Phenotypes of Airway Diseases in Adults and Variation by Socioeconomic Status

Department of internal medicine and clinical nutrition

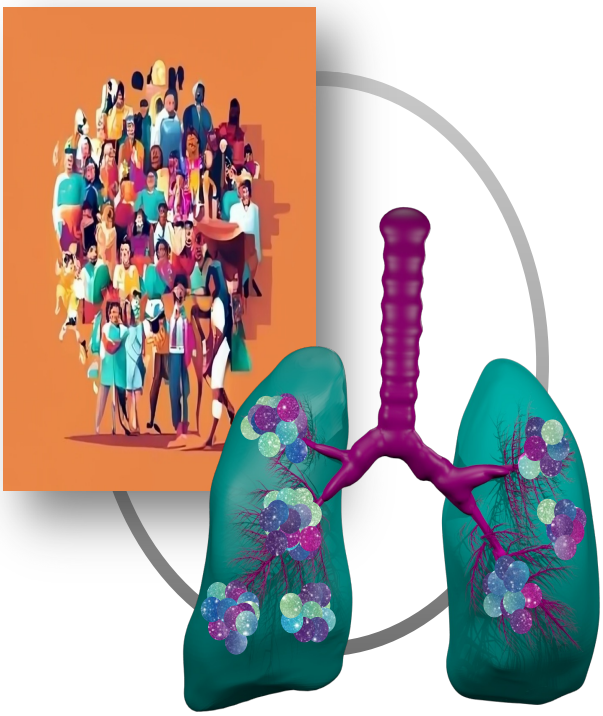
Institute of medicine

Sahlgrenska Academy, University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2024

Cover illustration: Self-produced by Artificial intelligence



Muwada.bashir@gu.se

ISBN 978-91-8069-744–6 (PRINT)
ISBN 978-91-8069-743-9 (PDF)
http://hdl.handle.net/2077/78883

Printed in Borås, Sweden 2024
Printed by Stema Specialtryck AB

”وقُلِ اعْمَلُوا فَسَيَرَى اللَّهُ عَمَلَكُمْ وَرَسُولُهُ وَالْمُؤْمِنُونَ ۖ وَسَتُرَدُّونَ إِلَىٰ عَالِمِ الْغَيْبِ وَالشَّهَادَةِ فَيُنَبِّئُكُم بِمَا كُنتُمْ تَعْمَلُونَ“

"Say, [O Muhammad], 'Do as you will, for Allah will see your deeds, and [so, will] His Messenger and the believers. And you will be returned to the Knower of the unseen and the witnessed, and He will inform you about what you used to do.”

Holy Quran, Surah At-Tawbah, verse 105

Phenotypes of Airway Diseases in Adults and Variation by Socioeconomic Status

Muwada Bashir Awad Bashir

Department of internal medicine and clinical nutrition , Institute of medicine
Sahlgrenska Academy, University of Gothenburg
Gothenburg, Sweden

ABSTRACT

Background: Lately, global prevalence of chronic obstructive airway diseases has risen, though some regions have seen stability. These diseases are heterogenous, have diverse manifestations and varying phenotypes. Ongoing research aims to identify and characterize phenotypes of obstructive airway diseases to enhance understanding of risk factors and improve treatments. The breadth and diversity of currently uncovered phenotypes of obstructive airway diseases are unclear. In the Nordic countries, there are social disparities in airway diseases, with those at lower socioeconomic status facing higher morbidity and mortality. Various socioeconomic indicators have been used to link socioeconomic status (SES) to airway diseases, but their impact on distinct asthma phenotypes remains unclear. Smoking is a risk factor for COPD, but its role in asthma remains controversial. While much is known about the independent impact of SES and smoking on obstructive airway diseases, their mutual synergism is unclear, particularly in relation to phenotypes of obstructive airway diseases. The goal of this thesis was to summarize the phenotypes of obstructive airway diseases and explore the connections between SES and phenotypes of airway diseases.

Objectives: The overarching aim of the thesis was to characterize phenotypes of obstructive airway diseases and determine their variation by SES. The thesis consisted of 6 papers: (1) a study of association of level of education and asthma control in adult-onset asthma; (2) a study on the effect modification between smoking and socioeconomic status towards risk of obstructive airway diseases; and (3) a systematic review of studies on computational phenotyping of airway disease using unsupervised computational methods; (4) a study on association between social status and forms/phenotypes of rhinitis; (5) a study

on phenotyping asthma in general population using machine learning approach and (6) a study on association between socioeconomic status and phenotypes of asthma.

Methods: In Paper 1, individuals with adult-onset asthma were examined in three studies: the Obstructive Lung Disease In Northern Sweden (OLIN) (n = 593), Seinäjoki Adult Asthma Study (SAAS) (n = 200), and West Sweden Asthma Study (WSAS) (n = 301) during 2009-2014. Education levels were primary, secondary, or tertiary, with uncontrolled asthma defined as asthma control test (ACT) score ≤19. Eight hundreds ninety-six participants with complete data were included (OLIN n = 511, SAAS n = 200, WSAS n = 185). Paper 2 analyzed data from WSAS (23,753 participants) and OLIN (6,519 participants) aged 20-75. Bayesian analysis explored smoking and socioeconomic status effect modification on respiratory outcomes. Paper 3 was a systematic review that synthesized studies on computational phenotyping of airway diseases published 2010-2020. In Paper 4 and 6, we estimated the association between education and occupational classes and risk of forms of rhinitis and asthma phenotypes, respectively. Paper 5 employed a deep clustering algorithm for clustering asthma, that was defined based on self-report from general population.

Results: In Paper 1, the odds ratio (OR) for primary education in relation to uncontrolled asthma was 1.92 (95% CI 1.15-3.20) in daily ICS users and 3.42 (95% CI 1.30-8.96) in non-atopic subjects. In Paper 2, SES and smoking had marginal synergistic effect on respiratory outcomes, with more pronounced effect-modification seen among lower occupational classes regarding smoking effect on allergic asthma, while among professionals, it affected non-allergic asthma. In Paper 3, there were variations in the literature regarding computational phenotyping of obstructive airway diseases, particularly in study design and methods, study settings, participant profiles, and variables used to perform the phenotyping exercise, altogether leading to differences in characterized phenotypes. In paper 4, higher education and occupation levels were identified to be risk factors for allergic and chronic rhinitis. Paper 5 identified four asthma phenotypes that were distinguishable by age of onset, severity, risk factors, and prognosis. Paper 6 showed an association between high and low education levels and presenting certain asthma phenotypes.

Conclusion: In affluent Nordic countries, classical phenotypes of asthma are present in the general population. Lower education poses a risk for uncontrolled adult-onset asthma, while high education levels were linked to

high risk of allergic and chronic forms of rhinitis, suggesting the significance of incorporating social aspect into management and prevention of airway diseases.

Keywords: airway diseases, asthma, COPD, asthma control, phenotypes, social determinants, adult-onset asthma, allergic asthma, non-allergic asthma, unsupervised phenotyping, education, occupational exposures, Bayesian analysis, rhinitis, allergic, chronic rhinitis, machine learning.

ISBN 978-91-8069-744–6 (PRINT)
ISBN 978-91-8069-743-9 (PDF)

SAMMANFATTNING PÅ SVENSKA

Bakgrund: Förekomsten av obstruktiva lungsjukdomar har ökat globalt under de senaste åren, om än stabiliserats i vissa regioner. Dessa sjukdomar uppvisar stora variationer, bland annat gällande symptombild, drabbade grupper samt relaterade hälsoproblem och risker. Pågående forskning syftar till att identifiera och karakterisera fenotyper av obstruktiva lungsjukdomar för att fördjupa förståelsen för riskfaktorer och förbättra behandlingar. Bredden och mångfalden av dessa fenotyper är i dagsläget oklar. I de nordiska länderna föreligger social ojämlikhet med högre sjuklighet och dödsfall i obstruktiva lungsjukdomar bland personer med lägre socioekonomisk status. Ett flertal socioekonomiska mått har använts för att länka socioekonomisk status med obstruktiv lungsjukdom, men dess inverkan på distinkta fenotyper av astma är fortsatt ofullständigt utredda. Likaså utgör rökning en riskfaktor för att utveckla kronisk obstruktiv lungsjukdom (KOL), men dess roll i astma är kontroversiell. Och fastän mycket redan är känt gällande den oberoende inverkan av socioekonomisk status och rökning på obstruktiva lungsjukdomar, så är deras synergiska effekt än så länge okänd, speciellt vad gäller specifika fenotyper av obstruktiv lungsjukdom. Målet med denna avhandling har varit att sammanfatta fenotyper av obstruktiva lungsjukdomar samt utforska kopplingen mellan socioekonomisk status och specifika fenotyper av obstruktiva lungsjukdomar.

Syfte: Det övergripande syftet med denna avhandling har varit att karakterisera fenotyper av obstruktiva lungsjukdomar och fastställa variationer inom dessa avseende socioekonomisk status. Avhandlingen består av sex artiklar: (1) en studie om associationen mellan utbildningsnivå och astmakontroll i vuxendebuterad astma; (2) en studie om effektmodifikationen mellan rökning och socioekonomisk status för risken av obstruktiv lungsjukdom; (3) en systematisk litteraturöversikt av studier som utforskat fenotyper av obstruktiva lungsjukdomar med artificiell intelligens (AI)-modeller; (4) en studie om associationen mellan social status och former/fenotyper av rinit; (5) en AI-baserad fenotypning av astma i den allmänna befolkningen; och (6) en studie om associationen mellan socioekonomisk status och fenotyper av astma.

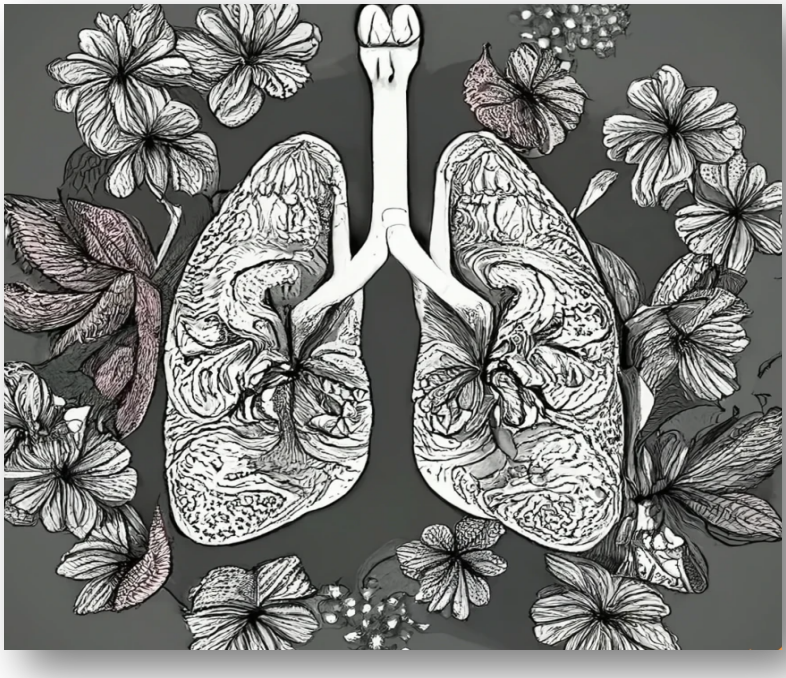
Metoder: I artikel 1 undersöktes individer från tre populationsbaserade studier från 2009–2014 i Sverige och Finland (totalt 1,094 individer) med astma som

debuterat i vuxen ålder. Utbildningsnivå definierades med tre nivåer (primär, sekundär och tertiär) och okontrollerad astma med ≤19 poäng i Astma Kontroll Test (AKT). Totalt 896 individer med fullständiga data inkluderades. I artikel 2 analyserades data från två svenska populationsbaserade studier (totalt 30,272 individer) i åldern 20–75 år. Med Bayesianska analysmetoder undersöktes effektmodifikationen av rökning och socioekonomisk status på utfall av obstruktiv lungsjukdom. Artikel 3 utgjordes av en systematisk litteraturöversikt av studier med AI-baserad fenotypning av lungsjukdomar publicerade 2010-2020. I artikel 4 och 6 kartlades associationen mellan utbildningsnivå/profession och risken för olika former av rinit och astmafenotyper. I artikel 5 utfördes en klusteranalys (fenotypning) av astma i ett populationsbaserat underlag med en så kallad *Deep learning* AI-algoritm.

Resultat: I artikel 1 var oddskvoten (OR) för utbildning på grundskolenivå i relation till okontrollerad astma 1.92 (95% konfidensintervall [95%CI] 1.15-3.20) bland de som dagligen nyttjar inhalationssteroider, och 3.42 (95%CI 1.30-8,96) bland individer utan atopi. I artikel 2 hade socioekonomisk status och rökning marginell synergisk effekt på respiratoriska utfallsmått, med mer uttalad effektmodifikation bland individer med lägre utbildning gällande röknings effekt på allergisk astma, medan det bland individer med högkvalificerade yrken påverkade icke-allergisk astma. I artikel 3 fann vi variationer i den publicerade litteraturen gällande AI-baserad utforskning av fenotyper av obstruktiva lungsjukdomar, i synnerhet gällande studiedesign och metoder, kontext, karakteristika hos studiedeltagare samt variabler som användes för att identifiera fenotyper; sammantaget bidrog dessa faktorer till skillnaderna i de identifierade/karakteriserade fenotyperna. I artikel 4 identifierades högre nivå av utbildning och profession som riskfaktorer för att utveckla allergisk och kronisk rinit. I artikel 5 fann vi fyra astmafenotyper som kunde särskiljas med avseende på debutålder av astma, sjukdomssvårighet, riskfaktorer och prognos. I artikel 6 visade vi på en association mellan hög/låg utbildningsnivå och specifika astmafenotyper.

Slutsats: I de välbeställda nordiska länderna är klassiska fenotyper av astma vanligt förekommande i den allmänna befolkningen. Lägre utbildningsnivå utgör en risk för okontrollerad vuxendebuterad astma, medan högre utbildningsnivå är associerad med ökad risk för allergisk och kronisk rinit, vilket indikerar betydelsen av att inkorporera socioekonomiska aspekter i behandling och prevention av lungsjukdomar.

TABLE OF CONTENTS



LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

I. Stridsman C, Ilmarinen P, Bashir MB, Tuomisto LE, Vähätalo I, Goksör E, Kankaanranta H, Backman H, Rönmark E, Lundbäck B, Nwaru BI. Level of education and asthma control in adult-onset asthma in Finland and Sweden-A report from the Nordic EpiLung Study.

Published in Journal of asthma, 22 Apr 7;59(4):840-9

II. Bashir MB, Basna R, Hedman L, Backman H, Ekerljung L, Andersén H, Wennergren G, Bhatta L, Lindberg A, Lundbäck B, Kankaanranta H, Rönmark. Nwaru BI. Interaction of smoking and social status on the risk of respiratory outcomes in a Swedish adult population: A Nordic Epilung study.

Published in Respiratory Medicine. 2023 May 1; 211:107192. Submitted correction text. **Status:** under review

III. Bashir MB, Basna R, Zhang GQ, Backman H, Lindberg A, Ekerljung L, Axelsson M, Hedman L, Vanfleteren L, Lundbäck B, Rönmark E, Kankaanranta H, Nwaru BI. Computational phenotyping of obstructive airway diseases: a systematic review.

Submitted to journal of Asthma and Allergy– 2024. **Status:** under review

IV. Bashir MBA, Pullerits T, Ekerljung L, Backman H, Wennergren G, Kankaanranta H, Nwaru BI. Socioeconomic Status and Different Forms of Rhinitis in Swedish Adults.

Submitted to Clinical and Translational Allergy. 2024. **Status:** under revision.

V. Bashir MBA, Lisik D, Ermis SSO, Basna R, Abohalaka R, Ercan S, et al. Machine learning-derived asthma phenotypes in a representative Swedish adult population.

Submitted to Thorax. 2024. **Status:** under review.

VI. Bashir MBA, Basna R, Wennergren G, Rådinger M, Backman H, Goksör E, et al. Level of education, but not occupation, is differentially associated with asthma phenotypes in adults.

Submitted to Clinical and Experimental Allergy. 2024. **Status:** under review.

ABBREVIATIONS

AA	Allergic Asthma
ACO	Asthma COPD Overlap
ARIA	Allergic Rhinitis and Its Importance In Asthma
AR	Allergic Rhinitis
ASA	Aspirin Sensitive Asthma
BMI	Body Mass Index
CRS	Chronic Rhinosinusitis
CRSnP	Chronic Rhinosinusitis with Nasal Polyposis
GERD	Gastroesophageal Reflux Diseases
GINA	The Global Initiative for Asthma
GOLD	The Global Initiative for Obstructive Lung Disease
HDM	High Density Molecules
LBRS	Lung Volume Reduction Surgery
LLN	Lower Limit of Normal
NAA	Non-Allergic Asthma
NAPT	Non-Allergic Provocative Test
NAR	Non-Allergic Rhinitis
NSAID	Non-Steroidal Anti-Inflammatory Drugs
OLIN	The Obstructive Lung Disease in Northern Sweden
RS	Rhino Sinusitis

SAAS	Seinäjoki Adult Asthma Study
SES	Socioeconomic Status
sIgE	Specific Immunoglobulin E
WSAS	West Sweden Asthma Study

1 INTRODUCTION

This thesis is built upon data obtained from three major population-based studies conducted in different regions Sweden and Finland: the West Sweden Asthma Study (WSAS) in western Sweden, the obstructive lung disease in Northern Sweden study (OLIN), and the Seinäjoki Adult Asthma Study (SAAS) in Finland. Data from WSAS was collected in two separate periods: WSAS1, conducted from 2008 to 2022, and WSAS2, an ongoing study initiated in 2016. The OLIN study was carried out from 1985 to 2016 and the data used for this thesis scope was collected in 2016. For SAAS, the data was collected over the study span from 1999 to 2013.

The primary objective of these three studies is to characterize obstructive airway diseases in adults both from clinical and population perspectives, particularly asthma, chronic obstructive pulmonary disease (COPD), and rhinitis. WSAS and OLIN data utilized in this thesis comprised information gathered from a postal questionnaire distributed among a representative sample of the population. Additionally, clinical investigations were conducted on a subset of randomly selected individuals who had responded to the postal questionnaires. For SAAS, information was gathered from clinical setting where patients were recruited from hospital settings.

In the following sections, this thesis will provide background information, including definitions and fundamental concepts related to the objectives of the conducted studies. It will also outline the methods employed, present the results obtained, and discuss the implications of these findings for future research.

1.1 HISTORY, DEFINITION AND EPIDEMIC OF AIRWAY DISEASES

1.1.1 ASTHMA

1.1.1.1 HISTORY

The term "asthma" originates from the Greek word "asthma" meaning "breathing difficulty with gasping.[1] " The earliest recorded mention of asthma in medical history comes from China, where it was described as "noisy

breathing." Over time, stories emerged of herbal remedies being used to alleviate episodes of labored breathing. Gaius Plinius Secundus called Pliny the Elder, a Roman writer from around 50 ADS, advocated the use of ephedra, a plant from which ephedrine was later derived, to treat breathing difficulties. Both Hippocrates and Pliny acknowledged the influence of environmental factors in triggering breathing difficulties, highlighting the role of activities like metalworking and exposure to pollen. [2]

As we delve into more recent history, Henry Hyde Salter emerges as a prominent figure who made significant contributions to our present-day comprehension of asthma [3]. Not only did he recognize the impact of external triggers, but he also emphasized the importance of neuronal and inflammatory pathways in the development of asthma [3]. During the period spanning from the 1920s to the 1970s, several investigations uncovered various agents that significantly impacted the airways of individuals afflicted with asthma. These included mediators like histamine, prostaglandins, and leukotrienes. Additionally, subsequent research emphasized the significance of mast cells, basophils, and eosinophils, leading to the introduction of the notion of asthma as an inflammatory disorder [1]. The 20th century ushered in a period of increased understanding of asthma, with numerous studies conducted between 1920 and 1970 shedding light on the role of key mediators within the airways of asthma patients. Additionally, a clearer distinction emerged between allergic and non-allergic components of asthma, leading to a deeper understanding of the inflammatory aspects contributing to the development of the condition.[4]

1.1.1.2 DEFINITION OF ASTHMA

Asthma is a heterogenous disease, characterized by chronic airway inflammation. It is characterized by respiratory symptoms such as wheeze cough, shortness of breath and chest tightness, that vary overtime and intensity, together with variable expiratory airflow limitation [5]. According the global initiative of asthma diagnosis (GINA), [5] asthma is defined based onset of symptoms that vary in presentation from one patients to another. Similarly, the triggers of such symptoms are several and various like exercise, allergens, irritants, weather changes or external factors like infections. Presentation could also vary in sense of some patients resolving spontaneously, others depending on medication, and some patients experiencing episodic flare ups in oppose to other who have more chronic continuous symptoms, or in term of severity and response to treatment.

1.1.1.3 CLINICAL APPROACHES TO ASTHMA EVALUATION, DIAGNOSIS AND CONTROL

Addressing obstructive airway diseases in clinical practice poses significant challenges due to their heterogeneous nature and the variability in how they present based on patient characteristics and disease attributes [5, 6]. Therefore, when dealing with these diseases in adult patients, a comprehensive approach involves considering the patient's medical history, presentation at the time of illness, and relevant biological tests.[5, 6]

The key criteria for defining asthma, according to GINA (Global Initiative for Asthma) guidelines, include symptoms characteristic of asthma and airflow limitation suggestive of asthma [5, 6]. To differentiate these symptoms from those of other common diseases, several criteria have been proposed. These include the episodic pattern of symptoms, exacerbation triggered by specific factors such as weather conditions, allergen exposure, or exercise. Additionally, asthma symptoms tend to worsen at night and in the early morning and are aggravated by viral infections. Asthma symptoms often coexist in the form of at least two symptoms rather than occurring in isolation [5, 6]. Some features, such as a persistent cough, exercise-induced dyspnea with wheezing, or shortness of breath accompanied by lightheadedness and fainting, are considered less indicative of asthma.

Biological assessment of asthma involves measuring the degree of airflow obstruction using spirometry, which assesses the extent of obstruction and its reversibility, a hallmark feature of asthma. Spirometry is conducted following standardized protocols in clinical settings[5, 6]. The criteria for defining asthma through spirometry include a reduction in the forced expiratory volume during the first second of expiration (FEV1) by 20% from the age- and gender-standardized lower limit of normal (LLN), along with documented excessive variability in lung function. Excessive variability can be confirmed by one or more of the following:

- A positive response to bronchodilator administration (FEV1 increase by >12% or >200ml).
- Excessive daily peak expiratory flowmeter (PEF) variability (>10%) over two weeks.
- Improved lung function (FEV1 increase by >=12% and >200ml from baseline) after four weeks of treatment.
- A positive response to exercise challenge testing (FEV1 fall by >12% or >200ml from baseline).

- A positive response to bronchial challenge testing (FEV1 fall of >=20% after methacholine or >=15% after mannitol).
- Variation in lung function measures (>12% of FEV1) between standard visits.

Patient medical history, clinical presentation, and the results of biological tests are all equally important in confirming an asthma diagnosis before initiating treatment [5, 6].

For patients with a confirmed asthma diagnosis, assessing asthma control is a crucial aspect of evaluation. Asthma control refers to the extent to which disease manifestations are observed or can be reduced or eliminated through treatment [5, 6]. Two main components characterize asthma control: symptom severity and the risk of future exacerbations. Various factors contribute to these aspects, including genetics, comorbid conditions, psychological factors, treatment adherence, and proper inhaler technique[5, 6].

Evaluating asthma control involves assessing symptoms over the past four weeks and identifying risk factors for exacerbations, along with regular lung function measurements. Standard tools, such as the Asthma Control Test (ACT) and the Asthma Control Questionnaire (ACQ), are used to assess symptoms[5, 6]. Evaluating risk factors for adverse outcomes and persistent airflow limitation involves assessing various aspects, including risk factors for adverse outcomes, persistent airflow limitation, and potential side effects of medication.[5, 6]

Lung function measurements play a critical role in assessing the risk of adverse outcomes, as they do not always correlate well with symptom presentation. FEV1 is a key indicator of the risk of adverse outcomes, particularly when it is measured at less than 60% of predicted values [5, 6]. Significant improvement in FEV1 (>12% and 200ml) under specific treatment conditions, such as the use of inhaled corticosteroids (ICS), short-acting beta-agonists (SABA) within 4 hours, long-acting beta-agonists (LABA) within 12 hours, or once-daily LABA within 24 hours, indicates uncontrolled asthma.[5, 6]

1.1.1.4 EPIDEMIOLOGY OF ASTHMA AND ITS SUBTYPES

Epidemic wise, the global prevalence of asthma ranges from 1-18% [5]. Countries of the developed world have higher reported prevalence of asthma in comparison to developing countries. In Europe, near 10 million < 45 years of age has asthma with a prevalence of 8. 2% in adults and 9.4% in children,

as per 2019. Out of yearly 25 billion lost years of DALYs, globally, 5.2 of those are lost in Europe. [7] However, the overall trend of asthma prevalence was found to be declining in the latest years compared to the last decade. Globally the incidence of asthma keeps rising in low- and middle-income countries, however in developed countries the rise in asthma appears to have plateaued in the latest decades [8, 9].

Across the last two decades in Sweden, according to report from large cohort study in Northern Sweden, the incidence of asthma was found to be stable around 2.4 per 1000 person-years [10]. Similar figure was reported from WSAS, a large cohort study in western Sweden [11]. Asthma prevalence, in turn, was reported as 11.6 % as per 2022 , according to the public health agency of Sweden [12], with women exhibiting the highest prevalence of the asthma in population 13.3 % compared to men 9.9 %. In addition, reports from the large cohort of West Sweden asthma Study in Southwest Sweden showed that the prevalence of asthma among adults, defined based on reporting from study questionnaires as ever asthma, physician diagnosed asthma, use of asthma medication and current asthma, has increased between 2008 and 2016: 9.6% to 11%, 8.3% to 10%, 8.6% to 9.8%, and 8.1% to 9.1%, respectively. With such notably high asthma prevalence, Sweden lies among the leading European countries when it comes to the number of individuals affected asthma, with a crude mortality rate of 0.94%. Economy and health system burden wise, asthma accounted for 125 272 hospital bed days in the Swedish health care system as per 2018. For Sweden being ranked second or third amongst European countries with the lowest number of hospital beds for population size, this is a relatively high figure [13].

Considering asthma subtypes by allergic status, the prevalence of allergic asthma in Sweden showed a proportionate increase from 5%, to 6% and 7.3 %, in 1996,2006 and 2016 , respectively [14], according to reports from the large cohort study from Northern Sweden: OLIN. Non- allergic asthma however was stable ranging from 3.4 to 3.8 % [14]. Overall, allergic asthma is the most common form of asthma showing amongst 52% of asthmatics compared to non-allergic asthma that presents 48% of total asthma, as reported in large population base study from Finland. The prevalence of allergic asthma was inversely proportionate with age, prevailing among youngsters compared to others. In turn, the incidence, of non-allergic asthma was highest among higher age groups compared to low incidence among children and middle agers [15].

1.1.2 SEVERE ASTHMA

1.1.2.1 DEFINITION OF SEVERE ASTHMA

The term severe asthma refers to the state of confirmed asthma based on diagnostic criteria in addition to certain requirement for labelling asthma as severe. According to GINA, severity of asthma requires treatment with high dose of inhaled corticosteroids plus additional use of controller and /or systematic corticosteroids for more than 50 % of the time of the year to prevent asthma from being uncontrolled or asthma remaining uncontrolled despite such regime. As a clinical subtype of asthma, severe asthma has its own variability in clinical presentation, response treatment, and prognostic outcome. A heterogeneity that is attributable in part to different biological and molecular interplay in the disease pathogenesis along with differences in patients' endogenous and exogenous characteristics.

1.1.2.2 EPIDEMIOLOGY OF SEVERE ASTHMA

Overall, 3-10 % of those with asthma have severe asthma [5]. In Scandinavian countries, similar range of severe asthma was reported among adult asthmatics in Sweden, Norway, and Finland: 3.5 %, 5.5% and 5.2%, Half of those is managed in specialist care [16]. At primary care level, 4.2% of asthmatics had severe asthma, based on large cohort study in Sweden. Out of those, more than half reported poor control status and one of every five of them required tertiary medical care at least once, and primary medical at least 3 times [17]. Another report from WSAS has found a higher prevalence of severe asthma among asthmatics: 9.5% and an overall prevalence of 1.1 % from the general population. Severe asthma patients presented higher symptoms, activity limitation, comorbidities, and adverse respiratory outcomes. Strikingly, one third of them having no annual contact with health care [18].

1.1.3 COPD

1.1.3.1 DEFINITION OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE (COPD)

When considering COPD, it is crucial to examine it within the context of other similar obstructive respiratory diseases like emphysema, chronic bronchitis, and asthma, as there are overlaps and close interrelations between them. COPD is defined as "heterogeneous lung condition characterized by chronic respiratory symptoms (dyspnea, cough, expectoration, exacerbations) due to abnormalities of the airway (bronchitis, bronchiolitis) and/or alveoli (emphysema) that cause persistent, often progressive, airflow obstruction "[19]

The global initiative for chronic obstructive lung diseases outlines the diagnostic criteria for COPD, which include the presence of relevant symptoms like cough, dyspnea, and sputum production, a clinical context indicating prolonged smoking, and evidence of airway obstruction.[19]

Hence, the definition of COPD encompasses a broad range of respiratory diseases that share a common characteristic of airflow limitation caused by abnormalities in the airway (such as bronchitis, bronchiolitis, or inflammation and hyperresponsiveness) or abnormalities in the alveoli (emphysema).

Airflow limitation refers to the inability to exhale sufficient air. It is assessed by measuring the forced expiratory volume in one second (FEV_1) and the ratio of FEV_1 to the total forced expiratory volume (FEV_1/FVC) [20]. The significance of this concept in defining chronic obstructive pulmonary disease (COPD) stems from the fact that airflow limitation can have varying characteristics. It can be triggered by exogenous factors such as environmental conditions, temperature, and emotional factors. Additionally, airflow limitation can exhibit different patterns, being either reversible or persistent over time. Such diversity in the manifestation of airflow limitation plays a crucial role in distinguishing between different diseases encompassed within the broader spectrum of COPD.

Asthma is distinguished from COPD by its characteristic reversible pattern of airflow limitation. In contrast, COPD patients experience an irreversible pattern of airflow limitation, often accompanied by pathognomonic features of chronic bronchitis or emphysema in airways and lungs. A third condition termed Asthma-COPD overlap (ACO) refers to a condition where asthma patients exhibit an irreversible pattern of airflow limitation, but with a different underlying cause and disease process compared to COPD. Such ACO patients commonly share certain universally recognized features, including being over 40 years of age, having a history of early-onset asthma, and demonstrating persistent irreversibility or partial reversibility after using bronchodilators. [21]

1.1.3.2 EPIDEMIOLGY OF CHRONIC OBSTRCUTIVE PULLOMONARY DISEASES (COPD)

COPD is the third leading cause of death globally, resulting in approximately 3.23 million deaths in 2019. Moreover, it ranks as the seventh leading cause of overall poor health worldwide [22]. Recent review has estimated the global prevalence of COPD to 10.3% [23], affecting approximately 391.9 million adults worldwide. The predominant risk factor for COPD remains tobacco smoking,

contributing to around 70% of cases in high-income countries [22, 24]. However, the prevalence of COPD varies across studies due to differences in the definition used [24]. Many studies rely solely on spirometry measures rather than considering symptoms, and the prevalence estimates are often self-reported physician diagnosis. Consequently, evidence suggest that less than 6% of the global population has been officially diagnosed with COPD [24].

