinical Practice Research Datalink) GOLD

The major limitations of the CPRD GOLD database are: firstly, the lack of standard definitions for specific diseases and conditions; secondly, the missing information from secondary care; thirdly the variability in completeness of the data; and, fourthly, the issue that some data is not accurately captured.(208) The validation study on asthma recording and a consistent asthma definition alleviated the issue of the lack of standard definition for asthma, but this remained a limitation for the covariates used in the last two studies. The linkages to HES, ONS and questionnaires mitigated the problem of missing information that originated in secondary care. The variability in completeness of the data remained an issue for the studies presented in Chapters 5 and 6. There are some characteristics that can be used to phenotype asthma patients that were not adequately recorded in the CPRD GOLD, including allergen exposure in childhood, family history of asthma, early life infections, maternal smoking and TH2 cytokines. As a result, the phenotypes that required knowledge on one of these characteristics could not be studied. The age of asthma onset is not available in the CPRD GOLD. The study in Chapter 6 used atopic asthma as one of the established phenotypes, while the original cluster analysis identified the early onset atopic asthma phenotype. Multiple imputation was considered for smoking status and BMI values,

but was decided against as the patterns of missingness were likely to not meet the assumptions required. Furthermore, some data that might have been useful as covariates, such as over-the-counter medication prescriptions, could not be extracted from the CPRD GOLD as the datasets do not contain these variables. Finally, the CPRD GOLD contains information on whether a treatment was prescribed, rather than whether it was administered. The adherence to treatment is, therefore, difficult to measure. This is a general limitation of studying medication in EHR databases.

(b) HES (Hospital Episode Statistics)

While the linkage to HES was advantageous to obtain information from secondary care, it also brought a few limitations with it. The data quality of inpatient data (HES APC) was found to be excellent in a systematic review as 96% of primary diagnoses were correct, but the accuracy varied according to hospital.(246) In addition, HES APC data do not include specific diagnosis dates for asthma exacerbations. Furthermore, the data originating from the accidents and emergency care and outpatient departments were not used as they did not add much information.

Information bias

(a) Misclassification of asthma

The identification of asthma patients from the CPRD GOLD was based on one of the algorithms tested by the validation study presented in Chapter 4. As a result, the limitations of this study are limitations that extend to the whole thesis. The limitations of this study are briefly summarised in this paragraph. First, the reference standard of the validation study (review of GP questionnaires and additional information) was not absolute and human error remained possible. Second, the contacted GPs might have consulted the same information available in the CPRD GOLD that led to the inclusion of a patient in one of the algorithms to fill in the questionnaires. Third, GPs connected to more complicated cases might be less likely to participate as the filling in of the questionnaires would require more effort, or the asthma diagnosis might be

inconclusive. Fourth, no questionnaires of deceased people were included, which may result in survivor bias. Finally, the methods of this study did not allow one to measure the sensitivity or specificity of a recording of an asthma diagnosis. An estimate of the sensitivity would allow to predict the total number of asthma patients in the CPRD GOLD and, by extension, the UK. The reliability of the study on concomitant asthma and COPD included in Chapter 5 depended on the results of the validation study of asthma recording presented in Chapter 4 and the results of a second validation study of COPD recording.(297) Both validation studies share the same limitations as the methods were similar. In addition, a PPV of 86.4% for the most practical algorithm to identify asthma patients is considerable, but also suggests misclassification of asthma status in the remaining 13.6% when the cohort of asthma patients was constructed for the cohort study presented in Chapter 6.

(b) Misclassification of covariates

When using the CPRD for observational research, the assumption must be made that people without a recording of a distinct diagnosis do not have this condition (for example, GORD), while this might not necessarily be true. Many diseases and covariates have a high PPV in the CPRD GOLD, but the sensitivity of those diagnoses is mostly unknown.(208,229) This can lead to underestimation of the adjusted risk or rate of the outcome. For example, in the cohort study presented in Chapter 6:, if a comorbidity is more likely to be recorded in patients with a phenotype and with frequent exacerbations, adjustment for this comorbidity would lead to an underestimation of the exacerbation rate. In addition, classifying measured factors such as eosinophil levels in a dichotomous variable can lead to misclassification as these factors can fluctuate in time.

Confounding

Confounding by unmeasured factors, including those which are truly unknown and those which are not measured or recorded in the CPRD GOLD, remains possible in the cohort study included in Chapter 6. Furthermore, imperfect measures of variables lead to incomplete adjustment.

Generalisability

As the CPRD GOLD is representative of the population of the UK, the generalisability of the results included in this thesis to the UK general population is presumed to be sound.(228) There are some exclusions from the CPRD GOLD, however, including migrants and practices/individuals who have opted out of having their data available for research, that hamper the generalisability to a certain extent. In the paper included in Chapter 6, the exclusion of people without a valid eosinophil count limited the study population to those with at least one full blood count in the study period. Similarly, there was a small percentage of people who were excluded as their smoking status and BMI were not recorded or deducible. In addition, I restricted the study population to those registered at GP practices which agreed to HES linkage. There is a risk that the patient characteristics and prescribing habits of GPs differ between general practices that do allow linkage and those that do not allow linkages to HES.

Sample size

The relatively large size of the CPRD GOLD was listed earlier as a strength of this thesis, but the statistical power to detect difference in exacerbation rates after stratification by asthma severity may still be somewhat limited, as the overlapping 95% confidence intervals indicate.

Changing or conflicting guidelines

For this thesis, the BTS/SIGN 2016 guidelines for the diagnosis and management of asthma were followed.(4) These guidelines came into effect after the start of this PhD project, and superseded the earlier BTS/SIGN 2014 guidelines. One of the main differences between these two guidelines was the change from a five-step asthma management programme which included SABA prescriptions in the main steps in the 2014 guidelines to a six-step asthma management programme in the 2016 guidelines. The 2016 guidelines do not include SABA in the maintenance treatment steps, but only as a rescue treatment. In addition, there are conflicting guidelines in the UK relating to the diagnosis and management of asthma. An example of a divergence in guidelines is the use of fractured exhaled nitric oxide (FeNo) in the diagnosis of asthma. The Global Initiative for Asthma (GINA) and the BTS/SIGN 2016 guidelines decided against the use of FeNO measurement in the diagnosis of asthma, while the National Institute for Health and Care Excellence (NICE) guidelines recommend the use of this test.(3,37,380) The difference in guidelines might appear remarkable, but is mainly a result of a difference in methodology. The methodology BTS/SIGN is based on critical appraisal of the available literature, multidisciplinary and clinically led. Its main aim is to provide clinically relevant recommendations. The NICE methodology looks at both the literature and health economic modelling, advised by a multidisciplinary guideline development group; as a result, it has a slightly different focus than the BTS/SIGN guidelines.(381)

7.4 Recommendations for practice

1: Asthma may be overdiagnosed in people with COPD

The study presented in Chapter 5 of this thesis suggests that asthma is over-recorded in the electronic health records of COPD patients. There was no indication that the reverse (over-recording of COPD in asthma patients) was likely. When a patient has a

presumed concomitant diagnosis of asthma and COPD, reversibility testing can be used to verify the diagnosis of asthma. A possible reason for the excess recording of asthma in COPD patients could be that COPD is more conservatively diagnosed as it is considered a more severe disease than asthma. Another possible explanation is that a COPD patient can be diagnosed with asthma in the years before first being diagnosed with COPD, after which no further recording of asthma is made. This would suggest misdiagnosis of asthma, as the previous asthma diagnosis might be either outdated or misdiagnosed. As a result, an asthma diagnosis seems to be less reliable in COPD patients. Incorrect management can expose COPD patients without asthma to adverse effects and incur additional costs for the patient and health system, for example through unnecessary medication regimens (such as the usage of montelukast in COPD patients).

2: Asthma patients are identifiable through EHR data

Accurate coding and a clearer definition of asthma exacerbations or asthma attacks is important for both clinical care and secondary users of the data. As exacerbation frequency is important for clinical management, clinicians should be able to access information on recent asthma exacerbations easily. A standard definition of asthma exacerbation would help greatly with this.(24) The relatively high PPV for the recording of asthma in the CPRD GOLD is reassuring, and this might motivate clinicians and contributors to the CPRD data to keep up the work needed to record high-quality health data. While the general recording of asthma in CPRD GOLD is good, with a PPV of 86.4%, there is still some room for improvement. Possible ways to attain an even better PPV would be to use training or incentives for GPs to improve coding or to reduce the total number of non-specific Read codes.

3: Established asthma phenotypes are identifiable to a certain extent in primary care

Clinicians should be aware that asthma phenotypes can be identified in a sizeable minority of asthma patients in EHR using stringent inclusion criteria, and that the exacerbation risk differs between patients with different phenotypes. Routine blood tests can be useful to categorise asthma patients according to their eosinophil counts as these eosinophil counts are an important piece of the puzzle when identifying these asthma phenotypes in primary care. However, a majority of asthma patients in primary care did not fit into one of the three predefined asthma phenotypes, indicating that a phenotype-based approach to asthma management in primary care is not yet attainable for all asthma patients using the measurements currently recorded in EHR. Possible solutions are to establish phenotypes that are more easily recognisable in primary care, or add additional tools and measurements in primary care. One of the main aims of asthma phenotyping could be the benefit to patients through precision medicine and the way to do this might be more attainable using a different approach such as the treatable traits strategy, which is discussed in the following subchapter.

7.5 Recommendations for research

1: Recommendations for future validation studies of asthma recording in other databases

Identifying asthma cases in different electronic health records databases is possible with high sensitivity, specificity or positive predictive value by combining multiple data sources, or by focussing on specific test measures. Attaining high PPVs (>80%) for specific algorithms is possible using one of three possible reference standards: manual validation, comparison with a second database, or using questionnaires. The studies retrieved by the systematic review that test a range of case definitions show a

wide variation in the validity of each case definition or algorithm. This suggests that testing different case definitions may be important to obtain asthma definitions with optimal validity for the pursued study question.

2: Asthma patients are identifiable from the CPRD GOLD

Future epidemiological studies using the CPRD GOLD on asthma should use a validated definition of asthma (included in the appendix). This validated definition will be useful for studies using asthma as an exposure, covariate or outcome. The algorithm consisting of only a specific asthma code algorithm alone is the most practical approach to identify patients with asthma in CPRD GOLD (PPV=0.86; 95% CI 0.77-0.95), as asthma diagnoses were confirmed in a high percentage of patients with specific asthma codes. This suggests that epidemiological studies on asthma using the CPRD GOLD can be conducted with reasonably high validity. The findings of this validation study can also help inform service planning and audits involving asthma patients. The inclusion of airflow measurements or asthma medication in the algorithm to identify asthma patients in EHR did not clearly improve accuracy in the asthma recording validation study and severely restricted the total identifiable population. Recently, a new primary care database managed by the CPRD has become available: CPRD Aurum.(226) This database contains routinely collected data from GP practices using EMIS-Web (Egton Medical Information Systems electronic patient record system). Currently, more registered practices contribute to CPRD Aurum than to CPRD GOLD. As the data structure and clinical coding between CPRD Aurum and CPRD GOLD differs, the validity of asthma recording in CPRD GOLD is not directly extendable to CPRD Aurum. In the future, a validation study on asthma recording in CPRD Aurum may be indicated.

3: Identifying patients with concomitant asthma and COPD in EHR

The findings from the study presented in Chapter 5 have implications on further research into concomitant asthma and COPD. Identifying potential concomitant asthma and COPD using electronic health records should be done cautiously. The prevalence of concomitant asthma and COPD in validated asthma and COPD patients were both around 15%. However, around half of all validated COPD patients had a Read code for asthma recorded, suggesting over-recording of asthma in COPD patients. The prevalence of COPD in validated asthma patients is similar to the percentage of asthma patients with a COPD code recorded. If the algorithm to identify both diseases consists of only a single code for each algorithm, the prevalence of concomitant diagnosis of asthma and COPD is likely to be overestimated.

4: Asthma phenotyping using primary care EHR

While it is possible to identify patients with different asthma phenotypes in primary care, most patients did not fit into one of three pre-specified asthma phenotypes that were identified in a primary care study using cluster analysis on a limited number of patients. The 62.7% of patients that were not included in one of the three predefined asthma phenotypes (benign asthma, obese non-eosinophilic asthma, and atopic asthma) were included in the asthma NOS group. Some of the patients included in this asthma NOS group could belong to one of the three predefined asthma phenotypes but were unable to be classified due to unrecorded or missing data, while some other patients could belong to other, yet unspecified phenotypes. Easier to define phenotypes or more complete records, including full blood counts, could help with the phenotyping of patient records in electronic health records.

The rationale for the conduct of the cohort study included in Chapter 6 was that the classification of asthma into phenotypes could help tailor the treatment for asthma patients. Asthma diagnosis is still based on clinical presentation and associated lung

function, which are both non-specific.(375) As a result, asthma is often treated similarly to COPD and not optimally.(382) Precision medicine aims to define treatments targeted to the needs of specific patients based on genetic, biomarker, phenotypic, or psychosocial characteristics that differentiate these patients from others with a similar presentation.(351) However, many methods to classify asthma patients into phenotypes have been insufficient to identify patients who are likely to benefit from a specific treatment. In theory, phenotyping could help to inform our understanding of the underlying asthma endotypes, but in practice this remains problematic as many phenotypic characteristics can be caused by several different disease mechanisms.(24,383,384)

While there has been progress on the endotyping of asthma (identifying asthma subgroups that share pathophysiologic processes),(383,385) how much of this explains the phenotypic heterogeneity of asthma remains unclear. Endotyping of asthma would help in the conduct of clinical trials on asthma treatment as many outcomes have been biased by the adoption of inclusion and exclusion criteria which can fail to address whether a particular asthma medication works equally well for all patients with the asthma syndrome.(386)

Of note is that the phenotyping of asthma patients in electronic health records should not be the ultimate goal in and of itself. If the phenotyping of patients in EHR does not help promote precision medicine or tailoring of asthma treatment, other strategies that do provide a clearer theoretical base to this end will be of greater concern.

For example, the treatable traits strategy has recently been proposed(24,351) and might represent a better conceptual framework towards precision medicine than phenotyping using primary care electronic records at this stage.(342,352,353) This strategy focuses asthma management on traits that are identifiable and treatable, such as eosinophilia or airflow limitation. This reductionist approach would use asthma

only as a descriptive label for a collection of symptoms without assuming any specific pathophysiology. Pavord et al. suggest the following treatable traits in order of decreasing importance and recognisability (based on earlier work by Hargreave et al.): Airflow limitation, airway inflammation, airway infection/impaired airway defences, and altered cough reflex sensitivity/efficacy.(24,155) The first treatable trait, airflow limitation, can be assessed using airflow measurements, while the risk of the second treatable trait, airway inflammation, can be assessed using biomarkers of eosinophilic airway inflammation such as sputum/blood eosinophils or FeNO measurements.

7.6 Future work

A possible next step in this field would be a de novo cluster analysis of asthma patients in the CPRD GOLD. The next paragraph outlines possible methodology of this cluster analysis. Previous cluster analyses on primary care asthma populations have been conducted in smaller populations.(387,388)

Study population

The study period could be April 2007 to the last extraction date. Individuals would be eligible for inclusion from 18 year onwards, they had been registered in a UTS primary care practice and had a previously validated asthma code, without an upper age limit. Patients would enter the study cohort when they met the inclusion criteria, and the index date for each participant would be the first asthma diagnosis recording while the patient is eligible. Patients would leave the cohort on the earliest of leaving the primary care practice, death or last practice data collection.

Exposure definition

The asthma population would consist of patients with an asthma diagnosis code validated in the fourth chapter of this thesis. Patients with a diagnosis of COPD would be excluded, as asthma diagnoses in COPD patients can be less reliable in CPRD (fifth chapter of this thesis).

Covariates

Variable selection would be chosen based on clinical value, recording in the CPRD, and avoid variables that could introduce issues with multicollinearity.

The covariates for generating the clusters would be defined from the CPRD GOLD, and include demographic characteristics, symptoms, atopy, body mass index, smoking status, gastro-oesophageal reflux disease, chronic rhinosinusitis, anxiety, depression, lung function defined by FEV1 % predicted and sex. Asthma therapy would be classified as defined by the BTS 2019 guidelines.

Outcomes

For the evaluation and interpretation of clusters, we would use an a priori set of clinical outcomes, based on the rate of moderate and severe asthma exacerbations, socio-economic status and mortality. Exacerbations would be measured in the same way as the study in the sixth chapter of this PhD thesis and SES will be divided into quintiles.

Statistical methods

We would split the data into a training and test set at random and use the training dataset to perform statistical analyses. Multiple correspondence analysis would be used on all covariates, and the numerical covariates will be transformed to categorical variables. The data clustering (unsupervised learning) are a set of techniques to identify subsets by grouping by similarity. K-means and hierarchical clustering algorithms would be used on complete cases. Hierarchical cluster analysis using Ward's method would be used to estimate the likely number of clusters in the population. K-clustering would be used as the main clustering technique. The stability and repeatability of results would be ensured by repeating the k-means clustering analysis at different starting points (potentially with different statistical software).

The Euclidian distance would be used for measurement in both methods. The silhouette coefficient measures clustering performance (both cohesion and separation). Another option is to use the training dataset to generate the clusters, and test the reproducibility of the resulting clusters on the testing dataset. Clusters are considered stable if they yield a similarity greater than 75%. Descriptive statistics would be used to describe and compare the separate cluster populations.

Use by general practitioners in practice

The classification of primary patients into different phenotypes could aid in identifying those patients who are most at risk of severe exacerbations. In addition, it could enable GPs to guide the treatment of patients with asthma, allocate asthma patients to secondary care if needed and adjust the frequency of asthma consultations. Differentiating asthma patients by phenotype in primary care would also allow the healthcare system to more efficiently allocate resources, by prioritising the phenotype group where an intervention can be the most beneficial. The use of resources in secondary and tertiary care for patients in phenotypes with a better prognosis could be limited. If this cluster analysis would produce meaningful phenotypes in primary care and a phenotype-based approach to asthma management would be implemented in primary care based on already recorded information, a further study would still be needed to assess the implications of the implementation.

Handling of multiple exacerbations per patients

Clinical interest in asthma lies in both the final outcome (death/survival time) and the dynamics of disease itself, as exacerbations lead to a state that is not always fully reversible. A standard regression model (Logistic, Cox or Poisson) may not be appropriate as the exacerbations are not independent of one another. Approaches to overcome this include marginal and multi-state models. In most marginal models the assumption is that all events are identical. Multi-state models can differentiate between different types of events, and are a stochastic process where a patients occupies one of several states at any time.

Regression analysis would be able to estimate the probability of readmission due to asthma. The timing of readmissions would be much harder to assess with most regression methodologies. Multistate models could help with modelling the process of readmissions. They are statistical tools in which a patient occupies one of several possible states. There are multiple standard structures, where the alternating model or recurrent events would provide the best fit. A simple multi-state model would include a patient during an exacerbation and the following 2 weeks (State 1), not having an exacerbation (State 2), and death or loss-to-follow up (State 3). A more complicated multi-state model could include more different states to include the number of exacerbations and the readmission, with first, second, 4th exacerbation and first, second, nth post-exacerbation period.

7.7 Overall conclusions

It is increasingly clear that electronic health records can play an important role in the future of asthma research. This thesis discusses approaches to identify asthma patients from these records and finds that the asthma status of patient can be established in electronic health records with a reasonably high reliability. This thesis includes a list of studies which validated asthma recording and their respective test measures obtained through a systematic review and the results of a validation study of asthma recording in the CPRD GOLD database.

As asthma and COPD share many characteristics and symptoms, the potential to differentiate between both diseases is of great concern to researchers aspiring to employ electronic health records to study these diseases. Concomitant asthma and COPD was present in 14.8% of validated asthma patients and in 14.5% of validated COPD patients. Asthma diagnoses may be less reliable in COPD patients however, as close to half of all COPD patients had ever received an asthma diagnostic Read code.

Previously suggested asthma phenotypes can, to a certain extent be distinguished in primary care electronic health records, but the majority of patients could not be classified into three predefined phenotypes with stringent inclusion criteria. There are two possible ways to improve the categorisation of primary care asthma patients into phenotypes: either by thorough recording of key variables in EHR, or by constructing new phenotypes that would be easier to identify in EHR. Many clinical variables used to phenotype patients in clinical trials are simply not available in routinely recorded electronic health records. The treatable traits strategy may be more likely to succeed in providing precision medicine for asthma patients in primary care than a phenotype-based approach.

REFERENCES

1. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention [Internet]. 2018. Available from: www.ginasthma.org
2. WHO. Asthma fact sheet [Internet]. 2017 [cited 2018 May 14]. Available from: http://www.who.int/en/news-room/fact-sheets/detail/asthma
3. GINA. Global Strategy for Asthma Management and Prevention 2016. Available from: www.ginasthma.org. 2016.
4. BTS/SIGN. British guideline on the management of asthma [Internet]. 2016. Available from: https://www.brit-thoracic.org.uk/document-library/clinical-information/asthma/btssign-asthma-guideline-2016/
5. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. J Allergy Clin Immunol. 2007 Nov;120(5 Suppl):S94-138.
6. Adams F. The genuine works of Hippocrates. Vol. 17. Sydenham society; 1849.
7. Keeney EL. The history of asthma from Hippocrates to Meltzer. J Allergy. 1964;35(3):215–26.
8. Marketos SG, Ballas CN. Bronchial asthma in the medical literature of Greek antiquity. J Asthma. 1982;19(4):263–9.
9. Karamanou M, Androutsos G. Aretaeus of Cappadocia and the first clinical description of asthma. Am J Respir Crit Care Med. 2011;184(12):1420–1.
10. Oberhelman SM. Aretaeus of Cappadocia: the pneumatic physician of the first century AD. Aufstieg und Niedergang der Römischen Welt II Berlin-New York Gruyter. 1997;941–96.
11. Rosner F. Moses Maimonides' treatise on asthma. Thorax. 1981;36(4):245–51.
12. Ramazzini B. De morbis artificum diatriba... apud J. Corona; 1743.
13. Salter HH. On asthma: its pathology and treatment. Churchill; 1868.
14. Owen R, Simon J, Bell T, Huxley TH. Henry Hyde Salter (1823-71): biographical sketch. Thorax. 1985;40:887–8.
15. Blackley CH. Experimental researches on the causes and nature of catarrhus aestivus (hay-fever or hay-asthma). Baillière, Tindall & Cox; 1873.
16. Cooke RA, Vander Veer A. Human sensitization. J Immunol. 1916;1(3):201–305.
17. Wolff-Eisner A. Das Heufieber, sein Wesen und seine Behandlung: mit 2 grossen Tabellen. Lehmann; 1906.
18. Rackemann FM. Intrinsic asthma. J Allergy. 1940;11(2):147–62.
19. Rackemann FM. A working classification of asthma. Am J Med. 1947;3(5):601–6.
20. Rackemann FM. Studies in Asthma: IX. The Incidence and Severity of Asthma According to the Season of the Year. Trans Am Climatol Clin Assoc. 1928;44:158.
21. Spriggs EA. The history of spirometry. Br J Dis Chest. 1978;72:165–80.
22. Tiffeneau R, Drutel P. Etude des facteurs alvéolaires et bronchiques de la ventilation pulmonaire. J Fr Med Chir Thorac. 1949;3:401–31.
23. Brown HM. Treatment of chronic asthma with prednisolone significance of eosinophils in the sputum. Lancet. 1958;272(7059):1245–7.
24. Pavord ID, Beasley R, Agusti A, Anderson GP, Bel E, Brusselle G, et al. After asthma: redefining airways diseases. Lancet. 2018;391(10118):350–400.
25. Grainger J, Woodman K, Pearce N, Crane J, Burgess C, Keane A, et al. Prescribed

fenoterol and death from asthma in New Zealand, 1981-7: a further case-control study. Thorax. 1991;46(2):105–11.
26. Crane J, Flatt A, Jackson R, Ball M, Pearce N, Burgess C, et al. Prescribed fenoterol and death from asthma in New Zealand, 1981-83; Case-control study. Lancet. 1989;333(8644):917–22.
27. Suissa S, Ernst P, Boivin JF, Horwitz RI, Habbick B, Cockroft D, et al. A cohort analysis of excess mortality in asthma and the use of inhaled beta-agonists. Am J Respir Crit Care Med. 1994;149(3):604–10.
28. Ebmeier S, Thayabaran D, Braithwaite I, Bénamara C, Weatherall M, Beasley R. Trends in international asthma mortality: analysis of data from the WHO Mortality Database from 46 countries (1993–2012). Lancet. 2017;390(10098):935–45.
29. Greening AP, Ind PW, Northfield M, Shaw G. Added salmeterol versus higher-dose corticosteroid in asthma patients with symptoms on existing inhaled corticosteroid. Lancet. 1994;344(8917):219–24.
30. Pauwels RA, Löfdahl C-G, Postma DS, Tattersfield AE, O'Byrne P, Barnes PJ, et al. Effect of inhaled formoterol and budesonide on exacerbations of asthma. N Engl J Med. 1997;337(20):1405–11.
31. Green RH. No Title. Am J Respir Crit Care Med.
32. Green RH, Brightling CE, McKenna S, Hargadon B, Neale N, Parker D, et al. Comparison of asthma treatment given in addition to inhaled corticosteroids on airway inflammation and responsiveness. Eur Respir J. 2006;27(6):1144–51.
33. Leckie MJ, ten Brinke A, Khan J, Diamant Z, O'connor BJ, Walls CM, et al. Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. Lancet. 2000;356(9248):2144–8.
34. Fingleton J, Weatherall M, Beasley R. Bronchodilator responsiveness: interpret with caution. BMJ Publishing Group Ltd; 2012.
35. Joos GF. [Indirect bronchial provocation tests: a new, specific group of bronchial provocation tests for asthma]. Verh K Acad Geneeskd Belg. 2005;67(3):153–67.
36. Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J. 2013;erj02020-2013.
37. NICE. National Institute for Health and Care Excellence. Asthma: diagnosis, monitoring and chronic asthma management. Nice Guidel 80 [Internet]. 2017; Available from: https://www.nice.org.uk/guidance/ng80/chapter/Recommendations#diagnosing-asthma-in-young-children
38. Pavord ID, Bafadhel M. Exhaled nitric oxide and blood eosinophilia: independent markers of preventable risk. J Allergy Clin Immunol. 2013;132(4):828–9.
39. Wilson SJ, Ward JA, Sousa AR, Corfield J, Bansal AT, De Meulder B, et al. Severe asthma exists despite suppressed tissue inflammation: findings of the U-BIOPRED study. Eur Respir J. 2016 Nov;48(5):1307–19.
40. Loymans RJB, Gemperli A, Cohen J, Rubinstein SM, Sterk PJ, Reddel HK, et al. Comparative effectiveness of long term drug treatment strategies to prevent asthma exacerbations: network meta-analysis. BMJ. 2014 May;348:g3009.
41. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet

Respir Med. 2017;5(9):691–706.
42. Asher MI, Montefort S, Björkstén B, Lai CKW, Strachan DP, Weiland SK, et al. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC Phases One and Three repeat multicountry cross-sectional surveys. Lancet. 2006;368(9537):733–43.
43. Beasley R. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. Lancet. 1998;351(9111):1225–32.
44. Burney UK, Chinn S, Luczynska C, Jarvis D, Neukirch F, Pin I, et al. Variations in the prevalence of respiratory symptoms, selfreported asthma attacks, and use of asthma medication in the European Community Respiratory Health Survey (ECRHS). Eur Respir J. 1996;9:687–95.
45. Leynaert B, Sunyer J, Garcia-Esteban R, Svanes C, Jarvis D, Cerveri I, et al. Gender differences in prevalence, diagnosis and incidence of allergic and non-allergic asthma: a population-based cohort. Thorax. 2012 Jul;67(7):625–31.
46. Raghavan D, Jain R. Increasing awareness of sex differences in airway diseases. Respirology. 2016 Apr;21(3):449–59.
47. Cabieses B, Uphoff E, Pinart M, Anto JM, Wright J. A systematic review on the development of asthma and allergic diseases in relation to international immigration: the leading role of the environment confirmed. PLoS One. 2014;9(8):e105347.
48. Hoskins G, McCowan C, Neville RG, Thomas GE, Smith B, Silverman S. Risk factors and costs associated with an asthma attack. Thorax. 2000;55(1):19–24.
49. Lloyd A, Price D, Brown R. The impact of asthma exacerbations on health-related quality of life in moderate to severe asthma patients in the UK. Prim Care Respir J. 2007;16(1):22–7.
50. Williams SA, Wagner S, Kannan H, Bolge SC. The association between asthma control and health care utilization, work productivity loss and health-related quality of life. J Occup Env Med. 2009;51(7):780–5.
51. Asthma UK. Asthma UK | Asthma facts and statistics [Internet]. Asthma Uk. 2015 [cited 2018 May 14]. Available from: https://www.asthma.org.uk/about/media/facts-and-statistics/
52. Mukherjee M, Gupta R, Farr A, Heaven M, Stoddart A, Nwaru BI, et al. Estimating the incidence, prevalence and true cost of asthma in the UK: secondary analysis of national stand-alone and linked databases in England, Northern Ireland, Scotland and Wales-a study protocol. BMJ Open. 2014 Nov;4(11):e006647.
53. Bayes HK, Thomson NC. Acute severe asthma in adults. Medicine (Baltimore) [Internet]. 44(5):297–300. Available from: http://dx.doi.org/10.1016/j.mpmed.2016.02.012
54. Gonzalez-Barcala FJ, Calvo-Alvarez U, Garcia-Sanz MT, Bourdin A, Pose-Reino A, Carreira JM, et al. Characteristics and prognosis of near-fatal asthma exacerbations. Am J Med Sci. 2015;350(2):98–102.
55. Chanez P, Wenzel SE, Anderson GP, Anto JM, Bel EH, Boulet LP, et al. Severe asthma in adults: what are the important questions? J Allergy Clin Immunol. 2007;119(6):1337–48.
56. ten Brinke A, Sterk PJ, Masclee AA, Spinhoven P, Schmidt JT, Zwinderman AH, et al. Risk factors of frequent exacerbations in difficult-to-treat asthma. Eur Respir J. 2005;26(5):812–8.
57. Martinez FD. Genes, environments, development and asthma: a reappraisal. Eur

Respir J. 2007;29(1):179–84.
58. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010 Sep;363(13):1211–21.
59. Kelly FJ, Fussell JC. Air pollution and airway disease. Clin Exp Allergy. 2011;41(8):1059–71.
60. Ramsey CD, Celedon JC. The hygiene hypothesis and asthma. Curr Opin Pulm Med. 2005;11(1):14–20.
61. Brooks C, Pearce N, Douwes J. The hygiene hypothesis in allergy and asthma: an update. Curr Opin Allergy Clin Immunol. 2013;13(1):70–7.
62. Papi A, Brightling C, Pedersen SE, Reddel HK. Asthma. Lancet (London, England). 2018 Feb;391(10122):783–800.
63. Boyd M, Lasserson TJ, McKean MC, Gibson PG, Ducharme FM, Haby M. Interventions for educating children who are at risk of asthma-related emergency department attendance. Cochrane database Syst Rev. 2009 Apr;(2):CD001290.
64. Melani AS, Bonavia M, Cilenti V, Cinti C, Lodi M, Martucci P, et al. Inhaler mishandling remains common in real life and is associated with reduced disease control. Respir Med. 2011 Jun;105(6):930–8.
65. Normansell R, Kew KM, Bridgman AL. Sublingual immunotherapy for asthma. Cochrane Database Syst Rev. 2014;(9).
66. Boyle RJ, Pedroletti C, Wickman M, Bjermer L, Valovirta E, Dahl R, et al. Nocturnal temperature controlled laminar airflow for treating atopic asthma: a randomised controlled trial. Thorax. 2011;thoraxjnl-2011.
67. Shay LA, Lafata JE. Where is the evidence? A systematic review of shared decision making and patient outcomes. Med Decis Making. 2015 Jan;35(1):114–31.
68. Chauhan BF, Ducharme FM. Anti-leukotriene agents compared to inhaled corticosteroids in the management of recurrent and/or chronic asthma in adults and children. Cochrane Libr. 2012;
69. Dahl R, Larsen BB, Venge P. Effect of long-term treatment with inhaled budesonide or theophylline on lung function, airway reactivity and asthma symptoms. Respir Med. 2002;96(6):432–8.
70. Rivington RN, Boulet L-P, Côté J, Kreisman H, Small DI, Alexander M, et al. Efficacy of Uniphyl, salbutamol, and their combination in asthmatic patients on high-dose inhaled steroids. Am J Respir Crit Care Med. 1995;151(2):325–32.
71. Guevara JP, Ducharme FM, Keren R, Nihtianova S, Zorc J. Inhaled corticosteroids versus sodium cromoglycate in children and adults with asthma. Cochrane Libr. 2006;
72. Kew KM, Dahri K. Long-acting muscarinic antagonists (LAMA) added to combination long-acting beta2-agonists and inhaled corticosteroids (LABA/ICS) versus LABA/ICS for adults with asthma. Cochrane Libr. 2016;
73. Normansell R, Walker S, Milan SJ, Walters EH, Nair P. Omalizumab for asthma in adults and children. Cochrane Libr. 2014;
74. Rodrigo GJ, Neffen H. Systematic review on the use of omalizumab for the treatment of asthmatic children and adolescents. Pediatr Allergy Immunol. 2015;26(6):551–6.
75. Haldar P, Brightling CE, Hargadon B, Gupta S, Monteiro W, Sousa A, et al. Mepolizumab and exacerbations of refractory eosinophilic asthma. N Engl J Med. 2009;360(10):973–84.
76. Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, et al. Mepolizumab

for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. Lancet. 2012;380(9842):651–9.
77. Castro M, Zangrilli J, Wechsler ME, Bateman ED, Brusselle GG, Bardin P, et al. Reslizumab for inadequately controlled asthma with elevated blood eosinophil counts: results from two multicentre, parallel, double-blind, randomised, placebo-controlled, phase 3 trials. Lancet Respir Med. 2015;3(5):355–66.
78. Nair P, Wenzel S, Rabe KF, Bourdin A, Lugogo NL, Kuna P, et al. Oral glucocorticoid–sparing effect of benralizumab in severe asthma. N Engl J Med. 2017;376(25):2448–58.
79. Lefebvre P, Duh MS, Lafeuille M-H, Gozalo L, Desai U, Robitaille M-N, et al. Acute and chronic systemic corticosteroid–related complications in patients with severe asthma. J Allergy Clin Immunol. 2015;136(6):1488–95.
80. Walsh LJ, Wong CA, Oborne J, Cooper S, Lewis SA, Pringle M, et al. Adverse effects of oral corticosteroids in relation to dose in patients with lung disease. Thorax. 2001;56(4):279–84.
81. Brusselle GG, VanderStichele C, Jordens P, Deman R, Slabbynck H, Ringoet V, et al. Azithromycin for prevention of exacerbations in severe asthma (AZISAST): a multicentre randomised double-blind placebo-controlled trial. Thorax. 2013;68(4):322–9.
82. Agbetile J, Bourne M, Fairs A, Hargadon B, Desai D, Broad C, et al. Effectiveness of voriconazole in the treatment of Aspergillus fumigatus–associated asthma (EVITA3 study). J Allergy Clin Immunol. 2014;134(1):33–9.
83. Bleecker ER, FitzGerald JM, Chanez P, Papi A, Weinstein SF, Barker P, et al. Efficacy and safety of benralizumab for patients with severe asthma uncontrolled with high-dosage inhaled corticosteroids and long-acting β2-agonists (SIROCCO): a randomised, multicentre, placebo-controlled phase 3 trial. Lancet. 2016;388(10056):2115–27.
84. Gonem S, Berair R, Singapuri A, Hartley R, Laurencin MFM, Bacher G, et al. Fevipiprant, a prostaglandin D2 receptor 2 antagonist, in patients with persistent eosinophilic asthma: a single-centre, randomised, double-blind, parallel-group, placebo-controlled trial. Lancet Respir Med. 2016;4(9):699–707.
85. Hanania NA, Korenblat P, Chapman KR, Bateman ED, Kopecky P, Paggiaro P, et al. Efficacy and safety of lebrikizumab in patients with uncontrolled asthma (LAVOLTA I and LAVOLTA II): replicate, phase 3, randomised, double-blind, placebo-controlled trials. Lancet Respir Med. 2016;4(10):781–96.
86. Stinson SE, Amrani Y, Brightling CE. D prostanoid receptor 2 (chemoattractant receptor–homologous molecule expressed on TH2 cells) protein expression in asthmatic patients and its effects on bronchial epithelial cells. J Allergy Clin Immunol. 2015;135(2):395–406.
87. Wenzel S, Castro M, Corren J, Maspero J, Wang L, Zhang B, et al. Dupilumab efficacy and safety in adults with uncontrolled persistent asthma despite use of medium-to-high-dose inhaled corticosteroids plus a long-acting β2 agonist: a randomised double-blind placebo-controlled pivotal phase 2b dose-ranging trial. Lancet. 2016;388(10039):31–44.
88. Dunican EM, Fahy J V. The Role of Type 2 Inflammation in the Pathogenesis of Asthma Exacerbations. Ann Am Thorac Soc. 2015 Nov;12 Suppl 2:S144-9.
89. Dusser D, Montani D, Chanez P, de Blic J, Delacourt C, Deschildre A, et al. Mild

asthma: an expert review on epidemiology, clinical characteristics and treatment recommendations. Allergy. 2007 Jun;62(6):591–604.
90. Nicholson KG, Kent J, Ireland DC. Respiratory viruses and exacerbations of asthma in adults. BMJ. 1993 Oct;307(6910):982–6.
91. Pauwels RA, Lofdahl CG, Postma DS, Tattersfield AE, O'Byrne P, Barnes PJ, et al. for the formoterol and corticosteroids establishing therapy (FACET) international study group. Eff inhaled formoterol budesonide exacerbations asthma N Engl J Med. 1997;337:1405–11.
92. Reddel HK, Taylor DR, Bateman ED, Boulet L-P, Boushey HA, Busse WW, et al. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. Am J Respir Crit Care Med. 2009;180(1):59–99.
93. Weiner ESC, Simpson JA, Oxford University P. The Oxford English dictionary. Oxford; Oxford; Oxford University Press: Clarendon Press ; Melbourne ; 2004.
94. Moore WC, Fitzpatrick AM, Li X, Hastie AT, Li H, Meyers DA, et al. Clinical heterogeneity in the severe asthma research program. Ann Am Thorac Soc. 2013;10 Suppl:S118-24.
95. Haldar P, Pavord ID. Noneosinophilic asthma: a distinct clinical and pathologic phenotype. J Allergy Clin Immunol. 2007;119(5):1043–52.
96. Tan DJ, Walters EH, Perret JL, Lodge CJ, Lowe AJ, Matheson MC, et al. Age-of-asthma onset as a determinant of different asthma phenotypes in adults: a systematic review and meta-analysis of the literature. Expert Rev Respir Med. 2015;9(1):109–23.
97. Koczulla AR, Vogelmeier CF, Garn H, Renz H. New concepts in asthma: clinical phenotypes and pathophysiological mechanisms. Drug Discov Today. 2017;22(2):388–96.
98. Spencer LA, Weller PF. Eosinophils and Th2 immunity: contemporary insights. Immunol Cell Biol. 2010;88(3):250–6.
99. Brooks CR, van Dalen CJ, Zacharasiewicz A, Simpson JL, Harper JL, Le Gros G, et al. Absence of airway inflammation in a large proportion of adolescents with asthma. Respirology. 2016;21(3):460–6.
100. Jones AC, Bosco A. Using Network Analysis to Understand Severe Asthma Phenotypes. Am J Respir Crit Care Med. 2017;195(11):1409–11.
101. Haldar P, Pavord ID, Shaw DE, Berry MA, Thomas M, Brightling CE, et al. Cluster analysis and clinical asthma phenotypes. Am J Respir Crit Care Med. 2008;178(3):218–24.
102. Boudier A, Curjuric I, Basagana X, Hazgui H, Anto JM, Bousquet J, et al. Ten-year follow-up of cluster-based asthma phenotypes in adults. A pooled analysis of three cohorts. Am J Respir Crit Care Med. 2013;188(5):550–60.
103. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li H, Li X, et al. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med. 2010;181(4):315–23.
104. Jo KW, Ra SW, Chae EJ, Seo JB, Kim NK, Lee JH, et al. Three phenotypes of obstructive lung disease in the elderly. Int J Tuberc Lung Dis. 2010;14(11):1481–8.
105. Kim TB, Jang AS, Kwon HS, Park JS, Chang YS, Cho SH, et al. Identification of asthma clusters in two independent Korean adult asthma cohorts. Eur Respir J. 2013;41(6):1308–14.
106. Loza MJ, Djukanovic R, Chung KF, Horowitz D, Ma K, Branigan P, et al. Validated

and longitudinally stable asthma phenotypes based on cluster analysis of the ADEPT study. Respir Res. 2016;17(1):165.
107. Heaney LG, Djukanovic R, Woodcock A, Walker S, Matthews JG, Pavord ID, et al. Research in progress: Medical Research Council United Kingdom Refractory Asthma Stratification Programme (RASP-UK). Thorax. 2016;71(2):187–9.
108. Ilmarinen P, Tuomisto LE, Niemela O, Tommola M, Haanpaa J, Kankaanranta H. Cluster Analysis on Longitudinal Data of Patients with Adult-Onset Asthma. J Allergy Clin Immunol Pr. 2017;5(4):967–978.e3.
109. Wardlaw AJ, Silverman M, Siva R, Pavord ID, Green R. Multi-dimensional phenotyping: towards a new taxonomy for airway disease. Clin Exp Allergy. 2005;35(10):1254–62.
110. Garcia-Aymerich J, Gomez FP, Benet M, Farrero E, Basagana X, Gayete A, et al. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. Thorax. 2011;66(5):430–7.
111. Kostikas K, Clemens A, Patalano F. The asthma–COPD overlap syndrome: do we really need another syndrome in the already complex matrix of airway disease? Int J Chron Obs Pulmon Dis. 2016;11:1297–306.
112. Bonten TN, Kasteleyn MJ, de Mutsert R, Hiemstra PS, Rosendaal FR, Chavannes NH, et al. Defining asthma-COPD overlap syndrome: a population-based study. Eur Respir J. 2017;49(5).
113. Chung KF. Asthma phenotyping: a necessity for improved therapeutic precision and new targeted therapies. J Intern Med. 2016;279(2):192–204.
114. Green RH, Pavord I. Stability of inflammatory phenotypes in asthma. Thorax. 2012;67(8):665–7.
115. Pavord ID. Eosinophilic phenotypes of airway disease. Ann Am Thorac Soc. 2013;10 Suppl:S143-9.
116. Haughney J, Morice A, Blyth KG, Lee AJ, Coutts A, McKnight E, et al. A retrospective cohort study in severe asthma describing commonly measured biomarkers: Eosinophil count and IgE levels. Respir Med. 2018;134:117–23.
117. Weatherburn CJ, Guthrie B, Mercer SW, Morales DR. Comorbidities in adults with asthma: Population-based cross-sectional analysis of 1.4 million adults in Scotland. Clin Exp Allergy. 2017;47(10):1246–52.
118. Burrows B, Martinez FD, Halonen M, Barbee RA, Cline MG. Association of asthma with serum IgE levels and skin-test reactivity to allergens. N Engl J Med. 1989;320(5):271–7.
119. Pearce N, Pekkanen J, Beasley R. How much asthma is really attributable to atopy? Thorax. 1999 Mar;54(3):268–72.
120. Del Giacco SR, Bakirtas A, Bel E, Custovic A, Diamant Z, Hamelmann E, et al. Allergy in severe asthma. Allergy. 2017 Feb;72(2):207–20.
121. Shaaban R, Zureik M, Soussan D, Neukirch C, Heinrich J, Sunyer J, et al. Rhinitis and onset of asthma: a longitudinal population-based study. Lancet. 2008;372(9643):1049–57.
122. Jeffery PK, Haahtela T. Allergic rhinitis and asthma: inflammation in a one-airway condition. BMC Pulm Med. 2006;6 Suppl 1:S5.
123. Abramson M, Matheson M, Wharton C, Sim M, Walters EH. Prevalence of respiratory symptoms related to chronic obstructive pulmonary disease and asthma among middle aged and older adults. Respirology. 2002;7(4):325–31.

124. Harding SM. The potential role of gastroesophageal reflux in asthma. Minerva Gastroenterol Dietol. 2001;47(2):75–83.
125. Field SK, Gelfand GA, McFadden SD. The effects of antireflux surgery on asthmatics with gastroesophageal reflux. Chest. 1999;116(3):766–74.
126. Havemann BD, Henderson CA, El-Serag HB. The association between gastro-oesophageal reflux disease and asthma: a systematic review. Gut. 2007;56(12):1654–64.
127. Boulet LP. Influence of comorbid conditions on asthma. Eur Respir J. 2009;33(4):897–906.
128. Lehrer P, Feldman J, Giardino N, Song HS, Schmaling K. Psychological aspects of asthma. J Consult Clin Psychol. 2002;70(3):691–711.
129. Cotton SJ, Belcher J, Rose P, K Jagadeesan S, Neal RD. The risk of a subsequent cancer diagnosis after herpes zoster infection: primary care database study. Br J Cancer. 2013 Feb;108(3):721–6.
130. Miles JF, Garden GM, Tunnicliffe WS, Cayton RM, Ayres JG. Psychological morbidity and coping skills in patients with brittle and non-brittle asthma: a case-control study. Clin Exp Allergy. 1997;27(10):1151–9.
131. Yigla M, Tov N, Solomonov A, Rubin AH, Harlev D. Difficult-to-control asthma and obstructive sleep apnea. J Asthma. 2003;40(8):865–71.
132. Sears MR, Greene JM, Willan AR, Wiecek EM, Taylor DR, Flannery EM, et al. A longitudinal, population-based, cohort study of childhood asthma followed to adulthood. N Engl J Med. 2003;349(15):1414–22.
133. Postma DS. Gender differences in asthma development and progression. Gend Med. 2007;4 Suppl B:S133-46.
134. Beuther DA, Sutherland ER. Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. Am J Respir Crit Care Med. 2007;175(7):661–6.
135. Boulet LP. Asthma and obesity. Clin Exp Allergy. 2013;43(1):8–21.
136. Matricardi PM, Bouygue GR, Tripodi S. Inner-city asthma and the hygiene hypothesis. Ann Allergy Asthma Immunol. 2002;89(6 Suppl 1):69–74.
137. Basagana X, Sunyer J, Kogevinas M, Zock JP, Duran-Tauleria E, Jarvis D, et al. Socioeconomic status and asthma prevalence in young adults: the European Community Respiratory Health Survey. Am J Epidemiol. 2004;160(2):178–88.
138. Burke W, Fesinmeyer M, Reed K, Hampson L, Carlsten C. Family history as a predictor of asthma risk. Am J Prev Med. 2003;24(2):160–9.
139. Whitrow MJ, Harding S. Asthma in Black African, Black Caribbean and South Asian adolescents in the MRC DASH study: a cross sectional analysis. BMC Pediatr. 2010;10(1):18.
140. Kogevinas M, Anto JM, Sunyer J, Tobias A, Kromhout H, Burney P. Occupational asthma in Europe and other industrialised areas: a population-based study. European Community Respiratory Health Survey Study Group. Lancet. 1999;353(9166):1750–4.
141. Feary J, Pinnock H, Cullinan P. Occupational asthma. Bmj. 2016;353:i2658.
142. Gergen PJ, Fowler JA, Maurer KR, Davis WW, Overpeck MD. The burden of environmental tobacco smoke exposure on the respiratory health of children 2 months through 5 years of age in the United States: Third National Health and Nutrition Examination Survey, 1988 to 1994. Pediatrics. 1998;101(2):E8.
143. Gilliland FD, Li YF, Peters JM. Effects of maternal smoking during pregnancy and environmental tobacco smoke on asthma and wheezing in children. Am J Respir Crit

Care Med. 2001;163(2):429–36.
144. Strachan DP, Butland BK, Anderson HR. Incidence and prognosis of asthma and wheezing illness from early childhood to age 33 in a national British cohort. Bmj. 1996;312(7040):1195–9.
145. Bloomfield SF, Rook GA, Scott EA, Shanahan F, Stanwell-Smith R, Turner P. Time to abandon the hygiene hypothesis: new perspectives on allergic disease, the human microbiome, infectious disease prevention and the role of targeted hygiene. Perspect Public Heal. 2016;136(4):213–24.
146. Simpson JL, Scott R, Boyle MJ, Gibson PG. Inflammatory subtypes in asthma: assessment and identification using induced sputum. Respirology. 2006 Jan;11(1):54–61.
147. Swedin L, Saarne T, Rehnberg M, Glader P, Niedzielska M, Johansson G, et al. Patient stratification and the unmet need in asthma. Pharmacol Ther. 2016;
148. Barnes PJ. Against the Dutch hypothesis: asthma and chronic obstructive pulmonary disease are distinct diseases. Am J Respir Crit Care Med. 2006;174(3):240–3.
149. Bousquet J, Schunemann HJ, Samolinski B, Demoly P, Baena-Cagnani CE, Bachert C, et al. Allergic Rhinitis and its Impact on Asthma (ARIA): achievements in 10 years and future needs. J Allergy Clin Immunol. 2012;130(5):1049–62.
150. Schneider L, Hanifin J, Boguniewicz M, Eichenfield LF, Spergel JM, Dakovic R, et al. Study of the Atopic March: Development of Atopic Comorbidities. Pediatr Dermatol. 2016;33(4):388–98.
151. Spergel JM. From atopic dermatitis to asthma: the atopic march. Ann Allergy Asthma Immunol. 2010;105(2):99,117.
152. Gergen PJ, Turkeltaub PC. The association of individual allergen reactivity with respiratory disease in a national sample: data from the second National Health and Nutrition Examination Survey, 1976–1980 (NHANES II). J Allergy Clin Immunol. 1992;90(4):579–88.
153. Sears MR, Burrows B, Flannery EM, Herbison GP, Hewitt CJ, Holdaway MD. Relation between airway responsiveness and serum IgE in children with asthma and in apparently normal children. N Engl J Med. 1991;325(15):1067–71.
154. Jarvis D, Burney P. ABC of allergies: the epidemiology of allergic disease. BMJ Br Med J. 1998;316(7131):607.
155. Hargreave FE, Parameswaran K. Asthma, COPD and bronchitis are just components of airway disease. Eur Respir J. 2006;28(2):264–7.
156. Kraft M. Asthma and chronic obstructive pulmonary disease exhibit common origins in any country! Am J Respir Crit Care Med. 2006;174(3):238–40.
157. Ghebre MA, Bafadhel M, Desai D, Cohen SE, Newbold P, Rapley L, et al. Biological clustering supports both "Dutch" and "British" hypotheses of asthma and chronic obstructive pulmonary disease. J Allergy Clin Immunol. 2015;135(1):63–72.
158. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: systematic review and meta-analysis. Eur Respir J. 2006;28(3):523–32.
159. Vonk JM, Jongepier H, Panhuysen CIM, Schouten JP, Bleecker ER, Postma DS. Risk factors associated with the presence of irreversible airflow limitation and reduced transfer coefficient in patients with asthma after 26 years of follow up. Thorax. 2003;58(4):322–7.
160. Calverley PMA, Rennard SI. What have we learned from large drug treatment trials in COPD? Lancet. 2007;370(9589):774–85.