In Scandinavian countries, the assessment of the burden of obstructive lung disease has been clouded by the use of different criteria for defining the disease [25]. These criteria vary from relying solely on spirometry measurements to incorporating symptoms as well. Furthermore, the confusion is exacerbated by the use of different criteria sets, such as the Global Initiative for Chronic Obstructive Lung Disease (GOLD) recommendation of a fixed FEV1/FVC ratio below 70% with or without an FEV1 below 80%, and an alternative method involving the FEV1/FVC ratio below the 5th percentile, also known as the lower limit of normal (LLN) [26].

This lack of consistency in diagnostic criteria, coupled with the limited use of spirometry measurements, contributes significantly to the underdiagnosis or misclassification of chronic obstructive pulmonary disease (COPD). In some cases, individuals are diagnosed based on symptoms alone without meeting the spirometry criteria. This complexity further hampers efforts to understand and address the disease from both clinical and epidemiological perspectives [26, 27].

Older reports (1950-2000) from community based surveys in Nordic countries on COPD that is defined based on various physician defined criteria showed that 4-6% of adults suffer from clinically relevant COPD [28]. Previous studies in Sweden between 1996 and 2009 [29-32] have reported a prevalence range of 16-8.5 %, using mostly fixed ratio criteria for defining COPD. More recent studies from Sweden [33] and Norway [34] have shown a noticeable decline in prevalence of COPD from 10.4% to 7.5% and from 16.7%% to 14.8%, respectively, when both were measured using fixed ratio criteria.

1.1.3.3 ADDRESSING COPD ON AN EPIDEMIOLOGICAL LEVEL FACE SEVERAL CHALLENGES.

Despite the increased awareness and efforts to combat COPD, including successful anti-smoking campaigns that have contributed to a decline in its prevalence, there are still significant challenges associated with the disease. One major standing challenge in the path of disease identification and proper

prevention is the lack of standardized diagnostic criteria. Additionally, spirometry, a crucial tool for diagnosing COPD, is underutilized [27].

The heterogeneity of COPD further complicates its identification, as it can manifest differently in patients based on their individual criteria and exposure history to factors such as tobacco smoking, indoor and outdoor pollution, as well as occupational exposures, all contribute to the complexity of identifying and diagnosing COPD accurately and at an early stage [35].

Globally, the underdiagnosis of COPD varies widely, ranging from 10% to 90% [26]. Furthermore, misclassification of COPD patients as having other obstructive lung diseases, most commonly asthma, is prevalent, as revealed by multinational studies. A population-wide study from Sweden identified risk factors for underdiagnosis or misclassification of COPD to be socially relevant, including low education, smoking, female gender, as well as clinical like experiencing high symptoms, and being diagnosed with another obstructive lung disease such as asthma [27].

1.1.4 RHINITIS

1.1.4.1 DEFINITION AND SUBTYPES

Rhinitis is a diverse airway condition that encompasses a range of disorders, all marked by mucosal inflammation that can cause symptoms such as nasal blockage, itchiness, sneezing, and diminished sense of smell [36]. The literature distinguishes various types of rhinitis by their clinical features, causes, comorbidity status, chronicity and inflammation patterns [37]. One common type is allergic rhinitis, which is set off by airborne allergens and involves an immunoglobulin E (IgE) mediated response [38]. Another type, chronic rhinitis, is identified by chronicity status as symptoms of nasal inflammation that persist for more than four weeks. Rhinosinusitis is a form of rhinitis that is characterized by symptoms of nasal congestion, discharge, facial pain, or a reduced ability to smell, and requires the presence of at least two sino-nasal symptoms for diagnosis. If the later symptoms continue for over 12 weeks, the condition is classified as chronic-rhinosinusitis[39-41].

In the past, allergic rhinitis (AR) was primarily regarded as a nasal disorder. However, more recently, there has been a shift in perspective, recognizing it as part of a broader pathological process that involves other related conditions such as atopy and asthma [42]. This realization has led to a deeper understanding of the complexities surrounding allergic rhinitis [43]. It can be categorized into

different classes based on its manifestation. Perennial allergic rhinitis is characterized by intermittent symptoms that persist throughout the year, while seasonal allergic rhinitis is marked by symptoms that occur during specific seasons [44]. Individuals affected by allergic rhinitis typically present with nasal symptoms, including congestion, itching, sneezing, and a runny nose. Additionally, they may experience symptoms associated with allergic conjunctivitis, such as itchy and watery eyes. Other manifestations can include a nonproductive cough, sinusitis, and ear-related symptoms [42]. Allergic rhinitis has two main mechanisms of pathogenesis: an IgE-mediated hypersensitivity reaction to inhaled allergens and a non-IgE mediated pathway. The IgE-mediated pathway involves the activation of T helper 2 cells and the release of inflammatory mediators like prostaglandins, IL-4, and IL-13. The non-IgE mediated pathway involves the infiltration and activation of eosinophils in response to irritants, leading to inflammation and symptom exacerbation [42].

Chronic rhinitis (CR) is defined by the continuous presence of symptoms for over four weeks. Animal studies have identified a link between chronic rhinitis and both acute and chronic mixed lymphocytic, as well as acute neutrophilic inflammatory responses [45]. This persistent form of rhinitis is often associated with infectious etiologies [45]. The prevalence of chronic rhinitis is a growing concern worldwide, leading to significant illness and disability [46]. The consequences of prolonged nasal inflammation include a higher incidence of related conditions such as asthma, rhinosinusitis, and otitis media (middle ear inflammation). Beyond the physical ailments, chronic rhinitis can impair emotional well-being and is associated with reduced productivity, increased absenteeism from work or school, and diminished performance.[46]

Chronic rhinosinusitis (CRS) is a specific subtype of rhinitis characterized by sustained inflammation localized to the mucosa of the nasal cavity and paranasal sinuses. Symptoms of CRS typically include facial pain, nasal discharge, and impaired olfaction. It frequently co-occurs with conditions such as allergic rhinitis and asthma [47]. The inflammatory process underlying CRS is categorized by type 1, type 3, and/or type 2 inflammatory patterns, with or without the presence of concurrent nasal polyps. Diagnosis of CRS is facilitated through endoscopic examination and computed tomography (CT) imaging. There are two distinct sub phenotypes of CRS: one with nasal polyps and one without. These sub phenotypes differ in symptomatology, disease severity, comorbidity profiles, and therapeutic approaches.[47]

1.1.4.2 EPIDEMIOLOGY OF FORMS OF RHINITIS

The estimated prevalence of allergic rhinitis based on physician diagnosis is approximately 17% [12]. However, when considering symptoms, the estimated prevalence increases to around 30% [48] . Among the population of adolescents and children, the prevalence is notably high, with rates reaching 14.6% among those aged 13-14 years and 8.6% among those aged 6-7 years. Seasonal allergic rhinitis tends to have a higher prevalence compared to perennial allergic rhinitis. [42] In Sweden, as of 2022, the reported prevalence of allergy-related problems, according to the Swedish Agency of Public Health, stands at 20.3% [12]. Specific studies focusing on allergic rhinitis in Sweden showed an increase in prevalence from 28% in 2008 to 32% in 2016 among adults aged 16-25 years [49]. Furthermore, a comprehensive study conducted in Sweden, Norway, Denmark, Iceland, and Estonia reported a similar figure of 25.9% prevalence of allergic rhinitis up until 2001[50]. A separate study revealed that allergic rhinitis has a prevalence of 24% among the entire working-age population in Sweden [51]. Despite the relatively high remission rates of allergic rhinitis, reaching 31%, the total annual estimated cost of allergic rhinitis incurred by the Swedish population is as significant as 1.3 billion euros [50, 51].

Research into chronic rhinitis is comparatively sparse relative to other forms of the condition, possibly due to its classification under a broad spectrum that encompasses various rhinitis types. A specific definition noting symptom duration of over four weeks is rarer. In Sweden, prior studies have identified a 16.1% prevalence of nasal congestion, 14.1% prevalence of rhinorrhea, and 19.1% of participants experiencing chronic nasal symptoms [52]. Similarly, research from the UK has reported prevalence rates of 16% for nasal obstruction and 19% for runny nose [53].

Regarding chronic rhinosinusitis (CRS), it is estimated to affect 10% of the adult population in Europe and America, exerting significant economic and health-related costs [47]. Analysis of 12 European studies comparing CRS prevalence, based on self-reported or physician-diagnosed criteria against the standards set by the European Position Paper on Rhinosinusitis and Nasal Polyps (EP3OS), showed a strong correlation with prevalence around 10% [54]. In Sweden, the reported incidence of CRS in individuals without prior nasal symptoms was 2% [55]. In children and adolescents, the prevalence based on self-reported symptoms was 1.5%, and 2% when applying EP3OS criteria. However, these figures dropped to 0.8% when clinical examination was used to diagnose CRS, and to 0.3% when an endoscopic definition was applied [56].

1.2 PHENOTYPES OF AIRWAY DISEASES

Asthma and airway disease has now long been acknowledged as heterogenous diseases, with varying presentations among patients who vary by their endogenous and exogenous characteristics. airway diseases present differently also due to verity in their physio-pathological mechanisms which results in various disease manifestation among patients of the same characteristics. The acknowledgement of airway diseases heterogeneity has resulted in the adoption of a phenotypic approach to address airway diseases.

1.2.1 HISTORICAL EVOLUTION OF THE CONCEPT OF AIRWAY DISEASE HETEROGENEITY

In 1989, Snider and colleagues made one of what could be named an early attempt to characterize airway diseases heterogeneity. They did so by identifying common traits shared among various obstructive airway diseases [57]. He introduced a non-proportional Venn diagram to represent the subtypes of these diseases. Moreover, his work highlighted that these subtypes exhibited diverse clinical presentations, encompassing symptoms such as productive cough, or wheezing, distinct pathophysiological features like the reversibility of airway obstruction, observable radiographic changes, varying degrees of airway inflammation and hyperresponsiveness, as well as the severity of obstruction measured by FEV_1. Additionally, he also noted differences in the frequency of exacerbations necessitating medical intervention. Although these conditions are now clearly defined as separate diagnoses, the concept of differentiating diseases based on their observed characteristics while all falling under the umbrella of obstructive airway patterns marked an introduction to the concept of phenotyping.

Subsequently, the concept of overlapping characteristics gained widespread acceptance among the clinical community, as it provided valuable insights into the challenges associated with patient classification for specific management or prognostic prediction groups. Further scholarly investigations were undertaken, focusing on the clinical, patho-physiological, and patient-specific aspects of these distinct subtypes.

Further, asthma emerged as more remarked syndrome, characterized by precise pathophysiological features such as airway inflammation, hyperresponsiveness, and a reversible pattern of airway obstruction. These attributes were intricately connected to specific pathophysiological

mechanisms, as proposed by Lötvall et al. [58] This link between asthma and underlying disease mechanisms marked the initiation of the era of airway diseases endotyping, where an "endotype" is precisely defined as a "subtype of a condition defined by a different pathophysiological mechanism" [37].

Furthermore, emphysema was postulated as a phenotype exhibiting more severe inflammation, evident through observable radiographic changes in the airways. In contrast, chronic bronchitis was characterized by an elevation in FEV1 (forced expiratory volume in one second), particularly noticeable among young adults [37].

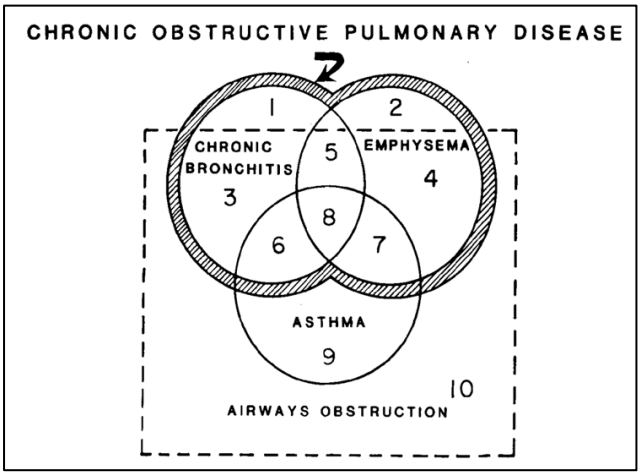


Figure 1: Scheme of chronic obstructive pulmonary disease. A nonproportional Venn diagram shows subsets of patients with chronic bronchitis, emphysema, and asthma in three overlapping circles [1].

1.2.2 AIRWAY DISEASE PHENOTYPES, ITS CLINICAL AND SCIENTIFIC UTILITY

A phenotype is defined as an observable variant of disease of specific demographic, clinical or pathophysiological characteristics, that are results of interaction between patient's endogenous characteristic (genotypes) and exogenous (environmental) Factors [5, 59, 60]. Hence, the same airway disease in presentation can have different characteristics based on the patients genetic and environmental features, further several pathophysiological mechanisms of the same disease can cause its various presentations among patients of the

[1] Snider G. Changes in COPD occurrence; chronic obstructive pulmonary disease. A definition and implication of structural determinants of airflow obstruction for epidemiology. Am Rev Respir Dis. 1989;**140**: s3-s8.

same characteristics. Further, the interaction between such pathophysiological, individual, and environmental factors can add up to the variety in disease prestation.

In attempts, of further understanding airway disease heterogeneity and phenotypes, focus had further shed into which features accurately characterize and differentiate phenotypes of airway diseases. Frequency and severity of exacerbation were proposed a key feature based on which phenotypes of COPD were characterized [61]. Clinical, demographic, and pathophysiological characteristics were also found to define subtypes of asthma [62-64].

The essence of such features identification is their utility in patients' identification and accurate diagnosis and classifying patients into subgroups that could be possibly targeted by specific pharmacological and non-pharmacological management approaches. Further, it assists better understanding of the disease pathognomonic, which further aid tailored actions for disease prevention and promotion of patients 'quality of life.

1.2.3 METHODS OF AIRWAY DISEASE PHENOTYPING, EVOLUTION AND STATE OF ART

Acknowledging the heterogeneity of airway diseases and the significance of gaining a deeper understanding of their complexities, scholarly efforts have taken diverse paths in the quest for subtyping these conditions over the past decades [65]. This pursuit has led to the proposal and implementation of variable approaches and methodologies in various research settings, generating a vast array of results [65].

During the early stages of the development of the concept of airway diseases, efforts to identify subtypes of these conditions were influenced by the philosophical paradigm of Sir William Osler, regarded as the father of modern medicine[66]. Osler's method involved identifying diseases based on the symptomatic portrayal of a physio-pathological process in specific organ anatomy or histology.[66]

Akin to the process of defining a disease status by linking symptoms to underlying pathophysiology, clinicians tended to correlate specific patients' characteristics (presumed to define a phenotype) with clinical, pathological, and physiological features to identify specific phenotypes [67]. In doing such,

clinicians rely on their knowledge of the disease, level of expertise, and presumptions. Thus, experts leaned on their presumed hypotheses of phenotype characterization.

While this approach proved valuable in confirming the taxonomy of airway diseases, it also had its limitations. It sometimes disregarded other contributing factors to a phenotype definition, such as individual susceptibility and various pre-clinical elements that play a role in shaping a phenotype [35, 67]. As a result, the produced phenotypes tended to be biased toward clinicians' assumptions and limited to the situational level of disease understanding.

In the airway disease literature, this approach is referred to as the hypothesis-driven approach of airway disease phenotyping[35, 67]. An alternative term that described this practice in phenotyping attempts pointes to the utilized approach of statistical modeling: the supervised mathematical modeling[35, 67]. Hence, in supervised phenotyping of airway diseases, patients are pre-labeled into specific phenotype group, and associations are learned by assessing how these phenotypes are linked to certain patients' clinical and pathophysiological features.

With the advancement of molecular medicine, network medicine, and breakthroughs in mathematical computational technologies for handling medical big data, there has been increasing interest in understanding disease pathogenesis from the ground up [68]. This involves delving into the finer details of individual factors contributing to disease causation and gradually progressing towards understanding major pathology. Hence, major shifting from the traditional Oslerian paradigm in understating disease definition and causation [35, 67].

To achieve this, computational approaches have been harnessed in the realms of network and biomolecular medicine, with the goal of uncovering hidden patterns within vast datasets related to airway diseases [35, 68]. Instead of relying on supervised methods that carry biases of the current state of knowledge, the focus has shifted to unsupervised computational unbiased approaches [35, 68, 69].

Unsupervised methods involve the application of computational algorithms to explore and identify inherent patterns within the data, potentially representing clinically relevant phenotypes of airway diseases [35, 67-70]. In doing such, the data is let to speaks itself and identify groups of subjects who have similar characteristics based on statistical parameters [70]. By adopting this approach,

the biases introduced by supervised methods can be minimized, while opening possibilities to discover new disease presentations and phenotypes that exist naturally [70, 71]. As a result, novel hypotheses about previously unrecognized subtypes of airway diseases can emerge, providing valuable insights for future research and clinical applications.

Supervised versus no- supervised methods of airway disease phenotyping.

Using the two of supervised and non-supervised approaches to study airway disease phenotypes, each has its own advantages and limitations. The supervised methods have proven useful, as they are based on clinical input and yield phenotypes that are clinically relevant and easily manageable for healthcare professionals. However, a significant drawback of this approach is the potential for bias, leading to a repetition of known information without venturing into new territories. [67]

On the other hand, unsupervised methods offer a solution to the originality issue, as they can uncover novel patterns in the data. However, their application introduces new challenges, particularly concerning technical difficulties and the need for specialized expertise in computational analysis. Clinicians may require assistance from those with technical knowledge to validate the clinical meaningfulness of the models' outputs [35, 70].

An important consideration when applying unsupervised methods to medical data is ensuring the quality of the data, in term of cleanings, preprocessing, completeness and representation of the general population [69, 72]. Failure to do so, and using data distorted by systematic bias or poor engineering, compromises the accuracy of the phenotyping process. Mathematical algorithms, by nature, can produce clusters in any data, regardless of data quality, making it crucial to use quality representative data for computational modelling.

Furthermore, achieving clinical meaningfulness of derived phenotypes is paramount. These phenotypes should be related to meaningful clinical outcomes, such as symptoms, exacerbations, therapy response, disease progression, or mortality [35, 69, 72-74]. Therefore, it is essential to validate the derived phenotypes against these disease parameters.

The inclusion or exclusion criteria for subjects in the data sample can also impact the accuracy of computational phenotyping [35, 69, 72, 73]. Specific criteria

may be employed to exclude certain groups, such as smokers in asthma studies, to focus solely on asthma subjects [75-81]. However, this approach may overlook asthma phenotypes with smoking-related features and hinder the identification of overlapping asthma-COPD phenotypes.

In defining the outcome for phenotypes, accuracy in diagnosing patients with the targeted disease, such as asthma, is crucial [35, 67, 72, 74]. Otherwise, the derived phenotypes may represent subtypes of other conditions rather than the intended disease.

To overcome these challenges, a careful and thoughtful application of unsupervised methods, along with thorough validation against clinical outcomes, is essential for the successful exploration and understanding of airway disease phenotypes.

1.3 A GLANCE ON COMMONLY REPORTED PHENOTYPES OF AIRWAY DISEASES

1.3.1 ASTHMA

Several clinical phenotypes of asthma have been reported in literature. Some of these are defined based on disease clinical presentation like, symptoms, their frequency and severity, status of accompanying atopy, time on onset [82-88]. Further, physiological parameters of disease aetiology like measures of lung function, level of inflammation and microanatomical pathological changes were also utilized [88-91]. Characteristics related to risk factors has also been used in attempt of extracting asthma phenotypes like sociodemographic factors of age, weight and gender, ethnicity[92-94]. Smoking status, exposure to specific triggers and indoor or outdoor pollutants were also considered under risk factor/etiological ration[86, 89, 95-98]. Another scope of asthma phenotyping included factors related to disease outcome, like mortality, morbidity, response to therapy, patterns of symptoms remission or progression or indices of disease impairment on health and quality of life [99-101]. Yet not all the extracted phenotypes were proven to be clinically relevant and of specific value in predicting disease clinical outcome, while only few presented clinical stabilities and proven to be sold and reapable across sufficient groups of asthma patients.

Allergic asthma: The allergic asthma phenotype stands as one of the earliest recognized asthma phenotypes and has consistently appeared in the literature on asthma phenotyping. It is typified by its onset during early childhood, often accompanied by concurrent allergic conditions such as allergic rhinitis, eczema, or allergies to specific foods or medications, and frequently exhibits a family history of allergic disorders [5, 62]. This phenotype also demonstrates a pattern of development from childhood into adulthood, initially manifesting as a mild form of asthma with manageable control and a favourable response to steroid treatment. The pathophysiology underlying this phenotype involves notable airway inflammation characterized by elevated levels of eosinophils in both the airways and blood[62].

Non-allergic asthma: this particular asthma phenotype stands apart in that it doesn't occur in conjunction with allergies. Noteworthy characteristics of this variant include its tendency to emerge at a later age than allergic asthma and its limited response to steroid treatments. Additionally, this phenotype often exhibits higher severity levels and reduced control compared to non-allergic asthma cases. Histopathological examinations suggest the involvement of various types of inflammatory cells, including neutrophils, eosinophils, or a limited number of other cells (Paucigranulocytic) [102]. Moreover, specific sociodemographic traits such as being female, and obesity have shown stronger associations with non-allergic asthma. This phenotype is also linked to a higher prevalence of comorbidities like gastroesophageal reflux disease (GERD) and chronic rhinitis and nasal inflammation [102].

Adults (late) onset asthma: The most recent version of the Global Initiative for Asthma Management and Control (GINA) has identified late-onset asthma as a well-established asthma phenotype. This phenotype is marked by the emergence of asthma in adulthood, particularly among women. Individuals with this phenotype tend to experience more severe symptoms and exhibit reduced responsiveness to steroid therapies. When defining this phenotype, it is crucial to rule out occupational asthma as the potential cause of adult-onset asthma. [5].

Obesity related asthma: the most recent release of the Global Initiative for Asthma Management and Control (GINA) has also recognized obesity-associated asthma as a distinct phenotype of asthma. This phenotype is characterized by a unique inflammatory reaction with low eosinophilic levels and higher severity. Numerous studies have outlined a consistent correlation between obesity and asthma, as well as an association with heightened

comorbidity rates [103]. This phenotype can manifest in childhood-onset and adult-onset asthma cases [86, 92-94, 104-106]. In the former, obesity is perceived as a complicating factor contributing to severe asthma, while in the latter, asthma is viewed as a consequent of life course obesity [107, 108]. For adults with obesity-related asthma, reports indicate the presence of both neutrophilic and pauci-granulocytic patterns of airway inflammation [108].

Non-steroidal anti-inflammatory drugs/aspirin sensitive asthma (ASA): This variant of asthma manifests in both children and adults. This type of asthma is specifically triggered by aspirin and NSAID derivative treatments. An additional characteristic of this variant is its association with nasal inflammatory conditions such as chronic rhinosinusitis, sinusitis, and nasal polyposis disease. Within this phenotype, various levels of severity have been observed, categorized as mild, moderate, and severe ASA. The severity is notably linked to factors like female gender and high eosinophilic levels. Interestingly, there are reports that indicate a correlation between higher levels of pro-inflammatory mediators and eosinophilia with severe ASA [98]. However, for cases categorized as moderate and mild, there appears to be a balanced pattern of pro and anti-inflammatory mediators, despite the presence of eosinophilia, implying that eosinophilia might not be a decisive factor for the severity of ASA [96].

Other phenotypes/endotypes of asthma

Other less reported asthma phenotypes include those characterized by distinct inflammatory statuses (eosinophilic, neutrophilic), the former are also referred to interchangeably as inflammatory asthma endotypes. Other asthma phenotypes are defined by specific triggers (such as exercise-induced and cold-triggered asthma), as well as variations in control status and severity [88, 92, 98, 104, 105, 109, 110]. However, it's important to note that the long-term stability of these phenotypes has not yet been firmly established. The fundamental question remains whether these individual traits, taken in isolation, are pivotal defining factors for a specific phenotype, or if these variations are integrated into more stable phenotypes that are delineated by a combination of different asthma characteristics. This question of phenotypic stability is still awaiting definitive answers. As the field of asthma phenotyping is still in its early stages, with research efforts continuously contributing to a growing wealth of data on asthma phenotypes, the process of synthesizing these findings into coherent and enduring subtypes of asthma, each with distinctive defining traits,

represents a ground-breaking milestone towards enhanced management and comprehension of airway diseases.

Asthma COPD overlap: this asthma phenotype, recognized by GINA, is characterized by airflow obstruction that exhibit a variable degree of bronchodilator reversibility [5]. The origins of this condition, known as Asthma-COPD overlap (ACO), are still under debate. Some perspectives suggest that ACO lies on a spectrum between asthma and COPD, where patients may initially experience exposure to environmental factors or possess genetic predispositions that contribute to the emergence of ACO phenotypes. Another viewpoint proposes that ACO represents a distinct phenotype resulting from different inflammatory processes, setting it apart from both asthma and COPD.

In cases of ACO, asthma presents with traits reminiscent of COPD, particularly in terms of refractory or incomplete reversibility of the airways. This overlap has been conventionally referred to as the asthma-COPD overlap (ACO) [111, 112]. Despite ongoing discussions surrounding its precise definition, a defining characteristic of this variant is the development of early-life asthma that leads to lung functional and structural abnormalities resembling those seen in COPD, including features of irreversible airflow obstruction with partial reversibility [113]. Typically associated with individuals aged over 40, this variant requires careful consideration to ensure that the factors contributing to the non-reversible COPD features are not connected to the aetiology of COPD itself. This includes factors such as genetics, early-life exposures, and irritant exposure, including smoking or exposure to pollutants [111-115].

1.3.2 COPD

Later research aimed to understand the diversity within COPD highlighted the importance of various factors. These factors encompassed disease characteristics such as genetic predisposition, abnormalities in the airway tissue, and exposure to irritants and pollutants, along with other lifelong adverse influences and the presence of other systemic diseases and comorbidities [65, 67, 74]. Furthermore, certain aspects related to how COPD presented and progressed were also found to be significant in identifying subtypes. Researchers measured these aspects using clinical parameters like symptoms, exacerbations, and specific indices that gauged the disease's impact on health and quality of life [65, 67, 74]. To date, a multitude of COPD phenotypes have been derived, considering the aforementioned criteria. A substantial number of these phenotypes have undergone clinical validation and demonstrated significant relevance when measured against meaningful clinical

outcomes. In this context, a review of the well-established COPD phenotypes documented in the existing scientific literature will be presented.

1.3.2.1 OLD WELL -ESTABLSIHED PHENOTYPES OF COPD

Efforts to categorize COPD have delved into diverse attributes that might facilitate the identification of patient groups sharing common biological underpinnings and clinical outcomes associated with the disease [116]. In the early stages of this endeavour, when COPD was defined as a broad condition characterized by restricted airflow due to abnormalities in the airways and lung tissue, researchers recognized the existence of subtypes: **chronic bronchitis** and **emphysema**. These subtypes were relatively discernible, with biomarkers or radiological scans effectively detecting emphysema, and symptomatology traits combined with patient characteristics aiding the identification of chronic bronchitis [117]. Consequently, these two variants have long been regarded as fulfilling the criteria to be designated as COPD phenotypes [65, 67].

α1-antitrypsin deficiency CODP: This stands as one of the firmly established and clinically substantiated variations within the realm of COPD. In essence, discerning this variant among COPD patients becomes feasible through the detection of the **α1-antitrypsin** serum biomarker, indicative of the underlying genetic anomalies driving the disease's pathophysiology. Moreover, this subgroup benefits more of **α1-antitrypsin replacement** therapy.[67, 118].

Upper lobe emphysema- COPD: this is a commonly observed COPD phenotype, distinguished by the occurrence of emphysema predominantly in the upper lung lobe, alongside diminished exercise capacity and specific genetic predisposition [67, 118]. The distinct anatomical attribute of this phenotype aligns with radiological evidence. Patients belonging to this category stand to gain substantial benefits from lung volume reduction surgical intervention (LVRS), as evidenced by improvements in exercise capacity, symptom relief, and oxygen saturation levels [59, 67, 118].However the effectiveness of such treatment approaches is yet of questionable utility in clinical practice.

1.3.2.2 NEW EMERGING PHENOTYPES OF COPD

Frequent COPD exacerbators: this is a distinct COPD phenotype that is defined by a particular aspect of the disease's manifestation: the frequency of exacerbations. This phenotype holds significance as it has been acknowledged in the GOLD recommendations due to its implication of specific therapeutic requirements [119]. The widely accepted criterion for identifying frequent

exacerbation in COPD is the occurrence of two or more exacerbations annually. An exacerbation is defined as an event wherein a patient's symptoms worsen beyond their daily variation, necessitating immediate medical intervention [19]. Beyond this definition, this phenotype is also marked by other clinical attributes, including rapid deterioration in lung function, pronounced obstruction, and significant impairments in physical health and quality of life. Additionally, it is closely associated with heightened levels of anxiety, depression, and mortality [59, 118]. Other research studies have highlighted increased instances of adverse radiographic changes and lower circulating eosinophil levels within this patient group [120], in contrast to those with infrequent exacerbations.

Comorbid (systematic) COPD: this particular COPD phenotype is closely linked to a wide array of comorbid systemic manifestations, including cardiovascular and metabolic diseases [59, 118]. Interestingly, this phenotype has been correlated with elevated mortality rates and a diminished quality of life, even though it is characterized by a moderate to mild level of airway obstruction. This disparity between the observed poor outcomes and the degree of obstruction suggests that the adverse consequences are more likely attributed to the associated comorbidities rather than the extent of airway obstruction [65, 118].

Physical Frailty COPD phenotype: While not extensively discussed in the literature, this particular COPD phenotype is occasionally referred to as the "cachexic COPD phenotype." It is characterized by extrapulmonary manifestations that align with three or more of the following criteria: weakness, slowness, low levels of physical activity, self-reported exhaustion, and unintentional weight loss. Research suggests that individuals with this phenotype often experience a loss of body fat-free mass, primarily in muscles, which subsequently contributes to reduced exercise capacity, compromised health status, and an elevated risk of COPD-related mortality associated with airway obstruction [121, 122]. This variant is notably intertwined with the emphysematous COPD phenotype demonstrating a closer connection [59, 118]. It is assumed to be linked with more pronounced airflow limitation, more severe symptoms, and an increased frequency of exacerbations. However, there is also a belief that this variant responds more favorably to pulmonary rehabilitation [118].

Emotional frailty COPD phenotype: a developing COPD phenotype has emerged wherein COPD is accompanied by emotional frailty, manifesting in

symptoms like anxiety, depression, and a heightened fear of breathlessness. These emotional aspects have been found to correlate with heightened disease morbidity, increased mortality rates, and prolonged hospital stays [118]. The emotional vulnerability present in this subset of COPD patients is believed to stem from the intricate interplay of multiple factors. These encompass the severity and biological attributes of the disease itself, the individual's inherent traits (such as emotional intelligence, personality, coping mechanisms, acceptance of illness, and resilience), and external factors (including psychological support and stressors). Effectively managing this phenotype may entail interventions that address psychosocial elements, possibly alongside targeted pharmacological or rehabilitation strategies [118].

Rapid decliner COPD: This emerging phenotype of COPD is proposed to be observed in young patients who experience high symptom frequency and severity, as well as a rapid decline in lung function. In the literature, this phenotype has been associated with higher mortality, despite low systemic morbidity. Although the underlying pathogenesis of this variant remains unclear, identifying this subgroup of patients could help facilitate early referral to specialized management and, if necessary, consideration for lung transplantation.

Non-smoking versus biomass smoking COPD: despite smoking being widely recognized as a risk factor for COPD, recent findings suggest that different types of smoking exposures, such as tobacco smoking and biomass smoke exposure, may lead to variations in the phenotypic presentation of COPD [65] . Reports indicate that smoking-associated COPD from tobacco tends to show emphysematous changes and higher markers of radiographic changes in CT scans, with fewer markers of airway entrapment compared to non-tobacco-related COPD [123-125]. Furthermore, tobacco-related COPD is more likely to present at a younger age and with a higher BMI than biomass smoking-induced COPD [5, 125]. Some studies even propose classifying wood biomass-induced COPD as a distinct variant from organic biomass sources like charcoal, grass, and crop dusts [126, 127]. On the other hand, biomass smoking-induced COPD shows more signs of structural changes, such as thickening of the basement membrane, lymphocytic infiltration, and evidence of anthracofibrosis [128]. However, studies characterizing differences in COPD phenotypes based on the type of smoking exposure are mostly limited by cross-sectional design and inclusion of unselected COPD populations. To enhance the understanding of phenotypic variations related to smoking exposure,

further research focusing on other defining factors of COPD and longitudinal study designs would be beneficial [65].