161. Calverley PMA, Burge PS, Spencer S, Anderson JA, Jones PW. Bronchodilator reversibility testing in chronic obstructive pulmonary disease. Thorax. 2003;58(8):659–64.
162. Marsh SE, Travers J, Weatherall M, Williams M V, Aldington S, Shirtcliffe PM, et al. Proportional classifications of COPD phenotypes. Thorax. 2008;63(9):761–7.
163. Dharmage SC Perret JL BJALCJJDPTPSGGGHJLAMJWEHMMC. Current asthma contributes as much as smoking to chronic bronchitis in middle age: a prospective population-based study. Int J COPD. 2016;Volume 201:Pages 1911—1920.
164. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. Am J Respir Crit Care Med. 2013;187(4):347–65.
165. To T, Stanojevic S, Moores G, Gershon AS, Bateman ED, Cruz AA, et al. Global asthma prevalence in adults: findings from the cross-sectional world health survey. BMC Public Health. 2012;12:204.
166. Bateman ED, Reddel HK, van Zyl-Smit RN, Agusti A. The asthma-COPD overlap syndrome: towards a revised taxonomy of chronic airways diseases? Lancet Respir Med. 2015 Sep;3(9):719–28.
167. Lavoie KL, Bacon SL, Barone S, Cartier A, Ditto B, Labrecque M. What is worse for asthma control and quality of life: depressive disorders, anxiety disorders, or both? Chest. 2006;130(4):1039–47.
168. Boulet LP, Boulay ME. Asthma-related comorbidities. Expert Rev Respir Med. 2011;5(3):377–93.
169. van Maanen A, Wijga AH, Gehring U, Postma DS, Smit HA, Oort FJ, et al. Sleep in children with asthma: results of the PIAMA study. Eur Respir J. 2013 Apr;41(4):832–7.
170. Bush A, Menzies-Gow A. Phenotypic differences between pediatric and adult asthma. Proc Am Thorac Soc. 2009;6(8):712–9.
171. Colak Y, Afzal S, Nordestgaard BG, Lange P. Characteristics and Prognosis of Never-Smokers and Smokers with Asthma in the Copenhagen General Population Study. A Prospective Cohort Study. Am J Respir Crit Care Med. 2015;192(2):172–81.
172. Aleman F, Lim HF, Nair P. Eosinophilic Endotype of Asthma. Immunol Allergy Clin North Am. 2016 Aug;36(3):559–68.
173. Kay AB. The role of eosinophils in the pathogenesis of asthma. Trends Mol Med. 2005;11(4):148–52.
174. Jacobsen EA, Helmers RA, Lee JJ, Lee NA. The expanding role (s) of eosinophils in health and disease. Blood. 2012;120(19):3882–90.
175. Barnes PJ. Th2 cytokines and asthma: an introduction. Respir Res. 2001;2(2):64–5.
176. Aaby P, Shaheen SO, Heyes CB, Goudiaby A, Hall AJ, Shiell AW, et al. Early BCG vaccination and reduction in atopy in Guinea-Bissau. Clin Exp allergy. 2000;30(5):644–50.
177. Matricardi PM, Rosmini F, Ferrigno L, Nisini R, Rapicetta M, Chionne P, et al. Cross sectional retrospective study of prevalence of atopy among Italian military students with antibodies against hepatitis A virus. Bmj. 1997;314(7086):999.
178. Ege MJ, Mayer M, Normand A-C, Genuneit J, Cookson WOCM, Braun-Fahrländer C, et al. Exposure to environmental microorganisms and childhood asthma. N Engl J Med. 2011;364(8):701–9.
179. Braun-Fahrländer CH, Gassner M, Grize L, Neu U, Sennhauser FH, Varonier HS, et al.

Prevalence of hay fever and allergic sensitization in farmer's children and their peers living in the same rural community. Clin Exp allergy. 1999;29(1):28–34.
180. Fuchs O, Genuneit J, Latzin P, Büchele G, Horak E, Loss G, et al. Farming environments and childhood atopy, wheeze, lung function, and exhaled nitric oxide. J Allergy Clin Immunol. 2012;130(2):382–8.
181. Devereux G, Litonjua AA, Turner SW, Craig LCA, McNeill G, Martindale S, et al. Maternal vitamin D intake during pregnancy and early childhood wheezing–. Am J Clin Nutr. 2007;85(3):853–9.
182. Camargo Jr CA, Rifas-Shiman SL, Litonjua AA, Rich-Edwards JW, Weiss ST, Gold DR, et al. Maternal intake of vitamin D during pregnancy and risk of recurrent wheeze in children at 3 y of age–. Am J Clin Nutr. 2007;85(3):788–95.
183. Dilts D, Khamalah J, Plotkin A. Using cluster analysis for medical resource decision making. Med Decis Making. 1995;15(4):333–47.
184. McLachlan GJ. Cluster analysis and related techniques in medical research. Stat Methods Med Res. 1992;1(1):27–48.
185. Clatworthy J, Buick D, Hankins M, Weinman J, Horne R. The use and reporting of cluster analysis in health psychology: a review. Br J Health Psychol. 2005 Sep;10(Pt 3):329–58.
186. Weir MR, Maibach EW, Bakris GL, Black HR, Chawla P, Messerli FH, et al. Implications of a health lifestyle and medication analysis for improving hypertension control. Arch Intern Med. 2000 Feb;160(4):481–90.
187. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. Knowl Inf Syst. 2008;14(1):1–37.
188. Rodriguez MZ, Comin CH, Casanova D, Bruno OM, Amancio DR, Costa L da F, et al. Clustering algorithms: A comparative approach. PLoS One. 2019;14(1):e0210236.
189. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. J R Stat Soc Ser C (Applied Stat. 1979;28(1):100–8.
190. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recognit Lett. 2010;31(8):651–66.
191. Steinley D. K-means clustering: a half-century synthesis. Br J Math Stat Psychol. 2006;59(1):1–34.
192. Sequeira K, Zaki M. ADMIT: anomaly-based data mining for intrusions. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM; 2002. p. 386–95.
193. Fraley C, Raftery AE. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput J. 1998;41(8):578–88.
194. Weatherall M, Travers J, Shirtcliffe PM, Marsh SE, Williams M V, Nowitz MR, et al. Distinct clinical phenotypes of airways disease defined by cluster analysis. Eur Respir J. 2009;34(4):812–8.
195. Sutherland ER, Goleva E, King TS, Lehman E, Stevens AD, Jackson LP, et al. Cluster analysis of obesity and asthma phenotypes. PLoS One. 2012;7(5):e36631.
196. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nat Med. 2012;18(5):716–25.
197. King J, Patel V, Jamoom EW, Furukawa MF. Clinical benefits of electronic health record use: national findings. Health Serv Res. 2014;49(1pt2):392–404.
198. Burton LC, Anderson GF, Kues IW. Using electronic health records to help coordinate care. Milbank Q. 2004;82(3):457–81.

199. Sittig DF, Singh H. Electronic Health Records and National Patient-Safety Goals. Vol. 367, The New England journal of medicine. 2012. p. 1854–60.
200. Blumenthal D. Stimulating the adoption of health information technology. N Engl J Med. 2009 Apr;360(15):1477–9.
201. Myers RB, Jones SL, Sittig DF. Review of Reported Clinical Information System Adverse Events in US Food and Drug Administration Databases. Appl Clin Inform. 2011;2(1):63–74.
202. Harrington L, Kennerly D, Johnson C. Safety issues related to the electronic medical record (EMR): synthesis of the literature from the last decade, 2000-2009. J Healthc Manag. 2011;56(1):31–4.
203. Warm D, Edwards P. Classifying health information technology patient safety related incidents - an approach used in Wales. Appl Clin Inform. 2012;3(2):248–57.
204. Suissa S. Immortal time bias in pharmacoepidemiology. Am J Epidemiol. 2007;167(4):492–9.
205. Suissa S. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf. 2007;16(3):241–9.
206. Lynge E, Sandegaard JL, Rebolj M. The Danish national patient register. Scand J Public Health. 2011;39(7_suppl):30–3.
207. Lo Re V. 3rd, Haynes K, Forde KA, Localio AR, Schinnar R, Lewis JD. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. Pharmacoepidemiol Drug Saf. 2009;18(9):807–14.
208. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015;44(3):827–36.
209. Abraha I, Giovannini G, Serraino D, Fusco M, Montedori A. Validity of breast, lung and colorectal cancer diagnoses in administrative databases: a systematic review protocol. BMJ Open. 2016;6(3):e010409.
210. Chen C, Garrido T, Chock D, Okawa G, Liang L. The Kaiser Permanente Electronic Health Record: transforming and streamlining modalities of care. Health Aff. 2009;28(2):323–33.
211. Cowan DC, Cowan JO, Palmay R, Williamson A, Taylor DR. The effects of steroid therapy on inflammatory cell subtypes in asthma. Thorax. 2009;thx-2009.
212. Jenkins C, Seccombe L, Tomlins R. Investigating asthma symptoms in primary care. Bmj. 2012;344:e2734.
213. Aaron SD, Vandemheen KL, Boulet L-P, McIvor RA, FitzGerald JM, Hernandez P, et al. Overdiagnosis of asthma in obese and nonobese adults. Can Med Assoc J. 2008;179(11):1121–31.
214. Pollart SM, Elward KS. Overview of changes to asthma guidelines: diagnosis and screening. Am Fam Physician. 2009;79(9).
215. van Wonderen KE, van der Mark LB, Mohrs J, Bindels PJE, van Aalderen WMC, ter Riet G. Different definitions in childhood asthma: how dependable is the dependent variable? Eur Respir J. 2009;
216. Gibson PG, Fujimura M, Niimi A. Eosinophilic bronchitis: clinical manifestations and implications for treatment. Thorax. 2002;57(2):178–82.
217. Balkissoon R, Kenn K. Asthma: vocal cord dysfunction (VCD) and other dysfunctional breathing disorders. In: Seminars in respiratory and critical care medicine. Thieme Medical Publishers; 2012. p. 595–605.

218. Thomas M, McKinley RK, Freeman E, Foy C. Prevalence of dysfunctional breathing in patients treated for asthma in primary care: cross sectional survey. Bmj. 2001;322(7294):1098–100.
219. Wilson NM, James A, Uasuf C, Payne DN, Hablas H, Agrofioti C, et al. Asthma severity and inflammation markers in children. Pediatr allergy Immunol. 2001;12(3):125–32.
220. PAYNE DNR, ADCOCK IANM, WILSON NM, Oates TIM, SCALLAN M, BUSH A. Relationship between exhaled nitric oxide and mucosal eosinophilic inflammation in children with difficult asthma, after treatment with oral prednisolone. Am J Respir Crit Care Med. 2001;164(8):1376–81.
221. Sont JK, Willems LNA, Bel EH, van KRIEKEN JHANJM, VANDENBROUCKE JANP, STERK PJ, et al. Clinical control and histopathologic outcome of asthma when using airway hyperresponsiveness as an additional guide to long-term treatment. Am J Respir Crit Care Med. 1999;159(4):1043–51.
222. Sont JK, Han J, Van Krieken JM, Evertse CE, Hooijer R, Willems LN, et al. Relationship between the inflammatory infiltrate in bronchial biopsy specimens and clinical severity of asthma in patients treated with inhaled steroids. Thorax. 1996;51(5):496–502.
223. CRIMI E, SPANEVELLO A, NERI M, IND PW, ROSSI GA, BRUSASCO V. Dissociation between airway inflammation and airway hyperresponsiveness in allergic asthma. Am J Respir Crit Care Med. 1998;157(1):4–9.
224. Hansell A, Hollowell J, Nichols T, McNiece R, Strachan D. Use of the General Practice Research Database (GPRD) for respiratory epidemiology: a comparison with the 4th Morbidity Survey in General Practice (MSGP4). Thorax. 1999;54(5):413–9.
225. Chen Y-C, Wu J-C, Haschler I, Majeed A, Chen T-J, Wetter T. Academic impact of a public electronic health database: bibliometric analysis of studies using the general practice research database. PLoS One. 2011;6(6):e21404.
226. CPRD. CPRD Aurum [Internet]. 2018 [cited 2018 Jul 25]. Available from: https://www.cprd.com/cprdaurum/
227. MHRA. CPRD welcome page [Internet]. [cited 2018 Jun 19]. Available from: https://www.cprd.com/intro.asp
228. Campbell J, Dedman DJ, Eaton SC, Gallagher AM, Williams TJ. Is the CPRD GOLD population comparable to the UK population? Pharmacoepidemiol Drug Saf. 2013;22:280–1.
229. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol. 2010;69(1):4–14.
230. Benson T. The history of the Read codes: the inaugural James Read Memorial Lecture 2011. J Innov Heal Informatics. 2011;19(3):173–82.
231. Leigh J. Medical Applications of Computers: the.
232. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform. 2006;121:279.
233. Crooks CJ, Card TR, West J. Excess long-term mortality following non-variceal upper gastrointestinal bleeding: a population-based cohort study. PLoS Med. 2013;10(4):e1001437.
234. Bhaskaran K, Forbes HJ, Douglas I, Leon DA, Smeeth L. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink

(CPRD). BMJ Open. 2013 Sep;3(9):e003389.
235. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. J Public Health (Oxf). 2014 Dec;36(4):684–92.
236. van Staa TP, Dennison EM, Leufkens HG, Cooper C. Epidemiology of fractures in England and Wales. Bone. 2001 Dec;29(6):517–22.
237. Meier CR, Napalkov PN, Wegmuller Y, Jefferson T, Jick H. Population-based study on incidence, risk factors, clinical complications and drug utilisation associated with influenza in the United Kingdom. Eur J Clin Microbiol Infect Dis. 2000 Nov;19(11):834–42.
238. NHS. QOF 2016/2017 results [Internet]. [cited 2018 Jun 17]. Available from: https://qof.digital.nhs.uk/
239. Grimaldi-Bensouda L, Rossignol M, Aubrun E, El Kerri N, Benichou J, Abenhaim L. Agreement between patients' self-report and physicians' prescriptions on cardiovascular drug exposure: the PGRx database experience. Pharmacoepidemiol Drug Saf. 2010 Jun;19(6):591–5.
240. Currie CJ, Peyrot M, Morgan CL, Poole CD, Jenkins-Jones S, Rubin RR, et al. The impact of treatment non-compliance on mortality in people with type 1 diabetes. J Diabetes Complications. 2013;27(3):219–23.
241. Gulliford MC, van Staa T, McDermott L, Dregan A, McCann G, Ashworth M, et al. Cluster randomised trial in the General Practice Research Database: 1. Electronic decision support to reduce antibiotic prescribing in primary care (eCRT study). Trials. 2011;12(1):115.
242. Wright AK. Examining Drug Utilisation Patterns and Optimal Treatment Pathways of Antidiabetic Medications. University of Manchester; 2014.
243. NHS. Hospital Episode Statistics [Internet]. [cited 2018 Jun 19]. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
244. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). Vol. 46, International Journal of Epidemiology. 2017. p. 1093–1093i.
245. Meacock R, Doran T, Sutton M. What are the Costs and Benefits of Providing Comprehensive Seven-day Services for Emergency Hospital Admissions? Health Econ. 2015 Aug;24(8):907–12.
246. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P, et al. Systematic review of discharge coding accuracy. J Public Health (Oxf). 2012 Mar;34(1):138–48.
247. ONS. Office of National Statistics homepage [Internet]. [cited 2018 Jun 19]. Available from: https://www.ons.gov.uk/
248. gov.uk. English indices of deprivation 2015 [Internet]. [cited 2018 Jun 19]. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
249. Burney PG, Luczynska C, Chinn S, Jarvis D. The European community respiratory health survey. Eur Respir J. 1994;7(5):954–60.
250. Pearce N, Aït-Khaled N, Beasley R, Mallol J, Keil U, Mitchell E, et al. Worldwide trends in the prevalence of asthma symptoms: phase III of the International Study of Asthma and Allergies in Childhood (ISAAC). Thorax. 2007;62(9):758–66.
251. Burney PG, Laitinen LA, Perdrizet S, Huckauf H, Tattersfield AE, Chinn S, et al. Validity and repeatability of the IUATLD (1984) Bronchial Symptoms Questionnaire:

an international comparison. Eur Respir J. 1989;2(10):940–5.
252. Peat JK, Salome CM, Toelle BG, Bauman A, Woolcock AJ. Reliability of a respiratory history questionnaire and effect of mode of administration on classification of asthma in children. Chest. 1992;102(1):153–7.
253. Sharifi M, Krishanswami S, McPheeters ML. A systematic review of validated methods to capture acute bronchospasm using administrative or claims data. Vaccine. 2013;31 Suppl 1:K12-20.
254. Mullooly JP, Schuler R, Mesa J, Drew L, DeStefano F. Wheezing lower respiratory disease and vaccination of premature infants. Vaccine. 2011;29(44):7611–7.
255. Sanders DL, Gregg W, Aronsky D. Identifying asthma exacerbations in a pediatric emergency department: a feasibility study. Int J Med Inf. 2007;76(7):557–64.
256. 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).
257. Pekkanen J, Pearce N. Defining asthma in epidemiological studies. Eur Respir J. 1999;14(4):951–7.
258. Blais L, Lemière C, Menzies D, Berbiche D. Validity of asthma diagnoses recorded in the Medical Services database of Quebec. Pharmacoepidemiol Drug Saf. 2006;15(4):245–52.
259. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32–5.
260. Langan SM, Benchimol EI, Guttmann A, Moher D, Petersen I, Smeeth L, et al. Setting the RECORD straight: developing a guideline for the REporting of studies Conducted using Observational Routinely collected Data. Clin Epidemiol. 2013;5:29–31.
261. Denney MJ, Long DM, Armistead MG, Anderson JL, Conway BN. Validating the extract, transform, load process used to populate a large clinical research database. Int J Med Inf. 2016;94:271–4.
262. Rimland JM, Abraha I, Luchetta ML, Cozzolino F, Orso M, Cherubini A, et al. Validation of chronic obstructive pulmonary disease (COPD) diagnoses in healthcare databases: a systematic review protocol. BMJ Open. 2016;6(6):e011777.
263. Lo Re V 3rd, Haynes K, Forde KA, Localio AR, Schinnar R, Lewis JD. Validity of The Health Improvement Network (THIN) for epidemiologic studies of hepatitis C virus infection. Pharmacoepidemiol Drug Saf. 2009 Sep;18(9):807–14.
264. Tang PC, Ralston M, Arrigotti MF, Qureshi L, Graham J. Comparison of Methodologies for Calculating Quality Measures Based on Administrative Data versus Clinical Data from an Electronic Health Record System: Implications for Performance Measures. J Am Med Inf Assoc. 2007;14(1):10–5.
265. Angier H, Gold R, Gallia C, Casciato A, Tillotson CJ, Marino M, et al. Variation in Outcomes of Quality Measurement by Data Source. Pediatrics. 2014;133(6):e1676-82.
266. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inf Assoc. 2013;20(1):144–51.
267. Afzal Z, Engelkes M, Verhamme KMC, Janssens HM, Sturkenboom MCJM, Kors JA, et al. Automatic generation of case-detection algorithms to identify children with asthma from large electronic health record databases. Pharmacoepidemiol Drug Saf. 2013;22(8):826–33.
268. Wu ST, Juhn YJ, Sohn S, Liu H. Patient-level temporal aggregation for text-based asthma status ascertainment. J Am Med Informatics Assoc. 2014;21(5):876–84.

269. Wing K, Douglas I, Bhaskaran K, Klungel OH, Reynolds RF, Pirmohamed M, et al. Development of predictive genetic tests for improving the safety of new medicines: the utilization of routinely collected electronic health records. Drug Discov Today. 2014;19(4):361–6.
270. Wing K, Bhaskaran K, Smeeth L, van Staa TP, Klungel OH, Reynolds RF, et al. Optimising case detection within UK electronic health records: use of multiple linked databases for detecting liver injury. BMJ Open. 2016;6(9):e012102.
271. Xi N, Wallace R, Agarwal G, Chan D, Gershon A, Gupta S. Identifying patients with asthma in primary care electronic medical record systems Chart analysis-based electronic algorithm validation study. Can Fam Physician. 2015;61(10):e474-83.
272. Cornish RP, Henderson J, Boyd AW, Granell R, Van Staa T, Macleod J. Validating childhood asthma in an epidemiological study using linked electronic patient records. BMJ Open. 2014;4(4):e005345.
273. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med. 2009;3(3):e123-30.
274. Rabe KF, Adachi M, Lai CK, Soriano JB, Vermeire PA, Weiss KB, et al. Worldwide severity and control of asthma in children and adults: the global asthma insights and reality surveys. J Allergy Clin Immunol. 2004;114(1):40–7.
275. Demoly P, Gueron B, Annunziata K, Adamek L, Walters RD. Update on asthma control in five European countries: results of a 2008 survey. Eur Respir Rev. 2010;19(116):150–7.
276. Nissen F, Quint JK, Wilkinson S, Mullerova H, Smeeth L, Douglas IJ. Validation of asthma recording in electronic health records: protocol for a systematic review. BMJ Open. 2017;7(5):e014694.
277. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
278. Engelkes M. Automated identification of asthma patients within an electronical medical record database using machine learning. Pharmacoepidemiol Drug Saf. 2012;21:12.
279. Dexheimer JW, Abramo TJ, Arnold DH, Johnson K, Shyr Y, Ye F, et al. Implementation and evaluation of an integrated computerized asthma management system in a pediatric emergency department: a randomized clinical trial. Int J Med Inform. 2014;83(11):805–13.
280. Wu ST, Sohn S, Ravikumar KE, Wagholikar K, Jonnalagadda SR, Liu H, et al. Automated chart review for asthma cohort identification using natural language processing: an exploratory study. Ann Allergy, Asthma Immunol. 2013;111(5):364–9.
281. Kozyrskyj AL, HayGlass KT, Sandford AJ, Pare PD, Chan-Yeung M, Becker AB. A novel study design to investigate the early-life origins of asthma in children (SAGE study). Allergy. 2009;64(8):1185–93.
282. Pacheco JA, Avila PC, Thompson JA, Law M, Quraishi JA, Greiman AK, et al. A highly specific algorithm for identifying asthma cases and controls for genome-wide association studies. In: AMIA Annual Symposium Proceedings. American Medical Informatics Association; 2009. p. 497.
283. Vollmer WM, O'Connor EA, Heumann M, Frazier EA, Breen V, Villnave J, et al. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol. 2004;57(4):392–7.

284. Donahue JG, Weiss ST, Goetsch MA, Livingston JM, Greineder DK, Piatt R. Assessment of asthma using automated and full-text medical records. J Asthma. 1997;34(4):273–81.
285. Premaratne UN, Marks GB, Austin EJ, Burney PGJ. A reliable method to retrieve accident & emergency data stored on a free-text basis. Respir Med. 1997;91(2):61–6.
286. Engeland A, Bjørge T, Daltveit AK, Vollset SE, Furu K. Validation of disease registration in pregnant women in the Medical Birth Registry of Norway. Acta Obstet Gynecol Scand. 2009;88(10):1083–9.
287. Coulter A, Brown S, Daniels A. Computer held chronic disease registers in general practice: a validation study. J Epidemiol Community Heal. 1989;43(1):25–8.
288. Ward DG, Halpin DMG, Seamark DA. How accurate is a diagnosis of asthma in a general practice database? A review of patients' notes and questionnaire-reported symptoms. Br J Gen Pr. 2004;54(507):753–8.
289. Moth G, Vedsted P, Schiøtz PO. Identification of asthmatic children using prescription data and diagnosis. Eur J Clin Pharmacol. 2007;63(6):605–11.
290. Hansell A, Hollowell J, McNiece R, Nichols T, Strachan D. Validity and interpretation of mortality, health service and survey data on COPD and asthma in England. Eur Respir J. 2003 Feb;21(2):279–86.
291. Leone FT, Grana JR, McDermott P, MacPherson S, Hanchak NA, Fish JE. Pharmaceutically-based severity stratification of an asthmatic population. Respir Med. 1999;93(11):788–93.
292. Pearson N, Brien JO, Thomas H, Ewings P, Gallier L, Bussey A. Collecting morbidity data in general practice: the Somerset morbidity project. BMJ. 1996;312(7045):1517–20.
293. Ehrenstein V, Petersen I, Smeeth L, Jick SS, Benchimol EI, Ludvigsson JF, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. Clin Epidemiol. 2016;8:49–51.
294. Gange SJ, Golub ET. From Smallpox to Big Data: The Next 100 Years of Epidemiologic Methods. Am J Epidemiol. 2016 Mar;183(5):423–6.
295. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. Vol. 9, Clinical Epidemiology. 2017. p. 245–50.
296. Manuel DG, Rosella LC, Stukel TA. Importance of accurately identifying disease in studies using electronic health records. BMJ. 2010 Aug;341:c4226.
297. Quint JK, Mullerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). BMJ Open. 2014;4(7):e005540.
298. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015;385(9963):117–71.
299. The Global Asthma Report 2014. 2014.
300. Levy M, Andrews R, Buckingham R, Evans H, Francis C, Houston R, et al. Why asthma still kills: the National Review of Asthma Deaths (NRAD). Royal College of Physcians; 2014.
301. James DR, Lyttle MD. British guideline on the management of asthma: SIGN Clinical Guideline 141, 2014. Arch Dis Child Educ Pr Ed. 2016;101(6):319–22.
302. Wang D, Ding W, Lo H, Morabito M, Chen P, Salazar J, et al. Understanding the spatial distribution of crime based on its related variables using geospatial

discriminative patterns. Comput Environ Urban Syst. 2013;39:93–106.
303. Pharmacovigilance. EN of C for P and. ENCePP Guide on Methodological Standards in Pharmacoepidemiology. Revision 5. 2017;
304. Bousquet J, Mantzouranis E, Cruz AA, Aït-Khaled N, Baena-Cagnani CE, Bleecker ER, et al. Uniform definition of asthma severity, control, and exacerbations: document presented for the World Health Organization Consultation on Severe Asthma. J Allergy Clin Immunol. 2010;126(5):926–38.
305. Sin DD, Miravitlles M, Mannino DM, Soriano JB, Price D, Celli BR, et al. What is asthma− COPD overlap syndrome? Towards a consensus definition from a round table discussion. Eur Respir J. 2016;48(3):664–73.
306. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv Drug Saf. 2012;3(2):89–99.
307. Thomas KH, Davies N, Metcalfe C, Windmeijer F, Martin RM, Gunnell D. Validation of suicide and self-harm records in the Clinical Practice Research Datalink. Br J Clin Pharmacol. 2013;76(1):145–57.
308. Rothnie KJ, Müllerová H, Hurst JR, Smeeth L, Davis K, Thomas SL, et al. Validation of the Recording of Acute Exacerbations of COPD in UK Primary Care Electronic Healthcare Records. PLoS One. 2016;11(3).
309. Ehrenstein V, Petersen I, Smeeth L, Jick SS, Benchimol EI, Ludvigsson JF, et al. Helping everyone do better: a call for validation studies of routinely recorded health data. Vol. 8, Clinical Epidemiology. 2016. p. 49–51.
310. Chew-Graham CA, Hunter C, Langer S, Stenhoff A, Drinkwater J, Guthrie EA, et al. How QOF is shaping primary care review consultations: a longitudinal qualitative study. BMC Fam Pract. 2013;14(1):103.
311. Quint JK, Müllerova H, DiSantostefano RL, Forbes H, Eaton S, Hurst JR, et al. Validation of chronic obstructive pulmonary disease recording in the Clinical Practice Research Datalink (CPRD-GOLD). Vol. 4, BMJ Open. BMA House, Tavistock Square, London, WC1H 9JR; 2014.
312. Hardin M, Silverman EK, Barr RG, Hansel NN, Schroeder JD, Make BJ, et al. The clinical features of the overlap between COPD and asthma. Respir Res. 2011 Sep;12:127.
313. Kauppi P, Kupiainen H, Lindqvist A, Tammilehto L, Kilpelainen M, Kinnula VL, et al. Overlap syndrome of asthma and COPD predicts low quality of life. J Asthma. 2011 Apr;48(3):279–85.
314. Soler-Cataluna JJ, Cosio B, Izquierdo JL, Lopez-Campos JL, Marin JM, Aguero R, et al. Consensus document on the overlap phenotype COPD-asthma in COPD. Arch Bronconeumol. 2012 Sep;48(9):331–7.
315. de Marco R, Pesce G, Marcon A, Accordini S, Antonicelli L, Bugiani M, et al. The coexistence of asthma and chronic obstructive pulmonary disease (COPD): prevalence and risk factors in young, middle-aged and elderly people from the general population. PLoS One. 2013;8(5):e62985.
316. Wurst KE, Kelly-Reif K, Bushnell GA, Pascoe S, Barnes N. Understanding asthma-chronic obstructive pulmonary disease overlap syndrome. Respir Med. 2016 Jan;110:1–11.
317. Gibson PG, McDonald VM. Asthma-COPD overlap 2015: now we are six. Thorax. 2015 Jul;70(7):683–91.

318. Bujarski S, Parulekar AD, Sharafkhaneh A, Hanania NA. The asthma COPD overlap syndrome (ACOS). Curr Allergy Asthma Rep. 2015;15(3):509.
319. Rodrigo GJ, Neffen H, Plaza V. Asthma-chronic obstructive pulmonary disease overlap syndrome: a controversial concept. Curr Opin Allergy Clin Immunol. 2017;17(1):36–41.
320. Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). BMJ Open. 2017;7(8):e017474.
321. Rothnie KJ, Chandan JS, Goss HG, Mullerova H, Quint JK. Validity and interpretation of spirometric recordings to diagnose COPD in UK primary care. Int J Chron Obs Pulmon Dis. 2017;12:1663–8.
322. Walters JA, Hansen E, Mudge P, Johns DP, Walters EH, Wood-Baker R. Barriers to the use of spirometry in general practice. Aust Fam Physician. 2005;34(3):201–3.
323. Bunker J, Hermiz O, Zwar N, Dennis SM, Vagholkar S, Crockett A, et al. Feasibility and efficacy of COPD case finding by practice nurses. Aust Fam Physician. 2009;38(10):826–30.
324. Zwar NA, Marks GB, Hermiz O, Middleton S, Comino EJ, Hasan I, et al. Predictors of accuracy of diagnosis of chronic obstructive pulmonary disease in general practice. Med J Aust. 2011;195(4):168–71.
325. Lucas AE, Smeenk FJ, Smeele IJ, van Schayck OP. Diagnostic accuracy of primary care asthma/COPD working hypotheses, a real life study. Respir Med. 2012;106(8):1158–63.
326. Levy ML, Quanjer PH, Booker R, Cooper BG, Holmes S, Small I. Diagnostic spirometry in primary care: Proposed standards for general practice compliant with American Thoracic Society and European Respiratory Society recommendations: a General Practice Airways Group (GPIAG)1 document, in association with the Associati. Prim Care Respir J. 2009;18(3):130–47.
327. Yawn BP. Optimizing chronic obstructive pulmonary disease management in primary care. South Med J. 2011;104(2):121–7.
328. Kaplan A, Thomas M. Screening for COPD: the gap between logic and evidence. Eur Respir Rev. 2017;26(143).
329. Walker PP, Mitchell P, Diamantea F, Warburton CJ, Davies L. Effect of primary-care spirometry on the diagnosis and management of COPD. Eur Respir J. 2006;28(5):945–52.
330. Alshabanat A, Zafari Z, Albanyan O, Dairi M, FitzGerald JM. Asthma and COPD Overlap Syndrome (ACOS): A Systematic Review and Meta Analysis. PLoS One. 2015;10(9):e0136065.
331. Melbye H, Drivenes E, Dalbak LG, Leinan T, Høegh-Henrichsen S, Østrem A. Asthma, chronic obstructive pulmonary disease, or both? Diagnostic labeling and spirometry in primary care patients aged 40 years or more. Int J Chron Obs Pulmon Dis. 2011;6:597–603.
332. Tinkelman DG, Price DB, Nordyke RJ, Halbert RJ. Misdiagnosis of COPD and asthma in primary care patients 40 years of age and over. J Asthma. 2006;43(1):75–80.
333. Inoue H, Nagase T, Morita S, Yoshida A, Jinnai T, Ichinose M. Prevalence and characteristics of asthma-COPD overlap syndrome identified by a stepwise approach. Int J Chron Obs Pulmon Dis. 2017;12:1803–10.
334. Keene ON, Calverley PM, Jones PW, Vestbo J, Anderson JA. Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. Eur Respir J.

2008;32(1):17–24.
335. Keene ON, Jones MR, Lane PW, Anderson J. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. Pharm Stat. 2007;6(2):89–97.
336. Agresti A. Categorical data analysis. Vol. 482. John Wiley & Sons; 2003.
337. Metcalfe C, Thompson SG. The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. Stat Med. 2006;25(1):165–79.
338. Calverley PMA, Spencer S, Willits L, Burge PS, Jones PW, Group IS. Withdrawal from treatment as an outcome in the ISOLDE study of COPD. Chest. 2003;124(4):1350–6.
339. Ho AM-H, Dion PW, Ng CSH, Karmakar MK. Understanding immortal time bias in observational cohort studies. Vol. 68, Anaesthesia. England; 2013. p. 126–30.
340. Zhou Z, Rahme E, Abrahamowicz M, Pilote L. Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. Am J Epidemiol. 2005 Nov;162(10):1016–23.
341. Calverley PMA, Anderson JA, Celli B, Ferguson GT, Jenkins C, Jones PW, et al. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. N Engl J Med. 2007;356(8):775–89.
342. Pavord ID. Asthma phenotypes. Semin Respir Crit Care Med. 2012;33(6):645–52.
343. Corren J. Asthma phenotypes and endotypes: an evolving paradigm for classification. Discov Med. 2013;15(83):243–9.
344. Soriano JB, Davis KJ, Coleman B, Visick G, Mannino D, Pride NB. The proportional Venn diagram of obstructive lung disease: two approximations from the United States and the United Kingdom. Chest. 2003;124(2):474–81.
345. Carr TF, Zeki AA, Kraft M. Eosinophilic and Noneosinophilic Asthma. Am J Respir Crit Care Med. 2018 Jan;197(1):22–37.
346. Bloom CI, Nissen F, Douglas IJ, Smeeth L, Cullinan P, Quint JK. Exacerbation risk and characterisation of the UK's asthma population from infants to old age. Thorax. 2017;
347. Price D, Wilson AM, Chisholm A, Rigazio A, Burden A, Thomas M, et al. Predicting frequent asthma exacerbations using blood eosinophil count and other patient data routinely available in clinical practice. J Asthma Allergy. 2016;9:1–12.
348. Zeiger RS, Schatz M, Li Q, Chen W, Khatry DB, Gossage D, et al. High blood eosinophil count is a risk factor for future asthma exacerbations in adult persistent asthma. J allergy Clin Immunol Pract. 2014;2(6):741–50.
349. Zemedkun K, Woldemichael K, Tefera G. Assessing control of asthma in Jush, Jimma, South West Ethiopia. Ethiop J Health Sci. 2014 Jan;24(1):49–58.
350. Kostikas K, Brindicci C, Patalano F. Blood eosinophils as biomarkers to drive treatment choices in asthma and COPD. Curr Drug Targets. 2018 Feb;
351. Agusti A, Bafadhel M, Beasley R, Bel EH, Faner R, Gibson PG, et al. Precision medicine in airway diseases: moving to clinical practice. Eur Respir J. 2017 Oct;50(4).
352. Agusti A, Bel E, Thomas M, Vogelmeier C, Brusselle G, Holgate S, et al. Treatable traits: toward precision medicine of chronic airway diseases. Eur Respir J. 2016;47(2):410–9.
353. Pavord ID, Wardlaw AJ. The A to E of airway disease. Clin Exp Allergy. 2010;40(1):62–7.
354. Lee JH, Rhee CK, Kim K, Kim JA, Kim SH, Yoo KH, et al. Identification of subtypes in subjects with mild-to-moderate airflow limitation and its clinical and socioeconomic

implications. Int J Chron Obs Pulmon Dis. 2017;12:1135–44.
355. Nadif R, Siroux V, Oryszczyn M-P, Ravault C, Pison C, Pin I, et al. Heterogeneity of asthma according to blood inflammatory patterns. Thorax. 2009 May;64(5):374–80.
356. Ortega H, Katz L, Gunsoy N, Keene O, Yancey S. Blood eosinophil counts predict treatment response in patients with severe eosinophilic asthma. J Allergy Clin Immunol. 2015;136(3):825–6.
357. Deeks ED. Mepolizumab: A Review in Eosinophilic Asthma. BioDrugs. 2016 Aug;30(4):361–70.
358. Fulkerson PC, Rothenberg ME. Targeting eosinophils in allergy, inflammation and beyond. Nat Rev Drug Discov. 2013 Feb;12(2):117–29.
359. Bel EH, Wenzel SE, Thompson PJ, Prazma CM, Keene ON, Yancey SW, et al. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. N Engl J Med. 2014 Sep;371(13):1189–97.
360. Hanania NA, Wenzel S, Rosen K, Hsieh H-J, Mosesova S, Choy DF, et al. Exploring the effects of omalizumab in allergic asthma: an analysis of biomarkers in the EXTRA study. Am J Respir Crit Care Med. 2013 Apr;187(8):804–11.
361. Busse W, Spector S, Rosen K, Wang Y, Alpan O. High eosinophil count: a potential biomarker for assessing successful omalizumab treatment effects. Vol. 132, The Journal of allergy and clinical immunology. United States; 2013. p. 485–6.e11.
362. Katz LE, Gleich GJ, Hartley BF, Yancey SW, Ortega HG. Blood eosinophil count is a useful biomarker to identify patients with severe eosinophilic asthma. Ann Am Thorac Soc. 2014 May;11(4):531–6.
363. Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, et al. Mepolizumab treatment in patients with severe eosinophilic asthma. N Engl J Med. 2014 Sep;371(13):1198–207.
364. FitzGerald JM, Bleecker ER, Nair P, Korn S, Ohta K, Lommatzsch M, et al. Benralizumab, an anti-interleukin-5 receptor alpha monoclonal antibody, as add-on treatment for patients with severe, uncontrolled, eosinophilic asthma (CALIMA): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet (London, England). 2016 Oct;388(10056):2128–41.
365. Corren J, Weinstein S, Janka L, Zangrilli J, Garin M. Phase 3 Study of Reslizumab in Patients With Poorly Controlled Asthma: Effects Across a Broad Range of Eosinophil Counts. Chest. 2016 Oct;150(4):799–810.
366. Bjermer L, Lemiere C, Maspero J, Weiss S, Zangrilli J, Germinaro M. Reslizumab for Inadequately Controlled Asthma With Elevated Blood Eosinophil Levels: A Randomized Phase 3 Study. Chest. 2016 Oct;150(4):789–98.
367. Brusselle G, Germinaro M, Weiss S, Zangrilli J. Reslizumab in patients with inadequately controlled late-onset asthma and elevated blood eosinophils. Pulm Pharmacol Ther. 2017 Apr;43:39–45.
368. Ortega HG, Yancey SW, Mayer B, Gunsoy NB, Keene ON, Bleecker ER, et al. Severe eosinophilic asthma treated with mepolizumab stratified by baseline eosinophil thresholds: a secondary analysis of the DREAM and MENSA studies. Lancet Respir Med. 2016 Jul;4(7):549–56.
369. Belda J, Giner J, Casan P, Sanchis J. Mild exacerbations and eosinophilic inflammation in patients with stable, well-controlled asthma after 1 year of follow-up. Chest. 2001;119(4):1011–7.
370. Ulrik CS. Peripheral eosinophil counts as a marker of disease activity in intrinsic and

extrinsic asthma. Clin Exp Allergy. 1995;25(9):820–7.
371. Bafadhel M, Peterson S, De Blas MA, Calverley PM, Rennard SI, Richter K, et al. Predictors of exacerbation risk and response to budesonide in patients with chronic obstructive pulmonary disease: a post-hoc analysis of three randomised trials. Lancet Respir Med [Internet]. 2018;6(2):117–26. Available from: http://www.sciencedirect.com/science/article/pii/S2213260018300067
372. Landis SH, Suruki R, Hilton E, Compton C, Galwey NW. Stability of Blood Eosinophil Count in Patients with COPD in the UK Clinical Practice Research Datalink. COPD. 2017 Aug;14(4):382–8.
373. Pascoe S, Pavord I, Hinds D, Locantore N, Barnes N. The association between blood eosinophils and risk and treatment outcome in COPD is not dichotomised. Lancet Respir Med [Internet]. 2018;6(5):e18. Available from: http://www.sciencedirect.com/science/article/pii/S2213260018301371
374. Hargreave FE, Nair P. The definition and diagnosis of asthma. Clin Exp Allergy. 2009;39(11):1652–8.
375. Reddel HK, Bateman ED, Becker A, Boulet L-P, Cruz AA, Drazen JM, et al. A summary of the new GINA strategy: a roadmap to asthma control. Eur Respir J. 2015;46(3):622–39.
376. Looijmans-Van den Akker I, van Luijn K, Verheij T. Overdiagnosis of asthma in children in primary care: a retrospective analysis. Br J Gen Pr. 2016;66(644):e152–7.
377. José BP de S, Camargos PAM, Cruz Filho ÁAS da, Corrêa R de A. Diagnostic accuracy of respiratory diseases in primary health units. Rev Assoc Med Bras. 2014;60(6):599–612.
378. Blakey JD, Price DB, Pizzichini E, Popov TA, Dimitrov BD, Postma DS, et al. Identifying Risk of Future Asthma Attacks Using UK Medical Record Data: A Respiratory Effectiveness Group Initiative. J allergy Clin Immunol Pract. 2017 Jul;5(4):1015–1024.e8.
379. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. Pharmacoepidemiol Drug Saf. 2013 Dec;22(12):1357–61.
380. Keeley D, Baxter N. Conflicting asthma guidelines cause confusion in primary care. BMJ Br Med J. 2018;360.
381. White J, Paton JY, Niven R, Pinnock H. Guidelines for the diagnosis and management of asthma: a look at the key differences between BTS/SIGN and NICE. Thorax. 2018;thoraxjnl-2017.
382. Holgate S, Agusti A, Strieter RM, Anderson GP, Fogel R, Bel E, et al. Drug development for airway diseases: looking forward. Nat Rev Drug Discov. 2015;14(6):367.
383. Anderson GP. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. Lancet. 2008;372(9643):1107–19.
384. Pavord ID. Complex airway disease: an approach to assessment and management. Lancet Respir Med. 2013;1(1):84–90.
385. Perlikos F, Hillas G, Loukides S. Phenotyping and Endotyping Asthma Based on Biomarkers. Curr Top Med Chem. 2016;16(14):1582–6.
386. Lotvall J, Akdis CA, Bacharier LB, Bjermer L, Casale TB, Custovic A, et al. Asthma endotypes: a new approach to classification of disease entities within the asthma

syndrome. J Allergy Clin Immunol. 2011 Feb;127(2):355–60.
387. Ortega H, Miller DP, Li H. Characterization of asthma exacerbations in primary care using cluster analysis. J Asthma. 2012;49(2):158–69.
388. Khusial RJ, Sont JK, Loijmans RJB, Snoeck-Stroband JB, Assendelft PJJ, Schermer TRJ, et al. Longitudinal outcomes of different asthma phenotypes in primary care, an observational study. npj Prim Care Respir Med [Internet]. 2017;27(1):55. Available from: https://doi.org/10.1038/s41533-017-0057-3

APPENDIX

This appendix contains the code lists that were used in the conduct of the research included in this thesis. The appendices of the previous chapters encompass materials specific to those chapters.

Included codelists:

- Codelist 1: Definite asthma
- Codelist 2: Possible asthma
- Codelist 3: Asthma medication
- Codelist 4: Oral corticosteroid codes
- Codelist 5: Antihistamines
- Codelist 6: COPD codes
- Codelist 7: Atopy
- Codelist 8: GORD
- Codelist 9: Anxiety
- Codelist 10: Depression

Codelist 1: Definite asthma

Medcode	Read Term
78	asthma
81	asthma monitoring
185	acute exacerbation of asthma
232	asthma attack
233	severe asthma attack
1555	bronchial asthma
2290	allergic asthma
3018	mild asthma
3366	severe asthma
3458	occasional asthma
3665	late onset asthma
4442	asthma unspecified
4606	exercise induced asthma
4892	status asthmaticus nos
5267	intrinsic asthma
5627	hay fever with asthma
5798	chronic asthmatic bronchitis

Medcode	Read Term
5867	exercise induced asthma
6707	extrinsic asthma with asthma attack
7058	emergency admission, asthma
7146	extrinsic (atopic) asthma
7191	asthma limiting activities
7378	asthma management plan given
7416	asthma disturbing sleep
7731	pollen asthma
8335	asthma attack nos
8355	asthma monitored
9018	number of asthma exacerbations in past year
9552	change in asthma management plan
9663	step up change in asthma management plan
10043	asthma annual review
10274	asthma medication review
10487	asthma - currently active
11370	asthma confirmed
12987	late-onset asthma
13064	asthma severity
13065	moderate asthma
13175	asthma disturbs sleep frequently
13176	asthma follow-up
14777	extrinsic asthma without status asthmaticus
15248	hay fever with asthma
16070	asthma nos
16667	asthma control step 2
16785	asthma control step 1
18223	step down change in asthma management plan
18224	asthma control step 3
18323	intrinsic asthma with asthma attack
19167	asthma monitoring by nurse
19519	asthma treatment compliance unsatisfactory
19520	asthma treatment compliance satisfactory
19539	asthma monitoring check done
20860	asthma control step 5
20886	asthma control step 4
21232	allergic asthma nec
22752	occupational asthma
24479	emergency asthma admission since last appointment
24506	further asthma - drug prevent.
24884	asthma causes daytime symptoms 1 to 2 times per week
25181	asthma restricts exercise
25791	asthma clinical management plan
26501	asthma never causes daytime symptoms
26503	asthma causes daytime symptoms most days

Medcode	Read Term
26504	asthma never restricts exercise
26506	asthma severely restricts exercise
26861	asthma sometimes restricts exercise
27926	extrinsic asthma with status asthmaticus
29325	intrinsic asthma without status asthmaticus
30458	asthma monitoring by doctor
30815	asthma causing night waking
31167	asthma night-time symptoms
31225	asthma causes daytime symptoms 1 to 2 times per month
38143	asthma never disturbs sleep
38144	asthma limits walking up hills or stairs
38145	asthma limits walking on the flat
38146	asthma disturbs sleep weekly
39478	wood asthma
39570	asthma causes night symptoms 1 to 2 times per month
40823	brittle asthma
41017	aspirin induced asthma
41020	absent from work or school due to asthma
42824	asthma daytime symptoms
45073	intrinsic asthma nos
45782	extrinsic asthma nos
46529	attends asthma monitoring
47337	asthma accident and emergency attendance since last visit
47684	detergent asthma
58196	intrinsic asthma with status asthmaticus
73522	work aggravated asthma
93353	sequoiosis (red-cedar asthma)
93736	royal college of physicians asthma assessment
98185	asthma control test
99793	patient has a written asthma personal action plan
100107	health education - asthma self management
100397	asthma control questionnaire
100509	under care of asthma specialist nurse
100740	health education - structured asthma discussion
102170	asthma review using roy colleg of physicians three questions
102209	mini asthma quality of life questionnaire
102301	asthma trigger - seasonal
102341	asthma trigger - pollen
102395	asthma causes symptoms most nights
102400	asthma causes night time symptoms 1 to 2 times per week
102449	asthma trigger - respiratory infection
102713	asthma limits activities 1 to 2 times per month
102871	asthma trigger - exercise
102888	asthma limits activities 1 to 2 times per week
102952	asthma trigger - warm air

Medcode	Read Term
103318	health education - structured patient focused asthma discuss
103321	asthma trigger - animals
103612	asthma never causes night symptoms
103631	royal college physician asthma assessment 3 question score
103813	asthma trigger - cold air
103944	asthma trigger - airborne dust
103945	asthma trigger - damp
103952	asthma trigger - emotion
103955	asthma trigger - tobacco smoke
103998	asthma limits activities most days
105420	asthma self-management plan review
105674	asthma self-management plan agreed
106805	chronic asthma with fixed airflow obstruction
107167	number days absent from school due to asthma in past 6 month

Codelist 2: Possible asthma

Medcode	Read Term
719	h/o: asthma
1208	childhood asthma
5138	patient in asthma study
7229	asthma prophylactic medication used
11022	asthma trigger
11387	refuses asthma monitoring
11673	excepted from asthma quality indicators: patient unsuitable
11695	excepted from asthma quality indicators: informed dissent
13066	asthma - currently dormant
13173	asthma not disturbing sleep
13174	asthma not limiting activities
16655	asthma monitoring admin.
18141	asthma monitoring due
18692	exception reporting: asthma quality indicators
18763	referral to asthma clinic
20422	asthma clinic administration
25705	asthma monitor 3rd letter
25706	asthma monitor 2nd letter
25707	asthma monitor 1st letter
25796	mixed asthma
26496	health education - asthma
29645	asthma control step 0
30308	dna - did not attend asthma clinic
30382	asthma monitoring admin.nos
31135	asthma monitor phone invite
35927	asthma leaflet given
37943	asthma monitor verbal invite
41554	asthma monitor offer default

Medcode	Read Term
43770	asthma society member
92109	asthma outreach clinic

Codelist 3: Asthma medication

prodcode	productname	groups
8	salbutamol 100micrograms/dose inhaler	SABA
17	salbutamol 100micrograms/dose inhaler cfc free	SABA
31	ventolin 100microgram/inhalation inhalation powder (glaxo wellcome uk ltd)	SABA
38	beclometasone 100micrograms/dose inhaler	ICS
44	prednisolone 5mg gastro-resistant tablets	OCS
95	prednisolone 5mg tablets	OCS
99	becotide 100 inhaler (glaxosmithkline uk ltd)	ICS
180	phyllocontin sup	THEOPH
218	aminophylline 100 mg cap	THEOPH
235	bricanyl 250micrograms/dose inhaler (astrazeneca uk ltd)	SABA
273	theophylline 200 mg cap	THEOPH
282	salbutamol 2mg/5ml oral solution sugar free	SABA
454	pulmicort 200microgram inhaler (astrazeneca uk ltd)	ICS
465	salmeterol 25micrograms/dose inhaler	LABA
510	ventolin 5mg/ml respirator solution (glaxosmithkline uk ltd)	SABA
534	atrovent 20micrograms/dose inhaler (boehringer ingelheim ltd)	SAMA
549	serevent 25micrograms/dose inhaler (glaxosmithkline uk ltd)	LABA
555	aminophylline 225mg modified-release tablets	THEOPH
556	combivent inhaler (boehringer ingelheim ltd)	SABA_SAMA
557	prednisolone 2.5mg gastro-resistant tablets	OCS
578	prednisolone 1mg tablets	OCS
590	phyllocontin continus 225mg tablets (napp pharmaceuticals ltd)	THEOPH
622	Montelukast 4mg chewable tablets sugar free	MONTELUKAST
638	seretide 250 accuhaler (glaxosmithkline uk ltd)	LABA_ICS
665	seretide 100 accuhaler (glaxosmithkline uk ltd)	LABA_ICS
674	ventolin 2.5mg nebules (glaxosmithkline uk ltd)	SABA
696	salbutamol 8mg modified-release capsules	SABA
719	salmeterol 50micrograms/dose dry powder inhaler	LABA
746	tiotropium 18 microgram capsule	LAMA
808	Montelukast 10mg tablets	MONTELUKAST
856	ventolin 2mg/5ml syrup (glaxosmithkline uk ltd)	SABA
860	salbutamol 4mg tablets	SABA
862	salbulin inhalation powder (3m health care ltd)	SABA
863	slo-phyllin 125mg capsule (lipha pharmaceuticals ltd)	THEOPH
879	theophylline 125mg modified-release capsules	THEOPH
880	theophylline 60mg modified-release capsules	THEOPH
881	salbutamol 2mg tablets	SABA
882	salbutamol 200microgram inhalation powder capsules	SABA
883	becodisks 200microgram disc (allen & hanburys ltd)	ICS
895	beclazone 100 easi-breathe inhaler (teva uk ltd)	ICS
896	becotide easi-breathe 100microgram/actuation pressurised inhalation (allen & hanburys ltd)	ICS

prodcode	productname	groups
898	ventolin evohaler 100 100microgram/inhalation pressurised inhalation (glaxo wellcome uk ltd)	SABA
907	bricanyl turbohaler 500 500microgram turbohaler (astrazeneca uk ltd)	SABA
908	pulmicort 400 turbohaler (astrazeneca uk ltd)	ICS
909	budesonide 200micrograms/dose inhaler	ICS
910	serevent diskhaler 50microgram inhalation powder (glaxo wellcome uk ltd)	LABA
911	flixotide accuhaler 250 250microgram/inhalation inhalation powder (allen & hanburys ltd)	ICS
947	budesonide 50micrograms/actuation refill canister	ICS
955	prednisolone 5mg soluble tablets	OCS
956	pulmicort 200 turbohaler (astrazeneca uk ltd)	ICS
957	salamol easi-breathe 100microgram/actuation pressurised inhalation (ivax pharmaceuticals uk ltd)	SABA
958	ventolin easi-breathe 100microgram/actuation pressurised inhalation (allen & hanburys ltd)	SABA
959	budesonide 50micrograms/dose inhaler	ICS
960	pulmicort 100 turbohaler (astrazeneca uk ltd)	ICS
987	ventolin 4mg tablet (allen & hanburys ltd)	SABA
1063	prednesol 5mg tablet (sovereign medical ltd)	OCS
1087	asmasal 95micrograms/dose clickhaler (focus pharmaceuticals ltd)	SABA
1093	salamol 100microgram/actuation inhalation powder (ivax pharmaceuticals uk ltd)	SABA
1097	slo-phyllin 60mg capsule (lipha pharmaceuticals ltd)	THEOPH
1100	beclazone 100 inhaler (teva uk ltd)	ICS
1236	becloforte 250micrograms/dose inhaler (glaxosmithkline uk ltd)	ICS
1242	beclometasone 250micrograms/dose inhaler	ICS
1243	beclazone 250 easi-breathe inhaler (teva uk ltd)	ICS
1258	becotide 200 inhaler (glaxosmithkline uk ltd)	ICS
1259	beclometasone 200micrograms/dose inhaler	ICS
1406	becotide 50 inhaler (glaxosmithkline uk ltd)	ICS
1409	ipratropium bromide 20micrograms/dose inhaler	SAMA
1410	ipratropium bromide 0.25mg/ml	SAMA
1411	ipratropium bromide 250micrograms/ml	SAMA
1412	flixotide 250microgram/actuation inhalation powder (allen & hanburys ltd)	ICS
1423	uniphyllin continus 200mg tablets (napp pharmaceuticals ltd)	THEOPH
1424	flixotide 250microgram disc (allen & hanburys ltd)	ICS
1426	flixotide 500microgram disc (allen & hanburys ltd)	ICS
1518	flixotide 50microgram/actuation inhalation powder (allen & hanburys ltd)	ICS
1537	becotide 200microgram rotacaps (glaxosmithkline uk ltd)	ICS
1551	beclazone 250 inhaler (teva uk ltd)	ICS
1552	becloforte easi-breathe 250microgram/actuation pressurised inhalation (allen & hanburys ltd)	ICS
1619	terbutaline 500micrograms/dose dry powder inhaler	SABA
1620	terbutaline 250micrograms/dose inhaler	SABA
1628	terbutaline 250micrograms/actuation refill canister	SABA
1635	salbuvent 2mg/5ml oral solution (pharmacia ltd)	SABA
1642	budesonide 400micrograms/dose dry powder inhaler	ICS