1.3.3 RHINITIS

Is an umbrella of disease that is characterised by upper airways inflammation that can be caused by various aetiologies. Several phenotypes of rhinitis are there based on caused an observable trait of these diseases.

1.3.3.1 RHINITIS PHENOTYPES BASED ON AETIOLGY

Allergic rhinitis (AR): an IgE mediated chronic inflammation of the nasal mucosa that is triggered by exposure to specific environmental aeroallergens. It's symptoms include nasal ones like nasal congestion/ obstruction, nasal itching, rhinorrhoea, sneezing and loss of smell [38, 129]. Ocular symptoms like tearing, itching and redness may coexist. Diagnosing allergic rhinitis is based on detection of sensitization to specific allergens tested using serum levels of specific IgE (sIgE) or skin prick test [129]. Allergic rhinitis could further be sub-phenotyped into **seasonal allergic rhinitis** which triggered by specific allergens form outdoor setting like trees, grass, wood pollens and mould. Other variant is the **perennial allergic rhinitis** triggered by allergens from indoor setting like house dust and mites (HDM) allergens [130].

Being well acknowledged as risk factor for asthma, the organisation of allergic rhinitis and its impact on asthma (ARIA)[131] was established aiming at evidence-based management of allergic rhinitis and asthma world-wide. It proposed further sub-phenotyping of allergic rhinitis based on frequency and severity of symptoms. **Intermittent AR** implies presence of symptoms for less than 4 day /a week or less than 4 consecutive weeks, compared to **persistent AR** with symptoms lasting for 4 days in a row or 4 consecutive weeks [129]. Severity classification of AR was based on the disease impairment on 4 domains including: work or school performance, sleep, daily activity, leisure or sport and troublesome symptoms. **Mild AR** affects none of the stated domains, **moderate** affects 1-3 and **severe** affects all the four domains [131, 132]. The term **local allergic rhinitis** has also been proposed describing a variant or allergic rhinitis that is characterized by no systematic atopy and rather local production of sIgE and a T2 pattern of inflammatory cell infiltration (eosinophils) and mediators (tryptase, eosinophil cationic protein) in the nasal mucosa [133, 134]. Hence, identification of high serum specific IgE or positive skin prick test is not possible [133, 135]. Such group of patients manifest nasal symptoms typical of allergic rhinitis only, mostly in response to trigger by indoor allergens of HDM or seasonal allergens[134]. Identification of such group

of patients is thought to be only possible using nasal allergen provocation test (NAPT). The clinical significance of such phenotype's identification lies in the great resemblance between it and non-allergic rhinitis, in the sense that these patients would exhibit negative serum sIgE testing, despite having a very good response to therapeutic management using inhaled steroid and antihistamine unlike those with non-allergic rhinitis [37].

Other phenotypes of allergic rhinitis include **occupational rhinitis** where nasal or ocular symptoms are triggered by allergen at workspace [136, 137]. **Mixed rhinitis** is another hypothesized phenotype of allergic rhinitis where symptoms are found to be triggered out of the pollen season among those with seasonal AR [138]. The underlying pathophysiology is thought to involve either of minimal persistent inflammation or neurogenic inflammatory response to irritants [139, 140]. The later results in activation of the trigeminal nerve endings which invokes and antidromic reflex of neuropeptides release. The later results in vasodilation, congestion, and oedema; hypersecretion and further release of inflammatory mediators[139, 140].

Non-allergic rhinitis: this is chronic rhinitis with at least two symptoms of (nasal obstruction/congestion, nasal dripping/rhinorrhoea, sneezing or nasal/ocular itching) that lasts for at least 1-hour daily form more than 12 weeks. The condition implies no evidence of atopy including high serum sIgE, positive skin prick test or positive NAPT [141]. According to aetiology many phenotypes of NAR can be identified as medication induced NAR, hormonal NAR, gustatory NAR and pregnancy related variants [37].

1.3.3.2 RHINITIS PHENOTYPES BASED ON INFLAMMATION CYTOLOGY

Another mode for classifying allergic rhinitis focus on nature of the dominating pattern of inflammation during the disease process. Specific cellular inflammatory profiles are differential for some endotypes of rhinitis. **Allergic rhinitis**, for instance, **is more characterised by predominance in eosinophils**, **mast cells, neutrophils, and lymphocytes**. High load of eosinophils in nasal cytology is highly suggestive for allergic AR compared to NAR [142].

Though, an endotype of **non-allergic rhinitis with high eosinophilia** is also present (NARES). This endotype is characterised by eosinophilic inflammation of the nasal mucosa in absence of systematic or local allergy. It presents in 10% of those with rhinitis. It is highly associated with comorbid

asthma and demonstrates good response to steroids therapy [143], in addition to carrying high susceptibility to upper respiratory tract infections [144]. Absence of local sIgE is a good way of differentiating this from LAR. **Neutrophil NAR** is the most prevalent cellular phenotype of rhinitis, presenting in 20% of patients with rhinitis. It is characterised by mild symptoms in absence of coexisting URTI [145, 146].

Other inflammatory endotypes of NAR rhinitis like **eosinophilic NAR out of pollen season**, **NAR with predominance mast cells** and **mixed pattern NAR** have also being reported and attributed to more disease severity and difficulty in management [147].

1.3.3.3 MULTIMORBID RHINITIS PHENOTYPES

Certain rhinitis phenotypes have been identified based on their co-occurrence with other health conditions such as asthma, COPD, eczema, atopic dermatitis, conjunctivitis, and specific food allergies. Notably, a phenotype of rhinitis characterized by simultaneous presence with asthma and conjunctivitis has been extensively documented among both children and adults. This particular phenotype is closely associated with severe disease, particularly of longer duration in asthmatic children [129]. This phenotype is believed to involve an IgE-mediated polysensitization pathway where rhino-conjunctivitis and asthma mutually trigger each other [148].

Another variant of rhinitis coexists with COPD. Here, COPD acts as an independent risk factor for rhinitis, alongside being atopic and a smoker. Conversely, COPD is also a risk factor for rhinitis [149]. Rhinitis also shows connections to other health issues, leading to various phenotypes. These include rhinitis combined with obstructive sleep apnoea, rhinitis accompanying gastroesophageal reflux disease (GERD), and rhinitis associated with eosinophilic inflammation of the oesophagus and food allergies [37].

1.3.3.4 OTHER RHINITIS PHENOTYPES AND DIFFERENTIAL DIAGNOSIS

Additional rhinitis variations that are clinically considered as differential diagnoses of rhinitis include rhinosinusitis in its chronic and acute forms, as well as rhinitis associated with NSAID-exacerbated respiratory diseases [150].

Rhinosinusitis (RS) shares symptoms with rhinitis, however necessitating the presence of two or more sino-nasal symptoms. One of these should include

nasal congestion/obstruction/blockage, anterior/posterior nasal discharge, facial pain/pressure, or reduction/loss of smell [151]. Acute RS is characterized by symptoms lasting under 12 consecutive weeks and is often tied to viral upper respiratory tract infections following viral or bacterial infections [151]. Conversely, chronic rhinosinusitis persists for more than 12 consecutive weeks. In the latter case, imaging verifies nasal mucosal thickening, oedematous changes, and nasal discharge [150, 151]. This form has two variations: one associated with nasal polyposis (CRSwNP) and another without [150, 151].

The rhinitis phenotypes linked to NSAID-exacerbated respiratory diseases indicate the presence of CRSwNP concurrent with asthma and sensitivity to aspirin. This sensitivity involves the inhibition of the cyclooxygenase 1 pathway and signifies a disease of heightened severity and challenging management [152, 153].

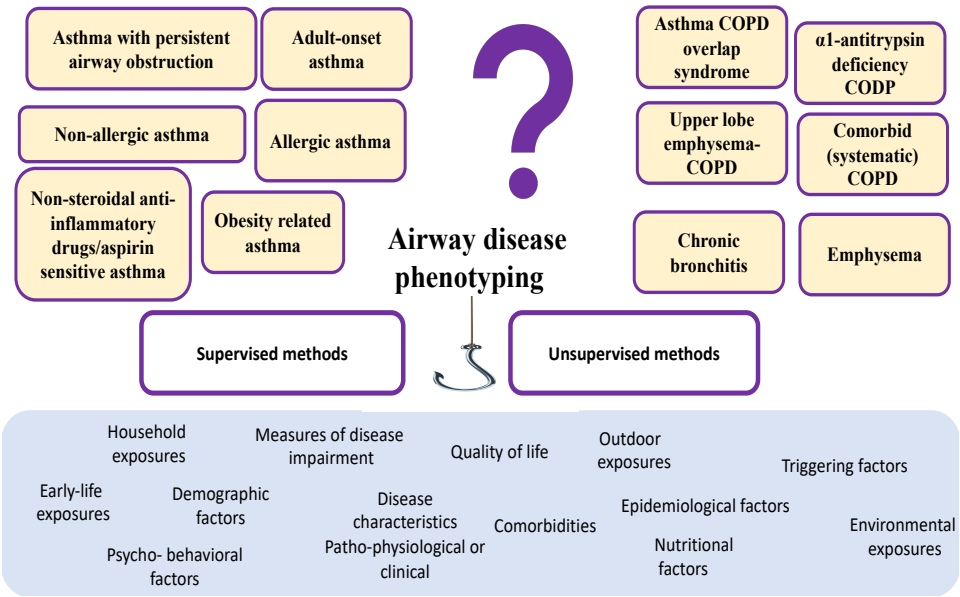


Figure 2 Airway disease phenotyping methods, phenotyping attributes and results.

1.4 SOCIAL DETERMINANTS OF AIRWAY DISEASES

1.4.1 DEFINITION

Social factors encompass the circumstances we are born into, where we grow up, work, and reside, which have a significant impact on our health, our perception of health, our access to healthcare, as well as our life expectancy and mortality rates [154, 155]. Socioeconomic status (SES) is a multifaceted phenomenon that implies aspects of income, education, and occupational status. Commonly used measures of socio-economic status in epidemiological research include education level, income, job titles and other factors that reflect material and cognitive assets or social standing that affect perception or access to health care.[156]

SES, as a measure of social status in relation to health, has inherent complexities because of the interrelated nature of the socio-economic factors that affect each other and their chronological effect that expand across different generations.[157] Measuring education quantitatively as number of years of obtained education does not consider the quality of education and the current status of income or employment. Further, income as a measure of socio-economic status excludes major social age groups like students and retirees and presents major differences by gender. Further complexity relates to the bidirectional nature of the association between health and SES, i.e. diseases can result in unemployability, negative impact on education and school dropping and poor professional performance, which are all characterizations of low SES and, in turn, causes of disease.[156-158]

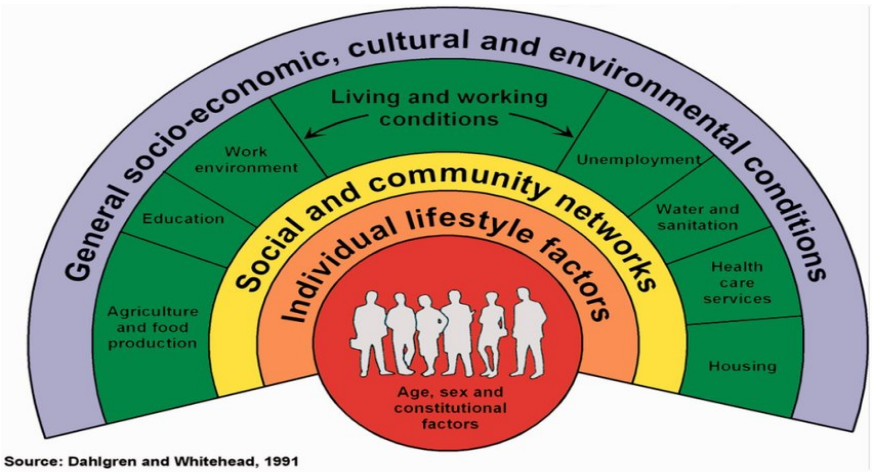


Figure 3 Illustration of social determinants of health, model by Dahlgren and Whitehead

1.4.2 SOCIAL DETERMINANTS AND AIRWAY DISEASES

With respect to airway diseases, low socioeconomic status is a known risk factor for asthma, rhinitis, respiratory symptoms, and increased morbidity and physical life impairment due to chronic airway diseases. Low income and education are risk factors for incident asthma, impaired lung function, and respiratory symptoms.[159, 160] In Sweden, low income was associated with increased prevalence of respiratory symptoms among women, while low education associated with poor control and excess asthma hospitalization [161-163]. Occupations like farmers, mechanics, iron and metal workers increased risk of asthma hospitalization among males contrary to nursing, social science and wood workers which increased risk among females [162]. Higher prevalence of respiratory symptoms was reported among blue collar workers in manual service and industrial jobs compared to white collar professionals [164]. Despite being long acknowledged as risk factor for respiratory diseases [159], precisely capturing the link between smoking and asthma in epidemiological studies has long been a challenge, especially in cross sectional designs where establishing the temporality of smoking preceding asthma is difficult. Though, smoking is not innocent of contributing to asthma overall [165]. Environmental tobacco exposures, active smoking and parental smoking were associated to asthma occurrence, symptoms, and poor control among, adults, adolescents, and children, respectively. [166-168]

1.4.3 SOCIAL FACTORS IN RELATION TO AIRWAY DISEASES PHENOTYPES

Social factors used to define specific asthma and COPD phenotypes such as obesity-related asthma, work-related asthma, and physically impaired COPD exacerbator and smoking asthma phenotype.[169-171] The consideration of such phenotypic definition of respiratory disease in epidemiological investigation could assist unmasking obscurity behind hiding or conflicting association paths, since they are affected differently, and could point opposite directions when measured using same risk factor [35]. Smoking was found to increase risk of non-allergic asthma but not allergic asthma in one study [14], occupational groups in certain job classes were of higher prevalence of non-allergic asthma but not allergic asthma [164]. Further, at pathophysiological level factor like obesity, smoking was found to operate differently in causing or aggravating diseases, according to the other demographic and clinical characteristics. [165] Further, obesity is associated with severe form of non-allergic asthma among females but not males.[172]

With regard to rhinitis and its forms, high prevalence of hay fever have been reportedly associated with industrialization and affluent social classes [173]. Early life exposure to low SES was found to be protective against atopy. Further, higher SES and high educational levels were associated with higher risk of atopy and allergic rhinitis [174, 175]. For later life exposures, in Korea, a large population-based study utilized a standard questionnaire that measured adolescents' SES using residence, family affluence (Family Affluence Scale; FAS), parental education levels, subjective academic achievement. The study also linked high affluent scale, high subjective academic achievement and Bachelor's parental education level to be linked to higher risk of allergic rhinitis [176].

Concerning chronic rhinitis (CR), the combined effect of smoking and occupational gases, dust and fumes was linked to higher risk of chronic rhinitis in population based study from Finland [177]. Environmental exposure to air pollution has been reportedly linked to exacerbation of CR. A unit increase in particulate matter air pollutants is linked to 1.89 times in need for surgical treatment among patients with CR [178]. At cellular level, studies have reported that exposure to air pollutants results in disruption of the Sino-nasal epithelial barrier, and hence plausibly result in worse CR outcomes [179]. lower SES is known to be linked to higher exposure to air pollution[180], hence the variation in risk of CR by SES is conceivably present. Yet, not much is reported and a

known about the direct association between SES and risk of CR as form of nasal inflammation.

For chronic rhino-sinusitis (CRS) as a form of rhinitis, a systematic review on social factors contribution to risk of chronic rhinosinusitis reported conflicting finding between SES and risk of chronic rhinosinusitis using different measures of SES [181]. Lower (household) income was significantly associated with higher risk of CRS both among adults and children of parental low SES. Education level, in turn, was significantly reciprocally linked to CRS. Lower educational attainment was associated with higher prevalence, symptoms, worse disease outcomes generally perceived health status among patients with CRS. A U-shaped link – where the risk of rhinitis increases among those with the lowest and the highest educational level- was reported in one study, where CRS was defined based on self-reported symptoms [182]. Occupation wise, higher risk of CRS was reported among blue-collar workers than white ones who were at higher exposure to dust fumes and gases at working environment [183, 184]. Another study reported, higher risk of CRS among manual worker in plant and machine workers, elementary occupations or craft and related trade workers [185]. However, in some studies there was no noted difference in risk of CRS by different in health care related jobs and jobs overall [186, 187].

1.4.4 SMOKING AND CHRONIC OBSTRUCTIVE AIRWAY DISEASES

Smoking is well acknowledged as significant risk factor for chronic obstructive pulmonary disease (COPD), with strong associations to increased mortality and unfavourable outcomes in COPD patients [188-190]. However, establishing a clear and definitive causal relationship between smoking and asthma has presented more challenges [189]. While numerous studies have indicated a notable prevalence of asthma among both current and former smokers, and longitudinal research has hinted at a correlation between smoking and a higher incidence of asthma, the overall strength of evidence is weakened by limitations in research design and the presence of systematic biases. These factors hinder the confidence to draw robust conclusions about the proposed association.

Furthermore, smoking has been linked to heightened asthma severity and an accelerated decline in lung function among individuals with asthma, alongside a challenge in achieving effective control over the condition [191, 192]. Parental smoking habits have also been implicated in an elevated asthma risk for their

offspring [158]. Both active and passive smoking exhibit connections to an increased asthma risk [193]. Notably, the prevalence of smoking in individuals with asthma mirrors that of the general population [194]. Moreover, individuals with asthma seem to have a heightened propensity for smoking, showing an increased vulnerability to nicotine dependence and encountering difficulties in quitting [195, 196]. Additionally, early-life smokers with asthma demonstrate an increased tendency to experience compromised lung function, structural airway alterations, and the development of irreversible obstruction due to smoking [196, 197]. This intricate interplay further complicates the differentiation between individuals with chronic obstructive pulmonary disease (COPD) directly attributed to smoking and asthmatics who initiated smoking early, developed dependence, and acquired an irreversible obstruction pattern akin to COPD [194].

Concerning allergies, the link between smoking and atopy remains a subject of ongoing debate. While certain studies have proposed an elevated risk of atopic asthma due to smoking, conflicting findings have suggested a reduced presence of atopy among smokers with severe asthma [190]. Moreover, research has revealed elevated levels of IgE in male smokers, and other studies have pointed to a robust connection between smoking and specific IgE levels related to house dust mites, both in in vivo and vitro models [194, 198, 199].

SMOKING EFFECT MODIFCATION WITH SES TOWARDS RISK OF RESPIRATORY DISEASES

Indeed, it is evident that smoking is intricately associated with various phenotypic attributes of airway diseases. This connection extends to the timing of disease manifestation and progression patterns. For instance, individuals with asthma tend to initiate smoking at an earlier age, and smokers generally exhibit accelerated disease onset, prolonged continuation, and escalated deterioration rates [200].

Furthermore, smoking exerts diverse effects on allergic responses, engaging in complex interactions, enhancing allergy and non -allergic outcomes in various settings [190]. In the context of chronic obstructive pulmonary disease (COPD), smoking serves as a direct causal factor. Notably, the duration and intensity of smoking are predictive of distinct phenotypic manifestations within this disease continuum.

Other phenotypic attributes that define airway diseases, are related to behavioural and environmental exposures, and are deeply integrated within the framework of an individual's socioeconomic status (SES). This context results in the intertwined nature of each obstructive airway disease trait with smoking, influenced by broader social dynamics. Since, smoking exhibits a conspicuous link with social standing, with a higher prevalence among individuals of lower SES, who often grapple with challenging social conditions [201]. Previous investigations have unveiled a twofold likelihood of smoking among laborers in manual occupations compared to those in managerial and high-professional roles [201].

Furthermore, it has been established that smoking within lower SES context is substantively linked to heightened exposure to suboptimal housing conditions, house dust mite (HDM) and pest-related allergens, alongside both indoor and outdoor pollutants [202]. Given the intricate and multifaceted nature of the association among smoking, SES, and airway diseases, extensive research endeavours have been directed towards disentangling this complexity.

Nonetheless, significant gaps persist in our understanding of how the interplay between smoking and SES mutually influences the risk and prevalence of obstructive airway diseases. Prior research has identified a correlation between low SES and smoking status, leading to an augmented risk of asthma and compromised lung function [203, 204]. Notably, limited attention has been directed towards elucidating whether the impact of either or both factors vary based on the distinctive phenotypes of respiratory diseases [205, 206.] For instance, one study has related low socioeconomic status defined by occupation and education level to non-allergic asthma[161]. Another one found low SES measured using occupational classification of manual job, which is indicative of low SES, is linked with allergic asthma [207]. Such discrepancy in results might not only be attributed to utilization of different measures reflection various facets of social phenomena but could also be explained by the variation due to the phenotypic trait of each outcome [161, 208, 209].

In this context, the incorporation of phenotypic attributes for identifying risk factors associated with airway diseases emerges as a pivotal approach in the realm of precision medicine. This approach transcends the consideration of individuals solely at the patient level, acknowledging the unique and distinct nature of each airway disease phenotype. By targeting risk factors through proactive measures, the potential for early intervention at the primary stages of disease management is rendered feasible [116].

2 AIM

The overall aim of this thesis was to characterize phenotypes of obstructive airway diseases, and to explore their variation by SES. The specific goals of this study are as follows:

1. To examine the potential correlation between the level of education and asthma control in cases of adult-onset asthma, and to explore whether this association varies atopic asthma status.

2. To determine whether socioeconomic status and smoking interact with/modify effect of each other to influence the risk of developing respiratory diseases, with a particular focus on how these effects might differ among allergic and non-allergic asthma.

3. To conduct a comprehensive analysis of existing literature that aims to classify airway diseases using unsupervised analytical methods, and to synthesize the narrative evidence derived from these studies.

4. To assess whether SES, based on education and occupation, is a determinant of different forms of rhinitis.

5. To identify and describe asthma phenotypes in an adult representative sample in western Sweden using a novel machine learning approach.

6. To assess the SES variations in machine learning derived asthma phenotypes.

PATIENTS AND METHODS

2.1.1 STUDY AREAS

2.1.1.1 WEST SWEDEN ASTHMA STUDY

The western region of Sweden extends from the northernmost part of the country's west coast to the central lakes of Vänern and Vättern. It ranks as the second-largest county in Sweden and is home to a population of 1,758,656 individuals, according to the national Swedish bureau of statistics data up to 2021. [210, 211] This population includes 887,897 males and 870,759 females, constituting approximately 17% of the nation's total. The largest municipality in the region is Gothenburg, serving as the metropolitan hub and the second-largest city in Sweden. [210, 211] With a mix of urban and rural areas, the region showcases a varied demographic profile in terms of age and gender. The average age of the population is 41.5 years, and women tend to outlive men, with an average life expectancy of 42.3 years for women and 40.6 years for men[210, 211].

The social landscape of the region leans toward a middle to high socio-economic status. The average annual earned income per person is 327,800 Swedish krona (SEK), with males earning more than females: 369,100 SEK for men compared to 286,200 SEK for women, as per 2021 statistics from Sweden. [212] The estimated annual household expenditure for single-living households averages at 330,870 SEK, with an average of 2 persons per household. Most of the population attained primary education, although there is an increasing trend in higher education attainment among the older population, particularly those aged 70 and above. In 2022, the employment rates were notably high, with 66% of the population being employed, 12% in education, 3.7% unemployed, and 13% retired. Among the employed, 87% are full-time workers, with women more likely to work part-time compared to men. [213]

The region's economy features thriving sectors, including manufacturing industries, healthcare, and social professions, as well as trade, scientific, and technical fields. This diverse economic landscape aligns with the metropolitan nature of the area. [213] The region experiences an oceanic climate, with temperatures ranging from around 15-16°C in Summer to -1 to -4°C in Winter.[211]

2.1.1.2 THE OBSTRUCTIVE LUNG DISEASE IN NORTHERN SWEDEN (OLIN) STUDIES

Situated in the northernmost part of Sweden, this region boasts a low population density of merely 2.6 inhabitants per square kilometer, with a total population of 249,177. The county encompasses a land area of 97,242 square kilometers and shares borders with Finland and Norway. [214] The coastal regions are the primary habitation zones, with notable concentration in the city of Lulea, a pivotal location for the sampling of OLIN subjects, alongside Piteå and Kiruna. [214]

Prominent economic sectors in the county encompass steel, mining, wood, and pulp production, in addition to emerging fields like tourism, wind power, and education and research. Norrbotten County's climate tends to be cold for the majority of the year, with temperature fluctuations spanning from 11-15 degrees Celsius during the summer months, down to -9 to -17 degrees in the winter season. [214]

2.1.1.3 SEINÄJOKI ADULT ASTHMA STUDY (SAAS)

Seinäjoki is the central city of the province of south Ostrobothnia in Finland with and area of 1,469.19 km². It has a population of approximately 63,000 people. Situated in western Finland, it is known for its flat terrain with agricultural fields, forests, rivers like Lapua River, and Lake Lappajärvi. The Seinäjoki city region boasts a diverse economy, encompassing industries such as agriculture, food processing, manufacturing, services, and education. Seinäjoki, a commercial and cultural hub, hosts the annual Provinssirock music festival and offers various cultural events throughout the year. The region experiences a temperate continental climate, with mild to warm summers (15°C to 25°C) and cold winters (-5°C to -15°C) featuring common snowfall.

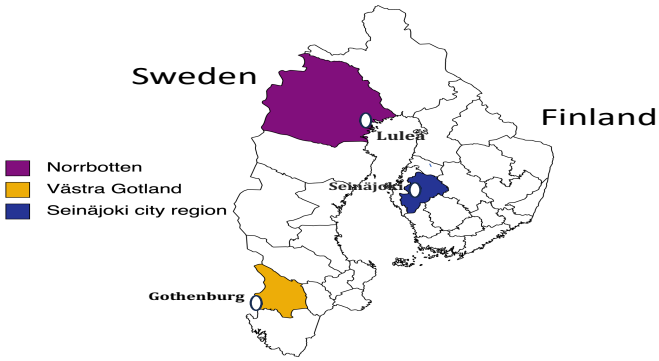


Figure 1. Map of Västra Götland and Norrbotten regions in Sweden and Southern Ostrobothnia a region in Finland, where West Sweden Asthma Study (WSAS), the Obstructive Lung Diseases In Northern Sweden (OLIN) study and Seinäjoki Adult Asthma Study (SAAS) was conducted, respectively. The main cities in each county are shown as white circles.

2.1.2 STUDY DESIGN

2.1.2.1 WEST SWEDEN ASTHMA STUDY PART 1 (WSAS 1)

In 2008, the initial phase of the West Sweden Asthma Study (WSAS1) was launched. This involved sending questionnaires to 30,000 individuals randomly chosen from the Västra Götaland region. Out of these, 18,087 individuals responded. Among the respondents, a subset of 2000 individuals were randomly selected and invited to undergo clinical assessments (random sample). Additionally, individuals who indicated a history of asthma in their questionnaire responses were also invited to participate in the clinical examination. In total, 2006 individuals took part in the clinical examinations, with 1172 from the random sample and 834 from the asthma sample. The analysis for **Paper 1 and paper 5** was based on these participants who underwent clinical examinations. Those who participated in WSAS 1 in 2008 were invited for a follow up survey that was conducted later in 2016. Out of the 18087 subjects who participated in WSAS 1 postal survey, 12447 participated in the follow up survey in 2016. The analysis in **paper 4** was based on the group of participants from WSAS 1 follow up postal survey conducted in 2016.

2.1.2.2 WEST SWEDEN ASTHMA STUDY PART 2 (WSAS 2)

The second phase of the West Sweden Asthma Study (WSAS) was initiated in 2016, involving a cohort of 50,000 individuals from the same age group as the

WSAS 1 participants. These individuals were invited to partake in a postal questionnaire survey. It's important to note that the sampling area and subjects included in WSAS 2 were distinct from those in WSAS 1. Out of the 24,534 respondents to the postal questionnaire, a subset of 2,448 reported experiencing current asthma, in accordance with the same criteria as in WSAS 1. Additionally, 5,000 subjects were randomly chosen to join the clinical examination phase of WSAS 2, along with the previously mentioned 2,448 asthma subjects. The content of **Paper 2 and paper 4** was grounded on the responses to the WSAS 2 postal questionnaire, while the content of **paper 5 and paper 6** was also based on the sample of subjects who underwent clinical survey within WSAS 2. For a comprehensive understanding of the study's design, refer to the detailed description available in the West Sweden Asthma Study documentation [215] .

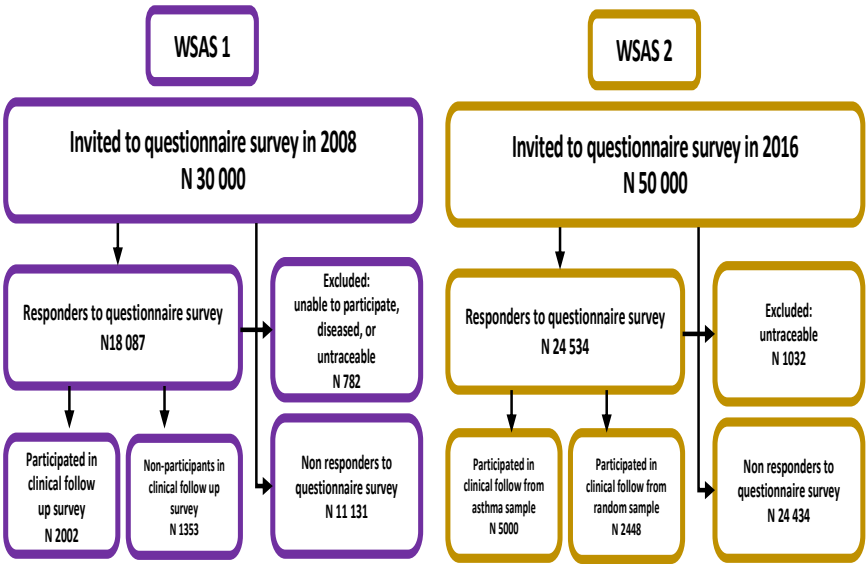


Figure 4. The study design of West Sweden Asthma Studies (WSAS) 1 and 2

2.1.2.3 THE OBSTRUCTIVE LUNG DISEASE IN NORTHERN SWEDEN STUDY (OLIN)

The Obstructive Lung Disease in Northern Sweden (OLIN) is an ongoing epidemiological research program established in 1985. It primarily investigates chronic obstructive pulmonary disease (COPD), asthma, allergies, and their associated health economics.

OLIN study sample in **paper 1** was utilized from OLIN's sample of 2,055 adults with asthma. These participants were identified from clinical examinations and detailed structured interviews conducted within the OLIN studies. The initial selection of participants came from five distinct population-based cohorts. This comprised an initial cohort of 5,697 individuals recruited in 1985, a second cohort of 7,735 individuals recruited in 1992, two random population samples of 4,851 and 7,420 individuals recruited in 1992 and 1996 respectively, and a case-control study involving 309 subjects with adult-onset asthma recruited between 1995 and 1999.

These 2,055 individuals were identified as having asthma across the previous five cohorts. Subsequently, between 2012 and 2014, a follow-up clinical examination was conducted for these 2,055 subjects, including those who were still alive and residing in the county. Out of the 2,055 adults with asthma who were identified and invited, a total of 1,425 subjects participated. This participation rate amounted to 71%, resulting in 1,006 participants who willingly underwent both structured interviews and clinical examinations.

The sample for **Paper 2** of the OLIN study was drawn from a pool of 11,755 subjects who were invited to take part in a postal questionnaire in 2016. Among these, 6,854 subjects provided responses. Importantly, this survey took place in the same year as WSAS 2 and encompassed individuals within a comparable age range of 20 to 79 years. For additional insights into the OLIN study's design, more details are available here.[164, 216, 217] .