prodcode	productname	groups
	flixotide 125microgram/actuation inhalation powder (allen & hanburys	
1676	ltd)	ICS
1680	pulmicort ls 50micrograms/dose inhaler (astrazeneca uk ltd)	ICS
1697	atrovent 20micrograms/dose autohaler (boehringer ingelheim ltd)	SAMA
1698	salbutamol 100micrograms/dose breath actuated inhaler	SABA
1725	beclazone 50 easi-breathe inhaler (teva uk ltd)	ICS
	becotide easi-breathe 50microgram/actuation pressurised inhalation	
1727	(allen & hanburys ltd)	ICS
1734	beclometasone 100micrograms/dose breath actuated inhaler	ICS
1741	salbutamol 100micrograms/dose breath actuated inhaler cfc free	SABA
	berotec 100microgram/actuation inhalation powder (boehringer	
1794	ingelheim ltd)	SABA
1801	ventide inhaler (glaxosmithkline uk ltd)	SABA_ICS
1832	theograd 350mg tablet (abbott laboratories ltd)	THEOPH
1833	theophylline 200mg modified-release tablets	THEOPH
1834	theophylline 400mg modified-release tablets	THEOPH
1861	aerobec 100 autohaler (meda pharmaceuticals ltd)	ICS
1882	ventodisks 200microgram/blister disc (allen & hanburys ltd)	SABA
1885	beclazone 200 inhaler (teva uk ltd)	ICS
1950	ventodisks 400microgram/blister disc (allen & hanburys ltd)	SABA
1951	becodisks 400microgram disc (allen & hanburys ltd)	ICS
1952	ventolin 400microgram rotacaps (glaxosmithkline uk ltd)	SABA
1956	pulmicort 1mg respules (astrazeneca uk ltd)	ICS
1957	ventolin 5mg nebules (glaxosmithkline uk ltd)	SABA
1959	pulmicort 0.5mg respules (astrazeneca uk ltd)	ICS
1960	volmax 8mg modified-release tablets (glaxosmithkline uk ltd)	SABA
1961	volmax 4mg modified-release tablets (glaxosmithkline uk ltd)	SABA
1974	oxis 12 turbohaler (astrazeneca uk ltd)	LABA
1975	oxis 6 turbohaler (astrazeneca uk ltd)	LABA
2020	berotec 200micrograms/dose inhaler (boehringer ingelheim ltd)	SABA
2044	prednisone 2.5 mg tab	OCS
2092	budesonide 200micrograms/dose dry powder inhaler	ICS
2124	pulmicort refil 200 mcg inh	ICS
2125	pulmicort 200microgram refill canister (astrazeneca uk ltd)	ICS
2147	theophylline 250mg modified-release capsules	THEOPH
2148	beclometasone 400microgram disc	ICS
2149	steri-neb salamol 2.5 mg inh	SABA
2152	ipratropium bromide with salbutamol 20mcg + 100mcg	SABA_SAMA
2159	aerobec 50 autohaler (meda pharmaceuticals ltd)	ICS
2160	beclometasone 50micrograms/dose breath actuated inhaler	ICS
2224	serevent 50micrograms/dose accuhaler (glaxosmithkline uk ltd)	LABA
2229	becodisks 100microgram disc (allen & hanburys ltd)	ICS
2282	fluticasone propionate 500micrograms/dose dry powder inhaler	ICS
2335	qvar 100 inhaler (teva uk ltd)	ICS
2368	prednisolone 2.5mg tablet	OCS
2390	prednisolone e/c 1 mg tab	OCS
2395	salbutamol 2 mg/5ml syr	SABA
2437	oxitropium bromide 100micrograms/dose inhaler	SAMA
	flixotide accuhaler 500 500microgram/inhalation inhalation powder (allen	
2440	& hanburys ltd)	ICS

prodcode	productname	groups
2600	beclometasone 250micrograms/dose breath actuated inhaler	ICS
2655	airomir 100micrograms/dose inhaler (teva uk ltd)	SABA
2704	prednisolone 25mg tablets	OCS
2722	duovent inhaler (boehringer ingelheim ltd)	SABA_SAMA
2723	fluticasone 25micrograms/dose inhaler	ICS
2757	slo-phyllin 250mg capsule (lipha pharmaceuticals ltd)	THEOPH
2758	bricanyl refill canister (astrazeneca uk ltd)	SABA
2799	prednisolone 10 mg tab	OCS
2850	salbutamol 400microgram inhalation powder capsules	SABA
2851	ventolin 200microgram rotacaps (glaxosmithkline uk ltd)	SABA
2862	duovent autohaler (boehringer ingelheim ltd)	SABA_SAMA
2869	salbutamol 8mg modified-release tablets	SABA
2892	becloforte 400microgram disks (glaxosmithkline uk ltd)	ICS
2893	beclometasone 200micrograms disc	ICS
2949	prednisone 5mg tablets	OCS
2951	fluticasone 250microgram/actuation pressurised inhalation	ICS
2978	salbutamol 200micrograms/dose dry powder inhaler	SABA
2992	beclazone 50 inhaler (teva uk ltd)	ICS
2994	atrovent aerocaps 40microgram inhalation powder (boehringer ingelheim ltd)	SAMA
2995	nuelin sa 175mg tablets (meda pharmaceuticals ltd)	THEOPH
3018	beclometasone 50micrograms/dose inhaler	ICS
3039	oxivent 100micrograms/dose inhaler (boehringer ingelheim ltd)	SAMA
3059	prednisolone 50 mg tab	OCS
3065	bextasol inhalation powder (allen & hanburys ltd)	ICS
3075	becotide 400microgram rotacaps (glaxosmithkline uk ltd)	ICS
3119	becloforte integra 250microgram/actuation inhaler with compact spacer (glaxo laboratories ltd)	ICS
3150	beclometasone 100micrograms/actuation extrafine particle cfc free inhaler	ICS
3163	salbutamol 200micrograms disc	SABA
3188	pulmicort complete 50 mcg inh	ICS
3189	salbuvent inh inh	SABA
3220	qvar 50 autohaler (teva uk ltd)	ICS
3254	salbulin 4mg tablet (3m health care ltd)	SABA
3289	flixotide 25micrograms/dose inhaler (glaxosmithkline uk ltd)	ICS
3297	salmeterol 50micrograms disc	LABA
3306	atrovent forte 40micrograms/dose inhaler (boehringer ingelheim ltd)	SAMA
3345	sintisone tablet (pharmacia ltd)	OCS
3363	becloforte 400microgram disks with diskhaler (glaxosmithkline uk ltd)	ICS
3388	theophylline 175mg modified-release tablets	THEOPH
3437	becotide rotahaler type 4 insufflator inhalation powder (allen and hanburys ltd)	ICS
3442	pulmicort complete 200 mcg inh	ICS
3443	salbutamol 100microgram/inhalation spacehaler (celltech pharma europe ltd)	SABA
3534	bricanyl 5mg tablets (astrazeneca uk ltd)	SABA
3546	qvar 50 inhaler (teva uk ltd)	ICS
3556	beclometasone 50micrograms with salbutamol 100micrograms/inhalation inhaler	SABA_ICS
3557	prednisone 1mg tablets	OCS

prodcode	productname	groups
3570	budesonide 200micrograms/actuation refill canister	ICS
3584	bricanyl 1.5mg/5ml syrup (astrazeneca uk ltd)	SABA
3666	seretide 500 accuhaler (glaxosmithkline uk ltd)	LABA_ICS
3743	filair 50 inhaler (meda pharmaceuticals ltd)	ICS
3753	flixotide diskhaler-community pack 250 mcg	ICS
3758	pulmadil inhalation powder (3m health care ltd)	SABA
3763	terbutaline respules inh	SABA
3764	bricanyl respules (5mg/2ml) 2.5 mg/ml inh	SABA
3786	fenoterol 100micrograms/dose / ipratropium 40micrograms/dose inhaler	SABA_SAMA
3838	salbutamol 400mcg/beclometh.100mcg r/cap inh	SABA
3850	oxivent 100micrograms/dose autohaler (boehringer ingelheim ltd)	SAMA
3927	filair 100 inhaler (meda pharmaceuticals ltd)	ICS
3947	becotide 100microgram rotacaps (glaxosmithkline uk ltd)	ICS
3988	flixotide diskhaler-community pack 100 mcg	ICS
3989	flixotide 100microgram disc (allen & hanburys ltd)	ICS
3993	filair forte 250micrograms/dose inhaler (meda pharmaceuticals ltd)	ICS
3994	salbutamol 4mg modified-release tablets	SABA
4055	salbulin 2mg/5ml oral solution (3m health care ltd)	SABA
4131	fluticasone 100microgram disc	ICS
4132	fluticasone 125microgram/actuation pressurised inhalation	ICS
4171	ventolin 2mg tablet (allen & hanburys ltd)	SABA
4222	bricanyl 10mg/ml respirator solution (astrazeneca uk ltd)	SABA
4268	ipratropium bromide 40micrograms/dose inhaler	SAMA
4365	beclometasone 100micrograms disc	ICS
4413	qvar 100 autohaler (teva uk ltd)	ICS
4497	ventolin accuhaler 200 200microgram/actuation inhalation powder (glaxo wellcome uk ltd)	SABA
4499	aerobec 250microgram/actuation pressurised inhalation (meda pharmaceuticals ltd)	ICS
4514	aminophylline 350mg modified-release tablets	THEOPH
4541	bricanyl sa 7.5mg tablets (astrazeneca uk ltd)	SABA
4545	pulmicort ls 50microgram refill canister (astrazeneca uk ltd)	ICS
4593	theophylline 125mg tablets	THEOPH
4601	asmabec 100 clickhaler (focus pharmaceuticals ltd)	ICS
4665	salbulin 100micrograms/dose inhaler (3m health care ltd)	SABA
4688	fluticasone 50microgram/actuation pressurised inhalation	ICS
4759	beclometasone 100microgram inhalation powder capsules	ICS
4803	beclazone 250microgram/actuation inhalation powder (actavis uk ltd)	ICS
4842	fenoterol 100microgram/actuation inhaler	SABA
4908	ventolin rotahaler (glaxosmithkline uk ltd)	SABA
4926	flixotide accuhaler 100 100microgram/inhalation inhalation powder (allen & hanburys ltd)	ICS
5143	seretide 50 evohaler (glaxosmithkline uk ltd)	LABA_ICS
5161	seretide 125 evohaler (glaxosmithkline uk ltd)	LABA_ICS
5170	salamol 100micrograms/dose inhaler cfc free (teva uk ltd)	SABA
5172	seretide 250 evohaler (glaxosmithkline uk ltd)	LABA_ICS
5185	fenoterol 200micrograms/dose inhaler	SABA
5223	fluticasone 50micrograms/dose inhaler cfc free	ICS
5261	nuelin sa 250 tablets (meda pharmaceuticals ltd)	THEOPH
5309	flixotide 50micrograms/dose evohaler (glaxosmithkline uk ltd)	ICS

prodcode	productname	groups
5453	uniphyllin continus 400mg tablets (napp pharmaceuticals ltd)	THEOPH
5490	deltacortril 5mg gastro-resistant tablets (alliance pharmaceuticals ltd)	OCS
5516	salamol 100micrograms/dose easi-breathe inhaler (teva uk ltd)	SABA
5521	beclometasone 200micrograms/dose dry powder inhaler	ICS
5522	beclometasone 100micrograms/dose dry powder inhaler	ICS
5551	flixotide 0.5mg/2ml nebules (glaxosmithkline uk ltd)	ICS
5558	salmeterol 50micrograms with fluticasone 500micrograms cfc free inhaler	LABA_ICS
5580	flixotide accuhaler 50 50microgram/inhalation inhalation powder (allen & hanburys ltd)	ICS
5683	flixotide 250micrograms/dose evohaler (glaxosmithkline uk ltd)	ICS
5718	flixotide 125micrograms/dose evohaler (glaxosmithkline uk ltd)	ICS
5740	airomir 100micrograms/dose autohaler (teva uk ltd)	SABA
5753	salbutamol 400micrograms disc	SABA
5804	beclometasone 250micrograms/dose dry powder inhaler	ICS
5822	fluticasone 250micrograms/dose inhaler cfc free	ICS
5864	salmeterol 25micrograms with fluticasone 250micrograms cfc free inhaler	LABA_ICS
5885	fluticasone propionate 100micrograms/dose dry powder inhaler	ICS
5889	salamol 100microgram/inhalation inhalation powder (kent pharmaceuticals ltd)	SABA
5913	deltacortril 2.5mg gastro-resistant tablets (alliance pharmaceuticals ltd)	OCS
5941	uniphyllin continus 300mg tablets (napp pharmaceuticals ltd)	THEOPH
5942	salmeterol 50micrograms with fluticasone 250micrograms cfc free inhaler	LABA_ICS
5957	Montelukast 5mg chewable tablets sugar free	MONTELUKAST
5975	fluticasone 125micrograms/dose inhaler cfc free	ICS
5992	beclometasone 50micrograms/dose dry powder inhaler	ICS
6050	spiriva 18 microgram capsule (boehringer ingelheim ltd)	LAMA
6081	ipratropium bromide 20micrograms/dose breath actuated inhaler	SAMA
6315	slo-phyllin 250mg capsules (merck serono ltd)	THEOPH
6325	symbicort 200/6 turbohaler (astrazeneca uk ltd)	LABA_ICS
6462	salbutamol 95micrograms/dose dry powder inhaler	SABA
6474	robinul 1mg tablet (idis world medicines)	LAMA
6512	atrovent 20micrograms/dose inhaler cfc free (boehringer ingelheim ltd)	SAMA
6522	ipratropium bromide 20micrograms/dose inhaler cfc free	SAMA
6526	formoterol 12microgram inhalation powder capsules with device	LABA
6569	salmeterol 25micrograms with fluticasone 125micrograms cfc free inhaler	LABA_ICS
6616	salmeterol 25micrograms with fluticasone 50micrograms cfc free inhaler	LABA_ICS
6746	budesonide 400micrograms/dose / formoterol 12micrograms/dose dry powder inhaler	LABA_ICS
6780	symbicort 400/12 turbohaler (astrazeneca uk ltd)	LABA_ICS
6796	budesonide 200micrograms/dose / formoterol 6micrograms/dose dry powder inhaler	LABA_ICS
6938	salmeterol 50micrograms with fluticasone 100micrograms dry powder inhaler	LABA_ICS
6988	aminophylline hydrate 100mg modified-release tablets	THEOPH
7013	symbicort 100/6 turbohaler (astrazeneca uk ltd)	LABA_ICS
7017	salbutamol 100micrograms/dose dry powder inhaler	SABA
7088	Montelukast 4mg granules sachets sugar free	MONTELUKAST
7133	formoterol 12micrograms/dose dry powder inhaler	LABA
7192	bambuterol 10mg tablets	SABA
7218	glycopyrronium bromide 1mg tablets	LAMA

prodcode	productname	groups
7268	serevent 25micrograms/dose evohaler (glaxosmithkline uk ltd)	LABA
7270	salmeterol 25micrograms/dose inhaler cfc free	LABA
7452	ventolin .25 mg inj	SABA
7550	Omalizumab 150mg powder and solvent for solution for injection vials	OMALIZUMAB
7584	prednisolone 4 mg tab	OCS
7597	glycopyrronium bromide 2mg tablets	LAMA
7602	fluticasone 50microgram disc	ICS
7638	fluticasone 250microgram disc	ICS
7653	beclometasone 400microgram inhalation powder capsules	ICS
7710	prednisolone 15 mg tab	OCS
7711	terbutaline 250micrograms/dose inhaler with spacer	SABA
7724	betamethasone valerate 100micrograms/actuation inhaler	ICS
7730	theo-dur 300mg tablets (astrazeneca uk ltd)	THEOPH
7731	theo-dur 200mg tablets (astrazeneca uk ltd)	THEOPH
7732	theophylline 300mg modified-release tablets	THEOPH
7733	theophylline 250mg modified-release tablets	THEOPH
7788	budesonide 100micrograms/dose dry powder inhaler	ICS
7832	choline theophyllinate 200mg tablets	THEOPH
7841	nuelin 125mg tablets (3m health care ltd)	THEOPH
7891	fluticasone 500microgram disc	ICS
7908	robinul 2mg tablet (wyeth pharmaceuticals)	LAMA
7934	prednisone 30 mg tab	OCS
7935	maxivent 100microgram/inhalation inhalation powder (ashbourne pharmaceuticals ltd)	SABA
7948	fluticasone propionate 250micrograms/dose dry powder inhaler	ICS
7953	terbutaline 1.5mg/5ml oral solution sugar free	SABA
7954	bricanyl 250micrograms/dose spacer inhaler (astrazeneca uk ltd)	SABA
8012	exirel 15mg capsule (3m health care ltd)	SABA
8056	aminophylline 100mg tablets	THEOPH
8057	aminophylline 100mg modified-release tablets	THEOPH
8111	becloforte vm 250microgram/actuation vm pack (allen & hanburys ltd)	ICS
8251	pulmicort refil 50 mg inh	ICS
8252	pirbuterol 15mg capsule	SABA
8267	sodium cromoglicate 1mg/dose / salbutamol 100micrograms/dose inhaler	SABA_CROMO
8333	ipratropium bromide 40microgram inhalation powder capsules	SAMA
8339	fenoterol hydrobromide complete unit inh	SABA
8365	moxisylyte 40mg tablets	LABA
8429	ventolin i/v 5 mg inj	SABA
8433	budesonide 100micrograms/actuation inhaler	ICS
8450	flixotide diskhaler-community pack 50 mcg	ICS
8470	aminophylline 225 mg sup	THEOPH
8504	exirel 15 mg tab	SABA
8522	terbutaline 7.5mg modified-release tablets	SABA
8572	rimiterol inhaler	SABA
8610	aminophylline 1 gm sup	THEOPH
8635	flixotide 50microgram disc (allen & hanburys ltd)	ICS
8636	ventolin s/r 8 mg spa	SABA
8653	aminophylline 360 mg sup	THEOPH
8806	phyllocontin continus 350mg tablet (napp pharmaceuticals ltd)	THEOPH
8955	theophylline 100 mg tab	THEOPH

prodcode	productname	groups
9092	theophylline 350mg modified release tablets	THEOPH
9164	fluticasone propionate 50micrograms/dose dry powder inhaler	ICS
9233	beclometasone 200microgram inhalation powder capsules	ICS
9270	ipratropium bromide with fenoterol hydrobromide 500micrograms + 1.25mg/4ml	SABA_SAMA
9356	becotide rotahaler insufflator inhalation powder (allen and hanburys ltd)	ICS
9384	salbutamol 4mg modified-release capsules	SABA
9477	asmabec 100microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
9571	beclometasone 250micrograms/actuation vortex inhaler	ICS
9577	asmabec 50 clickhaler (focus pharmaceuticals ltd)	ICS
9599	beclazone 50microgram/actuation inhalation powder (actavis uk ltd)	ICS
9651	asmasal 100microgram/inhalation spacehaler (celltech pharma europe ltd)	SABA
9658	oxitropium bromide 100micrograms/dose breath actuated inhaler	SAMA
9681	atrovent aerohaler 40microgram inhalation powder (boehringer ingelheim ltd)	SAMA
9711	formoterol 6micrograms/dose dry powder inhaler	LABA
9727	prednisolone 50mg tablets	OCS
9805	salbutamol 5mg/50ml solution for infusion vials	SABA
9921	beclometasone 100micrograms/dose breath actuated inhaler cfc free	ICS
10090	beclometasone 50micrograms/actuation extrafine particle cfc free inhaler	ICS
10218	budesonide 100micrograms/dose / formoterol 6micrograms/dose dry powder inhaler	LABA_ICS
10254	mometasone 400micrograms/dose dry powder inhaler	ICS
10289	aminophylline 200 mg sup	THEOPH
10321	budesonide 400microgram inhalation powder capsules	ICS
10331	nuelin 60mg/5ml liquid (3m health care ltd)	THEOPH
10353	salbuvent rondo	SABA
10360	aerocrom inhaler (castlemead healthcare ltd)	SABA_CROMO
10407	phyllocontin paediatric continus 100mg tablets (napp pharmaceuticals ltd)	THEOPH
10432	theophylline 300 mg sup	THEOPH
10433	theophylline 60mg/5ml oral solution	THEOPH
10458	ventolin cr 4mg tablet (allen & hanburys ltd)	SABA
10672	opilon 40mg tablet (concord pharmaceuticals ltd)	LABA
10723	theophylline 125mg/5ml syrup	THEOPH
10744	theophylline 80 mg eli	THEOPH
10825	terbutaline 5mg tablets	SABA
10831	biophylline 125mg/5ml oral solution (lorex synthelabo ltd)	THEOPH
10858	pulmadil auto inhalation powder (3m health care ltd)	SABA
10958	salbutamol .25 mg inj	SABA
10968	foradil 12microgram inhalation powder capsules with device (novartis pharmaceuticals uk ltd)	LABA
11046	ipratropium bromide with salbutamol 500micrograms + 2.5mg/2.5ml	SABA_SAMA
11149	betnelan 500microgram tablets (focus pharmaceuticals ltd)	ICS
11198	beclometasons 50 micrograms/actuation vortex inhaler	ICS
11307	salbutamol 100micrograms/dose / beclometasone 50micrograms/dose inhaler	SABA_ICS
11410	fluticasone propionate 500micrograms/dose / salmeterol 50micrograms/dose dry powder inhaler	LABA_ICS

prodcode	productname	groups
11497	beclometasone 400micrograms/dose dry powder inhaler	ICS
11588	fluticasone 125micrograms/dose / salmeterol 25micrograms/dose inhaler cfc free	LABA_ICS
11618	fluticasone 250micrograms/dose / salmeterol 25micrograms/dose inhaler cfc free	LABA_ICS
11719	slo-phyllin 60mg capsules (merck serono ltd)	THEOPH
11732	beclometasone 50micrograms/dose breath actuated inhaler cfc free	ICS
11779	ipratropium bromide 40microgram inhalation powder capsules with device	SAMA
11993	pro-vent 300mg capsule (wellcome medical division)	THEOPH
12042	ventolin cr 8mg tablet (allen & hanburys ltd)	SABA
12144	bambuterol 20mg tablets	SABA
12240	theophylline 300mg modified release capsules	THEOPH
12463	pirbuterol 15 mg tab	SABA
12486	bronchodil 500microgram/dose inhalation powder (viatris pharmaceuticals ltd)	SABA
12563	exirel inhalation powder (3m health care ltd)	SABA
12699	pecram 225mg modified-release tablet (novartis consumer health uk ltd)	THEOPH
12808	fenoterol 100micrograms/dose / ipratropium bromide 40micrograms/dose breath actuated inhaler	SABA_SAMA
12822	salbutamol 2.5mg with ipratropium bromide 500micrograms/2.5ml unit dose nebuilser solution	SABA_SAMA
12909	salbutamol 100micrograms/dose / ipratropium 20micrograms/dose inhaler	SABA_SAMA
12994	fluticasone 50micrograms/dose / salmeterol 25micrograms/dose inhaler cfc free	LABA_ICS
13037	pulvinal beclometasone dipropionate 200micrograms/dose dry powder inhaler (chiesi ltd)	ICS
13038	pulvinal salbutamol 200micrograms/dose dry powder inhaler (chiesi ltd)	SABA
13040	fluticasone propionate 250micrograms/dose / salmeterol 50micrograms/dose dry powder inhaler	LABA_ICS
13181	easyhaler salbutamol sulfate 100micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	SABA
13273	fluticasone propionate 100micrograms/dose / salmeterol 50micrograms/dose dry powder inhaler	LABA_ICS
13290	clenil modulite 100micrograms/dose inhaler (chiesi ltd)	ICS
13522	prednisolone 2 mg tab	OCS
13529	amnivent-225 sr tablets (ashbourne pharmaceuticals ltd)	THEOPH
13575	bambec 20mg tablets (astrazeneca uk ltd)	SABA
13615	prednisone 10 mg tab	OCS
13815	beclazone 100microgram/actuation inhalation powder (actavis uk ltd)	ICS
13996	salamol 100microgram/inhalation inhalation powder (sandoz ltd)	SABA
14294	qvar 50micrograms/dose easi-breathe inhaler (teva uk ltd)	ICS
14306	formoterol 12micrograms/dose inhaler cfc free	LABA
14321	beclometasone 200micrograms/dose inhaler cfc free	ICS
14482	bricanyl 2.5 mg inj	SABA
14524	bdp 250microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
14525	salbutamol 100micrograms/inhalation vortex inhaler	SABA
14527	bambec 10mg tablets (astrazeneca uk ltd)	SABA

prodcode	productname	groups
14561	salbutamol 400microgram / beclometasone 200microgram inhalation powder capsules	SABA_ICS
14567	asmabec 250 clickhaler (focus pharmaceuticals ltd)	ICS
14590	asmabec 250microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
14700	budesonide 400micrograms/actuation inhaler	ICS
14736	pulvinal beclometasone dipropionate 400micrograms/dose dry powder inhaler (chiesi ltd)	ICS
14739	norphyllin sr 225mg tablets (teva uk ltd)	THEOPH
14757	pulvinal beclometasone dipropionate 100micrograms/dose dry powder inhaler (chiesi ltd)	ICS
15025	aminophylline 25 mg sup	THEOPH
15075	bronchodil 20mg tablet (viatris pharmaceuticals ltd)	SABA
15165	reproterol 500micrograms/dose inhaler	SABA
15284	slo-phyllin 125mg capsules (merck serono ltd)	THEOPH
15326	beclometasone 100micrograms/dose inhaler cfc free	ICS
15365	theophylline 10mg/5ml sf elixir	THEOPH
15409	theophylline 3 mg sol	THEOPH
15441	fenoterol hydrobromide .5 % sol	SABA
15483	bricanyl oral solution (astrazeneca uk ltd)	SABA
15706	beclometasone 100 micrograms/actuation vortex inhaler	ICS
16018	mometasone 200micrograms/dose dry powder inhaler	ICS
16054	budesonide 200micrograms/actuation breath actuated powder inhaler	ICS
16148	clenil modulite 250micrograms/dose inhaler (chiesi ltd)	ICS
16151	clenil modulite 200micrograms/dose inhaler (chiesi ltd)	ICS
16158	clenil modulite 50micrograms/dose inhaler (chiesi ltd)	ICS
16236	pirbuterol acetate inhaler	SABA
16305	flixotide 2mg/2ml nebules (glaxosmithkline uk ltd)	ICS
16577	easyhaler salbutamol sulfate 200micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	SABA
16584	beclometasone 50micrograms/dose inhaler cfc free	ICS
16625	ventide rotacaps (glaxosmithkline uk ltd)	SABA_ICS
16724	prednisone 50 mg tab	OCS
16994	aminophylline hydrate 350mg modified-release tablets	THEOPH
17002	aminophylline hydrate 225mg modified-release tablets	THEOPH
17140	aminophylline 200mg tablets	THEOPH
17654	easyhaler beclometasone 200micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	ICS
17670	easyhaler budesonide 100micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	ICS
17696	ventmax sr 4mg capsules (chiesi ltd)	SABA
17874	monovent 1.5mg/5ml oral solution (lagap)	SABA
17875	terbutaline with guafenesin expectorant	SABA
17901	bricanyl nebule 2.5 ml	SABA
18140	respontin 500micrograms/2ml nebules (glaxosmithkline uk ltd)	SAMA
18288	choline theophyllinate 100mg tablets	THEOPH
18308	aminophylline 100 mg sup	THEOPH
18314	aerocrom syncroner with spacer (castlemead healthcare ltd)	SABA_CROMO
18394	bdp 50microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
18456	salbutamol 200microgram / beclometasone 100microgram inhalation powder capsules	SABA_ICS

prodcode	productname	groups
18484	ventide paediatric rotacaps (glaxosmithkline uk ltd)	SABA_ICS
18537	budesonide 200microgram inhalation powder capsules	ICS
18622	salbulin 2mg tablet (3m health care ltd)	SABA
18848	qvar 100micrograms/dose easi-breathe inhaler (teva uk ltd)	ICS
18937	sabidal sr 270 270 mg tab	THEOPH
18968	salbutamol 5mg/5ml solution for infusion ampoules	SABA
18988	choline theophyllinate 62.5mg/5ml oral solution	THEOPH
19031	bdp 100microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
19121	beclometasone 100micrograms with salbutamol 200micrograms inhalation capsules	SABA_ICS
19141	prednisolone 5mg soluble tablets (amco)	OCS
19350	aminophylline 62.5 mg sup	THEOPH
19376	beclometasone 200micrograms with salbutamol 400micrograms inhalation capsules	SABA_ICS
19389	asmabec 50microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
19401	beclometasone 250micrograms/actuation inhaler and compact spacer	ICS
19642	ventolin nebules	SABA
19649	ventolin rotahaler	SABA
19653	ventolin respirator	SABA
19726	ventolin s/r	SABA
19732	cobutolin inh	SABA
19735	uniphyllin continus	THEOPH
19736	becotide susp for nebulisation	ICS
19799	tulobuterol 2mg	LABA
19805	atrovent	SAMA
20095	precortisyl forte 25mg tablet (aventis pharma)	OCS
20171	aminophylline 180 mg sup	THEOPH
20225	aminophylline 500 mg inj	THEOPH
20670	prednisolone e/c	OCS
20675	salbutamol rotahaler complete unit	SABA
20707	becotide 100	ICS
20720	atrovent forte	SAMA
20763	becloforte	ICS
20781	salbutamol u.dose nebulising 2.5mg/2.5ml	SABA
20812	pulmicort refill	ICS
20825	spacehaler bdp 250microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
20838	salbuvent 2mg tablet (pharmacia ltd)	SABA
21005	beclometasone 250micrograms/dose inhaler cfc free	ICS
21102	salbutamol 2mg/5ml oral solution (lagap)	SABA
21417	prednisolone 5mg tablets (a a h pharmaceuticals ltd)	OCS
21482	beclometasone 100micrograms/dose inhaler (generics (uk) ltd)	ICS
21769	lasma 300mg tablet (pharmax ltd)	THEOPH
21833	decortisyl 5mg tablet (roussel laboratories ltd)	OCS
21859	asmaven 100microgram inhalation powder (berk pharmaceuticals ltd)	SABA
22080	aminophylline 20 ml inj	THEOPH
22225	beclomethasone /salbutamol	SABA
22313	ventmax sr 8mg capsules (chiesi ltd)	SABA
22430	spacehaler salbutamol 100microgram/inhalation spacehaler (celltech pharma europe ltd)	SABA

prodcode	productname	groups
22467	salbutamol respirator soln	SABA
22512	salbutamol inhaler	SABA
22550	duovent	SABA
22661	pirbuterol 10mg capsule	SABA
22663	respacal 2mg tablet (ucb pharma ltd)	LABA
22669	choline theophyllinate 270 mg tab	THEOPH
22790	reproterol 10mg/ml respirator solution	SABA
23512	precortisyl 5mg tablet (hoechst marion roussel)	OCS
23567	respontin 250micrograms/1ml nebules (glaxosmithkline uk ltd)	SAMA
23572	aminophylline sr 225mg modified-release tablet (ivax pharmaceuticals uk ltd)	THEOPH
23675	pulmicort l.s. refil	ICS
23688	ventolin rotacaps	SABA
23741	novolizer budesonide 200microgram/actuation pressurised inhalation (meda pharmaceuticals ltd)	ICS
23787	exirel 10mg capsule (3m health care ltd)	SABA
23961	ipratropium bromide 250microgram/ml inhalation vapour (galen ltd)	SAMA
24117	aminophylline 300 mg sup	THEOPH
24207	aminophylline paed 50 mg sup	THEOPH
24219	becotide rotacaps	ICS
24380	sodium cromoglicate 1mg/dose / salbutamol 100micrograms/dose inhaler with spacer	SABA_CROMO
24418	biophylline 350mg tablet (lorex synthelabo ltd)	THEOPH
24645	ventolin 5mg/5ml solution for infusion ampoules (glaxosmithkline uk ltd)	SABA
24660	betamethasone valerate	ICS
24674	biophylline 500mg tablet (lorex synthelabo ltd)	THEOPH
24716	prednisolone e/c	OCS
24898	spacehaler bdp 100microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
25020	ipratropium bromide (forte)	SAMA
25022	aminophylline 150 mg sup	THEOPH
25073	salbutamol	SABA
25093	theophylline s/r	THEOPH
25204	beclometasone 100micrograms/dose inhaler (a a h pharmaceuticals ltd)	ICS
25218	salbutamol cfc/free b/a	SABA
25272	precortisyl 1mg tablet (hoechst marion roussel)	OCS
25784	atimos modulite 12micrograms/dose inhaler (chiesi ltd)	LABA
25820	bronchodil 10mg/5ml oral solution (viatris pharmaceuticals ltd)	SABA
25821	exirel 7.5mg/5ml oral solution (3m health care ltd)	SABA
25829	pirbuterol 7.5mg/5ml oral solution	SABA
26063	beclometasone 100micrograms/dose inhaler (teva uk ltd)	ICS
26079	uniphyllin paediatric continus	THEOPH
26420	exirel 10 mg tab	SABA
26525	ventolin	SABA
26616	ipratropium bromide with fenoterol hydrobromide 0micrograms + 100micrograms/actuation	SABA_SAMA
26665	pulmicort complete	ICS
26716	airomir autohaler cfc free b/a	SABA
26829	brelomax 2mg tablet (abbott laboratories ltd)	LABA
26873	cobutolin 2mg tablet (actavis uk ltd)	SABA

prodcode	productname	groups
26987	bricanyl tablet (astrazeneca uk ltd)	SABA
27040	phyllocontin continus	THEOPH
27188	easyhaler budesonide 200micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	ICS
27505	ipratropium bromide with fenoterol hydrobromide 40micrograms + 100micrograms/actuation	SABA_SAMA
27525	becotide 50	ICS
27558	choledyl	THEOPH
27573	ventolin	SABA
27583	pulmicort	ICS
27593	aminophylline 350 mg sup	THEOPH
27679	beclometasone 100microgram/actuation pressurised inhalation (approved prescription services ltd)	ICS
27793	salbutamol cyclohaler type 5 insufflator inhalation powder (bristol-myers squibb pharmaceuticals ltd)	SABA
27842	aminophylline 2 ml inj	THEOPH
27889	prednisolone	OCS
27915	fluticasone prop disk refill	ICS
27959	prednisolone	OCS
27962	deltastab 1mg tablet (waymade healthcare plc)	OCS
28073	beclometasone 250microgram/actuation pressurised inhalation (approved prescription services ltd)	ICS
28375	prednisolone 2.5mg gastro-resistant tablets (a a h pharmaceuticals ltd)	OCS
28376	prednisolone 2.5mg gastro-resistant tablet (biorex laboratories ltd)	OCS
28508	salbutamol 100microgram/inhalation inhalation powder (ivax pharmaceuticals uk ltd)	SABA
28640	beclometasone 100microgram/actuation inhalation powder (actavis uk ltd)	ICS
28761	spacehaler bdp 50microgram/actuation spacehaler (celltech pharma europe ltd)	ICS
28859	deltastab 5mg tablet (waymade healthcare plc)	OCS
28881	salbutamol 2mg/5ml oral solution sugar free (a a h pharmaceuticals ltd)	SABA
29138	glycopyrronium bromide 1mg/5ml oral solution	LAMA
29267	salbuvent 4mg tablet (pharmacia ltd)	SABA
29273	aminophylline 225mg modified-release tablet (hillcross pharmaceuticals ltd)	THEOPH
29325	beclometasone 250micrograms/dose inhaler (generics (uk) ltd)	ICS
29333	prednisolone 5mg tablets (actavis uk ltd)	OCS
30118	salbutamol 100micrograms/dose inhaler cfc free (teva uk ltd)	SABA
30204	salbutamol 200micrograms inahalation capsules	SABA
30210	beclometasone 250micrograms/dose inhaler (teva uk ltd)	ICS
30212	salbutamol cyclohaler	SABA
30230	salbutamol 100micrograms/actuation breath actuated inhaler	SABA
30238	beclometasone 50microgram/actuation pressurised inhalation (approved prescription services ltd)	ICS
30390	deltastab 2 mg tab	OCS
30596	aminophylline 225mg modified-release tablet (actavis uk ltd)	THEOPH
30649	easyhaler budesonide 400micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	ICS
30971	decortisyl 25 mg tab	OCS

prodcode	productname	groups
31082	salbuvent 5mg/ml respirator solution (pharmacia ltd)	SABA
31290	salbulin cfc free	SABA
31327	prednisolone steaglate 6.65mg tablet	OCS
31532	prednisolone 5mg gastro-resistant tablets (a a h pharmaceuticals ltd)	OCS
31758	uniphyllin continus	THEOPH
31774	beclometasone 50micrograms/dose inhaler (generics (uk) ltd)	ICS
31845	salapin 2mg/5ml syrup (pinewood healthcare)	SABA
31933	salbutamol 100micrograms/dose inhaler (a a h pharmaceuticals ltd)	SABA
32050	salbutamol 400 cyclocaps (teva uk ltd)	SABA
32102	salbutamol 4mg tablets (a a h pharmaceuticals ltd)	SABA
32461	choline theophyllinate 90 mg tab	THEOPH
32803	prednisolone 5mg gastro-resistant tablets (actavis uk ltd)	OCS
32812	numotac 10mg tablet (3m health care ltd)	SABA
32835	prednisolone 5mg tablets (wockhardt uk ltd)	OCS
32874	beclometasone 50microgram/actuation inhalation powder (actavis uk ltd)	ICS
32893	theophylline 100mg/lysine 74mg mg tab	THEOPH
33089	salbutamol 100micrograms/dose inhaler (kent pharmaceuticals ltd)	SABA
33258	beclometasone 250micrograms/dose inhaler (a a h pharmaceuticals ltd)	ICS
33373	salbutamol 200 cyclocaps (teva uk ltd)	SABA
33588	salbutamol 100micrograms/dose inhaler (generics (uk) ltd)	SABA
33691	prednisolone 5mg gastro-resistant tablet (biorex laboratories ltd)	OCS
33817	salbutamol 100micrograms/dose inhaler cfc free (actavis uk ltd)	SABA
33849	beclometasone 100microgram/actuation inhalation powder (neo laboratories ltd)	ICS
33988	prednisolone 5mg tablet (co-pharma ltd)	OCS
33990	prednisolone 5mg tablet (ivax pharmaceuticals uk ltd)	OCS
34029	salbutamol 400micrograms inahalation capsules	SABA
34109	prednisolone 5 mg gastro-resistant tablet	OCS
34310	salbutamol 100micrograms/dose inhaler cfc free (a a h pharmaceuticals ltd)	SABA
34311	salbutamol 100microgram/inhalation inhalation powder (berk pharmaceuticals ltd)	SABA
34315	beclometasone 250microgram/actuation inhalation powder (actavis uk ltd)	ICS
34393	prednisolone 5mg gastro-resistant tablets (teva uk ltd)	OCS
34404	prednisolone 1mg tablets (actavis uk ltd)	OCS
34428	beclometasone 50microgram/actuation inhalation powder (neo laboratories ltd)	ICS
34452	prednisolone 1mg tablets (a a h pharmaceuticals ltd)	OCS
34461	prednisolone 2.5mg gastro-resistant tablets (actavis uk ltd)	OCS
34618	salbutamol 2mg tablets (actavis uk ltd)	SABA
34619	salbutamol 100microgram/inhalation inhalation powder (kent pharmaceuticals ltd)	SABA
34631	prednisolone 1mg tablet (co-pharma ltd)	OCS
34660	prednisolone 1mg tablets (kent pharmaceuticals ltd)	OCS
34702	salbutamol 100microgram/inhalation inhalation powder (c p pharmaceuticals ltd)	SABA
34739	beclometasone 50micrograms/dose inhaler (teva uk ltd)	ICS
34748	prednisolone 1mg tablets (teva uk ltd)	OCS
34781	prednisolone 5mg tablets (kent pharmaceuticals ltd)	OCS

prodcode	productname	groups
34794	beclometasone 200micrograms/dose inhaler (a a h pharmaceuticals ltd)	ICS
34859	beclometasone 250microgram/actuation inhalation powder (neo laboratories ltd)	ICS
34914	prednisolone 1mg tablet (celltech pharma europe ltd)	OCS
34919	beclometasone 50micrograms/dose inhaler (a a h pharmaceuticals ltd)	ICS
34938	salbutamol 4mg tablets (actavis uk ltd)	SABA
34978	prednisolone 1mg tablets (wockhardt uk ltd)	OCS
34995	spiriva 18microgram inhalation powder capsules with handihaler (boehringer ingelheim ltd)	LAMA
35000	spiriva 18microgram inhalation powder capsules (boehringer ingelheim ltd)	LAMA
35011	tiotropium bromide 18microgram inhalation powder capsules	LAMA
35014	tiotropium bromide 18microgram inhalation powder capsules with device	LAMA
35071	becodisks 200microgram (glaxosmithkline uk ltd)	ICS
35106	becodisks 100microgram with diskhaler (glaxosmithkline uk ltd)	ICS
35107	beclometasone 400microgram inhalation powder blisters with device	ICS
35113	beclometasone 200microgram inhalation powder blisters	ICS
35118	becodisks 400microgram with diskhaler (glaxosmithkline uk ltd)	ICS
35165	serevent 50microgram disks with diskhaler (glaxosmithkline uk ltd)	LABA
35225	flixotide 100microgram disks with diskhaler (glaxosmithkline uk ltd)	ICS
35288	beclometasone 400microgram inhalation powder blisters	ICS
35293	beclometasone 200microgram inhalation powder blisters with device	ICS
35299	becodisks 400microgram (glaxosmithkline uk ltd)	ICS
35374	flixotide 500microgram disks (glaxosmithkline uk ltd)	ICS
35392	flixotide 500microgram disks with diskhaler (glaxosmithkline uk ltd)	ICS
35408	becodisks 100microgram (glaxosmithkline uk ltd)	ICS
35430	becodisks 200microgram with diskhaler (glaxosmithkline uk ltd)	ICS
35461	flixotide 250microgram disks with diskhaler (glaxosmithkline uk ltd)	ICS
35503	salmeterol 50microgram inhalation powder blisters	LABA
35510	budesonide 200micrograms/dose dry powder inhalation cartridge with device	ICS
35542	salmeterol 50microgram inhalation powder blisters with device	LABA
35580	beclometasone 100microgram inhalation powder blisters with device	ICS
35602	budesonide 200micrograms/dose dry powder inhalation cartridge	ICS
35611	flixotide 250microgram disks (glaxosmithkline uk ltd)	ICS
35631	budelin novolizer 200micrograms/dose inhalation powder (meda pharmaceuticals ltd)	ICS
35638	fluticasone propionate 100microgram inhalation powder blisters with device	ICS
35652	beclometasone 100microgram inhalation powder blisters	ICS
35700	fluticasone propionate 500microgram inhalation powder blisters with device	ICS
35724	budelin novolizer 200micrograms/dose inhalation powder refill (meda pharmaceuticals ltd)	ICS
35725	formoterol easyhaler 12micrograms/dose dry powder inhaler (orion pharma (uk) ltd)	LABA
35772	fluticasone propionate 100microgram inhalation powder blisters	ICS
35825	serevent 50microgram disks (glaxosmithkline uk ltd)	LABA
35905	fluticasone propionate 250microgram inhalation powder blisters	ICS

prodcode	productname	groups
35986	flixotide 50microgram disks (glaxosmithkline uk ltd)	ICS
36021	fluticasone propionate 50microgram inhalation powder blisters with device	ICS
36090	flixotide 100microgram disks (glaxosmithkline uk ltd)	ICS
36290	flixotide 50microgram disks with diskhaler (glaxosmithkline uk ltd)	ICS
36401	fluticasone propionate 250microgram inhalation powder blisters with device	ICS
36462	fluticasone propionate 500microgram inhalation powder blisters	ICS
36677	reproterol 10mg/5ml oral solution	SABA
36864	tiotropium bromide 2.5micrograms/dose solution for inhalation cartridge with device cfc free	LAMA
36869	spiriva respimat 2.5micrograms/dose solution for inhalation cartridge with device (boehringer ingelheim ltd)	LAMA
37432	fostair 100micrograms/dose / 6micrograms/dose inhaler (chiesi ltd)	LABA_ICS
37447	fluticasone propionate 50microgram inhalation powder blisters	ICS
37470	beclometasone 100micrograms/dose / formoterol 6micrograms/dose inhaler cfc free	LABA_ICS
37791	ipratropium bromide 250microgram/ml	SAMA
38079	salbutamol 100micrograms/dose dry powder inhalation cartridge with device	SABA
38097	salbutamol cyclocaps 200microgram inhalation powder (dupont pharmaceuticals ltd)	SABA
38120	theophylline 500mg modified release tablets	THEOPH
38136	salbulin novolizer 100micrograms/dose inhalation powder (meda pharmaceuticals ltd)	SABA
38214	salbutamol 100micrograms/dose dry powder inhalation cartridge	SABA
38226	salbulin novolizer 100micrograms/dose inhalation powder refill (meda pharmaceuticals ltd)	SABA
38377	glycopyrronium bromide 2mg/5ml oral solution	LAMA
38407	prednisolone 20mg tablet	OCS
38416	salbutamol cyclocaps 400microgram inhalation powder (dupont pharmaceuticals ltd)	SABA
38419	terbutaline 1.5mg/5ml oral solution sugar free (a a h pharmaceuticals ltd)	SABA
38538	glycopyrronium bromide 2mg/5ml oral suspension	LAMA
39040	phyllocontin forte continus 350mg tablets (napp pharmaceuticals ltd)	THEOPH
39099	pulmicort 100micrograms/dose inhaler cfc free (astrazeneca uk ltd)	ICS
39102	budesonide 100micrograms/dose inhaler cfc free	ICS
39200	aerobec forte 250 autohaler (meda pharmaceuticals ltd)	ICS
39879	budesonide 200micrograms/dose inhaler cfc free	ICS
40057	pulmicort 200micrograms/dose inhaler cfc free (astrazeneca uk ltd)	ICS
40655	salbuvent 100microgram/actuation inhalation powder (pharmacia ltd)	SABA
41269	beclometasone 400 cyclocaps (teva uk ltd)	ICS
41412	beclometasone 400micrograms/actuation inhaler	ICS
41515	prednisolone 5mg tablets (teva uk ltd)	OCS
41548	salbutamol 2mg tablets (approved prescription services ltd)	SABA
41549	salbutamol 2mg tablet (c p pharmaceuticals ltd)	SABA
41691	salbutamol 2mg/5ml oral solution sugar free (sandoz ltd)	SABA
41745	prednisolone 25mg tablets (zentiva)	OCS
41832	monovent 1.5mg/5ml syrup (sandoz ltd)	SABA
42103	tulobuterol 1mg/5ml sugar free syrup	LABA

prodcode	productname	groups
42497	salbutamol 8mg tablet	SABA
42830	ventolin 100micrograms/dose evohaler (glaxosmithkline uk ltd)	SABA
42858	ventolin 200micrograms/dose accuhaler (glaxosmithkline uk ltd)	SABA
42867	terbutaline 1.5mg/5ml oral solution (sandoz ltd)	SABA
42886	bricanyl 500micrograms/dose turbohaler (astrazeneca uk ltd)	SABA
42928	flixotide 100micrograms/dose accuhaler (glaxosmithkline uk ltd)	ICS
42985	flixotide 50micrograms/dose accuhaler (glaxosmithkline uk ltd)	ICS
42994	flixotide 250micrograms/dose accuhaler (glaxosmithkline uk ltd)	ICS
43074	flixotide 500micrograms/dose accuhaler (glaxosmithkline uk ltd)	ICS
43085	bricanyl 5mg/2ml respules (astrazeneca uk ltd)	SABA
43090	atrovent 40microgram aerocaps (boehringer ingelheim ltd)	SAMA
43105	atrovent 40microgram aerocaps with aerohaler (boehringer ingelheim ltd)	SAMA
43544	prednisone 5mg tablet (knoll ltd)	OCS
43738	indacaterol 150microgram inhalation powder capsules with device	LABA
43764	opilon 40mg tablets (archimedes pharma uk ltd)	LABA
43893	onbrez breezhaler 150microgram inhalation powder capsules with device (novartis pharmaceuticals uk ltd)	LABA
44064	onbrez breezhaler 300microgram inhalation powder capsules with device (novartis pharmaceuticals uk ltd)	LABA
44380	prednisone 1mg modified-release tablets	OCS
44713	salbutamol 100microgram/inhalation inhalation powder (celltech pharma europe ltd)	SABA
44723	prednisone 5mg modified-release tablets	OCS
44802	lodotra 5mg modified-release tablets (napp pharmaceuticals ltd)	OCS
44803	lodotra 2mg modified-release tablets (napp pharmaceuticals ltd)	OCS
45302	prednisolone 5mg tablet (biorex laboratories ltd)	OCS
45610	indacaterol 300microgram inhalation powder capsules with device	LABA
46157	beclometasone 200 cyclocaps (teva uk ltd)	ICS
46214	glycopyrronium bromide 5mg/5ml oral solution	LAMA
46551	salbutamol 100microgram/inhalation inhalation powder (neo laboratories ltd)	SABA
46711	prednisone 2mg modified-release tablets	OCS
47142	prednisolone 5mg soluble tablet (amdipharm plc)	OCS
47269	glycopyrronium bromide 1mg/5ml oral suspension	LAMA
47638	neovent 25micrograms/dose inhaler cfc free (kent pharmaceuticals ltd)	LABA
47915	Omalizumab 150mg/1ml solution for injection pre-filled syringes	OMALIZUMAB
47943	beclazone easi-breathe (roi) 100microgram/actuation pressurised inhalation (ivax pharmaceuticals ireland)	ICS
48340	clenil modulite 100micrograms/dose inhaler (mawdsley-brooks & company ltd)	ICS
48484	theophylline 250mg/5ml oral suspension	THEOPH
48490	ventolin 100micrograms/dose evohaler (de pharmaceuticals)	SABA
48519	ventolin 100micrograms/dose evohaler (waymade healthcare plc)	SABA
48547	salamol 100micrograms/dose inhaler cfc free (arrow generics ltd)	SABA
48666	flutiform 250micrograms/dose / 10micrograms/dose inhaler (napp pharmaceuticals ltd)	LABA_ICS
48709	qvar 100micrograms/dose easi-breathe inhaler (sigma pharmaceuticals plc)	ICS
48739	seretide 250 evohaler (de pharmaceuticals)	LABA_ICS