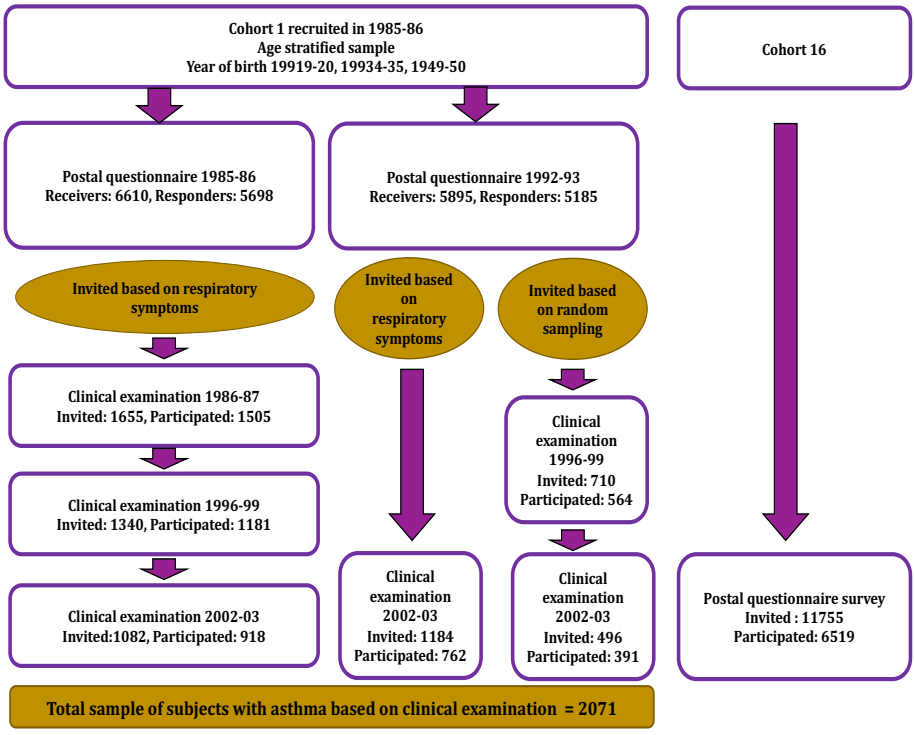


Figure 5. The study design and sample included from the obstructive lung disease in Northern Sweden (OLIN) study.

2.1.2.4 SEINÄJOKI ADULT ASTHMA STUDY (SAAS)

The Seinäjoki Adult Asthma Study (SAAS) is a longitudinal study spanning 12 years, focused on adult individuals newly diagnosed with asthma. The initial study sample consisted of 259 subjects who received a new diagnosis of asthma during adulthood. These individuals were invited to participate in the baseline investigation conducted between 1999 and 2002. Referrals for the study were made from primary care units in cases where the diagnosis of asthma was suspected. Two subjects were excluded due to the discovery of asthma in childhood, leaving a total of 257 subjects who were invited to take part. Ultimately, 203 subjects responded to the invitation, and all of them were confirmed to have adult-onset asthma.

The second follow-up investigation for the SAAS took place between 2012 and 2013. During this period, the subjects were invited for a follow-up visit, during which comprehensive information about their asthma status, comorbidities, medication usage, control, severity, and lung function was

collected. This investigation also incorporated additional information gathered from follow-up visits conducted in previous years. For a more detailed understanding of the SAAS study design, a comprehensive description is available here. [218] .

2.1.3 CLINICAL EXAMINIATION

WSAS and the OLIN study followed a similar clinical investigation protocol for subjects who underwent clinical examination. This protocol encompassed several components, including a structured and detailed interview administered by trained nurses. This interview aimed to gather comprehensive information about various aspects of the participants' health.

During this interview, subjects were asked about their symptoms related to airway diseases, allergies, rhinitis, potential risk factors, and environmental exposures. Additionally, data on medication use, healthcare utilization, and comorbid diseases were collected. Further details were also gathered specifically from subjects with asthma and COPD.

For the purpose of this thesis, for **paper 1** pertinent information about asthma control was obtained from the subjects through the use of a standard questionnaire for asthma control scores. Similarly, in the OLIN study, a structured interview was conducted as part of the clinical examination. This interview aimed to collect information regarding respiratory symptoms, smoking habits, educational levels, and pharmacological treatment.

As part of the clinical examination process, both studies also involved the collection of additional physical measurements. This included obtaining pre and post bronchodilator measures, as well as measurements of weight, height, and skin prick tests. Full version of clinical examination survey in Swedish is attached in the appendix.

For **paper 5**, a thorough information, on asthma key defining parameters were retrieved from both objective examination and clinical questionnaire interview conducted within the clinic setting from subjects in WSAS 1 (2008) and WSAS 2 (2016).

2.1.4 POSTAL QUESTIONNAIRE

The self-administered postal questionnaire employed in WSAS was adapted from similar questionnaires used in previous OLIN studies, the Global Allergy and Asthma European Network (GA2LEN) studies, the FinEsS studies conducted in Finland, Estonia, and Sweden, as well as the European Community Respiratory Health Survey (ECRHS). Drawing from questionnaires in OLIN and ECRHS, the survey encompassed a range of topics related to respiratory conditions like asthma, rhinitis, chronic bronchitis/COPD/emphysema, respiratory symptoms, asthma medication usage, and potential risk factors. These risk factors encompassed variables such as smoking habits, family history of respiratory ailments, occupation type, occupational and environmental exposures, co-morbidities, and socio-economic status. To gather more comprehensive insights into rhinitis and eczema, elements of the GA2LEN questionnaire were integrated. Prior to its inclusion in the study, the pertinent sections of the questionnaire were translated into Swedish. A publicly available sample ECRHS questionnaire can be accessed at http://www.ecrhs.org/Quests/ECRHSIImainquestionnaire.pdf. A Swedish version of the used postal survey questionnaire is attached to the appendix.

2.1.4.1 STUDY ON EDUCATION AND ASTHMA CONTROL IN ADULT-ONSET ASTHMA- PAPER 1

Paper 1 utilized data obtained from structured clinical interviews conducted through the clinical survey questionnaire in WSAS 1 and OLIN asthma cohort. Additionally, information collected during follow-up visits in SAAS was also incorporated. To define variables that represented characteristics of the study participants, such as age, BMI, skin prick test results, and smoking status, the definitions adopted from prior publications on the OLIN and WSAS studies were utilized as a foundation.

2.1.4.2 EFFECT MODIFICATION BETWEEN SMOKING AND SOCIAL STATUS ON THE RISK OF RESPIRATORY OUTCOMES IN A SWEDISH ADULT POPULATION: A NORDIC EPILUNG STUDY – PAPER 2

Paper 2 was based on information retrieved from the postal questionnaire survey adopted in WSAS 2 and OLIN cohort of 2016. Definitions of the study covariates was also based on previously published work from OLIN and WSAS.

2.1.4.3 SOCIOECONOMIC STATUS AND DIFFERENT FORMS OF RHINITIS IN SWEDISH ADULTS

Paper 4 was based on data from the postal survey questionnaire of WSAS 2 that was conducted in 2016 and the follow up postal survey questionnaire for WSAS 1 that was also conducted within the same year. The definition of the outcomes, exposures, and covariates, where available, was adopted from previous similar publication in WSAS. This included age, sex, body mass index (BMI), family history of allergy or asthma, smoking status, smoking exposure at home and work, and exposure to vapor, gas, dust, or fumes (VGDF) at work, growing up on a farm during childhood, or living in a rural area during childhood.

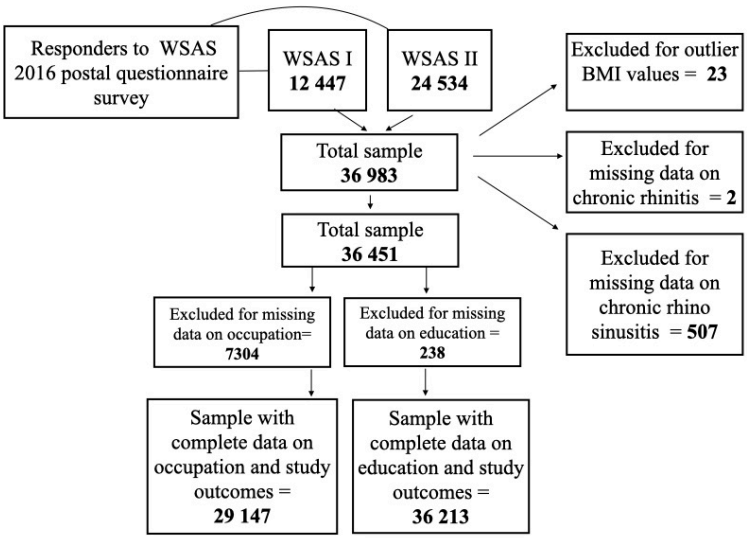


Figure 6 the study sample for **paper 4** and number of participants from WSAS 1 and WSAS 2

2.1.4.4 MACHINE LEARNING-DERIVED ASTHMA PHENOTYPES IN A REPRESENTATIVE SWEDISH ADULT POPULATION

Paper 5 was based on data from the clinical survey questionnaire and examination conducted in WSAS 1 and WSAS 2. Defining analysis parameters was based on previous publication, where possible, in addition to experts review and validation of parameters collected from clinical examination. Severe asthma was defined based on medication usage as self-

reported by subject in the clinical interview according to GINA 2021 guidelines [5], as being in either in GINA treatment step 4 or 5 in addition to using oral corticosteroids. Treatment steps was defined based on Global Initiative of Asthma (GINA) guidelines of 2021. Asthma control was also defined based in GINA guidelines 2021.[5]

2.1.4.5 LEVEL OF EDUCATION, BUT NOT OCCUPATION, IS DIFFERENTIALLY ASSOCIATED WITH ASTHMA PHENOTYPES IN ADULTS

Paper 6 was based on results from **paper 5** and the outcome was used in the analysis encompassed assignment to certain asthma phenotypes derived from **paper 5**. Definition for exposures and analysis covariates were also based on previous literature. These include data from postal survey on SES defined based on education and occupational classification and covariates like age, sex, body mass index, smoking status, occupational exposure to vapor, gas, or dust fumes, being raised on a farm, and urbanization level.

2.1.5 DEFINITIONS

2.1.5.1 STUDY COVARIATES AND POPULATION SAMPLE CHARACTERISTICS

Gender

Age: the age of participants at completion of study (paper 1 and 2).

Body mass index (BMI in kg/m2): classified as underweight (<18.5), normal weight (18.5–25), overweight (25.0 to < 30), and obese (30 or above).

Atopy status: a positive skin-prick test (SPT) (a mean wheal diameter >= 3 mm after 15 min) against any airborne allergen.

Vapor, gas, dust, and fumes exposure at work: have you been exposed to vapor, gas, or smoke at work?

Tobacco exposure at home: have you been exposed to smoke at home?

Tobacco exposure at work: have you been exposed to smoke at work?

Comorbidities: as ever receiving treatment for hypertension, diabetes, or sleeping disorders.

Family history of either asthma, allergy, or other lung disease: do any of your parents or siblings have asthma, allergic eye, or nose problem (hay fever), or other lung or respiratory disease?

Age at asthma onset: how old were you when your diagnosed with asthma?

Growing up on a farm during childhood: positive answer to the question Did your family farm during your first five years of life?

Living in a rural area during childhood: negative answer to the question Did you live in the countryside (i.e. not a city or town) during your first five years of life?

Urbanization level: as living in an area populated with 10000 or more.

2.1.5.2 EXPOSURE DEFINITIONS:

Education: categorized as primary school (usually 9 years), secondary school (usually 12 years), and tertiary education (more than 12 years).

Adult-onset asthma: asthma as asthma starting at age older than of 15 years.

Smoking status: Smoking status was defined as current smoker: those who reported smoking within the last 12 months of completing the questionnaire; former smoker: those who reported quitting smoking at least 12 months before completing the questionnaire; and never smokers.

Social and occupational groups

Previous studies have highlighted the variability in sensitivity among different systems when measuring occupational exposure in relation to respiratory diseases and symptoms [164]. The Swedish socio-economic classification system (SEI), which was published in 1982, categorizes individuals based solely on their occupation, without considering the educational requirements associated with different occupations. As a result, SEI primarily reflects the material aspect of socioeconomic status (SES). In contrast, the Swedish standard classification of occupations 2012 (known as "Svensk standard yrkesklassificering" or SSYK in Swedish) is an updated version of the international standard classification of occupations (ISCO)[219] system. It considers the educational attainment required for various occupations and provides a more accurate reflection of specific occupational exposures present in the workplace. For **paper 2**, socio-economic (SE) groups were determined

using job titles according to both the SEI and SSYK classification systems [220, 221].

Class	Group label
The Swedish socioeconomic classification system groups (SEI):	
Manual workers in service	
Non-manual employees - lower level	Manual work in service
Non-manual employees - intermediate	Assistant non-manual employee
Professionals and executives	Intermediate non-manual employee
Self-employed nonprofessionals	Professionals and executives
Students and housewives	Self-employed non-professionals
Non-classifiable individuals.	Students and housewives
The Swedish standard classification of occupations (SSYK):	
Occupations requiring advanced level	Professionals
Occupations requiring higher education	Mangers
Military personnel	Military
Administration and customer service	Administration
Service care and shops sales workers	Service
Agricultural horticultural forestry and	Agriculture
Building and manufacturing work	Building
Mechanical manufacturing and	Manufacturing
elementary occupations	Elementary
Unspecified with student's professional	Other

Figure 7. Socio-occupational classes definitions and labels as used in paper II

Occupational skill classification based on international classification of disease.

Another way for exposure definition that was adopted in **paper 4** was reducing occupational classification based on ISCO 8 into skill levels. The standard classification of occupation (ISCO)[219] is an international classification that categorize occupations in groups to facilitate communication and comparability. The system defines several key terms to assist in classifying occupations. These include the occupation itself, the skill needed to perform an occupation and the structure the occupation. Occupation is further specified based current, previous and future longest held job. Skills are defined based on degree of complexity, kind of task required to perform the job and hence they are defined based on factors like: the amount of the work needed in

relation to the nature of the task required in the occupation, the amount of formal education required to gain competency in conducting the task of the occupation, or the amount of in formal training or previous experience in related job that is required to gain the capacity required for performing the occupation tasks. The term structure, in turn, refers to the fact that each in the classification system occupation are structured in a way that for each occupation there is a major group consisting of related sub major groups that are consisting of minor groups and consecutively unite groups and each unit contains a single occupation. The ISCO classification system defined the following major groups and their related skill levels that we adopted in **paper 4** and **paper 6**. We further consolidated the original four-tier skill categorization into two distinct levels: Skill level 1, the reference category, now comprises the two highest original skill levels, 3 and 4. Meanwhile, Skill level 2 encompasses the two lower original skill levels, 1 and 2.

Major Groups	Sub Major Groups	Minor Groups	Unit Groups	Skill level
Managers	4	11	31	3 and 4
Professionals	6	27	92	4
Technicians and Associate Professionals	5	20	84	3
Clerical Support Workers	4	8	29	2
Service and Sales Workers	4	13	40	2
Skilled Agricultural, Forestry and Fishery Workers	3	9	18	2
Craft and Related Trades Workers	5	14	66	2
Plant and Machine Operators, and Assemblers	3	14	40	2
Elementary Occupations	6	11	33	2
Armed Forces Occupations	3	3	3	1, 2 and 4
Total number of groups	43	130	436	

Figure 8. Occupational groups and skill levels classified based on the Standard system of Classification of Occupation 2008

2.1.5.3 OUTCOMES' DEFINITION

Paper 1

ASTHMA:

- current asthma in subjects from SAAS study was defined as "physician diagnosis of asthma with at least one of the following: asthma medication use, symptoms (ACT <25), oral steroid burst within 2 years, pre-

bronchodilator FEV_1 ≤80% predicted, post-bronchodilator FEV_1/FVC ≤0.7, fractional exhaled nitric oxide (FeNO) >20 ppb or FEV_1 reversibility ≥12% and 200 m**.**

- Among subjects from WSAS 1, current asthma was defined as either "ever and/or physician-diagnosed asthma in combination with at least one of asthma medication use, attacks of shortness of breath and/or wheeze during the last 12 months."
- For OLIN subjects, current asthma was defined as "ever asthma or strong indications of undiagnosed asthma), based on predefined criteria, identified by clinical examinations of population-based samples performed between 1986 and 2001."

Uncontrolled asthma:

An asthma control score (ACT) less than 19.

Paper 2

Current asthma:

Defined as above.

Allergic asthma:

Allergic asthma was defined as having current asthma (defined above) and positive answer to having allergic rhino-conjunctivitis.

Non-allergic asthma:

Non-allergic asthma was defined as having current asthma (defined above) and negative answer to having allergic rhino-conjunctivitis.

Chronic bronchitis and COPD or either of them (chronic bronchitis and/or chronic obstructive pulmonary disease (COPD):

Affirmative answer to 1) chronic productive cough defined as coughing up mucus or having mucus in the chest that is difficult to be expectorated and 2) whether subjects have experienced mucus most days for periods lasting at least three months and 3) whether subjects had such periods for at least 2 years in row, and/or 4) whether subjects reported use of COPD medications.

Paper 4

Allergic rhinitis: Have you ever had allergic eye or nose problems (hay fever)?

Chronic rhinitis: Have you had any rhinitis and either of nasal blockage more or less constantly or having rhinitis more or less constantly.

Chronic rhinosinusitis: self-report of either of 3 or more of the following for more than 12 weeks in the last 12 months: nasal blockage; pain or pressure around the forehead, nose, or yes;
discoloured nasal discharge (snot) or discolored mucus in the throat or sense of smell being absent or reduced.

Paper 5

Asthma: self-reported history of ever having asthma or physician-diagnosed asthma. This definition was used to capture the aspect of timing of asthma onset. Thus, a follow-up question "at what age did your asthma start" was included.

Demographic:

1- Age
2- Gender
3- Body mass index (BMI). For graphical representation of proportions of obesity with each cluster in the radar graph, BMI of more than 25 was determined as defining cut off for obesity.

Age of asthma onset:

4- **Age at onset:** Self-report of first-time of having asthma symptoms as numeric variable collected from clinical interview questionnaire.

Symptoms

Symptoms were all self-reported from clinical interview questionnaire as subjects answer to the following questions:

5- **Cough**: "Have you had longstanding cough during the last year`".
6- **Breathlessness:** "ever having wheeze or have it in your chest any time during the last 12 months"
7- **Recurrent wheeze:** "Do you usually have wheezing or whistling in your chest when breathing".

8- **Dyspnea:** "Do you have to walk slower than your peers on smooth ground because of shortness of breath?".
9- **Chronic productive cough:** Sputum production for at least 3 months during two subsequent years.

Lung function measures:

Objectively measured in the clinic:

10- Spirometry:
FEV_1/FVC ratio

11- Reversibility in percent:
Reversibility, as numeric value, was tested at the same visit to the clinic as the methacholine challenge, meaning that some subjects performed the reversibility test without a prior methacholine challenge, and some performed it after a methacholine challenge. As a consequence, not all subjects have been reversibility tested in an optimal way. In cases where the subject first underwent a methacholine challenge, the subjects were given 4x0.1 mg of salbutamol (Ventolin®) followed by two capsules of 4 µg ipratropium bromide (Atrovent®) with the reversibility spirometry measured 30 minutes after. A spacer was used for both drugs. In cases where no methacholine was given, the subject was administered 4x0.1 mg of Ventolin and spirometry was performed after 15 minutes.

12- Methacholine test as PD20 value:
Reactivity to methacholine was determined using the Spira equipment (Spira Respiratory Care Center Ltd, Hämeenlinna, Finland) following a shortened protocol. The highest cumulative dose was 1.96 mg. The cumulative dose where a 20% decrease in FEV1 was reached was calculated using the following formula reported in previous documentation[222].

Diffusion capacity:

13- TLCO
14- KCO

Inflammation:

15- FeNO: the fractional exhaled nitric oxide levels (FeNO) were assessed with a NIOX device (Aerocrine AB, Solna, Sweden) at three different flow rates—50, 100, and 270 milliliters per second—to detect inflammation throughout various regions of the lung. For the purposes

of this thesis, only the readings obtained at the 50 ml/s flow rate are utilized. Participants were instructed to exhale twice at each flow rate, and the mean value of these exhalations was documented.

Blood levels of:

16- Eosinophils
17- Neutrophils
18- Lymphocytes
19- Monocytes
20- Basophiles

Allergy and atopy:

21- **Allergic sensitization:** this was measured objectively in the clinic where subjects underwent skin prick test (SPT) for sensitization to allergens. A positive sensitization status was defined as having least one SPT positivity to either of the following allergens: pteronyssinus, farina, Alternaria, Cladosporium, horse, cat, dog, timothy, mug wort or birch. Ful information on how Skin prick test was conducted in WSAS was reported in previous publication. A positive SPT was indicated by mean wheel diameter >= 3.
22- **Family history of asthma:** Have any of your parents of siblings gave asthma? (Postal questionnaire)
23- **Ever having rhinitis:** from clinical survey questionnaire as self-report of positive answer to the question: 'Have you ever had problems with sneezing, runny nose or nasal congestion without having a cold?" Or "in the past 5 years, have you used medicines for hay fever or other nasal problems such as runny nose or nasal congestion without having a cold at the same time."

NSAID related asthma

24- **Drug induced dyspnea**: as self-report of positive answer to having difficulty breathing when using pain killer.
25- **Nasal polyposis**: as self-report of positive answer to either of the following questions in the clinical survey questionnaire: I got nasal polyps only after the onset of asthma, I had nasal polyps in connection with (or something before or after) the onset of asthma, or I had nasal polyps even before the onset of asthma.

Treatment:

26- **GINA treatment steps categorized into 4 groups based on GINA 2021 guidelines for treatment [223]:**

– Step 1 or 2

- Step 3
- Step 4 or 5

Asthma control

27- GINA control: Asthma control based on GINA 2021 based on having day symptoms, night symptoms, activity limitation, and usage of SABA frequency [223], categorized as controlled partially controlled and uncontrolled.

Asthma severity

28- Severe asthma: As being in either in GINA 2021 treatment step 4 or 5 in addition to using oral corticosteroids [223] .

Hospitalization:

29- **Hospitalization last year:** Positive answer to being hospitalized due to breathing problem during the last 12 months.
30- **Emergency visit:** positive answer to having to visit emergency department during the last 12 months due to breathing problem.

Socioeconomic status:

31- **Education:** self-report of highest attained level of education as primary, secondary, and tertiary education.
32- **Urbanization level**: based on residents count whether 10000 or above =. Urbanized
33- **Smoking status:** as never former and current smoker
34- Pack years.
35- Smoking exposure at home
36- Smoking exposure at work
37- **Household damage:** as self-report of having either mold, carpet, or water damage at house.

Asthma triggers

38- **Exertion asthma:** as positive answer to either of:
- I get asthma symptoms on exertion but also on several other occasions.
- I get asthma symptoms on or shortly after exertion and only rarely on other occasion.
- I get asthma symptoms on exertion only if I have a cold.
- I get asthma symptoms on only if it's cold.

39- **Infection triggered asthma:** positive answer to either of
- I get asthma symptoms both during colds and (many) other times.
- I only get asthma symptoms when I have colds and almost never at other times.

- My asthma symptoms are in connection with colds often last more than two weeks in connection with or after colds.
- My asthma symptoms are not affected at all or only slightly by colds.
- I never have such a cold that I notice it in the trachea.

40- **Allergen triggered asthma** as positive answer either of asthma symptoms got exacerbated by:

- Pollen from trees such as birch, rowan, alder etc.
- Pollen from grass.
- Fur animals such as cat, dog, horse, rabbit etc.
- Mold"

41- Irritant triggered asthma as positive answer to asthma exacerbated by perfumes, smells or smoke.

42- **Weather triggered asthma** as positive answer to asthma symptoms exacerbated by either of:

- Cold air
- Cold moist air
- Warm moist air
- Temperature changes

43- **Emotion triggered asthma** as positive answer to asthma symptoms triggered by laughing.

44- <u>Comorbidities</u>

COPD based on self-report and clinical examination.

Paper 6

The outcomes were derived from **paper 5** as**:**

- Phenotype 1: troublesome late-onset, non-atopic asthma with more women and smokers
- Phenotype 2: Female-dominated early adult-onset asthma
- Phenotype 3: Adult-onset asthma with high inflammation
- Phenotype 4: Early-onset, mild, atopic asthma

2.1.6 COMPUTATIONAL PHENOTYPING OF OBSTRUCTIVE AIRWAY DISEASES: A SYSTEMATIC REVIEW

We developed a systematic review of published state of art on derivation of airway diseases phenotypes using computational unsupervised methods of phenotyping.

2.1.6.1 PROTOCOL AND REGISTRATION

We developed a protocol that outlined the review processes and methods before undertaking this work, which was registered in PROSPERO with the number (CRD42020164898) and published [224] .

2.1.6.2 ELIGIBILITY CRITERIA

The review encompassed population-based research employing computational methods to define phenotypes of chronic airway diseases, regardless of whether they were conducted within the general population or in clinical settings. Conversely, we excluded studies that utilized hypothesis-based approaches to characterize phenotypes of chronic airway diseases.

2.1.6.3 STUDY DESIGN

Our study comprised of observational studies on chronic airway diseases, conducted within the general population or clinical settings, including cohort, case control, and cross-sectional designs. However, we excluded studies that used computational phenotyping based on randomized clinical trials or other experimental designs. Additionally, case studies, case series, and ecological studies were also excluded from our analysis.

2.1.6.4 PARTICIPANTS

Studies that were conducted among both children and adults were included.

2.1.6.5 YEARS OF CONSIDERATION

Studies conducted in the last 10 years (2010–2020) were considered. The selected time window coincides with the evolution of the use of computational approaches in phenotyping of chronic obstructive airway diseases.

2.1.6.6 LANGUAGE

There were no language-based exclusions of studies, and where needed, we translated studied published in languages other than English.

2.1.6.7 INFORMATION SOURCE

To gather pertinent studies for our review, we conducted thorough searches on various databases, including PubMed, Embase, Web of Science, Scopus, and Google Scholar. We also extended our search to unpublished materials, such as conference proceedings, by exploring conference databases and gray literature databases like Open Grey. Furthermore, we reached out to experts in the field to inquire about any potentially overlooked papers that might not have appeared in our initial searches. Lastly, we meticulously examined the reference lists of the included studies to uncover any additional relevant papers.

2.1.6.8 SEARCH STRATEGY

We developed search strategies for all the databases to identify relevant studies for the review. The search strategies were first developed in PubMed and then adapted in searching the other databases. The full detailed search strategy for all databases is provided in the appendix.

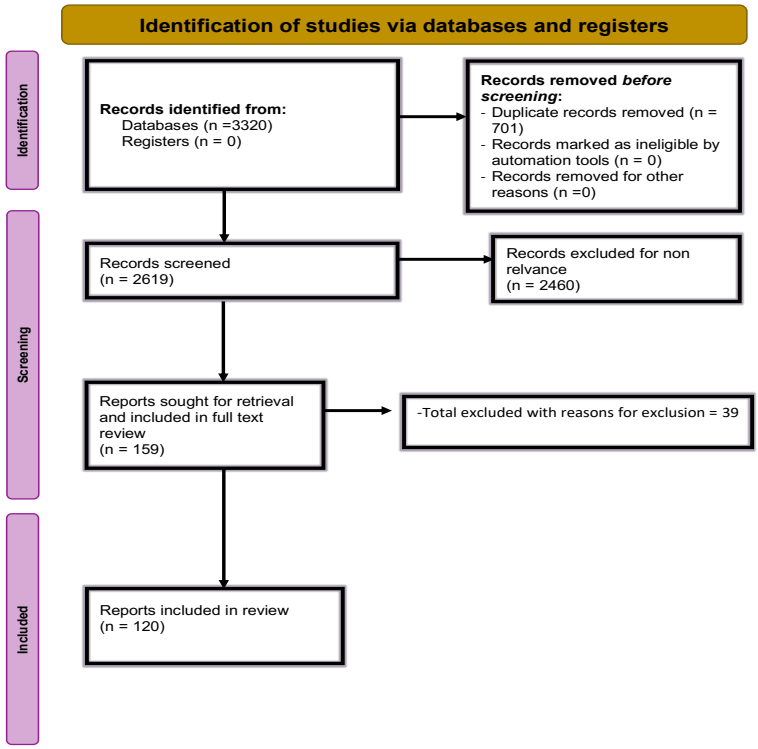


Figure 9. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram illustrating the studies' selection process.

2.1.6.9 DATA MANAGEMENT AND SELECTION PROCESS

The search results obtained from the different databases were exported to EndNote for the screening process. Two reviewers independently assessed the studies based on the predefined inclusion and exclusion criteria established for the review. Any disagreements or discrepancies between the reviewers were resolved through discussion, and in cases where a consensus could not be reached, a third reviewer acted as an arbitrator. The initial phase of the literature review involved removing duplicate studies identified through the database searches. Following this, a thorough screening of titles and abstracts was conducted to narrow down the selection of potentially relevant studies. In the final stage, full-text screening was performed on the studies that appeared to meet the eligibility criteria based on their titles and abstracts. Throughout the screening process, we meticulously documented each step using the

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart [225].

2.1.6.10 DATA COLLECTION PROCESS

Two reviewers worked in pairs to extract relevant data from the included studies using a data extraction form that was specifically created for the review. If any discrepancies arose, they were resolved through discussion, and if consensus could not be reached, a third reviewer arbitrated. Prior to using the form on all included studies, the form was piloted on three studies and any necessary amendments were made.

2.1.6.11 DATA ITEMS

Data from the included studies were extracted into a data extraction form, covering various information such as general study details, population characteristics, type of chronic obstructive airway disease, variables selected for phenotyping, computational approach used, and information about the derived phenotypes.

2.1.6.12 OUTCOME AND PRIORITIZATION

We included studies focusing on computational phenotyping of the following chronic obstructive airway diseases:

- Asthma
- COPD
- Rhinitis
- Emphysema

2.1.6.13 QUALITY ASSESSMENT OF INCLUDED STUDIES

We assessed the quality of the included studies using a preliminary checklist specifically developed for this review. The checklist enabled us to extract information on computational approaches used and compare them across studies. We evaluated various aspects, including subject selection, missing data handling, outcome definition, variables used in phenotyping, clinical relevance of derived phenotypes, and their reproducibility. Detailed information and the full quality assessment form can be found in the appendix.

2.1.7 STATISTICAL ANALYSIS AND DATA SYNTHESIS

For **paper 1**, Continuous data were presented as mean and standard deviation (SD) or median and interquartile range (IQR). One-way ANOVA and Kruskal-Wallis's test were used for comparisons of continuous variables, depending on their distribution. Categorical variables were analyzed using the chi-square test. Multivariable binary logistic regression was performed to assess the association between education level and uncontrolled asthma (ACT 19), adjusting for gender, age, BMI, smoking status, and study cohort. A sensitivity analysis was conducted by excluding subjects fulfilling COPD criteria, followed by another multivariable binary logistic regression with ACT 19 as the dependent variable and education, gender, age, BMI, smoking status, and study cohort as covariates. IBM SPSS Statistics software (version 26) was used for all statistical analyses, and a p-value below 0.05 was considered statistically significant.