prodcode	productname	groups
48741	ventolin 100micrograms/dose evohaler (mawdsley-brooks & company ltd)	SABA
48742	ventodisks 400microgram (glaxosmithkline uk ltd)	SABA
48809	ventodisks 400microgram with diskhaler (glaxosmithkline uk ltd)	SABA
49000	seretide 250 evohaler (waymade healthcare plc)	LABA_ICS
49114	symbicort 100/6 turbohaler (sigma pharmaceuticals plc)	LABA_ICS
49227	aclidinium bromide 375micrograms/dose dry powder inhaler	LAMA
49228	eklira 322micrograms/dose genuair (almirall ltd)	LAMA
49367	clenil modulite 50micrograms/dose inhaler (mawdsley-brooks & company ltd)	ICS
49368	ventodisks 200microgram with diskhaler (glaxosmithkline uk ltd)	SABA
49369	salbutamol 200microgram inhalation powder blisters	SABA
49370	ventodisks 200microgram (glaxosmithkline uk ltd)	SABA
49591	salbutamol 100micrograms/dose inhaler cfc free (sandoz ltd)	SABA
49711	pulmicort 200micrograms/dose inhaler (astrazeneca uk ltd)	ICS
49772	fluticasone 250micrograms/dose evohaler (sigma pharmaceuticals plc)	ICS
49868	fluticasone 250micrograms/dose / formoterol 10micrograms/dose inhaler cfc free	LABA_ICS
50036	flutiform 125micrograms/dose / 5micrograms/dose inhaler (napp pharmaceuticals ltd)	LABA_ICS
50037	pulmicort 0.5mg respules (waymade healthcare plc)	ICS
50047	glycopyrronium bromide 5mg/5ml oral solution	LAMA
50051	serevent 25micrograms/dose evohaler (waymade healthcare plc)	LABA
50103	spiriva 18microgram inhalation powder capsules with handihaler (waymade healthcare plc)	LAMA
50129	qvar 100micrograms/dose easi-breathe inhaler (de pharmaceuticals)	ICS
50287	qvar 100 inhaler (de pharmaceuticals)	ICS
50292	spiriva 18microgram inhalation powder capsules (sigma pharmaceuticals plc)	LAMA
50315	salbutamol 200microgram inhalation powder blisters with device	SABA
50503	ventolin 200micrograms/dose accuhaler (mawdsley-brooks & company ltd)	SABA
50557	ventolin 200micrograms/dose accuhaler (lexon (uk) ltd)	SABA
50560	seretide 250 accuhaler (sigma pharmaceuticals plc)	LABA_ICS
50577	spiriva 18microgram inhalation powder capsules with handihaler (de pharmaceuticals)	LAMA
50689	flutiform 50micrograms/dose / 5micrograms/dose inhaler (napp pharmaceuticals ltd)	LABA_ICS
50701	becotide rotahaler (glaxosmithkline uk ltd)	ICS
50739	symbicort 400/12 turbohaler (mawdsley-brooks & company ltd)	LABA_ICS
50810	atrovent 20micrograms/dose inhaler cfc free (de pharmaceuticals)	SAMA
50886	seretide 250 evohaler (stephar (u.k.) ltd)	LABA_ICS
50945	symbicort 100/6 turbohaler (mawdsley-brooks & company ltd)	LABA_ICS
50956	ventolin 200micrograms/dose accuhaler (de pharmaceuticals)	SABA
51027	seretide 125 evohaler (de pharmaceuticals)	LABA_ICS
51151	seretide 125 evohaler (lexon (uk) ltd)	LABA_ICS
51209	fluticasone 125micrograms/dose / formoterol 5micrograms/dose inhaler cfc free	LABA_ICS
51234	qvar 100 inhaler (waymade healthcare plc)	ICS

prodcode	productname	groups
	fluticasone 50micrograms/dose / formoterol 5micrograms/dose inhaler cfc	
51270	free	LABA_ICS
51394	seretide 500 accuhaler (waymade healthcare plc)	LABA_ICS
51415	qvar 50 inhaler (mawdsley-brooks & company ltd)	ICS
51430	theophylline 60mg/5ml oral suspension	THEOPH
51480	qvar 100 autohaler (de pharmaceuticals)	ICS
51570	symbicort 200/6 turbohaler (de pharmaceuticals)	LABA_ICS
51593	seretide 500 accuhaler (de pharmaceuticals)	LABA_ICS
51681	qvar 100 inhaler (sigma pharmaceuticals plc)	ICS
51753	prednisolone 1mg tablets (co-pharma ltd)	OCS
51759	symbicort 200/6 turbohaler (mawdsley-brooks & company ltd)	LABA_ICS
51815	flixotide 250micrograms/dose evohaler (waymade healthcare plc)	ICS
51861	seretide 500 accuhaler (mawdsley-brooks & company ltd)	LABA_ICS
51909	seretide 250 evohaler (necessity supplies ltd)	LABA_ICS
	spiriva 18microgram inhalation powder capsules (mawdsley-brooks &	
51967	company ltd)	LAMA
52410	bricanyl 500micrograms/dose turbohaler (necessity supplies ltd)	SABA
52543	salbutamol 400microgram inhalation powder blisters	SABA
52732	pulmicort 0.5mg respules (necessity supplies ltd)	ICS
52799	salbutamol 400microgram inhalation powder blisters with device	SABA
52806	qvar 100 autohaler (lexon (uk) ltd)	ICS
53019	ventolin 2.5mg nebules (mawdsley-brooks & company ltd)	SABA
53057	flixotide 50micrograms/dose evohaler (lexon (uk) ltd)	ICS
53230	seretide 250 accuhaler (de pharmaceuticals)	LABA_ICS
53237	symbicort 400/12 turbohaler (de pharmaceuticals)	LABA_ICS
53283	seretide 100 accuhaler (waymade healthcare plc)	LABA_ICS
53297	ventolin 200micrograms/dose accuhaler (sigma pharmaceuticals plc)	SABA
53313	prednisolone 20mg/5ml oral suspension	OCS
53336	prednisolone 25mg tablets (a a h pharmaceuticals ltd)	OCS
53480	qvar 100 autohaler (stephar (u.k.) ltd)	ICS
53491	symbicort 200/6 turbohaler (sigma pharmaceuticals plc)	LABA_ICS
	glycopyrronium bromide 55microgram inhalation powder capsules with	
53761	device	LAMA
	seebri breezhaler 44microgram inhalation powder capsules with device	
53982	(novartis pharmaceuticals uk ltd)	LAMA
54118	prednisolone 25mg/5ml oral suspension	OCS
54151	glycopyrronium bromide 600micrograms/5ml oral suspension	LAMA
54207	qvar 50 inhaler (de pharmaceuticals)	ICS
54399	qvar 100 autohaler (sigma pharmaceuticals plc)	ICS
54432	lodotra 1mg modified-release tablets (napp pharmaceuticals ltd)	OCS
54434	prednisolone 2.5mg/5ml oral suspension	OCS
54742	salmeterol 25micrograms/dose inhaler cfc free (a a h pharmaceuticals ltd)	LABA
55024	prednisolone 5mg/5ml oral solution	OCS
55480	prednisolone 2.5mg gastro-resistant tablets (alliance pharmaceuticals ltd)	OCS
55677	seretide 500 accuhaler (lexon (uk) ltd)	LABA_ICS
55794	glycopyrronium bromide 5mg/5ml oral suspension	LAMA
55795	glycopyrronium bromide 500micrograms/5ml oral suspension	LAMA
55911	glycopyrronium bromide 500micrograms/5ml oral solution	LAMA
56262	glycopyrronium bromide 200micrograms/5ml oral solution	LAMA
56462	becodisks 400microgram (waymade healthcare plc)	ICS

prodcode	productname	groups
56471	becodisks 200microgram (mawdsley-brooks & company ltd)	ICS
56474	flixotide 125micrograms/dose evohaler (de pharmaceuticals)	ICS
56475	flixotide 50micrograms/dose accuhaler (sigma pharmaceuticals plc)	ICS
56477	flixotide 100micrograms/dose accuhaler (waymade healthcare plc)	ICS
56478	serevent 50micrograms/dose accuhaler (de pharmaceuticals)	LABA
56482	oxis 12 turbohaler (waymade healthcare plc)	LABA
56484	flixotide 250micrograms/dose accuhaler (waymade healthcare plc)	ICS
56493	qvar 50micrograms/dose easi-breathe inhaler (sigma pharmaceuticals plc)	ICS
56498	pulmicort 200 turbohaler (waymade healthcare plc)	ICS
56499	flixotide 500micrograms/dose accuhaler (waymade healthcare plc)	ICS
56604	Montelukast 4mg chewable tablets sugar free (Actavis UK Ltd)	MONTELUKAST
56756	Montelukast 4mg granules sachets sugar free (A A H Pharmaceuticals Ltd)	MONTELUKAST
56891	prednisolone 1mg tablets (waymade healthcare plc)	OCS
57249	asmavent 100micrograms/dose inhaler cfc free (kent pharmaceuticals ltd)	SABA
57524	ventolin 200micrograms/dose accuhaler (dowelhurst ltd)	SABA
57525	flixotide 250micrograms/dose accuhaler (stephar (u.k.) ltd)	ICS
57544	serevent 50micrograms/dose accuhaler (waymade healthcare plc)	LABA
57555	flixotide 125micrograms/dose evohaler (dowelhurst ltd)	ICS
57557	atrovent 20micrograms/dose inhaler cfc free (lexon (uk) ltd)	SAMA
57558	oxis 6 turbohaler (lexon (uk) ltd)	LABA
57579	flixotide 50micrograms/dose accuhaler (de pharmaceuticals)	ICS
57589	becloforte 250micrograms/dose inhaler (dowelhurst ltd)	ICS
57694	vertine 25micrograms/dose inhaler cfc free (teva uk ltd)	LABA
58000	prednisolone 5mg tablets (almus pharmaceuticals ltd)	OCS
58061	prednisone 50mg tablets	OCS
58234	prednisolone 10mg/5ml oral solution	OCS
58269	airsalb 100micrograms/dose inhaler cfc free (sandoz ltd)	SABA
58369	prednisolone 5mg tablets (boston healthcare ltd)	OCS
58384	prednisolone 1mg tablets (almus pharmaceuticals ltd)	OCS
58987	prednisolone 5mg gastro-resistant tablets (phoenix healthcare distribution ltd)	OCS
59173	glycopyrronium bromide 200micrograms/5ml oral suspension	LAMA
59229	dilacort 5mg gastro-resistant tablets (auden mckenzie (pharma division) ltd)	OCS
59263	Montelukast 10mg tablets (Teva UK Ltd)	MONTELUKAST
59283	dilacort 2.5mg gastro-resistant tablets (auden mckenzie (pharma division) ltd)	OCS
59327	relvar ellipta 92micrograms/dose / 22micrograms/dose dry powder inhaler (glaxosmithkline uk ltd)	LABA_ICS
59338	prednisolone 1mg/5ml oral solution	OCS
59409	salbutamol 100micrograms/dose inhaler cfc free (waymade healthcare plc)	SABA
59439	fluticasone furoate 92micrograms/dose / vilanterol 22micrograms/dose dry powder inhaler	LABA_ICS
59573	relvar ellipta 184micrograms/dose / 22micrograms/dose dry powder inhaler (glaxosmithkline uk ltd)	LABA_ICS
59638	spiriva 18microgram inhalation powder capsules with handihaler (sigma pharmaceuticals plc)	LAMA
59819	Montelukast 10mg tablets (Actavis UK Ltd)	MONTELUKAST

prodcode	productname	groups
59899	fluticasone furoate 184micrograms/dose / vilanterol 22micrograms/dose dry powder inhaler	LABA_ICS
59912	prednisolone 5mg gastro-resistant tablets (waymade healthcare plc)	OCS
59968	Montelukast 5mg chewable tablets sugar free (Teva UK Ltd)	MONTELUKAST
60331	Montelukast 10mg tablets (Ranbaxy (UK) Ltd)	MONTELUKAST
60421	prednisolone 5mg tablets (co-pharma ltd)	OCS
60920	atrovent 20micrograms/dose inhaler cfc free (sigma pharmaceuticals plc)	SAMA
60923	salamol 100micrograms/dose easi-breathe inhaler (de pharmaceuticals)	SABA
60937	pulmicort 200 turbohaler (dowelhurst ltd)	ICS
61132	prednisolone 1mg tablets (boston healthcare ltd)	OCS
61162	prednisolone 5mg tablets (waymade healthcare plc)	OCS
61176	anoro ellipta 55micrograms/dose / 22micrograms/dose dry powder inhaler (glaxosmithkline uk ltd)	LABA_LAMA
61280	seretide 250 accuhaler (waymade healthcare plc)	LABA_ICS
61490	umeclidinium bromide 65micrograms/dose / vilanterol 22micrograms/dose dry powder inhaler	LABA_LAMA
61582	spiriva respimat 2.5micrograms/dose solution for inhalation cartridge with device (waymade healthcare plc)	LAMA
61591	salbutamol 100micrograms/dose inhaler cfc free (phoenix healthcare distribution ltd)	SABA
61644	fostair nexthaler 100micrograms/dose / 6micrograms/dose dry powder inhaler (chiesi ltd)	LABA_ICS
61664	clenil modulite 250micrograms/dose inhaler (waymade healthcare plc)	ICS
61666	duoresp spiromax 320micrograms/dose / 9micrograms/dose dry powder inhaler (teva uk ltd)	LABA_ICS
61689	prednisolone 5mg soluble tablets (a a h pharmaceuticals ltd)	OCS
61782	duoresp spiromax 160micrograms/dose / 4.5micrograms/dose dry powder inhaler (teva uk ltd)	LABA_ICS
61879	incruse ellipta 55micrograms/dose dry powder inhaler (glaxosmithkline uk ltd)	LAMA
62030	beclometasone 100micrograms/dose / formoterol 6micrograms/dose dry powder inhaler	LABA_ICS
62109	umeclidinium bromide 65micrograms/dose dry powder inhaler	LAMA
62126	seretide 100 accuhaler (de pharmaceuticals)	LABA_ICS
62410	Montelukast 10mg tablets (Alliance Healthcare (Distribution) Ltd)	MONTELUKAST
62490	Montelukast 10mg tablets (A A H Pharmaceuticals Ltd)	MONTELUKAST
63457	Montelukast 5mg chewable tablets sugar free (Accord Healthcare Ltd)	MONTELUKAST
64648	Montelukast 10mg tablets (Milpharm Ltd)	MONTELUKAST
65038	Montelukast 10mg tablets (Accord Healthcare Ltd)	MONTELUKAST

Codelist 4: Oral corticosteroid codes

prodcode	productname
5913	Deltacortril 2.5mg gastro-resistant tablets (Alliance Pharmaceuticals Ltd)
5490	Deltacortril 5mg gastro-resistant tablets (Alliance Pharmaceuticals Ltd)
27962	Deltastab 1mg Tablet (Waymade Healthcare Plc)
28859	Deltastab 5mg Tablet (Waymade Healthcare Plc)
59283	Dilacort 2.5mg gastro-resistant tablets (Auden McKenzie (Pharma Division) Ltd)
59229	Dilacort 5mg gastro-resistant tablets (Auden McKenzie (Pharma Division) Ltd)
25272	Precortisyl 1mg Tablet (Hoechst Marion Roussel)
23512	Precortisyl 5mg Tablet (Hoechst Marion Roussel)
20095	Precortisyl forte 25mg Tablet (Aventis Pharma)
58234	Prednisolone 10mg/5ml oral solution
34914	Prednisolone 1mg Tablet (Celltech Pharma Europe Ltd)
34631	Prednisolone 1mg Tablet (Co-Pharma Ltd)
578	Prednisolone 1mg tablets
34452	Prednisolone 1mg tablets (A A H Pharmaceuticals Ltd)
34404	Prednisolone 1mg tablets (Actavis UK Ltd)
58384	Prednisolone 1mg tablets (Almus Pharmaceuticals Ltd)
61132	Prednisolone 1mg tablets (Boston Healthcare Ltd)
51753	Prednisolone 1mg tablets (Co-Pharma Ltd)
34660	Prednisolone 1mg tablets (Kent Pharmaceuticals Ltd)
34748	Prednisolone 1mg tablets (Teva UK Ltd)
56891	Prednisolone 1mg tablets (Waymade Healthcare Plc)
34978	Prednisolone 1mg tablets (Wockhardt UK Ltd)
59338	Prednisolone 1mg/5ml oral solution
28376	Prednisolone 2.5mg Gastro-resistant tablet (Biorex Laboratories Ltd)
557	Prednisolone 2.5mg gastro-resistant tablets
28375	Prednisolone 2.5mg gastro-resistant tablets (A A H Pharmaceuticals Ltd)
34461	Prednisolone 2.5mg gastro-resistant tablets (Actavis UK Ltd)
55480	Prednisolone 2.5mg gastro-resistant tablets (Alliance Pharmaceuticals Ltd)
2368	Prednisolone 2.5mg tablet
54434	Prednisolone 2.5mg/5ml oral suspension
38407	Prednisolone 20mg tablet
53313	Prednisolone 20mg/5ml oral suspension
2704	Prednisolone 25mg tablets
53336	Prednisolone 25mg tablets (A A H Pharmaceuticals Ltd)
41745	Prednisolone 25mg tablets (Zentiva)
54118	Prednisolone 25mg/5ml oral suspension
34109	Prednisolone 5 mg gastro-resistant tablet
9727	Prednisolone 50mg tablets
33691	Prednisolone 5mg Gastro-resistant tablet (Biorex Laboratories Ltd)
44	Prednisolone 5mg gastro-resistant tablets
31532	Prednisolone 5mg gastro-resistant tablets (A A H Pharmaceuticals Ltd)
32803	Prednisolone 5mg gastro-resistant tablets (Actavis UK Ltd)
58987	Prednisolone 5mg gastro-resistant tablets (Phoenix Healthcare Distribution Ltd)
34393	Prednisolone 5mg gastro-resistant tablets (Teva UK Ltd)
59912	Prednisolone 5mg gastro-resistant tablets (Waymade Healthcare Plc)
45302	Prednisolone 5mg Tablet (Biorex Laboratories Ltd)
33988	Prednisolone 5mg Tablet (Co-Pharma Ltd)
33990	Prednisolone 5mg Tablet (IVAX Pharmaceuticals UK Ltd)

prodcode	productname
95	Prednisolone 5mg tablets
21417	Prednisolone 5mg tablets (A A H Pharmaceuticals Ltd)
29333	Prednisolone 5mg tablets (Actavis UK Ltd)
58000	Prednisolone 5mg tablets (Almus Pharmaceuticals Ltd)
58369	Prednisolone 5mg tablets (Boston Healthcare Ltd)
60421	Prednisolone 5mg tablets (Co-Pharma Ltd)
34781	Prednisolone 5mg tablets (Kent Pharmaceuticals Ltd)
41515	Prednisolone 5mg tablets (Teva UK Ltd)
61162	Prednisolone 5mg tablets (Waymade Healthcare Plc)
32835	Prednisolone 5mg tablets (Wockhardt UK Ltd)
55024	Prednisolone 5mg/5ml oral solution
1063	Prednesol 5mg Tablet (Sovereign Medical Ltd)
47142	Prednisolone 5mg Soluble tablet (Amdipharm Plc)
955	Prednisolone 5mg soluble tablets
61689	Prednisolone 5mg soluble tablets (A A H Pharmaceuticals Ltd)
19141	Prednisolone 5mg soluble tablets (AMCo)
31327	Prednisolone steaglate 6.65mg tablet
3345	Sintisone Tablet (Pharmacia Ltd)
21833	Decortisyl 5mg Tablet (Roussel Laboratories Ltd)
54432	Lodotra 1mg modified-release tablets (Napp Pharmaceuticals Ltd)
44803	Lodotra 2mg modified-release tablets (Napp Pharmaceuticals Ltd)
44802	Lodotra 5mg modified-release tablets (Napp Pharmaceuticals Ltd)
44380	Prednisone 1mg modified-release tablets
3557	Prednisone 1mg tablets
46711	Prednisone 2mg modified-release tablets
58061	Prednisone 50mg tablets
44723	Prednisone 5mg modified-release tablets
43544	Prednisone 5mg Tablet (Knoll Ltd)
2949	Prednisone 5mg tablets
30971	DECORTISYL 25 MG TAB
30390	DELTASTAB 2 MG TAB
27889	PREDNISOLONE
27959	PREDNISOLONE
2799	PREDNISOLONE 10 MG TAB
7710	PREDNISOLONE 15 MG TAB
13522	PREDNISOLONE 2 MG TAB
7584	PREDNISOLONE 4 MG TAB
3059	PREDNISOLONE 50 MG TAB
20670	PREDNISOLONE E/C
24716	PREDNISOLONE E/C
2390	PREDNISOLONE E/C 1 MG TAB
13615	PREDNISONE 10 MG TAB
2044	PREDNISONE 2.5 MG TAB
7934	PREDNISONE 30 MG TAB
16724	PREDNISONE 50 MG TAB

Codelist 5: Antihistamines

prodcode	productname
56798	Lloydspharmacy Allergy Relief 2mg/5ml syrup (Lloyds Pharmacy Ltd)
1549	Chlorphenamine 2mg/5ml oral solution
23054	Ephedrine HCl with Chlorphenamine 4mg with 1mg/5ml oral solution sugar free
1436	Haymine tablets (Chemidex Pharma Ltd)
5732	Chlorphenamine 2mg/5ml oral solution sugar free
28967	Expulin sugar free Oral solution (Shire Pharmaceuticals Ltd)
9815	Pseudoephedrine 30mg/5ml / Chlorphenamine 2mg/5ml oral solution sugar free
31525	Chlorphenamine 2mg/5ml oral solution sugar free (A A H Pharmaceuticals Ltd)
8341	Galpseud Plus linctus (Thornton & Ross Ltd)
46914	Chlorphenamine 4mg tablets (Vantage)
961	Piriton 2mg/5ml Oral solution (Stafford-Miller Ltd)
43415	Chlorphenamine 50mg/5ml oral solution
16478	Hayleve 4mg tablets (Genesis Pharmaceuticals Ltd)
30928	Chlorphenamine 10mg/1ml solution for injection ampoules (Kyowa Kirin Ltd)
32962	Chlorphenamine 4mg tablets (Teva UK Ltd)
27812	Chlorphenamine 4mg tablets (Sussex Pharmaceutical Ltd)
13956	Ephedrine hydrochloride 15mg / Chlorphenamine 10mg tablets
32240	Chlorphenamine 4mg tablets (Actavis UK Ltd)
36589	Pollenase Allergy 2mg/5ml syrup (E M Pharma)
71045	Chlorphenamine 4mg tablets (Bristol Laboratories Ltd)
34136	Chlorphenamine 12mg modified-release tablets
4423	Dichlorphenamide 50mg tablets
58808	Chlorphenamine 4mg tablets (Waymade Healthcare Plc)
61263	Boots Allergy Relief Antihistamine 4mg tablets (The Boots Company Plc)
66342	Pseudoephedrine with chlorphenamine & pholcodine oral solution sugar free
22801	Chlorphenesin 1% powder
23076	Allerief 2mg/5ml oral solution (Orbis Consumer Products Ltd)
22337	Pseudoephedrine with chlorphenamine & pholcodine oral solution
29872	Chlorphenamine 4mg tablets (A A H Pharmaceuticals Ltd)
18490	Phenylpropanolamine 50mg / Chlorphenamine 4mg modified-release capsules
533	Piriton 4mg Tablet (Stafford-Miller Ltd)
21435	Contac 400 modified-release capsules (GlaxoSmithKline Consumer Healthcare)
55536	Chlorphenamine 4mg tablets (Strides Shasun (UK) Ltd)
16181	Tixylix Cough & Cold oral solution (Novartis Consumer Health UK Ltd)
12062	Daranide 50mg Tablet (MSD Thomas Morson Pharmaceuticals)
884	Chlorphenamine 4mg tablets
43521	Chlorphenamine 4mg Tablet (Family Health)
11985	Piriton Allergy 4mg tablets (GlaxoSmithKline Consumer Healthcare)
23875	Alunex 4mg Tablet (M A Steinhard Ltd)
15757	Pholcodine 2mg with chlorphenamine 1mg/5ml oral solution sugar free
686	Piriton 4mg tablets (GlaxoSmithKline Consumer Healthcare)
1305	Chlorphenamine 10mg/1ml solution for injection ampoules
37093	ALLERcalm Allergy Relief 4mg tablets (Actavis UK Ltd)

prodcode	productname
31639	Unichem allergy relief 4mg Tablet (Unichem)
32443	Pollenase Antihistamine 4mg tablets (E M Pharma)
28554	Expulin children's cough sugar free Oral solution (Shire Pharmaceuticals Ltd)
60280	Chlorphenamine 4mg tablets (Alliance Healthcare (Distribution) Ltd)
168	Piriject 10mg/ml Injection (Link Pharmaceuticals Ltd)
61110	Numark Antihistamine 4mg tablets (Numark Ltd)
59556	Chlorphenamine 4mg tablets (Sigma Pharmaceuticals Plc)
27117	Calimal 4mg tablets (Sussex Pharmaceutical Ltd)
36522	Expulin sugar free Oral solution (Shire Pharmaceuticals Ltd)
42009	Chlorphenamine 4mg Tablet (Genesis Medical Ltd)
2604	Piriton 12mg Spandets (Stafford-Miller Ltd)
26744	Expulin decongestant sugar free Oral solution (Shire Pharmaceuticals Ltd)
629	Piriton 2mg/5ml syrup (GlaxoSmithKline Consumer Healthcare)
70953	Chlorphenamine 4mg tablets (Kent Pharmaceuticals Ltd)
31818	Chlorphenamine 2mg/5ml oral solution sugar free (Sandoz Ltd)
37834	Chlorphenamine 4mg tablets (Almus Pharmaceuticals Ltd)
70310	Chlorphenamine 10mg/1ml solution for injection ampoules (Alliance Healthcare (Distribution) Ltd)
12590	PSEUDOEPHEDRINE 30MG/CHLORPHENIRAMINE2MG
1849	CHLORPHENIRAMINE MALEATE S/R 8 MG TAB
27227	CHLORPHENESIN .5 % OIN
3555	CHLORPHENIRAMINE MALEATE S/R 12 MG TAB
52138	Chlorphenesin 0.5% ointment
2189	Acrivastine 8mg capsules
10087	Pseudoephedrine 60mg / Acrivastine 8mg capsules
5671	Benadryl Allergy Relief 8mg capsules (McNeil Products Ltd)
9782	Benadryl Allergy Relief Plus Decongestant capsules (McNeil Products Ltd)
3525	Semprex 8mg capsules (GlaxoSmithKline UK Ltd)
46684	Bilastine 20mg tablets
57004	Ilaxten 20mg tablets (A. Menarini Farmaceutica Internazionale SRL)
40150	Cetirizine 1mg/ml Oral solution (Lagap)
39743	Cetirizine 1mg/ml Oral solution (Ratiopharm UK Ltd)
48222	Cetirizine 10mg capsules
45266	Cetirizine 10mg tablets (Dexcel-Pharma Ltd)
15946	Cetirizine 10mg tablets (A A H Pharmaceuticals Ltd)
59329	Cetirizine 10mg tablets (Alliance Healthcare (Distribution) Ltd)
29666	Cetirizine 10mg tablets (Sandoz Ltd)
45253	Cetirizine 10mg tablets (Sterwin Medicines)
40783	Levocetirizine 5mg tablets (Teva UK Ltd)
5730	Levocetirizine 5mg tablets
43133	Cetirizine 1mg/ml oral solution sugar free (Pinewood Healthcare)
58379	Cetirizine 10mg tablets (Chanelle Medical UK Ltd)
29459	Cetirizine 10mg Tablet (Niche Generics Ltd)
25782	Cetirizine 1mg/ml oral solution sugar free (Teva UK Ltd)
56903	Cetirizine 1mg/ml oral solution sugar free (Sandoz Ltd)