In **Paper 2**, proportions were used to present descriptive statistics. The statistical differences between variable categories were assessed using Pearson's chi-squared test. To handle missing data, multiple imputation was performed using the delta adjustment method [28]. Sensitivity analysis was conducted to examine the assumption of missingness. Further in paper 2, A Bayesian network model was constructed to estimate the probability of the presumed effect modification between socioeconomic status (SES) and smoking on the outcomes. The model utilized Directed Acyclic Graphs (DAGs) to represent the dependency structure among variables. The bnlearn package [226] was used for all computational aspects of the Bayesian Network analysis. The network structure was learned using a Hillclimbing algorithm with BIC-CG score, and bootstrap aggregation and model averaging were employed to refine the structure. The Bayesian network model was then fitted to estimate the parameters, and conditional probabilities were computed by sampling realizations of the model variables under specific conditions. Model validation was performed using cross-validation and comparing statistical characteristics of simulated data with the original data. Assessment for effect modification was done by calculation of Interaction contrast ratios (ICR). The calculation of the ICR is based on the formula: p11 + p10 + p01 + p00. If net ICR equals 0, there is no evidence of modification of smoking across levels of SES; if ICR > 0, it is a positive effect modification, suggesting that the combined influence of two variables is greater than the sum of their individual effects, while ICR< 0 indicates a negative effect modification [227, 228].

P11 = probability of effect among subjects with combined exposures to smoking and a specific category of SES

P01 = probability of effect among subjects without exposure to smoking but exposed to the specific category of SES

P10 = probability of effect among subjects with exposure to smoking but not to the specific category of SES

P00 = probability effect among subjects with neither exposure to smoking nor to the specific category of SES

The 95% confidence interval for the ICR was calculated using the estimated coefficients and standard errors of ICR estimates as follows: CI limit (L)=ICR ± z×SE ICR[229]. The likelihood weighting sampling method was employed for approximate inference of conditional probabilities, and the results were presented as probabilities with 95% confidence intervals. Additional details on the analysis method can be found in the supplementary appendix.

In **paper 3**, The data collected was synthesized narratively, summarized, and presented in tabular forms. These tables included information about the study characteristics, phenotyping methods, variables utilized in phenotyping, the number and description of phenotypes, and the quality of reporting on the assessed aspects.

In **paper 4**, For descriptive analysis, we performed Pearson chi-square test to examine differences in distribution of covariates by categorical groups of exposures. We also examined the variation in the prevalence of each form of rhinitis by SES groups within strata of age groups, sex, BMI levels, and family history of asthma or allergy and whether being raised on a farm or not. We employed logistic regression to model the associations between SES and each form of rhinitis. Each model was adjusted for age, sex, BMI, family history of allergy or asthma, smoking status, smoking exposure at home or work, and exposure to VGDF at work, being raised on a farm or being raised in rural area as a child. We estimated the associations between SES and outcomes using odds ratios and their corresponding 95% confidence interval. We reported P-values for the comparison of the unadjusted model with exposure only versus the adjusted model including the full list of covariates, with separate models with education and occupation. Data on covariates, such as age, sex, body mass

index (BMI), family history of allergy or asthma, smoking status, smoking exposure at home and work, and exposure to vapor, gas, dust or fumes (VGDF) at work, growing up on a farm during childhood, or living in a rural area during childhood, were collected. Interaction and moderation analysis in **paper 4** aimed at assessing whether education and occupation exhibited interaction effects with covariates in relation to each form of rhinitis. We conducted model comparisons, contrasting a model incorporating interaction terms between SES and each covariate with one lacking such terms. In cases where the difference in model coefficients reached significance, we conducted further analyses, comparing the simple effects of each interacting factor across exposure groups and vice versa. To mitigate the risk of obtaining statistically significant results by chance alone, we applied the Bonferroni adjustment method during this later step, as recommended [230]. Our presentation of the interaction analysis results included odds ratios and their 95% confidence intervals. Additionally, we visually depicted the interaction effect through graphical representation, using predicted probabilities derived from the logarithms of estimated values, along with 95% confidence intervals.

For **paper 5**, missingness in the data set was retrieved using random forest method of multiple imputation. When used in multiple imputation Random Forest algorithm presents ability to leverage complexity in the data structure in form of non-linearity and complex interaction to present missing information based on the available one. The imputation implied defining variables in the data which are auxiliary, i.e. are likely explanatory of the other missing variables. Here we assessed whether missing in the data was correlated to the status of having asthma as asthma was the main outcome though to determine definition of clusters. Among those we further determined to be those with less than 60% missingness to be used for predicting the missing information. Further, these variables were used utilized in modelling the missing ones. Additionally, the mechanism of missingness was tested for all missing variables whether it was missing at random (MAR) missing not at random (MNAR) or missing completely at random (MCAR). Although some variables were found to be missing not at random, still our implementation of MICE imputation algorithm is thought to produce results that are not biased to such assumption violation [231]. Hundred imputed sets were generated. The result from the imputation solution were validated visually and convergence of the imputation model was insured. One set of the 100 generated imputation sets was randomly selected to run the final clustering analysis on. R package miceRanger [232] was used to conduct the imputation analysis using R software version 4.2.1. Cluster derivation was done using deep embedded (DEC)

clustering algorithm. DEC is a clustering algorithm that combines clustering ability with deep machine learning. It deemed superior to traditional clustering methods due its ability to learn from lower dimensional representation of the data using autoencoders. This lower dimension feature spaces are more suitable for clustering due to compact representation at lower dimensionality. Secondly, DEC's iterative optimization process that utilizes distance metrices to optimize both the feature representation and cluster assignments in a way that traditional methods, such as k-means or spectral clustering, cannot [233]. These qualities make DEC particularly effective for complex datasets, offering improved clustering accuracy, efficiency in handling large datasets. After the data was processes, The R package NbClust was used to decide the optimal number of clusters using voting consensus methods [234]. Additionally, the optimal number of cluster was confirmed using Monte Carlo reference based consensus clustering approach [235], implemented through M3C R package [236]. The output was further fed into the DEC algorithm to perform the clustering. The cluster were later validated using prediction strength approach. The final numbers proposed by such metrics were then evaluated in conjunction with clinical experience before a final determination of the optimal number of clusters were decided to represent the data. The cluster solution determined were then named based on their distribution with regards to the variables used to derive the clusters. For other statistical analysis, continuous data were expressed as means and standard deviations (SD). Group comparisons were performed by 1-way analysis of variance with the Tukey post hoc test, the Kruskal- Wallis's test, or the chi-square, for categorical and continuous variables as suitable. Graphical representation of variation between cluster was performed through radar plots, where categorical variables were represented as proportions.

For **paper 6,** We fitted a multinomial logistic regression model, one separately for each exposure, to study these associations, using the no asthma group (n=1,206) as the outcome reference group, adjusting for age, sex, body mass index, smoking status, occupational exposure to vapor, gas, or dust fumes, being raised on a farm, and urbanization level. The first categories of education (primary education) and occupation (lowest skill level) were used as the reference exposure categories. The estimates of the associations are presented as odds ratios (OR), accompanied by their 95% confidence intervals (95%CI).

3 RESULTS

3.1 PAPER ONE: LEVEL OF EDUCATION AND ASTHMA CONTROL IN ADULTONSET ASTHMA

3.1.1 BASELINE CHARACTERISTICS

The study encompassed combined data from three study cohorts: WSAS, OLIN, and SAAS. In the OLIN cohort, subjects tended to be more frequently obese and older in age. They were less likely to use inhaled corticosteroids and to be current smokers compared to individuals in the other two cohorts. Subjects in WSAS and OLIN were more often found to have atopic tendencies compared to those in SAAS.

In the analysis involving the pooled data, nearly half of the subjects (61%) were females. Approximately 56% exhibited atopic characteristics, and 60% were regular users of inhaled corticosteroids (ICS). Approximately 10% of the subjects were current smokers. The data revealed that 19% of the subjects had uncontrolled asthma. Regarding education levels, 35% had primary education, 36% had secondary education, and 29% had tertiary education.

Among those with primary education, individuals tended to be older than those with secondary or tertiary education. Higher levels of education were associated with a greater proportion of females than males. Additionally, current smoking was less prevalent among those with tertiary education compared to individuals with secondary or primary education. This trend was consistent across individual study samples as well as when the data was pooled.

Subjects with tertiary education displayed better-preserved lung function when contrasted with individuals in the lower education groups. The Asthma Control Test (ACT) score was lower among those with primary education across all study samples, including the pooled data, compared to those with high education, The percentage of individuals with uncontrolled asthma, as defined by an ACT score greater than 19, was notably higher among the lower education group in WSAS and SAAS samples, compared to the highly educated, but this difference was not significant within the OLIN sample.

3.1.2 ASSOCIATION BETWEEN EDUCATIONAL LEVEL AND ASTHMA CONTROL

Association between education level and asthma control were studied through binary unadjusted model and multiple adjusted models. For the association between levels of education and asthma control score of 19, the analysis was run over a pooled sample including subjects from the three study cohorts, in addition to 3 subgroups of the study sample: the ICS users, atopic subjects, non-atopic subjects.

3.1.2.1 BINARY REGRESSION ANALYSIS

Across the entire combined dataset, there was a clear connection between having a lower education level and inadequate asthma control, with an odds ratio of 1.76 (confidence interval: 1.22–2.56). Moreover, among individuals who were daily users of inhaled corticosteroids (ICS), a lower education level was linked with suboptimal asthma control: 2.33 (1.45–3.72). This relationship persisted even among the subgroup of individuals without atopic tendencies: 3.10 (1.36–7.08). However, for individuals with atopic tendencies, the statistical significance of the results did not last: 1.47 (0.67–3.24).

MULTIPLE REGRESSION ANALYSIS

Multiple logistic regression model was employed, adjusting for age up to a maximum limit of 60 years, gender, body mass index (BMI), smoking status, and study cohort. The adjusted analysis revealed that within individuals using inhaled corticosteroids (ICS), those with low education had a higher risk of uncontrolled asthma: 1.92 (1.15–3.20). This association was also observed among non-atopic subjects: 3.42 (1.30–8.96). However, among the subset of atopic subjects, the relationship between low education and uncontrolled asthma remained statistically insignificant after accounting for these factors: 1.01 (0.42–2.42). Further, among the group of non-ICS users, the association between low education and asthma control was not statistically significant: 0.67 (0.32-1.38). full results in (**Table 1**).

3.1.2.2 SENSITIVITY ANALYSIS

Subjects suspected to have COPD were intentionally excluded from the analysis to ensure that the subset of individuals using inhaled corticosteroids (ICS) daily solely represented those with asthma or severe asthma. Despite taking this precaution, a reliable connection between lower education levels and uncontrolled asthma remained evident among the subgroups of individuals using ICS: 2.24 (1.26–3.97).

Table 1. Associations between educational level and uncontrolled asthma (ACT ≤ 19) in the pooled sample and in subgroups of the sample presented as odds ratios and 95% confidence intervals.

Exposure groups	Crude OR (95% CI)	Adjusted OR (95% CI)
ICS users, n = 538		
Tertiary education	1	1
Secondary education	1.26 (0.77–2.04)	1.18 (0.71–1.96)
Primary education	2.33 (1.45–3.72)	1.92 (1.15–3.20)
Atopic n = 298		
Tertiary education	1	1
Secondary education	0.72 (0.38–1.34)	0.65 (0.33–1.26)
Primary education	1.47 (0.67–3.24)	1.01 (0.42–2.42)
Non-atopic n = 237		
Tertiary education	1	1
Secondary education	1.72 (0.76–3.90)	2.05 (0.82–5.11)
Primary education	3.10 (1.36–7.08)	3.42 (1.30–8.96)
All subjects n = 894		
Tertiary education	1	1
Secondary education	0.97 (0.66–1.43)	0.92 (0.61–1.40)
Primary education	1.76 (1.22–2.56)	1.33 (0.89–2.01)

3.2 SMOKING EFFECT MODIFICATION BY SOCIAL STATUS AND THE RISK OF RESPIRATORY OUTCOMES

3.2.1 BASELINE CHARACTERISTICS OF THE STUDY SUBJECTS BY OUTCOME AND EXPOSURE GROUPS

Subjects with current asthma had higher proportion of females than males, they were of higher education and had more comorbidities compared to healthy subjects without any disease condition. Those with allergic asthma were mostly manual workers in service, with high prevalence of hereditary lung disease and exposure to home smoking compared to healthy ones. Non-allergic asthma subjects were most often either non-manual workers or manual workers in service, with higher levels of education compared to heathy subjects. Subjects with COPD and/or Chronic bronchitis were more comorbid, obese, older in age, had lower level of education and mostly worked in manual industry jobs, compared to healthy ones.

Individuals with current asthma had higher proportion of females compared to males. They were also more likely to be highly than low educated and have greater number of comorbidities in comparison to healthy subjects.

Among those with allergic asthma, a significant portion held positions as manual workers in service roles than other jobs. They exhibited a higher prevalence of hereditary lung disease and exposure to smoking within their homes, in contrast to healthy individuals.

In the case of non-allergic asthma subjects, they were of higher proportion of non-manual worker or manual worker in service roles than other jobs. This group were also of higher levels of education compared to healthy subjects. On the other hand, individuals diagnosed with COPD or chronic bronchitis had more comorbid conditions, obesity, older age, and lower education levels. They were primarily engaged in manual industry jobs, all distinguishing them from their healthy counterparts.

3.2.2 SMOKING AND RISK OF RESPIRATORY DISEASES

Both former and current smoking were associated with higher probability of respiratory disease. Former smokers compared to never smokers had higher proportions of current asthma, allergic asthma, non-allergic asthma, and COPD and/or chronic bronchitis. Current smoking was associated with higher probability of only non-allergic asthma and COPD and/or chronic bronchitis.

3.2.3 EDUCATION, OCCUPATION, SOCIOECONOMIC CLASSES, AND RISK OF RESPIRATORY DISEASES.

3.2.3.1 EDUCATION

Education levels were assessed in relation to primary education, which served as the reference group. Tertiary education was linked to a heightened probability of allergic asthma, measuring 5.86% (5.76-5.98), compared to individuals with primary education at 4.59% (4.5-4.7). However, tertiary education predicted a lower likelihood of non-allergic asthma, at 2.9% (2.83-2.98), when compared to those with low primary education at 4.12% (4.02-4.24). There was no observed association between education levels and the probability of current asthma and COPD and/or chronic bronchitis.

3.2.3.2 THE SWEDISH SOCIOECONOMIC CLASSIFICATION SYSTEM (SEI)

Based on the SEI classification, individuals in manual service jobs exhibited a probability of 5.70% (5.61-5.81), and intermediate employees showed a probability of 5.73% (5.65-5.85), both indicating a higher likelihood of having allergic asthma compared to professionals and executives at 5.47% (5.39-5.59).

However, for non- allergic asthma, manual workers in service displayed a lower probability of non-allergic asthma at 2.95% (2.88-3.04). In reverse, for manual workers in the industry, the probability of non-allergic asthma was higher at 3.43% (3.33-3.55), compared to professionals and executives at 3.20% (3.13-3.29). (**Figure 10**) shows the probability of respiratory disease by SEI socioeconomic groups.

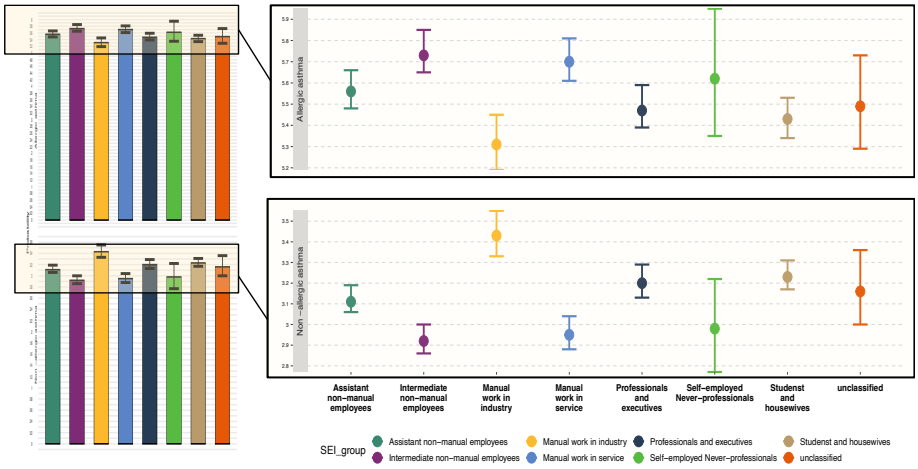


Figure 10.The probabilities and the 95% credible interval of allergic and non-allergic asthma by SEI socioeconomic groups.

3.2.3.3 THE SWEDISH STANDARD CLASSIFICATION OF OCCUPATIONS 2012 (SSYK)

When investigating the association between SSYK and respiratory outcomes, none of the respiratory outcomes varied significantly by SSYK occupational classification.

3.2.4 EFFECT MODIFCATION OF SES AND SMOKING IN RELATION TO RESPIRATORY OUTCOMES

3.2.4.1 SMOKING AND EFFECT MODIFCATION BY SEI GROUOPS

Former smoking increased the likelihood of both allergic and non-allergic asthma. The impact of smoking on probability was more pronounced for allergic asthma and less prominent for non-allergic asthma, with variations across different occupational classes.

Among professionals and executives, the probability of allergic asthma among never smokers was 5.45% (CI 5.37-5.57), while it was 5.68% (CI 5.58-5.79) among former smokers. Compared to professionals and executives, the probability of allergic asthma among manual workers in service was 2.90% (CI 2.83-2.98) among never smokers and 5.93% (CI 5.84–6.05) among former smokers, giving an ICR of 0.03% (CI 0.02-0.03). The corresponding probability among intermediate never manual employees (compared to

professionals and executives) was 2.88% (CI 2.81-2.95) among never smokers and 5.95% (CI 5.84-6.06) among former smokers, giving an ICR of 0.03% (CI 0.02-0.03).

Among professionals and executives, the probability of non-allergic asthma among never smokers was 3.13% (CI 3.08-3.21), while it was 3.38% (CI 3.32-3.47) among former smokers. Compared to professionals and executives, the probability of non-allergic asthma was 2.9% (CI 2.83-2.98) among never smokers and 3.11% (CI 3.05–3.19) among former smokers, giving an ICR of -0.04% (-0.02 to -0.06). The corresponding probability among intermediate never manual employees (compared to professionals and executives) was 2.88% (CI 2.81-2.95) among never smokers and 3.09 (CI 3.02-3.17) for former smokers, giving an ICR of -0.04% (CI -0.02 to -0.06).

Regarding current asthma and COPD and/or chronic bronchitis, the effect of smoking on respiratory diseases did not show significant modification by socio-economic index (SEI) occupational classes. See (**Table 2**) and (**Figure 11**) for full results.

Table 2 the probabilities and 95% credible intervals for the association between smoking and respiratory diseases and its effect modification by SEI occupational groups.

Allergic asthma								Non-Allergic asthma							
(SEI) group	**Prob.**	**Lower**	**Upper**	**Smoking**	**Prob**	**Lower**	**Upper**	**(SEI) group**	**Prob**	**Lower**	**Upper**	**Smoking**	**Prob**	**Lower**	**Upper**
Intermediate non-manual employees	**5.73**	**5.65**	**5.85**	Never smoker	5.69	5.58	5.79	**Intermediate non-manual employees**	2.92	2.86	3	Never smoker	2.88	2.81	2.95
				Former smoker	**5.95**	**5.84**	**6.06**					Former smoker	3.09	3.02	3.17
				Current smoker	**5.72**	**5.62**	**5.83**					Current smoker	2.98	2.92	3.05
Students and housewives	5.43	5.34	5.53	Never smoker	5.42	5.34	5.54	**Students and housewives**	3.23	3.17	3.31	Never smoker	3.15	3.09	3.23
				Former smoker	5.64	5.55	5.75					Former smoker	**3.4**	**3.33**	**3.48**
				Current smoker	5.37	5.3	5.46					Current smoker	3.31	3.25	3.39
Professionals And executives	5.47	5.39	5.59	Never smoker	**5.55**	**5.47**	**5.57**	**Professionals And executives**	3.2	3.13	3.29	Never smoker	**3.13**	**3.08**	**3.21**
				Former smoker	5.68	5.58	5.79					Former smoker	3.38	3.32	3.47
				Current smoker	5.41	5.34	5.51					Current smoker	3.3	3.24	3.37
Assistant Non-manual employees	5.56	5.48	5.66	Never smoker	5.54	5.45	5.63	**Assistant Non-manual employees**	3.11	3.06	3.19	Never smoker	3.04	2.98	3.13
				Former smoker	**5.77**	**5.69**	**5.87**					Former smoker	3.29	3.22	3.37
				Current smoker	5.5	5.43	5.61					Current smoker	3.21	3.14	3.28
Manual work in service	**5.7**	**5.61**	**5.81**	Never smoker	5.67	5.58	5.77	**Manual work in service**	**2.95**	**2.88**	**3.04**	Never smoker	2.9	2.83	2.98
				Former smoker	**5.93**	**5.84**	**6.05**					Former smoker	3.11	3.05	3.19
				Current smoker	**5.69**	**5.6**	**5.81**					Current smoker	3	2.93	3.08
Unclassified	5.49	5.29	5.73	Never smoker	5.47	5.27	5.72	**Unclassified**	3.16	3	3.36	Never smoker	3.11	2.95	3.28
				Former smoker	5.69	5.47	5.95					Former smoker	3.36	3.21	3.54
				Current smoker	5.45	5.24	5.67					Current smoker	3.24	3.1	3.42
Self-employed non-professionals	5.62	5.35	5.95	Never smoker	5.58	5.3	5.92	**Self-employed non-professionals**	2.98	2.77	3.22	Never smoker	2.95	2.78	3.19
				Former smoker	5.85	5.56	6.21					Former smoker	3.17	2.96	3.43
				Current smoker	5.59	5.23	6.04					Current smoker	3.06	2.81	3.34
Manual work in industry	5.31	5.19	5.45	Never smoker	5.28	5.16	5.45	**Manual work in industry**	**3.43**	**3.33**	**3.55**	Never smoker	3.35	3.24	3.47
				Former smoker	5.52	5.41	5.66					Former smoker	**3.58**	**3.48**	**3.71**
				Current smoker	5.26	5.16	5.39					Current smoker	**3.49**	**3.39**	**3.61**

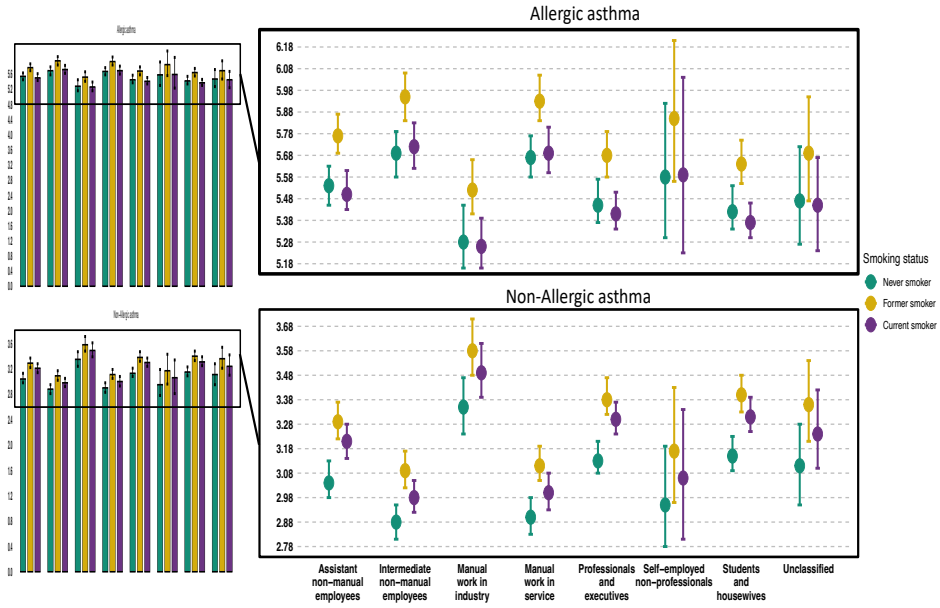


Figure 11. The probability and 95% credible interval of allergic and non-allergic asthma by smoking status and the effect modification by SEI occupational groups.

3.2.4.2 SMOKING AND EFFECT MODIFICATION BY EDUCATION LEVEL

Among those with primary education, the probability of allergic asthma among never smokers was 4.55% (CI 4.44-4.67), while it was 4.76% (CI 4.65-4.88) among former smokers. Among those with secondary education (compared to those with primary education), the probability of allergic asthma was 5.72% (CI 5.62-5.82) among never smokers and 6.0% (CI 5.91-6.11) among former smokers, giving an ICR of 0.07% (CI 0.03-0.09). The corresponding probabilities for those with tertiary education (compared to those with primary education) was 5.7% (CI 5.62-5.81) among never smokers and 5.97% (CI 5.88-6.08) among former smokers, giving an ICR of 0.06% (CI 0.02-0.07).

Among those with tertiary education, the probability of non-allergic asthma among never smokers was 2.85% (CI 2.78-2.93), while it was 3.05% (CI 2.98-3.15) among former smokers. Among those with secondary education (compared to those with tertiary education), the probability of non-allergic asthma was 2.88% (CI 2.82-2.96) among never smokers and 3.09% (CI 3.02-3.19) among former smokers, giving an ICR of 0.01% (CI 0.00-0.05). The corresponding probabilities for those with primary education (compared to those with tertiary education) was 4.04% (CI 3.94-4.16) among never smokers, and 4.33% (CI 4.25-4.45) among former smokers, giving an ICR of 0.09% (CI 0.03-0.10). full results in (**Table 3** and **Figure 12**)

Table 3: The probabilities and 95% credible intervals for the association between smoking and respiratory diseases and its effect modification by SEI occupational groups

Allergic asthma								Non-Allergic asthma							
Education level	Prob.	Upper limit	Lower limit	Smoking status	Prob.	Upper limit	Lower limit	Education level	Prob.	Upper limit	Lower limit	Smoking status	Prob.	Lower limit	Upper limit
Primary	4.59	4.5	4.7	**Never smoker**	4.55	4.44	4.67	**Primary**	4.12	4.02	4.24	**Never smoker**	4.04	3.94	4.16
Primary				**Former smoker**	4.76	4.65	4.88	**Primary**				**Former smoker**	4.33	4.25	4.45
Primary				**Current smoker**	4.58	4.47	4.68	**Primary**				**Current smoker**	4.17	4.06	4.29
Secondary	5.77	5.71	5.85	**Never smoker**	5.72	5.62	5.82	**Secondary**	2.93	2.86	3.02	**Never smoker**	2.88	2.82	2.96
Secondary				**Former smoker**	6	5.91	6.11	**Secondary**				**Former smoker**	3.09	3.02	3.19
Secondary				**Current smoker**	5.76	5.67	5.88	**Secondary**				**Current smoker**	2.97	2.9	3.04
Tertiary	5.86	5.76	5.98	**Never smoker**	5.7	5.62	5.81	**Tertiary**	2.9	2.83	2.98	**Never smoker**	2.85	2.78	2.93
Tertiary				**Former smoker**	5.97	5.88	6.08	**Tertiary**				**Former smoker**	3.05	2.98	3.15
Tertiary				**Current smoker**	5.74	5.66	5.85	**Tertiary**				**Current smoker**	2.93	2.86	3.01

Current asthma								COPD and/or chronic bronchitis							
Education level	Prob.	Upper limit	Lower limit	Smoking status	Prob.	Upper limit	Lower limit	Education level	Prob.	Upper limit	Lower limit	Smoking status	Prob.	Lower limit	Upper limit
Primary	8.91	8.74	9.09	**Never smoker**	8.78	8.61	8.95	**Primary**	4.76	4.65	4.88	**Never smoker**	4.05	3.95	4.15
Primary				**Former smoker**	9.3	9.16	9.47	**Primary**				**Former smoker**	4.16	4.06	4.29
Primary				**Current smoker**	8.95	8.8	9.13	**Primary**				**Current smoker**	4.32	4.23	4.44
Secondary	8.87	8.75	9.01	**Never smoker**	8.78	8.61	8.95	**Secondary**	4.81	4.71	4.92	**Never smoker**	2.88	2.82	2.96
Secondary				**Former smoker**	9.25	9.1	9.38	**Secondary**				**Former smoker**	2.97	2.91	3.06
Secondary				**Current smoker**	8.9	8.8	9.03	**Secondary**				**Current smoker**	3.09	3.03	3.17
Tertiary	8.85	8.73	8.99	**Never smoker**	8.77	8.66	8.91	**Tertiary**	4.85	4.75	4.97	**Never smoker**	2.85	2.78	2.93
Tertiary				**Former smoker**	8.89	8.78	9.03	**Tertiary**				**Former smoker**	2.94	2.86	3.03
Tertiary				**Current smoker**	8.89	8.78	9.03	**Tertiary**				**Current smoker**	3.06	2.99	3.15

To elaborate, the excess probability of allergic asthma resulting from former smoking was more notable among individuals with secondary education: an increase of 0.28%, with former smoking at 6.0% (5.91-6.11) compared to never smoking at 5.72% (5.62-5.82), and tertiary education: an increase of 0.27%, with former smoking at 5.97% (5.88-6.08) compared to never smoking at 5.70% (5.62-5.81), in comparison to those with only primary education: an increase of 0.21%, with former smoking at 4.76% (4.65-4.88) compared to never smoking at 4.55% (4.44-4.67).

On the other hand, the probability of non-allergic asthma due to former smoking was higher among individuals with lower education levels, particularly those with primary education: an increase of 0.29%, with former smoking at 4.33% (4.25-4.45) compared to never smoking at 4.04% (3.94-4.16), in contrast to those with tertiary education: an increase of 0.20%, with former smoking at 3.05% (2.98-3.15) compared to never smoking at 2.85% (2.78-2.93).Figure table 2 and figure 10 for full results showing the effect of smoking on allergic and non-allergic asthma and the effect modification by education levels.

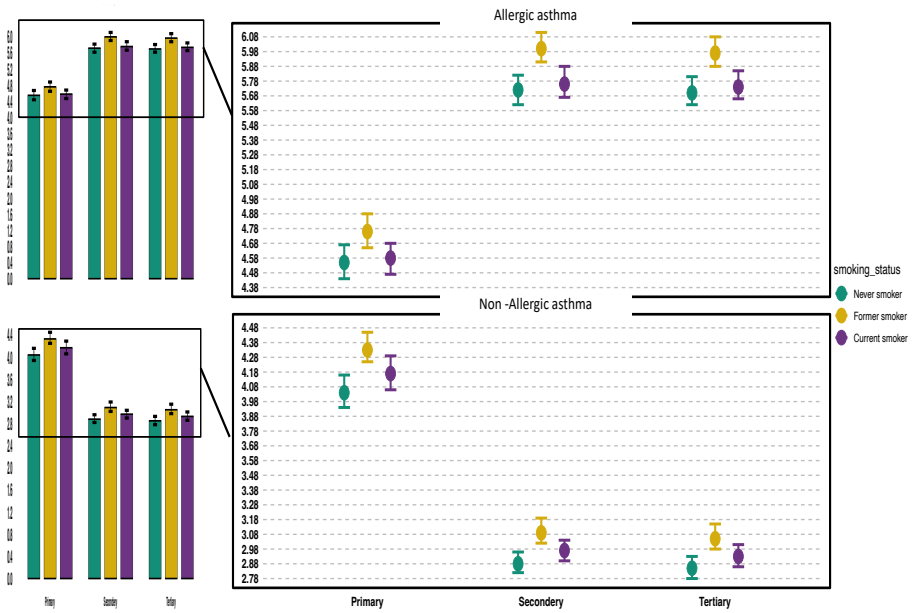


Figure 12. The probabilities and 95% credible intervals for the effect of smoking on allergic and non-allergic asthma and its modification by education level.

3.2.4.3 SMOKING AND EFFECT MODIFICATION SSYK OCCUPATIONAL GROUPS

With respect to the SSYK occupational classification, this system was insensitive to variation in the probability of current asthma, allergic or non-allergic asthma as well as COPD and/or chronic bronchitis due to occupational classification. Hence, the effect of smoking on respiratory outcomes showed no variation across this class.

3.3 COMPUTATIONAL PHENOTYPING OF OBSTRUCTIVE AIRWAY DISEASES: A SYSTEMATIC REVIEW

3.3.1 STUDY SELECTION

A total of 3320 records were identified in literature search. After removal of duplicates, 2619 records were screened by title and/or abstract, of which 2460 records were excluded for non-relevance. A total of 159 records were considered for full-text review, of which 39 were excluded for different reasons and summarized with reason of exclusion in table S1 in the supple tray material. Finally, 120 studies were included in this review analysis.

3.3.2 STUDY CHARACTERISTICS

3.3.2.1 ASTHMA

A total of 60 studies were conducted on asthma, with an average number of subjects of 1251 per study. The range of participants per study varied from 50 to 9651. Among these studies, the majority (n = 43) focused on adults [80, 81, 83-101, 105, 110, 237-255]. Conversely, a smaller number of studies (n = 17) were conducted among children and adolescents [73, 104, 109, 256-268]. Out of the total studies, the majority (n = 35) utilized a cohort design, while the remaining studies (n = 18) were cross-sectional. Furthermore, most studies (n = 37) were conducted in clinical settings, involving patients recruited from hospitals, pulmonary rehabilitation centers, and primary or tertiary care respiratory or general clinics [73, 80-82, 86-89, 91-93, 95, 97-100, 106, 110, 237, 238, 241, 244-246, 251, 252, 254, 255, 257, 258, 261, 262, 265, 267, 268]. Among the studies, 18 selected subjects from the general population, while 7 studies [82, 86, 95, 241, 243, 247, 260, 263] did not report on the source of their participants. Full information on characteristics of studies in children

and adults using unsupervised computational methods to phenotype asthma are presented in (**Table 4**) below.

3.3.2.2 COPD

A total of 28 studies were conducted on COPD [269-291]. The average number of subjects per study was 5218, with a range of 46 to 104143 subjects per study. The majority of these studies (n = 19) [269-275, 277, 279-282, 285-290] were conducted within the clinical setting. Among the reported study designs, cohort studies were the most common (n = 14), while the remaining studies (n = 8) were cross-sectional in nature. For full information regarding the characteristics of studies utilizing unsupervised computational methods to phenotype COPD and Asthma-COPD Overlap (ACO), please refer to (**Table 5**) below.

3.3.2.3 SEVERE ASTHMA

A total of 19 studies were conducted on severe asthma, as referenced by studies 92-106. The average number of participants included in each study was 230, with the range varying from 40 to 1424 subjects per study. Most of these studies (n = 13) were conducted in a clinical setting [292-304]. One study was conducted in a general population setting [305] , while another study did not clearly indicate the setting of the research [306]. Among the reported study designs, most were cohort studies (n = 12). The remaining three studies were cross-sectional in nature [292, 295, 306]. For detailed information on the characteristics of studies focusing on severe asthma in both children and adults, utilizing unsupervised computational methods for phenotyping, please refer to (**Table 4**) blow.

3.3.2.4 RHINITIS

A total of 9 studies were conducted on rhinitis [307-316]. The average number of participants included in each study was 516, ranging from 115 to 1831 participants. Among these studies, the majority (n = 6) took place in a clinical setting, specifically mentioned in studies [307-309, 311, 315, 316]. Three studies were conducted in the general population [310, 312, 313], while one report did not provide clear information about the setting.

In terms of study design, most of the studies (n = 8) were cohort studies [307, 309, 311-316]. One study followed a cross-sectional design [308], and one study employed a case-control design [310]. For a comprehensive understanding of the characteristics of studies utilizing unsupervised computational methods to phenotype rhinitis, please consult (**Table 6**) below.

3.3.2.5 ASTHMA AND COPD OVERLAP

Four of the included studies focused on asthma and COPD overlap, as indicated by references [317-320]. The average number of participants included in these studies was 255, with a range of 47 to 435 participants per study. All these studies were cross-sectional in nature. Among the four studies, three [317, 319, 320] were conducted in a clinical setting, while one study was conducted in a general population setting. For a comprehensive understanding of the characteristics of studies utilizing unsupervised computational methods to phenotype COPD and asthma-COPD overlap (ACO) in adults, please refer to (**Table 5**).

Table 4. Characteristics of studies in children and adults using unsupervised computational methods to phenotype asthma and severe asthma

Reference and country	Study design	Population and participants	Population type	Study setting	Method used for phenotyping
ASTHMA[1]					
Amaral et al. 2019 US	Cohort study	Adults = 1059	Current asthma	General population	Latent class analysis
Amelink et al.2013. Netherlands	Cross sectional	Adults n= 200	Patients with adult-onset asthma	Clinical setting	Nonhierarchical clustering
Bhargava et al.2018. India	Cohort	Adults n = 113	Asthma based on GINA followed up for 6 months	Clinical setting	Hierarchical clustering
Benton et al.2010. USA	Cross sectional	Children and adolescents n =154	Children with physician diagnosed asthma.	General population	Hierarchical and non- hierarchical clustering
Bochenek et al.2014. Poland	Cross sectional	Adults n=201	Patients diagnosed with AERD by physicians	Clinical setting	Latent class analysis
Boudier et al. 2013. Multicounty	Cohort	Adults n=3320	Patients with asthma	General population	Latent class analysis
Cabral et al.2017. Brazil	Cohort	Children and adolescents n =306	Children and nonsmoking adolescents (6-18 years of age) with a clinical diagnosis of asthma.	Clinical setting	Hierarchical clustering
Celejewska-Wójcik et al. 2020. Poland	Cross sectional	Adults n= 95	Patients with NERD based on provocation test and asthma based on GINA.	Not indicated	Latent class analysis
Chanoine et al.2017. France	Cohort	Adults n= 4328	(20–75) years old women with reported ever having asthma.	General population	Hierarchical and non- hierarchical
Couto et al. 2015. Portugal	Cross sectional	Adults n=150	Athletes with asthma diagnosis based on International Olympic	General population	Latent class analysis
Cruz et al. 2018. Brazil	Cohort	Adults n=966	Adults with asthma	Clinical setting	Nonhierarchical clustering
Damiens K et al.2013. Canada	Cohort	Adults n=272	Work related asthma.	Clinical setting	Cluster analysis
Deliu et al. 2016. Turkey	Cross sectional	Children n = 201	Asthma patients, children aged 6 – 17 years.	Clinical setting	Hierarchical clustering

Deliu et al. 2018. Turkey	Cross sectional	Children n = 613	Children with asthma aged 6-18 years.	General population	Hierarchical clustering
Depner et al. 2013; Multi-country	Cohort	Children n = 953	Children from rural areas	Clinical setting	Latent class analysis
Dudchenko et al. 2018. Russia	Not indicated	Adults n=300	Patients with asthma	Clinical setting	Cluster analysis
Folz et al.2018. USA	Cohort	Adults n= 136	Asthma patients	General population	Hierarchical clustering
Fontanella et al. 2018. UK	Cross sectional	Children = 461	Children with available component resolved diagnostic data	General population	Hierarchical clustering
Gonem S et al.2012. UK	Not indicated	Adults n=114	Asthma patients and healthy controls.	Not indicated	Hierarchical and non- hierarchical
Gower WA et al. 2013. USA	Cohort	Children n = 942	Mild, moderate, or severe asthma.	Not indicated	Hierarchical clustering
Hilvering et al.2015; Netherland	Cross sectional	Adults n=115	Adult asthma patients.	Not indicated	Nonlinear Principal Component Analysis
Hsiao HP et al.2018; Taiwan	Cohort	Adults n=720	Age > 20 years, diagnosed with stable physician diagnosed, mild-to-severe asthma.	Clinical setting	Orthogonal varimax factor analysis
Ilmarinen P et al. 2017. Finland	Cohort	Adults n=171	White patients with new onset adult-onset asthma diagnosed at age > 15 years.	Clinical setting	Hierarchical clustering
Jeong A et al.2017. Switzerland	Cohort	Adults n=959	Self-reported asthma.	General population	Latent class analysis
Just J et al. 2012. France	Cohort	Children n = 315	Aged 6– 12 years at the time of exploration; a history of asthma	Clinical setting	Hierarchical clustering
Just J et al. 2014. France	Cross sectional	Children=125	Children aged between 6 to 12 years, having allergic asthma	Clinical setting	Hierarchical clustering
Kaneko Y et al.2013. Japan	Cross sectional	Adults n=880	Physician diagnosed adult asthma based on either symptoms or spirometry measures or both.	Clinical setting	Hierarchical and non- hierarchical clustering

Kim HJ et al.2018. Korea	Cohort	Adults n=1679	Asthma patients aged 18 to 79 years	General population	Hierarchical clustering- Ward's
Kim MA et al., 2017. Korea	Cohort	Adults n=259	Physician diagnosed asthma patients according to GINA.	Clinical setting	Hierarchical and non- hierarchical
Kim TB et al. 2013. Korea	Cohort	Adults n=2567	Subjects with dyspnea, cough, sputum production or wheezing for >3 months, positive BHR, and stable conditions with regular medications	Clinical setting	Hierarchical and non- hierarchical clustering
Koike et al. 2018. (abstract) Not indicated.	Not indicated	Adults n = 458	Subjects with asthma presenting with cough and no other microbial infection.	Clinical setting	Hierarchical cluster analysis
Kwon et al. 2012. Korea	Cross sectional	Children n = 193	Children with physician diagnosed asthma	General population	Hierarchical and non- hierarchical
Lee E et al. 2017. South Korea	Cohort	children n=235	Children 6-8 years with parent reported physician diagnosed asthma in lifetime	Clinical setting	Latent class analysis
Li et al. 2016. USA	Cross sectional	Adults n= 2081	GOLD stage 0-4, smoking≥20 packs/year	Not indicated	Cluster analysis
Loureiro et al. 2015. Portugal	Cross sectional	Adults n=57	Asthma patients, aged >18 years; diagnosed according to the GINA	Clinical setting	Hierarchical cluster analysis
Loza et al. 2016. Multicounty	Cross sectional	Adults n=238	Not indicated	General Population	Non- hierarchical clustering
Mahut et al. 2011. France	Cross sectional	Children n = 169	Children who satisfy criteria of clinical and functional asthma.	Clinical setting	Nonhierarchical clustering
Mäkikyrö et al.2017. Finland	Cross sectional	Adults n=1995	Asthma patients according to the Social Insurance Institution of Finland criteria who received reimbursement right for asthma medication.	General population	Latent class analysis
Mason et al. 2018. Italy	Cohort	Adults n=187	Subjects with diagnosis of occupation asthma based on (SIC) result to diisocyanate.	General population	Hierarchical and non- hierarchical clustering

Mastalerz1 et al.2015. Poland	Cross sectional	Adults n=137	Patients with asthma, aspirin tolerant asthma, severe asthma according to ATS/ERS recommendation of 2013, atopic asthma, and asthma with rhinosinusitis.	Not indicated	Latent class analysis
Nadif et al 2018. France	Cohort	Adults n = 318	Adult with current asthma	Clinical setting	Mixture models
Nagasaki et al. 2014. Japan	Cohort	Adults n=224	Patients with stable asthma.	Clinical setting	Hierarchical clustering
Nasreen et al. 2019. Canada	Cohort	Children n = 403	Asthma patients based on parents' report.	General population	Latent class analysis
Qui et al. 2018. China	NC	Adults n=218	Asthma based on symptoms or physician diagnosis according to GINA. Severity defined based on need for SCS.	Clinical setting	Hierarchical and non- hierarchical clustering
Sakagami et al.2011. (abstract) Japan	Not indicated	Adults = 591	Patients from primary, secondary, and tertiary care centers.	Clinical setting	Cluster analysis
Schatz et al, 2013. USA	Cohort	Children and adolescents n	Asthma patients aged ≥15 years according to ATS.	Clinical setting	Hierarchical clustering
Schimdlin et al. 2015. (Abstract) USA	Cohort	Children n = 72	Asthma patients based on parents' report and lung function test.	General population	Hierarchical clustering
Seino et al. 2018. Japan	Cohort	Adults n= 2273	Adults with asthma diagnosis according to GINA.	Clinical setting	Hierarchical clustering
Sendín-Hernández et al. 2018. Spain	Retrospective record-based study	Adults n= 225	Allergic asthma patients based on lung function test and symptoms ages 18-65 years (ATS) criteria.	Clinical setting	Hierarchical clustering
Seys et al. 2017. Belgium	Cohort	Adults, asthma =205, healthy =80	Asthma based on symptoms and reversibility test at age 18-65 years.	Clinical setting	Hierarchical clustering
Siroux V et al. 2011. Multi country	ECRHSII = Cohort	Adults n= 1895	Age > 14 years; asthma diagnosed following GEMA 2009,	General population	Latent class analysis

	EGEA2 = Case–control and	Adults n= 641	Adults who had ever had asthma at EGEA2.	General population	Multivariate exploratory data clustering preceded
Tay et al. 2019. Singapore	Cohort	Adults n=630	Asthma based on objective measure of airflow limitation or physician diagnosis in addition to asthmatics	Clinical setting	Latent class analysis
Tsukioka et al. 2017. Japan	Cohort	Adults n=104	Elite athletes with asthma.	Clinical setting	Hierarchical clustering
Wang LL et al. 2017; (abstract) China	Cohort	Adults n=284	Clinically stable asthma patients.	Clinical setting	Hierarchical and non- hierarchical clustering
Watanabe et al. 2016 (abstract). Japan	Japan	Adults n=120	Non-smokers, aged ≥60 with diagnosis of asthma.	Clinical setting	Hierarchical and non- hierarchical clustering
Wisnivesky et al;2019 USA	cohort	Adults = 330	Physician diagnosed asthmatic among from population of exposed to 11th/9 disaster.	General population	Kamila Algorithm
Wu et al.2017. China	Cohort	Adults n=110	Patients with nasal polyposis and comorbid asthma; 18-65 years of	Clinical setting	Non- hierarchical clustering
Zaihra T et al.2016; Canada	Cohort	Adults N =125, Severe asthmatic =77, moderate =48	Asthmatic patients according to GINA.	Clinical setting	Non- hierarchical clustering
Zhang X et al. 2019 (abstract); China	Retrospective case control	Adults n= 825	Patients with NPcA diagnosed by pathology after biopsy.	Clinical setting	Hierarchical clustering
Zoratti E et al. 2018. USA	Cohort	Children n = 717	Asthma and high dose or systematic steroids or symptoms, lung	Clinical setting	Boruta feature selection algorithm
Brinkman et al, 2011. Multi-countries	Cross sectional	Adults n = 77	Severe asthma based on the IMI criteria.	Clinical	Hierarchical cluster analysis
Desai et al.2011. UK	Not indicated	Adults n = 164	Patients attending difficult asthma clinic	Clinical	Non-hierarchical clustering

Diver et al.2018. UK	Cohort	Adults n = 63	Severe asthma, moderate to severe COPD. maintenance oral corticosteroid therapy; subjects with sputum samples adequate for microbiome sequencing.	Clinical	Mixtur modelling
Fitzpatrick et al.2018. USA	Cohort	Children n = 161	6 to 17 years of age, never smoked, physician diagnosed asthma based on spirometry at baseline or during exacerbation.	Clinical	Hierarchical cluster analysis
Freitas PD et al.2018. (Abstract). Brazil	Cross sectional	Adults = 119	Patients with moderate to severe asthma under optimal treatment according to GINA.	Not indicted	Hierarchical cluster analysis
Gomez et al. 2017. USA	Cohort	Adults n = 156	Subjects with severe asthma	General population	Hierarchical clustering
Jang et al, 2017. Korea	cohort	Adults n = 86	Patients with refractory asthma	Clinical	Hierarchical and non-hierarchical clustering
Konstantellou et al. 2015. Greece	Cohort	Adults n = 170	Patients with asthma diagnosis followed up in 1sth and 2nd centers based on GINA.	Clinical	Hierarchical and non-hierarchical clustering
Lau et al, 2017: Singapore	Cohort	Adults n = 55	Patients with asthma exacerbation of status asthmatics requiring ICU admission and intubation	Clinical	Cluster analysis
Moore et al.2010. USA	Cohort	Adults n = 304	Refractory asthma.	General population	Hierarchical cluster analysis
Newby et al.2018. UK	Cohort	Adultsn = 349	Severe asthma as OCS 50% of the year or high dose ICS plus add on medications either at baseline or follow-up.	Clinical	Mixture models
Raherson et al, 2018. France	Cohort	Adults n = 1424	Not indicated	Clinical	Hierarchical and non-hierarchical clustering

Sekiya et al. 2015. Japan	Cohort	Adults n = 190	Stable, on maintenance asthma therapy 4 weeks prior to study, asthma diagnosis as per GINA and severity status as per (ERS) and (ATS).	Clinical	Non-hierarchical clustering
Serrano Pariente et al. 2015. Spain	Cohort	Adults n = 84	Asthma as per ATS criteria, Aged > 15 years	Clinical	Non-hierarchical clustering
Simpson et al (abstract);2017. Multi-country	Cross sectional	Adults = 421	Severe asthma based on U-BIOPRED consensus criteria indicated in detail here [321].	Clinical	Not indicated
Taniguchi et al; 2014. Japan	Cohort	Adults = 127	Severe refractory asthma based on ATS criteria without exclusion based on smoking history	Clinical l setting	Unbiased cluster analysis
Wu et al; 2014. USA	Cohort	Adults = 378	Severe asthma based on ATS criteria	General population	Non-hierarchical clustering
Weng-Jing Ye et al. 2017. China	Cohort	Adults n = 203	Severe asthma as per the IMI-criteria.	Clinical	Hierarchical and non-hierarchical clustering
Youroukova et al. 2017. Bulgaria	Cohort	Adults n = 40	Severe refractory asthma as use of OCS 50% of the year or high dose ICS plus add on medications at baseline or follow-up.	Clinical	Hierarchical cluster analysis

[1] **Abbreviations used in Table 1:** AERD: aspirin exacerbated respiratory diseases, GINA: Global Initiative for Asthma, ATS/ERS: American Thoracic society/European respiratory society, SCS: systematic corticosteroids, NSAID: non-steroidal anti-inflammatory diseases, GEMA: Spanish asthma management guidelines, NPcA: nasal polyposis and comorbid asthma, ICS: inhaled corticosteroids, IMI: innovative medicine initiative, OCS: oral corticosteroids.

Table 5. Characteristics of studies in children and adults using unsupervised computational methods to phenotype COPD and asthma COPD overlap.

Reference and country	Study design	Population And participants	Population type	Study setting	Method used for phenotyping
COPD[2]					
Augustin et al.; 2018 Netherland	Cohort	n =518	Clinically stable COPD	Clinical population	Self- organising maps.
Bafadhel et al (abstract). 2011. UK	Cohort	n =86	Physician diagnosis COPD based on spirometry, age> 40, GOLD stage I–IV and current sever exacerbation.	Clinical population	Cluster analysis
Bertini et al; 2013. Italy	Cross sectional	n =62	Clinically confirmed exacerbation within 3 months prior to study.	Clinical population	Non- hierarchical clustering
Burgel et al; 2017. France	Cohort	n =2409	Subjects with COPD diagnosed based on spirometry whether stable or on exacerbation.	Clinical population	Classification and regression tree
Burgel et al.; 2010. France	Cross sectional	n =322	Stable COPD diagnosed based on spirometry.	Clinical population	Hierarchical clustering
Burgel et al.; 2012 Belgium	Cross sectional	n =527	Smoking history ≥15 pack-years, age > 50 years, and 154 patients diagnosed of COPD based on spirometry.	General and clinical population	Hierarchical clustering
Chen et al; 2014 Taiwan and China	Cohort	n =332	Men diagnosed with COPD, ≥40 years old, based on symptoms and spirometry.	Clinical population	Cluster analysis
Chubachi et al (abstract); 2016 Japan	Cohort	n =311	COPD and completed data.	Not Indicated	Hierarchical clustering
De Torres et al; 2017 Spain	Cohort	n =521	Active and former smokers with COPD.	Clinical population	Non- hierarchical clustering

[2] **Abbreviations used in table 2**: GOLD: global initiative for chronic obstructive lung disease, GINA: global initiative of asthma.

Divo et al.(abstract); 2016 Not indicated	Cohort	n =120	Patients with COPD.	Not Indicated	Hierarchical clustering
Fens et al.; 2013. Netherland	Cross sectional	n =157	Smoking history of at least 15 pack-years, COPD and chronic bronchitis based on GOLD.	General population	Hierarchical clustering
Guillamet RV et al.; 2018 USA	Cohort	n =3144	>40 years, COPD, and ever hospital admission	Clinical population	sphere exclusion method
Harrison SL et al.; 2014 UK	Cohort	n =92	Asthmatics based on spirometry and COPD as per signs and symptoms, age > 40 years, and smoking history.	Clinical population	Hierarchical and non-hierarchical clustering
Haghighi et al, 2019. USA	Cross sectional	n= 406	Former smokers	General population	Non- hierarchical clustering
Kim W.J et al; 2017 Asia	Cohort	n =1676	Asian, age 40 years and with COPD based on post bronchodilator spirometry.	General population	Hierarchical clustering
Kim S et al 2017; Korea	Cohort	n = 272	age > 40 years and (FEV1/FVC) <0.7	General population	Hierarchical clustering
Kukol et al. 2019. (abstract). Russia	Not indicated	Not indicated	Elderly patients with COPD	Not indicated	Cluster analysis
Lee et al, 2019. Korea	Cohort	N = 1195	Patients with available follow up data on the first acute exacerbation of COPD.	Clinical	Non- hierarchical clustering
li et al. abstract;2016. USA	Not Indicated	n =2081	GOLD stage 0-4, smoking ≥20 packs/year	Not Indicated	Hierarchical clustering
Liang et al, 2019. Korea	Prospective cross sectional	n = 102 patients n = 18 controls	age > 40, COPD diagnosis based on GOLD, post BD EFV1 <0.7.	Clinical population	Hierarchical clustering
Lopes et al, 2019. Brazil	Cross sectional	n =150	Stable COPD diagnosis within the last 30 days, based on GOLD	Clinical population	Hierarchical and non-hierarchical clustering
Ning et al; 2016, (abstract). China	Not indicated	Not indicated	Subjects with wheeze for more than 12 months.	General population	Hierarchical clustering

Peters et al; 2016. Netherlands	Cohort	n =160	COPD based on spirometry and GOLD grades 2-3, receiving treatment as usual.	Clinical population	Hierarchical clustering
Pikoula et al; 2019 UK	Cohort	n =30961	Age >35 year, with at least one diagnostic code of COPD based on GOLD and complete data.	Clinical population	Hierarchical and non- hierarchical clustering
Rodrigues et al.; 2018. Brazil	Cohort	n =141	Stable COPD, with no sever cardia or musculoskeletal comorbidity	Clinical population	Non- hierarchical clustering
Scarlata et al; 2018. Italy	Cross sectional	n =50	COPD diagnosed based on spirometry.	Clinical population	Non- hierarchical clustering
Xavier F et al;2019 Barizal	Cross sectional	n =152	Stable COPD diagnosed according to GOLD.	Clinical Population	Non- hierarchical clustering
Yoon et al.; 2019. Korea	Cohort	n =1195	Physician diagnosed COPD and age ≥ 40 years.	Clinical population	Non- hierarchical clustering
Asthma COPD overlap					
De Vries et al. 2018. Netherlands	Cross sectional	Adults n = 435	Patients with asthma and COPD	Clinical setting	Hierarchical clustering.
Fingleton et al. 2017. New Zealand and China	Cross sectional	Adults n = 345	Caucasian, Chinese, complete data, aged 40–75 who agreed to participate.	General population	Hierarchical clustering.
Górska K t al. 2017. Poland	Cross sectional	Adults n = 47	Asthma or COPD defined by symptoms, spirometry, and history of smoking.	Clinical setting	Hierarchical clustering.
Rootmensen et al. 2016. Netherlands	Cross sectional	Adults n = 191	> 18 years old, physician diagnosed COPD, understood Dutch, never had consulted a pulmonary nurse.	Clinical setting	Non- hierarchical clustering

Table 6. Characteristics of studies using unsupervised computational methods to phenotype rhinitis

Reference and country	Study design	Population and participants	Population type	Study setting	Method used for phenotyping
Rhinitis [3]					
Adnane et al.2017. Morocco	Cohort	Adults n = 131	Patients with medical refractory chronic rhinosinusitis (CRS).	Clinical setting	Discriminant analysis
Agache et al. 2010. Romania	Cohort	Adults and children n = 115	Adults and children (aged 5–11 years), physician diagnosed SAR, asthma as per symptoms and spirometry test. SAR as per symptoms SPT and at least 1 of seasonal allergens.	General population	Non-hierarchical clustering.
Bousquet et al. 2015. France	Case control and family study	adults n = 825	Age >18 years, diagnosis of AR and symptoms at time of examination.	General population	Hierarchical clustering
Burte et al. 2015. France	Cohort	Adults n = 983	Age ≥16 years, complete data, allergic sensitization, and asthma	Clinical setting	Mixture models
Herr M et al.2012. France	Cohort	Infants n = 1831	Singleton full-term newborns, normal birth weight and an uncomplicated birth and neonatal period with parents who can commit to and participate in study.	Clinical setting	Hierarchical clustering
Kurukulaaratchy et al. 2015. UK	Cohort	Adolescents n = 468	Self-report of current rhinitis at age 18.	Clinical setting	Non-hierarchical clustering.

[3] Abbreviations in table 3: SAR: seasonal allergic rhinitis, SPT: skin prick test, AR: allergic rhinitis, CT: Computed Tomography, CRSnP: rhinosinusitis with nasal polyposis.

Lee E.L. et al 2016. Korea	Cohort	Children n = 512	Children with parental-reported, physician- diagnosed rhinitis and symptoms of rhinitis in the previous 12 months.	General population	Latent class analysis
T.A Nakayama et al. 2012. Japan	Cohort	Adults n = 425	CRSnP as per the European position paper and non-response to medical treatment and undergone surgery.	Clinical setting	Non-hierarchical clustering.
ZM Soler et al.2015. USA and Canada	Cohort	Adults n = 382	Consensus criteria: 3 months of at least 2 cardinals' symptoms and evidence of inflammation on sino-nasal endoscopy and CT scan, and active symptoms after initial treatment.	Clinical setting	Hierarchical clustering

3.3.3 METHODS OF PHENOTYPING

Various methods of unsupervised computational phenotyping of respiratory diseases were used across the reported studies. The most frequently implemented and reported unsupervised approaches for phenotyping of chronic airway diseases were hierarchical and non-hierarchical clustering [85, 94, 106, 237, 244, 248, 263, 271, 277, 287-289, 291, 293, 301, 302, 320] with some records (n = 19) reported the implementation of the two approaches in the same study [85, 89, 95, 239, 241, 242, 258, 275, 281, 285, 290, 298-300, 303, 322-324]. In addition, latent class modelling [80, 83, 84, 86, 91, 93, 96-98, 237, 245, 247, 250, 257, 259, 264, 311, 313] was also frequently used. Other non-model-based methods of dimensionality reduction, such as factor analysis, principal component analysis, discriminant analysis and multiple correspondence analysis were also reported as methods for deriving phenotypes, albeit less frequently. Over years, hierarchical and non-hierarchical clustering were common particularly between 2010-2018. However, from 2015 and later, there was increase in the use of other methods such as mixture-based model [292, 314], structural equational modelling [269], factor analysis with latent class modeling [325]. See (**Figure 22**) in the appendix.

3.3.4 QUALITY ASSESSMENT OF THE INCLUDED STUDIES

Overall, the comprehensiveness of variables included in deriving the phenotypes was the best quality aspect reported in majority of studies on asthma [80, 83, 84, 86, 92, 93, 96, 97, 99, 100, 104, 110, 237, 240, 242, 246, 247, 250, 254, 257, 258, 260, 264, 322, 323, 326, 327], COPD[244, 269, 272, 274, 275, 277, 278, 285-288, 291, 328, 329], severe asthma [295, 302, 303, 305, 330]; ACO [318-320] and rhinitis [308, 310, 312-314]. Random sampling of study subjects, however, was less frequently performed among studies on asthma (n = 8) [83, 93, 94, 110, 248, 259, 264, 331], severe asthma (n = 2) [301, 305], COPD (n = 1) [284], ACO (n = 2)[318, 320] and rhinitis (n = 1) [313]. Majority of studies excluded subjects based on either clinical, social, or demographic characteristics. With respect to method of outcome definition, the most reported approach was usage of physician diagnosis assisted by clinical and biomarkers, which was reported in 33 of studies on asthma [83, 85-93, 95, 96, 98, 99, 106, 110, 237, 238, 240, 242, 245, 247, 248, 250-253, 257, 258, 263, 267, 326, 327], 16 of studies on COPD [244, 269-272, 275-277, 279, 286-289, 291, 328, 329], 12 of studies on severe asthma [292, 294, 295, 298, 300-303, 305, 306, 324, 330], four of studies on rhinitis [307, 308, 314, 315], two of studies on ACO [319, 320]. Overall reporting on reproducibility practices was uncommon, including how investigators handled noise and variation in data, rationale for selecting statistical methods for phenotyping, visualization techniques, and utilization of available tools for

implementing reproducibility. With respect to clinical, biological or scientific relevance of the derived phenotypes, most studies reported on this aspect: 26 studies on asthma [84, 86-88, 90, 92, 94-98, 104, 237-242, 257-260, 326, 327, 332]; 18 studies on COPD [244, 269, 272, 274-281, 285-287, 289, 290, 328, 329]; 11 studies on severe asthma [292-295, 298, 301, 305, 306, 324, 330]; three studies [317-319] on ACO and nine studies on rhinitis [307-315]. Full information on the quality assessment results can be found in table S2 in supplementary material. See (**Figure 23**) in the appendix.

3.3.5 Endotypes and phenotypes of airways diseases

Efforts to define phenotypic subtypes of airway diseases involved utilizing various approaches, including the assessment of serum and sputum-induced inflammatory cells as well as other biomarkers associated with inflammation and related processes. These biomarkers include cytokines, airway-inducible inflammatory mediators, and the composition of the airway microbiome.
Mastalerz et al.[250] and Linag et al.[245] conducted studies investigating the role of airway-induced pro and anti-inflammatory lipid eicosanoids mediators in asthma. In their research, Mastalerz et al.[250] identified three distinct asthma phenotypes based on high levels of anti-inflammatory mediators. One phenotype exhibited chronic rhinosinusitis (CRS), good control, and mixed inflammation. The second phenotype showed atopy, no CRS, good control, and mixed inflammation. The third phenotype had poor control, aspirin sensitivity, and eosinophilia. Additionally, Mastalerz et al. [250] identified a phenotype characterized by high levels of anti-inflammatory mediators among obese women with early-onset, atopic, and severe asthma.

Liang et al. [245] derived a phenotype that is characterized by a mixture of pro and anti-inflammatory mediators. This phenotype exhibited low basophils, high functional activity, and poor control. Liang et al. [245] also identified a phenotype with low basophils, high anti-inflammatory mediators, and low control. Similarly, Caljwaska et al.[96] utilized induced sputum supernatants for phenotyping of non-steroidal anti-inflammatory drug-exacerbated respiratory disease (NERD). Seys et al.[91] focused on the expression of inflammatory cytokines in airway sections to further subtype asthma patients. Their work reported an unexpected pattern of cytokine predominance among Th2-high asthmatics. It, further, identified subclusters with high levels of IL-5, IL-10, IL-25, and IL-17, associated with low lung function, high eosinophils, neutrophils, and fractional exhaled nitric oxide (FeNO). Another cluster consisted of IL-5 and/or IL-10 high asthmatics, while a separate cluster showed high levels of IL-6. A fifth cluster exhibited a normal pattern of high

Th2 cytokines but with high eosinophils and low neutrophils. These findings indicate further heterogeneity among individuals with Th2-high asthma. However, this phenotyping approach did not include other clinical parameters such as symptoms, lung function, and outcomes. Instead, the derived clusters were modeled against clinical outcomes for further evaluation.

Nagasaki et al.[80] identified phenotypes characterized by high serum periostin. One phenotype exhibited high eosinophilia, early-onset disease with good control, while the other phenotype showed high periostin, mixed inflammation, severe disease with poor control, high IL-6, mixed inflammation, and multimorbidity.