prodcode	productname
2916	Cetirizine 1mg/ml oral solution sugar free
61275	Cetirizine 10mg tablets (Waymade Healthcare Plc)
55492	Cetirizine 10mg tablets (Almus Pharmaceuticals Ltd)
59752	Cetirizine 1mg/ml oral solution sugar free (Actavis UK Ltd)
36093	Levocetirizine 500micrograms/ml oral solution sugar free
33709	Cetirizine 1mg/ml oral solution sugar free (A A H Pharmaceuticals Ltd)
59210	Cetirizine 10mg tablets (Bristol Laboratories Ltd)
34538	Cetirizine 10mg tablets (Mylan)
45239	Cetirizine 10mg tablets (Fannin UK Ltd)
61064	Cetirizine 1mg/ml oral solution sugar free (Waymade Healthcare Plc)
61174	Cetirizine 10mg tablets (Wockhardt UK Ltd)
29297	Cetirizine 10mg tablets (Teva UK Ltd)
20886	Cetirocol 10mg tablets (Teva UK Ltd)
70993	Cetirizine 10mg tablets (DE Pharmaceuticals)
36284	Cetirizine 1mg/ml Oral solution (Hillcross Pharmaceuticals Ltd)
70	Cetirizine 10mg tablets
62117	Cetirizine 10mg tablets (Dr Reddy's Laboratories (UK) Ltd)
41965	Cetirizine 10mg tablets (Actavis UK Ltd)
6103	Zirtek Allergy 10mg tablets (UCB Pharma Ltd)
22759	Zirtek Allergy Relief for Children 1mg/ml oral solution (UCB Pharma Ltd)
57584	Zirtek 5mg/5ml oral solution (Dowelhurst Ltd)
19174	Benadryl allergy 1mg/ml Oral solution (Pfizer Consumer Healthcare Ltd)
45627	Pollenshield Hayfever 10mg tablets (Actavis UK Ltd)
5901	Xyzal 5mg tablets (UCB Pharma Ltd)
33235	Hayfever and allergy relief 10mg Tablet (Herbal Concepts Ltd)
38307	Pollenshield Hayfever Relief 10mg tablets (Actavis UK Ltd)
9950	Piriteze Allergy 1mg/ml syrup (GlaxoSmithKline Consumer Healthcare)
45710	Histease allergy relief 10mg Tablet (Dr Reddy's Laboratories (UK) Ltd)
44121	Benadryl One A Day 10mg tablets (McNeil Products Ltd)
10047	Hayfever and allergy relief 10mg Tablet (A A H Pharmaceuticals Ltd)
50608	Zirtek 5mg/5ml oral solution (Sigma Pharmaceuticals Plc)
31891	Galpharm Hayfever and Allergy Relief 10mg tablets (Galpharm International Ltd)
36828	Xyzal 0.5mg/ml oral solution (UCB Pharma Ltd)
30508	Zirtek Allergy Relief 10mg tablets (UCB Pharma Ltd)
10097	Benadryl One A Day Relief 10mg tablets (McNeil Products Ltd)
6348	Piriteze Allergy 10mg tablets (GlaxoSmithKline Consumer Healthcare)
39668	Benadryl Allergy 1mg/ml oral solution (McNeil Products Ltd)
25783	Cetec 10mg tablets (Herbal Concepts Ltd)
35139	Benadryl Allergy Children's 6+ 1mg/ml oral solution (McNeil Products Ltd)
6427	Zirtek Allergy 1mg/ml oral solution (UCB Pharma Ltd)
67658	Allacan 10mg tablets (Bristol Laboratories Ltd)
42470	AllerTek 10mg tablets (Ratiopharm UK Ltd)
2734	Zirtek 1mg/ml Oral solution (UCB Pharma Ltd)
63948	Benadryl Allergy Children's 1mg/ml oral solution (McNeil Products Ltd)
51976	Benadryl Allergy Liquid Release 10mg capsules (McNeil Products Ltd)

prodcode	productname
69217	Benadryl Allergy One A Day 10mg tablets (McNeil Products Ltd)
71144	Zirtek 5mg/5ml oral solution (Waymade Healthcare Plc)
9780	Benadryl one a day 10mg Tablet (Pfizer Consumer Healthcare Ltd)
68007	BecoAllergy 10mg tablets (Omega Pharma Ltd)
1443	Zirtek 10mg tablets (UCB Pharma Ltd)
54261	Midetorin 2.5mg/5ml oral solution (Actavis UK Ltd)
39657	Neoclarityn 2.5mg/5ml oral solution (Merck Sharp & Dohme Ltd)
5380	Desloratadine 5mg tablets
66466	Desloratadine 5mg tablets (Actavis UK Ltd)
42393	Desloratadine 2.5mg/5ml oral solution sugar free
57636	Desloratadine 5mg tablets (A A H Pharmaceuticals Ltd)
5362	Neoclarityn 5mg tablets (Merck Sharp & Dohme Ltd)
5934	Desloratadine 2.5mg/5ml oral solution
5910	Neoclarityn 2.5mg/5ml syrup (Schering-Plough Ltd)
42298	Fexofenadine 120mg tablets (A A H Pharmaceuticals Ltd)
68650	Fexofenadine 120mg/5ml oral suspension
2602	Fexofenadine 120mg tablets
37785	Fexofenadine 180mg tablets (Zentiva)
67198	Fexofenadine 180mg tablets (Actavis UK Ltd)
69999	Fexofenadine 180mg/5ml oral solution
60103	Fexofenadine 120mg tablets (Kent Pharmaceuticals Ltd)
55383	Fexofenadine 180mg tablets (Alliance Healthcare (Distribution) Ltd)
9983	Fexofenadine 30mg tablets
62861	Fexofenadine 180mg tablets (Dr Reddy's Laboratories (UK) Ltd)
43978	Fexofenadine 180mg tablets (A A H Pharmaceuticals Ltd)
65780	Fexofenadine 120mg tablets (Zentiva)
69580	Fexofenadine 180mg tablets (PLIVA Pharma Ltd)
63109	Fexofenadine 120mg tablets (Actavis UK Ltd)
2161	Fexofenadine 180mg tablets
64159	Fexofenadine 120mg tablets (Mylan)
1084	Telfast 180mg tablets (Sanofi)
2740	Telfast 120mg tablets (Sanofi)
49112	Telfast 120mg tablets (Lexon (UK) Ltd)
52190	Telfast 180mg tablets (Lexon (UK) Ltd)
17036	Telfast 30mg tablets (Sanofi)
51673	Telfast 120mg tablets (Necessity Supplies Ltd)
52957	Telfast 180mg tablets (Necessity Supplies Ltd)
13700	Terfinax 120mg Tablet (Ashbourne Pharmaceuticals Ltd)
13330	Terfinax 60mg Tablet (Ashbourne Pharmaceuticals Ltd)
1437	Triludan Forte 120mg tablets (Hoechst Marion Roussel)
38112	Terfenor 30mg Tablet (IVAX Pharmaceuticals UK Ltd)
944	Terfenadine 30mg/5ml suspension
154	Terfenadine 60mg tablets
21818	Histafen 60mg Tablet (Berk Pharmaceuticals Ltd)
25091	Terfenor forte 120mg Tablet (IVAX Pharmaceuticals UK Ltd)

prodcode	productname
44870	Aller-eze clear 60mg Tablet (Novartis Consumer Health UK Ltd)
36715	Seldane 120mg Tablet (Hoechst Marion Roussel)
1444	Terfenadine 120mg tablets
1405	Triludan 30mg/5ml sugar free Oral suspension (Hoechst Marion Roussel)
63651	Terfenadine 120mg Tablet (Approved Prescription Services Ltd)
30397	Histafen 120mg Tablet (Berk Pharmaceuticals Ltd)
28639	Terfenadine 60mg Tablet (Lagap)
32749	Terfenor 60mg Tablet (IVAX Pharmaceuticals UK Ltd)
233	Triludan 60mg Tablet (Hoechst Marion Roussel)
3027	Terfenadine 30mg tablets
63647	Loratadine 10mg tablets (Almus Pharmaceuticals Ltd)
34755	Loratadine 5mg/5ml oral solution (A A H Pharmaceuticals Ltd)
45227	Loratadine 10mg tablets (Sandoz Ltd)
55150	Clarityn Rapide Allergy 10mg tablets (Bayer Plc)
1077	Clarityn 5mg/5ml syrup (Schering-Plough Ltd)
30101	Loratadine 10mg tablets (Actavis UK Ltd)
46501	Loratadine 10mg tablets (Ranbaxy (UK) Ltd)
34304	Loratadine 10mg tablets (Mylan)
70041	Loratadine 10mg tablets (Alliance Healthcare (Distribution) Ltd)
51215	Loratadine 10mg oral lyophilisates sugar free
68597	Loratadine 5mg/5ml oral solution (Mylan)
26707	Clarityn Allergy 5mg/5ml syrup (Bayer Plc)
92	Loratadine 10mg tablets
26646	Galpharm Non-Drowsy Allergy Relief 10mg tablets (Galpharm International Ltd)
34082	Non-drowsy allergy relief 5mg/5ml Oral solution (A A H Pharmaceuticals Ltd)
69039	Loratadine 5mg/5ml oral solution (Mawdsley-Brooks & Company Ltd)
34752	Loratadine 10mg tablets (Zentiva)
33893	Loratadine 10mg tablets (Teva UK Ltd)
1015	Clarityn 10mg Tablet (Schering-Plough Ltd)
1554	Loratadine 5mg/5ml oral solution
34753	Loratadine 10mg tablets (A A H Pharmaceuticals Ltd)
34262	Loratadine 5mg/5ml oral solution (Teva UK Ltd)
31364	Non-drowsy allergy relief Tablet (A A H Pharmaceuticals Ltd)
36201	Hay-Rite Allergy 10mg tablets (Teva UK Ltd)
34333	Loratadine 10mg Tablet (Niche Generics Ltd)
67296	Loratadine 5mg/5ml oral solution (Kent Pharmaceuticals Ltd)
57547	Clarityn 5mg/5ml syrup (Waymade Healthcare Plc)
31177	Loratadine 5mg/5ml oral solution (Actavis UK Ltd)
20162	Clarityn Allergy 10mg tablets (Bayer Plc)
69887	Lorapaed Allergy Relief 5mg/5ml oral solution (Pinewood Healthcare)
4311	Mizolastine 10mg modified-release tablets
11417	Mistamine 10mg modified-release tablets (Galderma (UK) Ltd)
9461	Mizollen 10mg modified-release tablets (Sanofi)
47915	Omalizumab 150mg/1ml solution for injection pre-filled syringes

prodcode	productname
60113	Xolair 75mg/0.5ml solution for injection pre-filled syringes (Novartis Pharmaceuticals UK Ltd)
53289	Xolair 150mg/1ml solution for injection pre-filled syringes (Novartis Pharmaceuticals UK Ltd)
7550	Omalizumab 150mg powder and solvent for solution for injection vials
69598	Omalizumab 75mg/0.5ml solution for injection pre-filled syringes
41287	Xolair 150mg powder and solvent for solution for injection vials (Novartis Pharmaceuticals UK Ltd)

Codelist 6: COPD codes

medcode	term
794	Emphysema
998	Chronic obstructive airways disease
1001	Chronic obstructive pulmonary disease
4084	Airways obstructn irreversible
5710	Chronic obstructive airways disease NOS
9520	Chronic obstructive pulmonary disease monitoring
9876	Severe chronic obstructive pulmonary disease
10802	Moderate chronic obstructive pulmonary disease
10863	Mild chronic obstructive pulmonary disease
10980	Centrilobular emphysema
11287	Chronic obstructive pulmonary disease annual review
12166	Other specified chronic obstructive airways disease
14798	Emphysematous bronchitis
18476	COPD follow-up
18621	Chronic obstructive pulmonary disease follow-up
18792	Chronic obstructive pulmonary disease monitoring admin
23492	Chronic bullous emphysema NOS
26018	Chronic obstructive pulmonary disease monitoring by nurse
26306	Chronic bullous emphysema
28755	Chronic obstructive pulmonary disease monitoring 1st letter
33450	Emphysema NOS
34202	Chronic obstructive pulmonary disease monitoring 2nd letter
34215	Chronic obstructive pulmonary disease monitoring 3rd letter
37247	Chronic obstructive pulmonary disease NOS
37371	Chronic obstructive pulmonary disease monitoring due
38074	Chronic obstructive pulmonary disease monitor phone invite
42258	Chronic obstructive pulmonary disease monitoring verb invite
42313	Health education - chronic obstructive pulmonary disease
44525	Obstructive chronic bronchitis NOS
45770	Chronic obstructive pulmonary disease disturbs sleep
45771	Chronic obstructive pulmonary disease does not disturb sleep
45777	Chronic obstructive pulmonary disease clini management plan
45998	Chronic obstructive pulmonary disease monitoring by doctor
93568	Very severe chronic obstructive pulmonary disease

Codelist 7: GORD

medcode	readterm
19470	reflux cough
7577	gastric reflux

medcode	readterm
592	oesophagitis
2535	reflux oesophagitis
15054	acid reflux
7104	gastro-oesophageal reflux with oesophagitis
16605	oesophageal reflux with oesophagitis
15579	peptic oesophagitis
16450	regurgitant oesophagitis
4614	barrett's oesophagus
14760	oesophagitis nos
5596	barrett's ulcer of oesophagus
25610	oesopheal reflux without mention of oesophagitis
1327	oesophageal reflux
984	gastro-oesophageal reflux
2281	acid reflux

Codelist 8: Atopy

medcode	readterm
175	allergic rhinitis
230	eczema nos
334	allergic contact dermatitis
610	infantile eczema
619	skin:type 1 immediate reaction
620	allergy, unspecified
768	desensitisation to allergens
774	chronic rhinitis
775	allergic rhinitis due to unspecified allergen
788	allergic conjunctivitis
805	chronic catarrhal rhinitis
964	allergic rhinitis nos
1095	discoid eczema
1240	flexural eczema
1243	o/e - allergic rash
1275	allergic reaction
1424	infected eczema
1468	perennial rhinitis
1674	Chronic ethmoidal sinusitis
1741	atopic dermatitis/eczema
1838	allergic rhinitis due to pollens
1930	house dust mite allergy
1973	allergic drug reaction nos
2011	acute atopic conjunctivitis
2257	Chronic sinusitis
2290	allergic asthma
2372	allergic rhinitis due to other allergens
3162	house dust allergy
3699	hand eczema
3798	hay fever - unspecified allergen
4425	egg allergy
4433	Chronic maxillary sinusitis
4684	discoid eczema

medcode	readterm
4861	cat allergy
4882	peanut allergy
5437	Chronic sinusitis NOS
5869	allergic (intrinsic) eczema
6180	atopic dermatitis nos
6274	allergic pharyngitis
6399	contact dermatitis and other eczemas
6400	allergic urticaria
7146	extrinsic (atopic) asthma
7179	cow's milk allergy
7309	[v]personal history of aspirin allergy
7426	allergic contact dermatitis due drugs in contact with skin
7530	allergic reaction to bee sting
7796	latex allergy
10182	nut allergy
10546	Chronic rhinosinusitis
10840	neurodermatitis - atopic
11132	allergic contact dermatitis due to adhesives
11148	allergic reaction to venom
11306	seen by clinical allergist
11352	h/o: aspirin allergy
12239	h/o: multiple allergies
12382	allergy to animal
13223	atopic dermatitis and related conditions
13377	dander (animal) allergy
13378	perfume contact dermatitis
13401	allergic reaction to insect bite
13408	allergic reaction to wasp sting
13409	h/o: cat allergy
14645	chronic rhinitis nos
14688	other chronic allergic conjunctivitis
15163	Chronic frontal sinusitis
15722	allergy drug side effect
15795	allergic enteritis
16134	hay fever - other allergen
16556	allergic purpura
16676	acute allergic conjunctivitis
16685	allergic dermatitis - eyelid
16832	contact or allergic eyelid dermatitis
17173	Recurrent sinusitis
18207	allergic bronchitis nec
18572	allergic rhinosinusitis
19862	allergy skin test positive
20023	allergic enterocolitis
21232	allergic asthma nec
22763	bronchial allergy challenge
22764	[x]exacerbation of eczema
28589	chronic simple rhinitis
29458	allergy test positive
29845	allergic otitis media nos

medcode	readterm
30157	under care of clinical allergist
30375	feather allergy
30664	allergic contact dermatitis due to food in contact with skin
33820	allergic gastritis
33959	allergic parotitis
35086	allergic purpura nos
37597	chronic allergic otitis media
38383	allergic contact dermatitis due to dyes
39501	Chronic pansinusitis
41618	allergic contact dermatitis due to plants, except food
46977	allergic alveolitis and pneumonitis nos
47599	[x]other allergic rhinitis
48703	Chronic sphenoidal sinusitis
49548	Other chronic sinusitis
53095	allergic alveolitis and pneumonitis nos
53414	allergic eosinophilia
54375	Other chronic sinusitis NOS
59742	[x]allergic contact dermatitis due to other agents
62442	allergic extrinsic alveolitis nos
63733	[X]Other chronic sinusitis
63780	acute allergic sanguinous otitis media
70788	acute allergic serous otitis media
72240	allergic arthritis of multiple sites
72490	[x]other seasonal allergic rhinitis
73453	[x]personal history of allergy to other antibiotic agents
73749	allergic arthritis of other specified site
73880	[x] adverse reaction to antiallergic and antiemetic drugs
91301	[v]personal history of vitamin d3 allergy
94213	oral allergy syndrome
95938	dog allergy
104056	acute allergic mucoid otitis media
105338	allergen specific ige antibody level
108904	Atopy

Codelist 9: Anxiety

medcode	readterm
131	Anxiousness
462	Panic attack
514	Tension - nervous
636	Anxiety states
655	Anxiety with depression
791	Nervous breakdown
962	[X]Anxiety neurosis
1582	Nervous exhaustion
1758	Chronic anxiety
2509	[D]Nervousness
2524	Worried
3076	Agoraphobia with panic attacks
3328	General nervous symptoms
4069	Panic disorder
4081	[X]Panic state
4534	Anxiety state NOS
4634	Recurrent anxiety
4659	Generalised anxiety disorder
5385	[X]Other anxiety disorders
5902	Anxiousness - symptom
6221	Separation anxiety disorder
6408	[X]Panic attack
6939	Anxiety state unspecified
7749	[X]Mild anxiety depression
7999	Anxiety counselling
8205	[X]Panic disorder [episodic paroxysmal anxiety]
8424	[X]Anxious [avoidant] personality disorder
8725	O/E - nervous
10344	[X]Generalized anxiety disorder
10390	Fear of death
10723	[D]Nervous tension
11890	C/O - panic attack
11913	[X]Mixed anxiety and depressive disorder
11940	Acute panic state due to acute stress reaction
12838	Agoraphobia without mention of panic attacks
13124	O/E - anxious
14890	[X]Panic disorder with agoraphobia
16729	[X]Agoraphobia without history of panic disorder
17687	[X]Dream anxiety disorder
19000	O/E - panic attack
20089	General nervous symptom NOS
20163	Apprehension
23838	[X]Anxiety disorder, unspecified
24066	[X]Other specified anxiety disorders
25638	[X]Anxiety NOS
26331	O/E - fearful mood
28167	[X]Anxiety hysteria
28381	Alleviating anxiety

medcode	readterm
28925	Referral for guided self-help for anxiety
29608	
34064	[X]Phobic anxiety disorder, unspecified
35825	[X]Anxiety reaction
38155	O/E - afraid
40431	Cries easily
44321	[X]Other mixed anxiety disorders
50191	[X]Anxiety state
56924	Adjustment reaction with anxious mood
93401	Anxious
101422	Feeling low or worried

Codelist 10: Depression

medcode	readterm
2716	H/O: depression
1996	Depressed
4824	C/O - feeling depressed
9796	Symptoms of depression
10438	Depressive symptoms
19439	Depression resolved
18702	Postnatal depression counselling
44848	Depression management programme
30483	Patient given advice about management of depression
12399	Depression annual review
12122	Depression medication review
30405	Depression interim review
42931	On depression register
44936	Removed from depression register
48970	Exception reporting: depression quality indicators
28970	Excepted from depression quality indicators: Patient unsuita
43239	Excepted from depression quality indicators: Informed dissen
30583	Depression - enhanced services administration
65435	Depression - enhanced service completed
96995	On full dose long term treatment depression - enh serv admin
51258	Depression monitoring administration
71009	Depression monitoring first letter
72966	Depression monitoring second letter
91105	Depression monitoring third letter
88644	Depression monitoring verbal invite
85852	Depression monitoring telephone invite
2560	Depressive psychoses
10610	Single major depressive episode
5879	Agitated depression
6546	Endogenous depression first episode
6950	Endogenous depression first episode
595	Endogenous depression
34390	Single major depressive episode, unspecified
16506	Single major depressive episode, mild
15155	Single major depressive episode, moderate
15219	Single major depressive episode, severe, without psychosis

medcode	readterm
32159	Single major depressive episode, severe, with psychosis
43324	Single major depressive episode, partial or unspec remission
57409	Single major depressive episode, in full remission
7011	Single major depressive episode NOS
15099	Recurrent major depressive episode
6932	Endogenous depression - recurrent
35671	Recurrent major depressive episodes, unspecified
29342	Recurrent major depressive episodes, mild
14709	Recurrent major depressive episodes, moderate
25697	Recurrent major depressive episodes, severe, no psychosis
24171	Recurrent major depressive episodes, severe, with psychosis
56273	Recurrent major depressive episodes,partial/unspec remission
55384	Recurrent major depressive episodes, in full remission
6482	Recurrent depression
25563	Recurrent major depressive episode NOS
9183	Masked depression
8478	Reactive depressive psychosis
17770	Psychotic reactive depression
1055	Agitated depression
655	Anxiety with depression
1131	Neurotic depression reactive type
2639	Postnatal depression
10455	Depressive personality disorder
1533	Brief depressive reaction
36246	Brief depressive reaction NOS
16632	Prolonged depressive reaction
324	Depressive disorder NEC
2972	Postviral depression
4323	Chronic depression
20785	[X]Post-schizophrenic depression
11055	[X]Schizoaffective disorder, depressive type
35274	[X]Schizoaffective psychosis, depressive type
41022	[X]Schizophreniform psychosis, depressive type
4639	[X]Depressive episode
9055	[X]Single episode of depressive reaction
18510	[X]Single episode of psychogenic depression
7604	[X]Single episode of reactive depression
11717	[X]Mild depressive episode
9211	[X]Moderate depressive episode
9667	[X]Severe depressive episode without psychotic symptoms
41989	[X]Single episode agitated depressn w'out psychotic symptoms
22806	[X]Single episode major depression w'out psychotic symptoms
59386	[X]Single episode vital depression w'out psychotic symptoms
12099	[X]Severe depressive episode with psychotic symptoms
24117	[X]Single episode of major depression and psychotic symptoms
52678	[X]Single episode of psychogenic depressive psychosis
24112	[X]Single episode of psychotic depression
28863	[X]Single episode of reactive depressive psychosis
10667	[X]Mild depression
98346	[X]Major depression, mild

medcode	readterm
98252	[X]Major depression, moderately severe
98414	[X]Major depression, severe without psychotic symptoms
98417	[X]Major depression, severe with psychotic symptoms
101054	[X]Single major depr ep, severe with psych, psych in remiss
101153	[X]Recurr major depr ep, severe with psych, psych in remiss
6854	[X]Other depressive episodes
10720	[X]Atypical depression
56609	[X]Single episode of masked depression NOS
2970	[X]Depressive episode, unspecified
543	[X]Depression NOS
3291	[X]Depressive disorder NOS
28248	[X]Prolonged single episode of reactive depression
5987	[X] Reactive depression NOS
3292	[X]Recurrent depressive disorder
8851	[X]Recurrent episodes of depressive reaction
19696	[X]Recurrent episodes of psychogenic depression
8902	[X]Recurrent episodes of reactive depression
28756	[X]Seasonal depressive disorder
8826	[X]SAD - Seasonal affective disorder
29784	[X]Recurrent depressive disorder, current episode mild
29520	[X]Recurrent depressive disorder, current episode moderate
33469	[X]Recurr depress disorder cur epi severe without psyc sympt
11329	[X]Endogenous depression without psychotic symptoms
11252	[X]Major depression, recurrent without psychotic symptoms
29451	[X]Manic-depress psychosis,depressd,no psychotic symptoms
73991	[X]Vital depression, recurrent without psychotic symptoms
47009	[X]Recurrent depress disorder cur epi severe with psyc symp
23731	[X]Endogenous depression with psychotic symptoms
28677	[X]Manic-depress psychosis,depressed type+psychotic symptoms
32941	[X]Recurr severe episodes/major depression+psychotic symptom
31757	[X]Recurr severe episodes/psychogenic depressive psychosis
16861	[X]Recurrent severe episodes of psychotic depression
37764	[X]Recurrent severe episodes/reactive depressive psychosis
22116	[X]Recurrent depressive disorder, currently in remission
47731	[X]Other recurrent depressive disorders
44300	[X]Recurrent depressive disorder, unspecified
36616	[X]Monopolar depression NOS
8584	[X]Depressive neurosis
10290	[X]Depressive personality disorder
7737	[X]Neurotic depression
15220	[X]Persistant anxiety depression
19054	[X]Recurrent brief depressive episodes
11913	[X]Mixed anxiety and depressive disorder
7749	[X]Mild anxiety depression
13307	[X]Postnatal depression NOS
4979	[X]Postpartum depression NOS
32845	[X]Depressive conduct disorder