In the context of COPD, Bafadhel et al.[270] identified five distinct phenotypes of COPD exacerbations based on biological biomarkers. These phenotypes were characterized by different predominant factors, including bacteria, viruses, eosinophils, paucigranulocytemia, and elevated levels of Sputum IL-1b and serum CXCL10. Britani et al.[271] employed exhaled breath condensate analysis to assess inflammatory biomarkers in COPD, revealing unique metabolite profiles such as the low proline phenotype and high serine, valine, lysine, acetate, alanine, isoleucine, and other metabolite phenotypes.

In severe asthma, Diver et al.[294] focused on airway microbiology and identified clusters of severe asthma with varying levels of *Haemophilus* and *Moraxella* sputum communities, as well as different ratios of Gammaproteobacterial (G) to Firmicutes (F). Gomez and colleagues[296] explored severe asthma phenotypes characterized by varying levels of the chitinase-like protein YKL-40 as an inflammatory mediator which was inversely associated with disease severity, control, and treatment response.

An endotype of high eosinophilic chronic rhinosinusitis with high nasal polyposis and low CT and endoscopy score was identified by Adnane et al.[307], while Nakayma ae al.[315] reported a high eosinophil and basophiles rhinitis that is characterized by comorbid asthma, high symptoms, high CT and endoscopy score. Low eosinophilic rhinitis reported by Adnane et al. [307] were variants among male subjects, and others among females but with high CRSnP and CT and endoscopy score. Nakayma and colleagues [315] variants of low eosinophilic rhinitis were non differential in symptoms, endoscopy and CT score.

3.3.6 PHENOTYPES OF RESPIRATORY DISEASES

3.3.6.1 ASTHMA

In total, a comprehensive analysis of studies on asthma revealed the identification of 256 phenotypes, highlighting a significant overlap among them. The characterization of asthma phenotypes primarily focused on several key features. The most commonly reported feature in characterizing asthma phenotypes was atopy, which was included in the majority of studies. This differentiation led to the categorization of atopic and non-atopic asthma phenotypes [73, 82-87, 89, 92, 93, 95, 97, 98, 100, 101, 109, 110, 240, 246, 247, 249, 250, 252, 257, 260, 261, 263, 264, 267, 268].

Atopy status was mainly determined through skin prick tests, serum IgE levels, or self-reported familial atopy. Specifically, atopic asthma phenotypes were reported in 30 studies [73, 82-87, 89, 92, 93, 95, 97, 98, 100, 109, 110, 240, 244, 246, 247, 249, 250, 252, 257, 260, 261, 263, 264, 267, 268], while non-atopic asthma phenotypes were reported in 23 studies [73, 82-85, 87, 88, 92, 95, 97, 100, 109, 110, 240, 246, 247, 252, 257, 258, 263, 264, 267, 268]. Another prominent feature considered in the characterization of asthma phenotypes was lung function measures, which were featured in 85 of the reported phenotypes. This feature was considered in 34 studies [73, 80, 84, 85, 87-90, 92, 93, 96, 97, 100, 109, 110, 237, 240-242, 245-247, 251, 257, 258, 260, 261, 263, 264, 267, 268]. The time of asthma onset was also an important feature, with 74 phenotypes being described based on this factor. This feature was reported in 27 studies [73, 80, 82, 83, 85-88, 92, 93, 95, 98, 100, 104, 110, 240, 244, 246, 247, 250, 251, 257-260, 267, 268], distinguishing between early-onset asthma [73, 80, 82, 83, 85, 87, 93, 95, 100, 110, 240, 244, 246, 247, 250, 251, 257, 258, 260, 267, 268] and late-onset asthma [73, 80, 82, 83, 85-87, 92, 93, 95, 98, 100, 104, 240, 244, 246, 258, 259, 267, 268].

The level of asthma control was commonly reported, occurring in 45 phenotypes and reported in 17 studies [80, 82, 90, 93, 98, 99, 104, 240, 242, 245-248, 250, 258, 261, 265]. Well-controlled asthma phenotypes were reported in 11 studies [82, 93, 99, 242, 245-248, 250, 258, 265], while uncontrolled asthma phenotypes were reported in 16 studies [80, 82, 90, 93, 98, 99, 104, 240, 242, 245-248, 250, 258, 261]. Gender was a notable factor, featuring in 65 of the reported phenotypes. Female asthma phenotypes were reported in 22 studies [73, 82, 86, 92, 93, 95-98, 104, 105, 110, 237, 240, 245, 246, 250, 251, 253, 265, 267, 268], while male asthma phenotypes were reported in 21 studies [73, 82, 92, 93, 97, 98, 104, 105, 110, 237, 240, 245, 246, 249, 251, 253, 260, 263, 265, 267, 268]. Twelve studies specifically reported on obesity-related asthma phenotypes. Disease severity was

characterized based on symptom frequency, exacerbation rates, or direct classification of asthma severity. Severe asthma phenotypes were reported in 12 studies, while 21 studies focused on asthma phenotypes with high symptoms or exacerbation rates.

Inflammation was a crucial consideration in deriving asthma phenotypes, with indicators such as inflammatory cell counts, fractional exhaled nitric oxide (FeNO), or inflammatory cytokine measurements. Specifically, 36 phenotypes were described based on high or low levels of eosinophilic inflammatory cells. Additionally, variants of neutrophilic asthma phenotypes were less commonly reported in 10 studies. For detailed information on the number of derived phenotypes and their descriptions within studies on asthma and severe asthma, please refer to appendix.

3.3.6.2 COPD

A comprehensive analysis of studies on COPD identified a total of 57 reported phenotypes. Among these phenotypes, the most frequently reported feature for defining COPD was lung function, which was measured by spirometry and differentiated 44 phenotypes. Several other features were commonly reported in the characterization of COPD phenotypes. Age was a prominent feature, featured in 26 phenotypes. Symptoms and the frequency of exacerbations were also commonly reported, featuring in 24 phenotypes. Gender was a notable factor, featured in 17 phenotypes. Additionally, cardiovascular, metabolic, and psychiatric comorbidities were reported in 14 to 17 phenotypes. These findings indicate the importance of lung function measurements, age, symptoms, exacerbation frequency, gender, and comorbidities in the characterization of COPD phenotypes.

COPD phenotypes were categorized as mild, moderate, or severely obstructed based on spirometry lung function measures. The studies included in the analysis revealed that severe to moderately obstructed COPD phenotypes were reported in 12 studies [333], while mild obstructed COPD phenotypes were observed in 5 studies [272, 275, 278, 283, 289]. Additionally, other measures of lung function, such as lung diffusion capacity for carbon monoxide (DLCO) [269], and computed tomography (CT) measurements of lung density and airway wall thickness [271, 273, 279, 284], were employed to determine COPD phenotypes and assess accompanying emphysematous changes. Among these studies, 4 studies [271, 273, 279, 284] identified COPD phenotypes with varying levels of emphysematous changes, ranging from high to moderate to low.

Demographic and social characteristics, including age, gender, body mass index (BMI), and smoking status, were utilized to define COPD phenotypes. Among the studies examined, an elderly-related COPD phenotype was reported in 9 studies [269, 272, 274, 275, 277, 278, 280, 283, 290], while 7 studies described COPD phenotypes characterized by a young age [269, 272, 275, 277, 278, 280, 290]. Female-related variants of COPD phenotypes were reported in 3 records [269, 272, 279], while male gender-related COPD was reported in 4 studies [269, 272, 278, 283]. Body mass index (BMI) played a role in defining COPD phenotypes, with both overweight and underweight categories being associated. Obesity-related COPD was reported in two studies by Burgel et al. [272, 273],, while underweight and low weight-related COPD phenotypes were reported in 5 studies [273, 275, 277, 288, 290].

Different patterns of smoking were also linked to specific COPD phenotypes. Heavy, persistent, high-rate, or long-duration smoking-related COPD phenotypes were reported in 3 studies [272, 279, 284], while 2 studies described COPD phenotypes associated with low levels of smoking [272, 283]. In studies focused on COPD phenotyping, disease activity and severity were assessed using various measures, such as the frequency of symptoms, exacerbations, and the level of treatment. COPD phenotypes characterized by a high frequency of symptoms and exacerbations were reported in 11 studies [272, 273, 275, 277, 279, 281, 283, 285, 286, 288, 289]. On the other hand, COPD phenotypes with low symptom burden were described in 8 studies [272, 273, 275, 279, 281, 283, 285, 289]. Furthermore, four studies reported on COPD phenotypes that required high treatment doses [272, 275, 279, 284], while two studies discussed COPD phenotypes associated with low dosage treatment [272, 279]. These findings provide insights into the variability of disease activity and the level of treatment required among different COPD phenotypes.

Comorbidities played a significant role in differentiating COPD phenotypes, with cardiovascular diseases and diabetes/metabolic diseases being the most frequently reported ones [273, 274, 276, 277, 280]. Additionally, comorbidities such as depression and anxiety were considered in characterizing COPD phenotypes [272, 276, 280, 281, 285]. Various features were taken into account to characterize COPD phenotypes, including the impact on physical and daily activities, respiratory health, quality of life, and mortality. COPD phenotypes associated with impaired quality of life were reported in 3 studies [279, 286, 290], while high mortality rates were observed in 5 studies [274, 275, 278, 289, 290]. For a comprehensive

overview of the number of derived COPD phenotypes and their descriptions, please refer to appended material.

3.3.6.3 SEVERE ASTHMA

Among the reported severe asthma phenotypes, a total of 61 were identified, showcasing a notable degree of overlap between them. The features that emerged most frequently in differentiating these severe asthma phenotypes were varied and informative. Atopy, the presence of allergic tendencies, was a distinguishing factor in 28 of the phenotypes. The time at which the disease onset occurred played a significant role in 25 phenotypes, providing insights into the progression and development of severe asthma. Treatment, encompassing factors such as medication dosage or treatment steps, emerged as another crucial feature for differentiation. Additionally, inflammation measures, reflecting the underlying inflammatory processes in the airways, were influential in 14 of the phenotypes. Disease activity, as indicated by the frequency of symptoms and exacerbations, further contributed to the differentiation of severe asthma phenotypes in 14 instances. Lastly, age and gender, intrinsic characteristics of individuals, were prominent factors in 13 phenotypes. By considering these diverse features, researchers aimed to capture the multifaceted nature of severe asthma and provide a comprehensive understanding of its phenotypic variations.

Ten studies reported phenotypes of atopic severe asthma, highlighting the allergic status of individuals [295-301, 303-305]. On the other hand, non-atopic severe asthma phenotypes were reported in eight studies, indicating the absence of allergic tendencies in these cases [296-300, 303-305]. Regarding the time at disease onset, early-onset severe asthma phenotypes were described in eight studies, suggesting the condition developed at a younger age [293, 295-297, 299, 303-305].Conversely, late-onset severe asthma phenotypes were identified in seven studies, indicating that the disease manifested later in life [293, 295, 296, 299, 303-305]. These findings contribute to a better understanding of the different subtypes of severe asthma based on allergic status and the timing of disease onset.

Phenotypes of severe asthma were differentiated based on disease activity and severity, particularly in terms of symptoms. High symptom presentation was observed in six studies [293, 296, 299, 301, 302, 305], while four studies reported phenotypes with low symptoms [293, 296, 299, 301]. Medication usage was also utilized to distinguish phenotypes of severe asthma. High treatment requirements were described in six studies [292, 296, 298, 299, 301, 305], while phenotypes with low to moderate medication usage or requirements were

reported in five studies [292, 296, 298, 299, 301]. While lung function values measured with spirometry were not as commonly reported as other indicators of disease activity, they were still informative. Highly obstructed variants of severe asthma phenotypes were reported in seven studies [295-297, 299, 303-305], while moderate to mild obstructed severe asthma phenotypes were identified in five studies [292, 295, 296, 299, 305]. These findings contribute to the characterization and understanding of different phenotypes of severe asthma based on symptom presentation, medication usage, and lung function measures.

In terms of demographic characteristics, female severe asthma phenotypes were described in four studies [296, 297, 301, 305], while male severe asthma phenotypes were reported in a similar number of records [296, 297, 300, 301]. Severe asthma phenotypes related to the elderly population were described in four studies [296, 301, 302, 305], whereas phenotypes associated with young age were reported in three records [297, 302, 305]. Obesity-related variants of severe asthma phenotypes were reported in five studies [293, 296, 299, 305, 306], indicating a connection between obesity and the development or manifestation of severe asthma. For a comprehensive overview of the number of derived severe asthma phenotypes and their descriptions, please refer to appended material.

3.3.6.4 RHINITIS

A total of 45 rhinitis phenotypes were reported, and several key features were considered to differentiate between them. Gender was a prominent factor, featured in 19 of the phenotypes. Disease severity was also a significant consideration, appearing in 18 phenotypes. The impact on the quality of life was another important distinguishing feature, present in 14 phenotypes. Disease activity, specifically measured by symptoms, featured in 10 phenotypes. These features played a crucial role in characterizing and differentiating the various phenotypes of rhinitis.

Socio-demographic characteristics such as gender, age, and SES played a role in identifying various rhinitis phenotypes. Female-related rhinitis phenotypes and male-related phenotypes were reported in approximately half of the studies (n = 5) [307-309, 312, 314, 316]. Old age-related rhinitis phenotypes were described in three studies [309, 314, 316], while young age-related phenotypes were reported in the same two studies [314, 316]. SES was an important factor in differentiating rhinitis phenotypes. High SES-related rhinitis phenotypes and low SES-related phenotypes were identified in a study by Lee et al. [313]. Additionally, alcohol intake was associated with high intake-related phenotypes of rhinitis, as reported by Soler et al. [316]. These socio-demographic characteristics provided

valuable insights into the diversity of rhinitis phenotypes and their associations with gender, age, SES, and alcohol intake.

Disease activity and severity, as measured by the frequency of symptoms, classification of disease based on severity status, and medication intake, were frequently employed to differentiate rhinitis phenotypes. High symptom presentation phenotypes of rhinitis were reported in three studies [309, 311, 313]. Similarly, severe rhinitis phenotypes were identified in three studies [308, 309, 311]. Additionally, Lee et al. [313] reported rhinitis phenotypes that require high treatment doses. By considering the frequency of symptoms, severity classification, and medication intake, researchers were able to distinguish between different rhinitis phenotypes. These findings shed light on the varying degrees of disease activity and severity within the realm of rhinitis.

Various measures of respiratory function were utilized to characterize rhinitis, including CT scanning for diagnostic purposes, endoscopy scores, FeNO (fractional exhaled nitric oxide), as well as lung function spirometry, bronchodilator reversibility, and bronchial hyperresponsiveness in individuals with accompanying asthma [312, 313]. Two studies described rhinitis phenotypes with high levels of airway obstruction [312, 313]. Adnane et al. [307] and Soler et al. [314] reported rhinitis phenotypes with elevated endoscopic and CT scores. Kurukulaaratchy et al. [312] reported rhinitis phenotypes among individuals with asthma, characterized by high levels of inflammation indicated by FeNO. Furthermore, among individuals with asthma and rhinitis, Kurukulaaratchy et al. [312] and Li et al. [313] reported rhinitis phenotypes with low to moderate bronchodilator reversibility and bronchial hyperresponsiveness. These respiratory function measures provided valuable insights into the different characteristics of rhinitis phenotypes, including airway obstruction, inflammation, and their associations with asthma.

Based on disease characteristics such as time of onset, seasonality, atopy status, and accompanying nasal polyposis, different variants of rhinitis phenotypes were identified. Two studies [309, 312], reported both early and late onset variants of rhinitis phenotypes. Seasonal rhinitis phenotypes were identified in two studies [310, 312], while nasal polyposis-related phenotypes were reported in an equal number of studies [315] [307].

Atopic rhinitis variants were described in three studies [311-313], indicating the involvement of allergic factors in these phenotypes. Additionally, polysensitization rhinitis phenotypes were reported in three studies [308, 310, 311],

indicating a sensitivity to multiple allergens in these cases. These findings highlight the diverse nature of rhinitis phenotypes, considering factors such as disease onset, seasonality, atopy status, and nasal polyposis. The identification of specific variants contributes to a better understanding of the underlying mechanisms and potential treatment approaches for different rhinitis subtypes.

The impact of rhinitis on quality of life (QOL) and the presence of comorbidities were important factors considered in characterizing rhinitis phenotypes. Four studies [309, 314-316] reported rhinitis phenotypes associated with impaired QOL, highlighting the negative effect of the condition on daily functioning and well-being.

Furthermore, two studies [314, 316] identified rhinitis phenotypes with comorbidities such as depression, fibromyalgia, diabetes, and dermatitis. These comorbidities were found to be associated with specific rhinitis subtypes. For a detailed overview of the derived phenotypes and their descriptions, please refer to appended material.

3.3.6.5 ASTHMA AND COPD OVERLAP (ACO)

A total of 21 phenotypes of asthma-COPD overlap (ACO) were identified, indicating the coexistence of features from both asthma and COPD. The distinguishing features used to differentiate ACO phenotypes included smoking status, inflammation status, atopy, lung function spirometry measures, and disease activity/severity based on symptoms. Smoking status was a prominent feature in defining ACO phenotypes, with seven phenotypes reported. Inflammation status, indicative of the immune response, was considered in nine phenotypes. Atopy, which refers to an allergic predisposition, featured in seven phenotypes. Lung function evaluated by spirometry, assessing respiratory capacity, were used to categorize five phenotypes. Additionally, disease activity/severity, based on symptom presentation, differentiated five phenotypes of ACO.

In terms of socio-demographic aspects, two studies [318, 319] reported ACO phenotypes related to smoking. Female ACO phenotype and obesity related ACO were each reported in one study [317].

Lung function measures revealed the presence of highly obstructed ACO phenotypes, as reported in two studies [317, 319]. In terms of symptoms, a high symptom phenotype of ACO was identified by De Vreis et al. [317]. The inflammation status of ACO showed variations, with both eosinophilic

variants reported in two studies [317-319], along with the presence of neutrophilic ACO phenotypes in one study [317]. Regarding other disease characteristics, early onset ACO phenotypes were reported in one study [318], suggesting the manifestation of ACO at a younger age. Additionally, three records reported the existence of atopic ACO phenotypes [317-319]. For a comprehensive overview of the derived phenotypes of COPD and ACO and their descriptions, please refer to appended material.

3.4 SOCIOECONOMIC STATUS AND DIFFERENT FORMS OF RHINITIS IN SWEDISH ADULTS

3.4.1 STUDY PARTICIPANTS' CHARACTERISTICS

The study's demographic breakdown indicates that the majority of participants were 30 or older, predominantly female, with 14% obese, 36% overweight, and the rest 50% having normal weight. Roughly one-third reported a family history of allergy or asthma, with a similar number having spent their childhood in rural areas, and 12% raised on a farm. Smoking habits varied, with 25% former smokers, 11% current smokers, and some exposed to secondhand smoke at home (16%) or work (9%). Participants with primary education tended to be older males, less obese, and less exposed to secondhand smoke or vapors, gases, dusts and fumes (VGDF) compared to those with higher education levels. Those in the lowest occupational skill categories were often females aged between 30-60, with a lower family history of allergies and were typically never smokers.

3.4.2 PREVALENCE OF FORMS OF RHINITIS BY SES INDICATORS

The study found that 28% of participants had allergic rhinitis, 19% had chronic rhinitis, and 2.7% had chronic rhinosinusitis.

3.4.2.1 STRATIFICATION BY SEX, EDUCATION, AGE, AND CHILDHOOD UPBRINGING:

Education

Education plays a significant role in the prevalence of allergic rhinitis, with those holding secondary and tertiary education levels experiencing higher rates

than those with only primary education. This pattern holds true across genders. Males with secondary education and females with secondary or tertiary levels show higher chronic rhinitis prevalence. Chronic rhinosinusitis is more common in females with primary or secondary education but not in males. Occupational status influences the prevalence of rhinitis in males, with higher rates in those with more prestigious jobs. In contrast, females show no such correlation. For both genders, individuals in higher occupational roles experience less chronic rhinitis and rhinosinusitis than those in lower roles. Full results are shown in (**figure 13 and 14**) below.

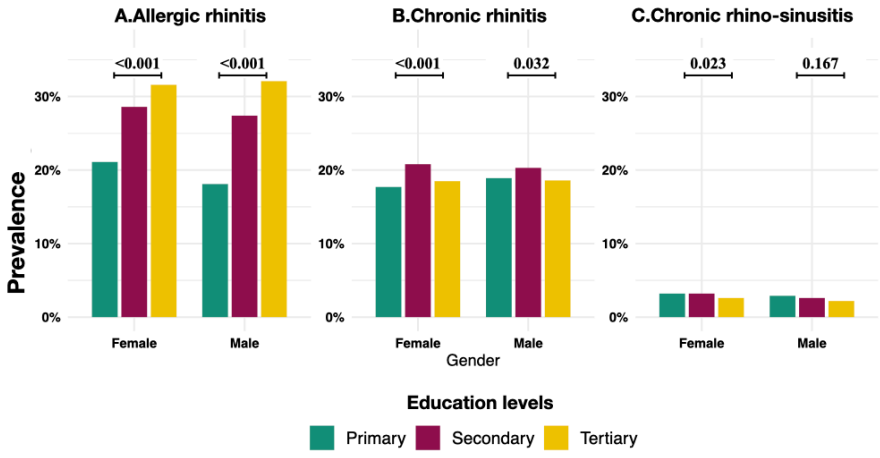


Figure 13. Variation in the prevalence of forms of rhinitis by education levels among males and females presented as percentages and p values for comparison between education groups.

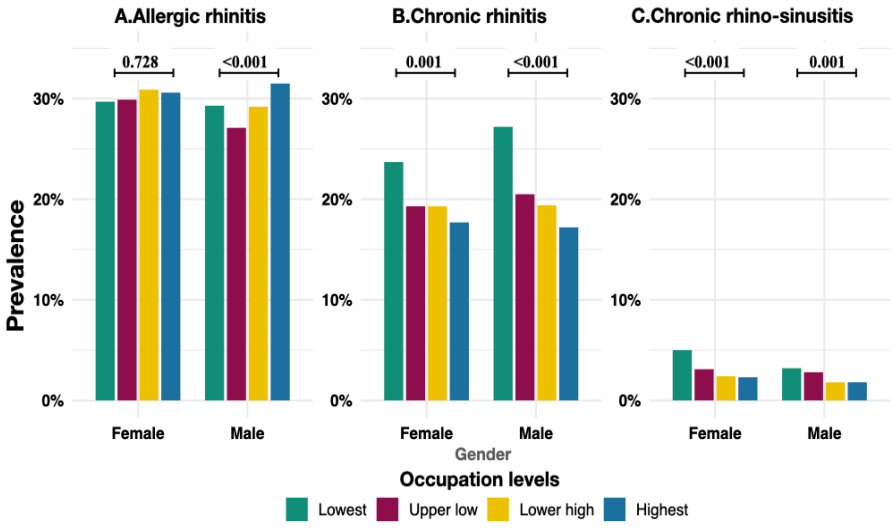


Figure 14. Prevalence of forms of rhinitis by occupation levels among males and females with p value for between occupation groups comparison.

Allergic rhinitis is more common in individuals with higher education and occupational status across different age groups. Those under 60 with tertiary education have less chronic rhinitis, and those aged 30 to 60 with higher education also have less chronic rhinosinusitis. Being raised on a farm affects the prevalence of rhinitis conditions. Higher education levels are associated with increased allergic rhinitis regardless of farm upbringing, while chronic rhinitis is more prevalent in those with higher education but not necessarily tied to occupational status or farm upbringing. Chronic rhinosinusitis rates are higher among those with the lowest educational level and occupational status, particularly for those not raised on farms.

These findings are detailed in **Figure 15** (**panels A, B, and C**) in the main text and further supported by (**Tables S1, S2 and S3**) in the appendix.

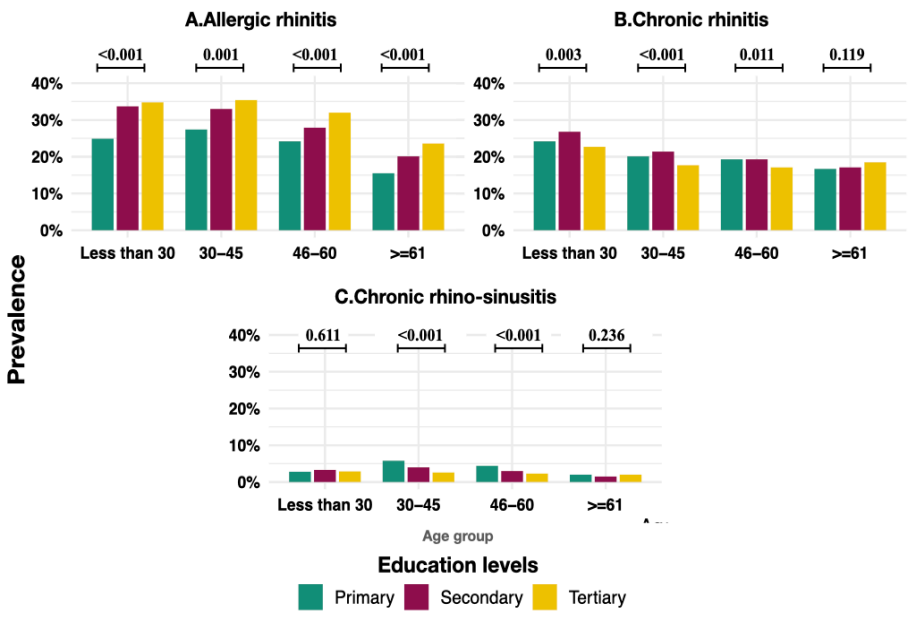


Figure 15. Prevalence of forms of rhinitis by education levels among age groups with the P value for between education groups comparison.

3.4.3 ASSOCIATION BETWEEN SES AND FORMS OF RHINITIS

3.4.3.1 EDUCATION

After adjusting for covariates, compared to primary education, secondary and tertiary education were associated with an increased risk of allergic and chronic rhinitis, but not chronic rhinosinusitis (Figure 16).

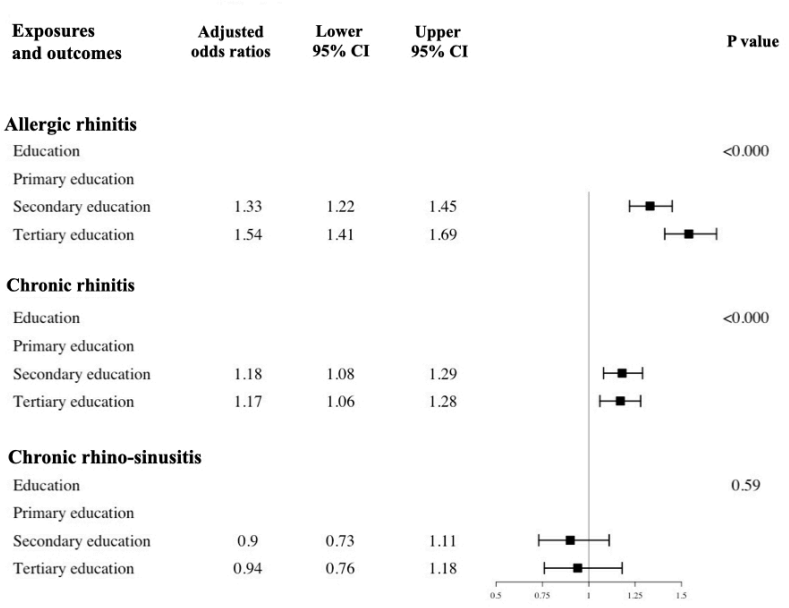


Figure 16. The odds ratios and their 95% confidence intervals for the association between education and forms of rhinitis

3.4.3.2 OCCUPATION

After accounting for various covariates, individuals in the two highest occupational skill groups were found to have a higher risk of allergic rhinitis compared to those in the lowest occupational skill group. However, no significant relationship was observed between occupational skill level and the incidence of chronic rhinitis or chronic rhinosinusitis, as depicted in (**Figure 17).**

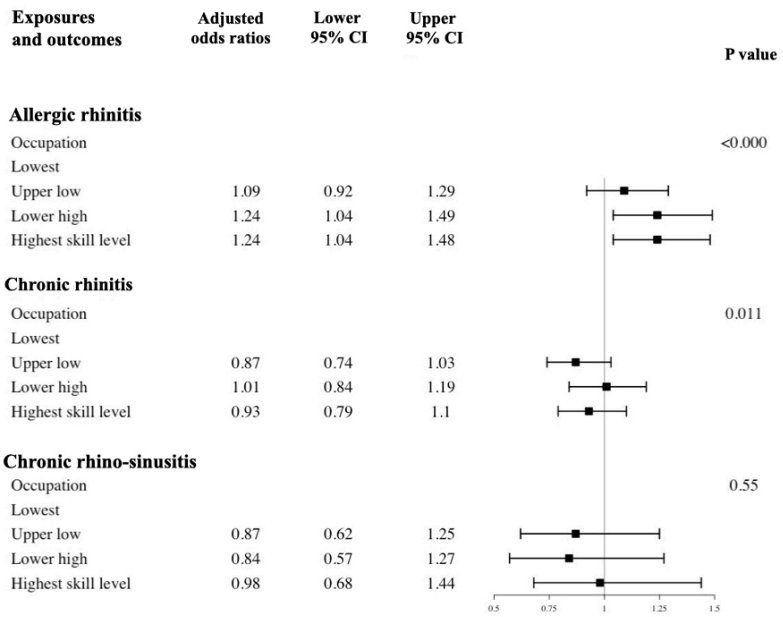


Figure 17. The odds ratios and their 95% confidence intervals for the association between occupation, and forms of rhinitis.

3.4.4 INTERACTION BETWEEN SOCIOECONOMIC STATUS AND SUBJECTS' CHARACTERISTICS IN RELATION TO THE RISK OF FORMS OF RHINITIS

Interactions between education level and factors like age, sex, and BMI were noted in their effects on the risk of certain types of rhinitis. Specifically, males with tertiary education exhibited a more pronounced increased risk for allergic rhinitis (Odds Ratio [OR] 2.03, 95% Confidence Interval [CI] 1.92-2.14) compared to females with the same level of education (OR 1.56, 95% CI 1.37-1.76), as shown in Table S4 and Figure S1. The increased risk associated with tertiary education for chronic rhinitis was significant in individuals over 60 years old (OR 1.38, 95% CI 1.21-1.58) and those aged 46-60 (OR 1.16, 95% CI 1.00-1.36), but not in the 30-45 age group (OR 1.01, 95% CI 0.76-1.36) nor in those under 30 (OR 0.92, 95% CI 0.74-1.16), detailed in Table S4 and Figure S1. Moreover, the increased risk of chronic rhinitis for those with tertiary education was evident only in obese (OR 1.22, 95% CI 1.00-1.42) and

overweight individuals (OR 1.32, 95% CI 1.15-1.52), not in those with normal weight (OR 1.03, 95% CI 0.95-1.11), as outlined in **Tables S4** and **S5** and **Figure S1** in the appendix.

3.5 MACHINE LEARNING-DERIVED ASTHMA PHENOTYPES IN A REPRESENTATIVE SWEDISH ADULT POPULATION

3.5.1 BASELINE CHARACTERISTICS OF ASTHMA VERSUS NON-ASTHMA SUBJECTS IN THE SAMPLE

Out of 3,101 participants in the WSAS clinical studies before the COVID-19 pandemic, 1,206 had never experienced asthma, while the other 1,895, who had a history of asthma, were selected for phenotype analysis. The characteristics comparing individuals with and without a history of asthma are detailed in (**Table 7**). Those with a history of asthma were generally younger, had a higher female representation, poorer lung function, more inflammation, increased respiratory symptoms, and a higher incidence of COPD. They also had a greater family history of asthma/allergy but less rhinitis and lower levels of atopy than those without a history of asthma.

Table 7. Study characteristics comparing subjects with ever asthma and non-asthma subjects.

Variables	Ever asthma	No asthma	P value
n	1895	1206	
Demographics			
Age, years, mean (SD)	49.72 (15.72)	51.68 (15.41)	0.001
Women, n (%)	1131 (59.7)	636 (52.7)	<0.001
BMI, kg/m^2, mean (SD)	27.06 (4.88)	25.91 (4.21)	<0.001
Social aspect			
Smoking status, n (%)			0.972
Never smoker	1029 (54.3)	653 (54.1)	
Former smokers	699 (36.9)	449 (37.2)	
Current smoker	167 (8.8)	104 (8.6)	
Smoking exposure at home, n (%)	1176 (62.1)	741 (61.4)	0.76
Smoking exposure at work, n (%)	746 (39.4)	494 (41.0)	0.397
Indoor household damage due to damp, n (%)	238 (12.6)	131 (10.9)	0.172
Lung function			
FEV_1/FVC ratio (post-BD), mean (SD)	0.77 (0.10)	0.80 (0.07)	<0.001
FEV_1 reversibility (%), mean (SD)	7.3 (7.7)	3.8 (3.9)	<0.001
TLCO (% predicted), mean (SD)	100.1 (18.2)	99.30(23.0)	0.255
KCO (% predicted), mean (SD)	99.0 (15.4)	94.9 (14.0)	<0.001
Measures of inflammation			
FeNO, ppb, mean (SD)	25.5 (22.4)	19.2 (11.4)	<0.001
Blood eosinophil count(x10^9/L), mean (SD)	0.22 (0.16)	0.17 (0.11)	<0.001
Blood neutrophil count(x10^9/L), mean (SD)	3.86 (1.38)	3.60 (1.28)	<0.001
Respiratory symptoms			
Breathlessness, n (%)	1277 (67.4)	155 (12.9)	<0.001
Recurrent wheeze, n (%)	413 (21.8)	40 (3.3)	<0.001
Cough, n (%)	638 (33.7)	175 (14.5)	<0.001
Dyspnea, n (%)	413 (21.8)	52 (4.3)	<0.001
Chronic productive cough, n (%)	421 (22.2)	85 (7.0)	<0.001
Atopy			
Allergic sensitization, n (%)	1248 (65.9)	909 (75.4)	<0.001
Family history of asthma or allergy, n (%)	929 (49.0)	335 (27.8)	<0.001
Comorbidities			
Ever rhinitis, n (%)	1293 (68.2)	917 (76.0)	<0.001
Nasal polyposis with rhinitis, n (%)	822 (43.4)	546 (45.3)	0.317
COPD [a], n (%)	240 (12.7)	34 (2.8)	<0.001

*BMI = Body mass index, BD= Bronchodilator, Chronic obstructive pulmonary disease, FeNO = Fractional exhaled nitric oxide measured in (ppb), FEV_1= Forced expiratory volume in 1 second, FVC= Forced vital capacity, KCO = Carbon monoxide transfer coefficient, SD= Standard deviation, TLCO =Transfer factor of carbon monoxide uptake measured in (mmol/min/kPa), COPD = Chronic obstructive pulmonary disease.

[a] COPD based on self-report and spirometry measures.

3.5.2 DERIVED ASTHMA PHENOTYPES AND THEIR NAMES

Utilizing DEC computational metrics, the data initially suggested the optimal categorization was into three clusters. However, after a thorough clinical assessment, a four-cluster model was deemed to provide a more clinically and scientifically accurate and diverse representation of asthma phenotypes. Therefore, the decision was made to adopt the four-cluster categorization. The clusters have been described as follows:

1. Phenotype 1 (comprising 24.2% of the sample, with 458 individuals): late-onset asthma, predominantly in non-atopic women and individuals who smoke.

2. Phenotype 2 (28.7% of the sample, with 545 individuals): Female-dominated early adult-onset asthma

3. Phenotype 3 (18.9% of the sample, with 358 individuals): Adult-onset asthma with high inflammation

4. Phenotype 4 (28.2% of the sample, with 534 individuals): Early-onset, mild, atopic asthma

The distinctions between these clusters are further elucidated in (**Table 8**).

Table 8. Comparison of demographics and clinical aspects between the derived asthma phenotypes

	Phenotype 1[1]	Phenotype 2[2]	Phenotype 3[3]	Phenotype 4[4]	P value
n	458	545	358	534	
Demographics					
Age, years, mean (SD)	65.7 (9.2)	44.6 (12.2)	48.7(14.6)	41.7(13.9)	<0.001
Sex, n (%)					<0.001
Men, n (%)	168 (36.7)	158 (29.0)	191 (53.4)	247 (46.3)	
Women, n (%)	290 (63.3)	387 (71.0)	167 (46.6)	287 (53.7)	
BMI, kg/m^2, mean (SD)	28.5 (4.9)	27.5 (5.1)	26.4(4.3)	25.7 (4.4)	<0.001
Obesity, n (%)	151 (33.0)	146 (26.8)	70 (19.6)	83 (15.5)	<0.001
Age at asthma onset, years, mean (SD)	35.1(19.3)	17.7(12.5)	21.4(17.3)	13.6 (10.7)	<0.001
Allergy and atopy					
Family history of allergy or asthma, n (%)	123 (26.9)	326 (59.8)	188 (52.5)	292 (54.7)	<0.001
Allergic sensitization, n (%)	180 (39.3)	369 (67.7)	277 (77.4)	422 (79.0)	<0.001
Social status					
Smoking status (%)					<0.001
Never smoker	151 (33.0)	342 (62.8)	212 (59.2)	324 (60.7)	
Former smokers	230 (50.2)	174 (31.9)	122 (34.1)	173 (32.4)	
Current smoker	77 (16.8)	29 (5.3)	24 (6.7)	37 (6.9)	
Smoking exposure at home, n (%)	346 (75.5)	324 (59.4)	229 (64.0)	277 (51.9)	<0.001
Smoking exposure at work, n (%)	267 (58.3)	180 (33.0)	140 (39.1)	159 (29.8)	<0.001
Drug induced dyspnea, n (%)	38 (8.3)	33 (6.1)	13 (3.6)	24 (4.5)	0.017
Asthma triggers, n (%)					
Exertion triggered symptoms	282 (61.6)	513 (94.1)	282 (78.8)	110 (20.6)	<0.001
Infection triggered symptoms	294 (64.2)	512 (93.9)	273 (76.3)	76 (14.2)	<0.001
Allergen triggered symptoms	105 (22.9)	243 (44.6)	148 (41.3)	58 (10.9)	<0.001
Irritant triggered symptoms	59 (12.9)	143 (26.2)	67 (18.7)	14 (2.6)	<0.001
Weather triggered symptoms	127 (27.7)	315 (57.8)	146 (40.8)	17 (3.2)	<0.001

Respiratory symptoms, n (%)					
Cough, n (%)	235 (51.3)	204 (37.4)	102 (28.5)	97 (18.2)	<0.001
Recurrent wheeze, n (%)	159 (34.7)	106 (19.4)	103 (28.8)	45 (8.4)	<0.001
Breathlessness, n (%)	328 (71.6)	415 (76.1)	315 (88.0)	219 (41.0)	<0.001
Dyspnea, n (%)	205 (44.8)	109 (20.0)	52 (14.5)	47 (8.8)	<0.001
Chronic productive cough, n (%)	180 (39.3)	122 (22.4)	67 (18.7)	52 (9.7)	<0.001
Lung function					
KCO (% predicted) (mean (SD))	85.05 (14.30)	105.5 (12.86)	103.2 (13.62)	101.34 (12.27)	<0.001
TLCO (% predicted) (mean (SD))	82.81 (14.55)	106.5 (17.81)	105.5 (14.27)	104.91 (13.75)	<0.001
Severe asthma [a], n (%)	123 (26.9)	96 (17.6)	68 (19.0)	20 (3.7)	<0.001
GINA* control level (%)					<0.001
Controlled	120 (26.2)	178 (32.7)	113 (31.6)	306 (57.3)	
Partially controlled	236 (51.5)	292 (53.6)	181 (50.6)	183 (34.3)	
Uncontrolled	102 (22.3)	75 (13.8)	64 (17.9)	45 (8.4)	
GINA* treatment steps (%)					<0.001
Only SABA or non-users	299 (65.3)	315 (57.8)	214 (59.8)	455 (85.2)	
Step 1 or 2	69 (15.1)	130 (23.9)	83 (23.2)	55 (10.3)	
Step 3	17 (3.7)	39 (7.2)	22 (6.1)	18 (3.4)	
Step 4 or 5	73 (15.9)	61 (11.2)	39 (10.9)	6 (1.1)	
Burden of asthma					
Hospitalization last year, n (%)	56 (12.2)	15 (2.8)	10 (2.8)	10 (1.9)	<0.001
Emergency visit, n (%)	130 (28.4)	81 (14.9)	46 (12.8)	49 (9.2)	<0.001
Co-existing COPD [b], n (%)	154 (33.6)	43 (7.9)	26 (7.3)	17 (3.2)	<0.001

[1] Troublesome late-onset, non-atopic asthma with more women and smokers.
[2] Female-dominated early adult-onset asthma.
[3] Adult-onset asthma with high inflammation.
[4] Early-onset, mild, atopic asthma.
*BMI = Body mass index, BD= Bronchodilator, Chronic obstructive pulmonary disease, FeNO= Fractional exhaled nitric oxide, FEV_1= Forced expiratory volume in 1 second, FVC= Forced vital capacity, KCO = Carbon monoxide transfer coefficient, SD= Standard deviation, TLCO = Transfer factor of carbon monoxide uptake.
[a] Severe asthma defined based on medication intake.
[b] COPD based on self-report and spirometry measures.

3.5.3 CHARACTERISTICS OF THE DERIVED ASTHMA PHENOTYPES

3.5.3.1 DERIVED ASTHMA PHENOTYPES BY DEMOGRAPHICS, RISK FACTORS AND TRIGGERS

Demographic details, risk factors, and asthma triggers varied across the identified asthma phenotypes. Cluster 1 members had the latest average onset age of asthma at 35 and the highest calendar age and BMI, with a notable proportion being smokers with a significant pack-year history, and more exposure to smoking both at home and work compared to other clusters. Cluster 2, with a relatively young average onset age, predominantly consisted of women and reported more household mold, plastic-carpet, or water damage, as well as the highest instances of exertion and infection-induced asthma. Cluster 3, with the second-oldest onset age, had a higher male composition. In contrast, Cluster 4 individuals had the earliest onset at age 13, the lowest calendar age and BMI, and were least exposed to smoking, with the lowest instances of exertion and infection as triggers. (**Table 8** and **Figure 18: A**).

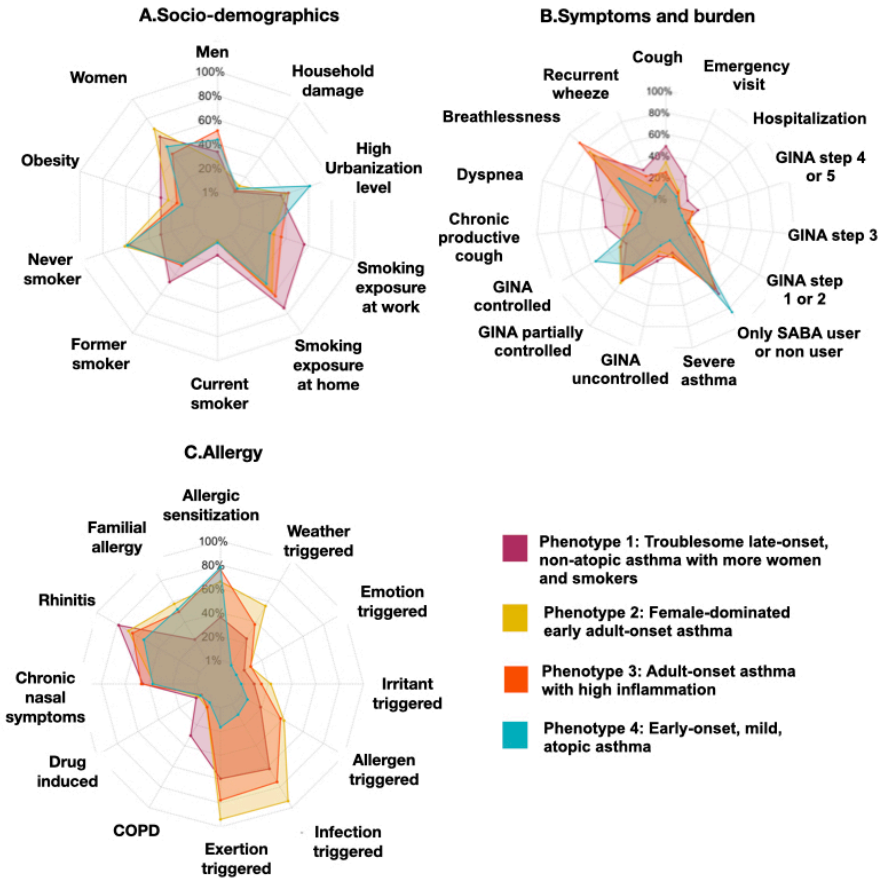


Figure 18. Variations in demographic and clinical variables between derived asthma phenotypes

3.5.3.2 DERIVED ASTHMA PHENOTYPES BY MARKERS OF INFLAMMATION

Concerning markers of inflammation, Cluster 3 had the highest levels of FeNO and eosinophils, while Cluster 1 had the highest neutrophil counts. Cluster 4 presented the lowest FeNO levels and neutrophil counts, with no substantial differences in FeNO levels among Clusters 1, 2, and 4. (**Figure 19)**.

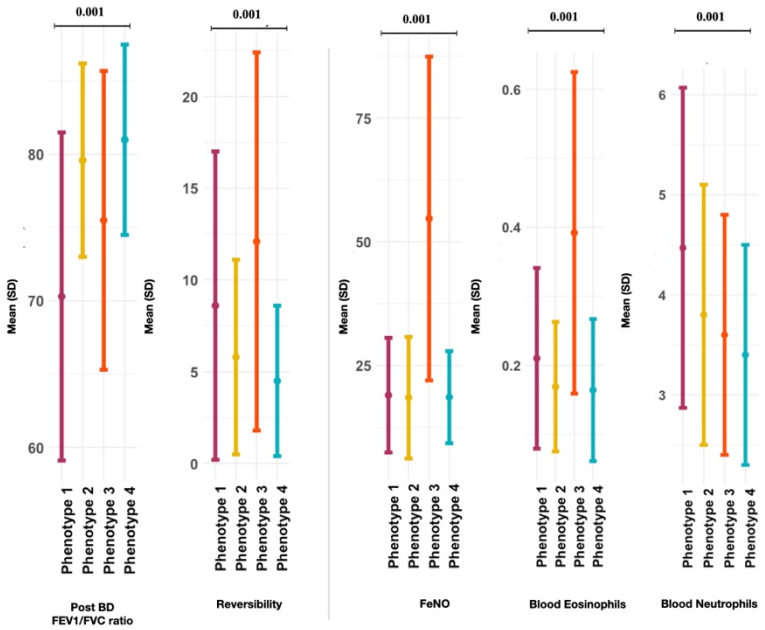


Figure 19. Distribution of lung function and inflammatory measures between derived asthma phenotypes

3.5.3.3 DERIVED ASTHMA PHENOTYPES BY SYMPTOM PROFILES, ALLERGIC STATUS, AND COMORBIDITIES

Symptom profiles, allergic status, and comorbidities also differed; Cluster 1 had the most respiratory symptoms, nasal polyps, rhinitis, and COPD cases but the least allergic sensitization. Cluster 4 experienced fewer respiratory symptoms had the highest allergic sensitization, and the fewest COPD cases, with Clusters 2 and 3 falling in between. (**Table 8** and **Figure 18: B and C**).

3.5.3.4 DERIVED ASTHMA PHENOTYPES BY LUNG FUNCTION

Lung function was poorest in Cluster 1, with the lowest FEV1/FVC ratios, whereas Cluster 4 had the best lung function and the lowest reversibility after bronchodilator. Clusters 2 and 3 had intermediate lung function measures. (**Figure 19**).

3.5.3.5 DERIVED ASTHMA PHENOTYPES BY DISEASE PROGNOSIS AND SEVERITY

In terms of disease prognosis and severity, Cluster 1 had the highest number of individuals with uncontrolled asthma and those at advanced GINA treatment stages, while Cluster 4 had the lowest. (**Table 8** and **Figure 18: B**).

3.5.3.6 DESCRIPTION OF DERIVED ASTHMA PHENOTYPES BY CURRENT ASTHMA

The prevalence of asthma in Cluster 1 was almost universal among members, contrasting with less than 80% in Cluster 4, indicating that current asthma was more common in clusters with later-onset ages. **(Figure 20).**

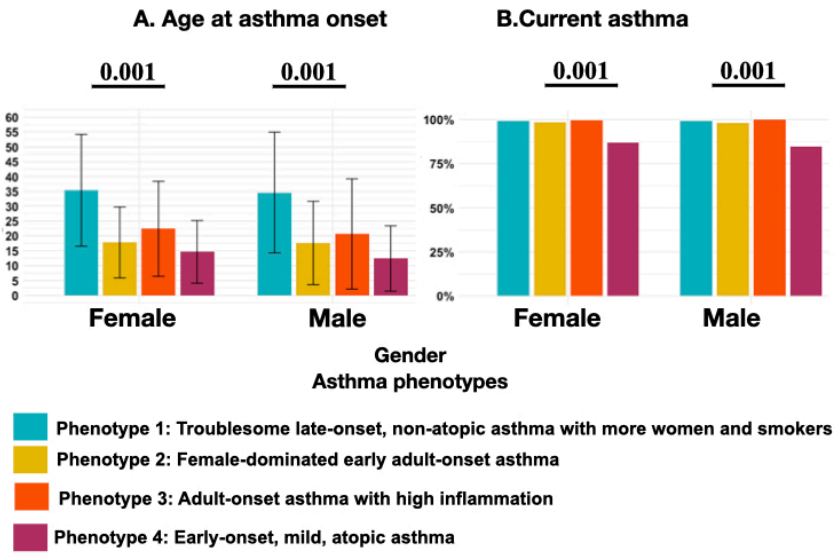


Figure 20. Prevalence of current asthma by asthma phenotypes and mean age at asthma onset, both among men and women.

3.6 LEVEL OF EDUCATION, BUT NOT OCCUPATION, IS DIFFERENTIALLY ASSOCIATED WITH ASTHMA PHENOTYPES IN ADULTS

3.6.1 PREVALENCE OF DIFFERENT ASTHMA PHENOTYPES ACROSS EDUCATIONAL AND OCCUPATIONAL LEVELS FOR BOTH MALES AND FEMALES

Among the study sample of all subjects who underwent clinical investigation, for males, the prevalence of troublesome late-onset, non-atopic asthma with smoking increases with educational level, being highest in the tertiary group. In contrast, the prevalence of early-onset, mild, atopic asthma is highest in secondary education. Those with no asthma mostly fall into the primary and tertiary education categories. For females, a similar pattern is observed for troublesome late-onset, non-atopic asthma with smoking, with the highest prevalence in the tertiary group. Early adult-onset asthma is most common in those with secondary education, and the proportion of females with no current asthma is fairly even across all education levels.

Looking at occupation, for males, the highest proportion of troublesome late-onset asthma is in the upper low occupational group, while the highest proportion of those with no asthma is in the highest occupational group. For females, the late-onset asthma is more prevalent in the lowest occupational group, and the no asthma group is more evenly distributed, with the highest prevalence in the lower high occupational category. (Table **9**)

Table 9. Prevalence of different asthma phenotypes and no asthma status across educational and occupational levels for both males and females and p values for between education and occupation groups comparison.

	Male				Female			
	Primary	Secondary	Tertiary	P	Primary	Secondary	Tertiary	P
n	573	573	188		918	649	200	
Phenotype				<0.001				<0.001
Troublesome late-onset, non-atopic asthma with	45 (7.9)	60 (10.5)	63 (33.5)		108 (11.8)	103 (15.9)	79 (39.5)	
Female-dominated early adult-onset asthma	79 (13.8)	75 (13.1)	4 (2.1)		228 (24.8)	140 (21.6)	19 (9.5)	
Adult-onset asthma with high inflammation	9 3 (16.2)	79 (13.8)	19 (10.1)		93 (10.1)	59 (9.1)	15 (7.5)	
Early-onset, mild, atopic asthma	108 (18.8)	120 (20.9)	19 (10.1)		156 (17.0)	113 (17.4)	18 (9.0)	
No asthma	248 (43.3)	239 (41.7)	83 (44.1)		333 (36.3)	234 (36.1)	69 (34.5)	

	Male					Female				
	Lowest	Upper low	Lower high	Highest	p	Lowest	Upper low	Lower high	Highest	p
n	24	428	216	276		56	608	216	439	
Phenotype					0.012					0.013
Troublesome late-onset, non-atopic asthma with smoking	2 (8.3)	65 (15.2)	17 (7.9)	28 (10.1)		13 (23.2)	107 (17.6)	15 (6.9)	53 (12.1)	
Female-dominated early adult-onset asthma	3 (12.5)	58 (13.6)	20 (9.3)	41 (14.9)		11 (19.6)	135 (22.2)	50 (23.1)	109 (24.8)	
Adult-onset asthma with high inflammation	3 (12.5)	70 (16.4)	39 (18.1)	30 (10.9)		3 (5.4)	60 (9.9)	22 (10.2)	43 (9.8)	
Early-onset, mild, atopic asthma	5 (20.8)	73 (17.1)	44 (20.4)	38 (13.8)		8 (14.3)	99 (16.3)	32 (14.8)	65 (14.8)	
No asthma	11 (45.8)	162 (37.9)	96 (44.4)	139 (50.4)		21 (37.5)	207 (34.0)	97 (44.9)	169 (38.5)	

3.6.2 ASSOCIATION BETWEEN EDUCATION AND OCCUPATION GROUPS AND ASTHMA PHENTYPES

Individuals with secondary education had a higher likelihood (Odds Ratio [OR] 1.61, 95% Confidence Interval [CI] 1.26-2.07) of being classified within the first asthma phenotype compared to those without asthma, relative to individuals with primary education. In contrast, individuals with either

secondary education (OR 0.44, 95% CI 0.29-0.66) or tertiary education (OR 0.44, 95% CI 0.29-0.66) were less likely to be classified within the second asthma phenotype compared to the non-asthma group, as opposed to those with primary education. Moreover, those with tertiary education were also less likely (OR 0.76, 95% CI 0.59-0.97) to belong to the fourth asthma phenotype compared to the non-asthma group than those with primary education. There was no significant educational correlation with the third asthma phenotype. Additionally, occupational status showed no association with any asthma phenotypes. (**Figure 21**)

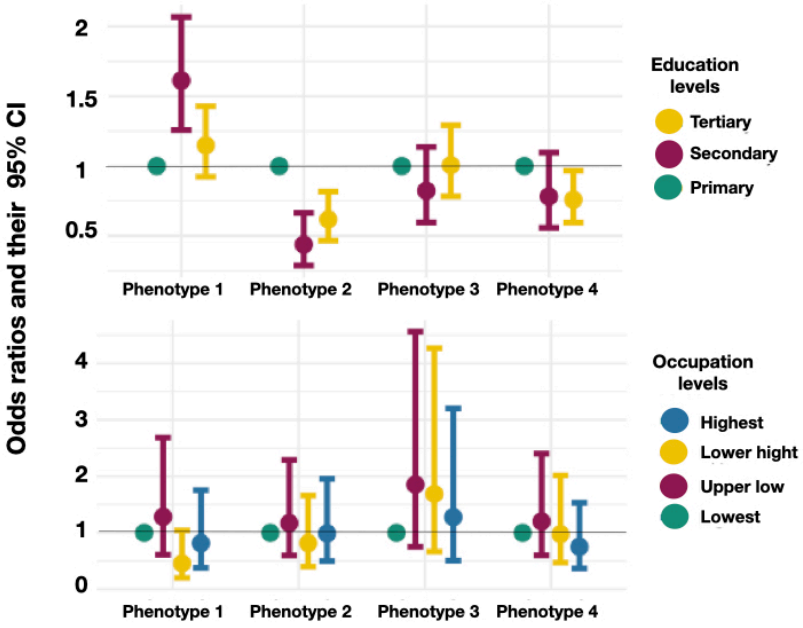


Figure 21. Association of education and occupation levels with risk of asthma phenotypes

4 DISCUSSION

4.1 DISCUSSION OF METHODS

4.1.1 VALIDITY AND RELIABILITY OF THE STUDY RESULTS

In research, the term "validity" pertains to the accuracy of the results obtained in measuring what they are intended to measure. Within this context, two types of validity are recognized. Internal validity refers to the extent to which the results derived from the study samples accurately reflect and represent the broader study population. On the other hand, external validity concerns the applicability of the study's results to other populations, indicating its potential for generalizability. "Reliability" refers to the consistency of the study's outcomes with those of other studies. Several factors impact these dimensions within a study setting, including representativeness, biases, confounding variables, consideration of medications, and other factors related to data processing [334].

Regarding the representativeness of our study samples, the West Sweden Asthma Study (WSAS) employed an initial postal cross-sectional survey questionnaire that covered a substantial area of Southwest Sweden. Both WSAS1 data in Paper 1,5 and 6, and WSAS 2 data in Paper 2 and 4 encompassed large sample sizes that were carefully stratified, thus ensuring a high degree of internal validity in the study findings. Furthermore, WSAS 1 and 2 utilized validated standard questionnaires that have been adopted and employed in other studies as well [14, 335-337]. The utilization of these standardized questionnaires not only facilitates the comparison of results within the study itself but also enhances the comparability of the study's outcomes with those of other research endeavors.

4.1.2 SELECTION BIAS

Bias represents a systematic error that can permeate study design and undermine result accuracy. Various forms of bias exist, with one prevalent type being selection bias. This occurs when the sampling process inadvertently favors the inclusion or exclusion of specific subject groups, leading to skewed estimates of the targeted research question.

In the context of our study, WSAS1 data utilized in Paper 1,5 and 6 stemmed from an initial postal survey questionnaire. Impressively, the response rate for the postal questionnaire reached a substantial 62%. Within this respondent group, a further subset was invited to undergo clinical examinations. An analysis of non-responders in the WSAS 1 postal survey indicated a trend toward non-response among younger males, smokers, individuals residing in the Gothenburg metropolitan area, and those employed full-time. Participants in the clinical survey also shared similarities with postal survey responders, despite no observed differences in the prevalence of respiratory symptoms and diseases [338]. However, potential underrepresentation of factors like age, smoking, and occupational exposures might be a risk in our study sample due to this bias [339, 340].

Regarding WSAS 2 data used in Paper 2 and 4, a non-response analysis was not conducted. Nevertheless, a robust response rate of 50% among those invited to answer the postal questionnaire implies a high level of representation in the study sample. Additionally, in WSAS1, the study sample for Paper 1 consisted of subjects from a randomly selected group of 2000 individuals, combined with 1524 individuals with asthma identified through postal survey responses. Response rates were 59% for the random sample and 55% for the asthmatics group, totaling 2006 subjects in the clinical survey cohort. Notably, caution should be exercised when extrapolating results from this group, as asthmatics might be more represented than the general population. For instance, the WSAS 1 clinical sample of adults with asthma comprises a blend of randomly selected individuals and asthmatics not randomly chosen from postal survey responders.

Turning to OLIN 2016 cohort data utilized in Paper 2, a non-responding study within the same cohort revealed a similar trend to the WSAS 1 postal cohort—non-responders were more likely to be young non-smoking males. However, analyses from occupational exposure literature show that such non-response patterns are common and represent a homogenous group, thus not introducing bias into the study sample [339, 340].

Considering data from the SAAS study in Paper 1, it is important to acknowledge the distinction that this study design wasn't grounded in the general population; subjects were recruited from clinical hospital settings. Consequently, there could be an overrepresentation of subjects with poor control and severe asthma. Despite variations in residual confounders across

cohorts, the effect remained consistent across all three study samples. Refer to (**Table 14**) for comprehensive results.

Table 14 Baseline characteristics of subjects with adult-onset asthma in each study sample and the pooled study sample.

Characteristic	SAAS	OLIN	WSAS	P value	Pooled sample
Subjects (n)	200	511	185		896
Females, n (%)	117(58.5)	308(60.3)	118 (63.8)	0.555	543 (60.6)
Age, mean (SD)	58(13)	61(12)	53(13)	**<0.001**	59 (13)
Age, range	28–90	32–92	23–76		23–92
BMI, n (%)				**0.006**	
<25	58 (29.0)	102 (20.0)	57 (30.8)		217 (24.2)
25–29.99	71 (35.5)	211 (41.3)	76 (41.1)		358 (40.0)
≥30	71 (35.5)	198(38.7)	52 (28.1)		321 (35.8)
Smoking status, n (%)				**0.010**	
Nonsmoker	95 (47.5)	248(48.5)	96 (52.5)		439 (49.1)
Ex-smoker	76 (38.0)	219(42.9)	58 (31.7)		353 (39.5)
Current smoker	29 (14.5)	44 (8.6)	29 (15.8)		102 (11.4)
Uncontrolled asthma (ACT ≤ 19)	56 (28.0)	136(26.6)	51 (27.6)	0.922	243 (27.1)
Prebronchodilator FEV1 (% predicted) a	85 (18)	86 (17)	93 (16)	**<0.001**	87 (17)
Daily ICS user, n (%)	155(77.5)	239(46.8)	146 (78.9)	**<0.001**	540 (60.3)
Atopic, n (%) b	67 (33.5)	139(62.9)	93 (68.9)	**<0.001**	299 (55.8)
Postbronchodilator FEV1/FVC < 0.7 and ≥10 pack years, n (%)	34 (17.2)	78(15.6)	19 (10.3)	0.127	131 (14.8)
Education, n (%)				**<0.001**	
Tertiary	30 (14.8)	149(29.2)	79 (42.7)		258 (28.8)
Secondary	95 (47.5)	168(32.9)	62 (33.5)		325 (36.3)
Primary	75 (37.5)	194(38.0)	44 (23.8)		313 (4.9)

4.1.2.1 HEALTHY WORKER EFFECT

Regarding **Paper 2 and 4**, where occupational exposures were considered, it is essential to underscore the significant impact of the healthy worker effect as a potential bias source. The healthy worker effect (HWE) is a phenomenon where the health status of individuals influences the distribution of both healthy and diseased individuals within an occupational population differently from how they are distributed in the broader general population. In simpler terms, the HWE leads to an occupational population that is generally healthier than the general population due to those who are unwell transitioning out of the workforce through retirement or other means.

Various aspects contribute to the HWE phenomenon. Firstly, there is the notion of "health employment," where certain occupations have higher health requirements for entry, resulting in a workforce comprised of individuals who are healthier and better able to handle physically demanding tasks. Consequently, these jobs exhibit a healthier employee population than the general populace. Secondly, survival HWE arises when individuals afflicted by disease opt to change careers, shift jobs, or retire prematurely due to their inability to cope with the working conditions. This choice further skews the representation of health within the occupational group. Lastly, the concept of workplace advantage is another facet, where specific occupations implement regular health assessments or provide health-related perks, leading to a healthier workforce compared to the norm.

In the context of **Paper 2 and 4**, the samples analyzed were drawn from the general population, which could potentially mitigate the influence of the HWE. Nevertheless, caution should be exercised, as the exclusion of very young and older age groups might inadvertently introduce bias, as these groups might be excluded due to unemployability, thus exacerbating the HWE bias.

4.1.3 RECALL AND INFORMATION BIAS

Recall bias pertains to discrepancies in accurately remembering past exposures or activities, constituting a systematic error that can result from participants or investigators either overemphasizing or downplaying their responses to questions relating to the study's subject. This concern is particularly pertinent when self-report questionnaires and interview methods are employed, as observed in both Paper 1,5 and 6, using clinical survey data , and Paper 2 and 4 [334] using postal survey data. In the context of Paper 1 where the outcome was asthma control status, this issue assumes heightened importance. The study sample was derived from a clinical setting (SAAS), in contrast to one study cohort (WSAS and OLIN) drawn from the general population. This discrepancy introduces the potential for a greater risk of encountering recall bias, with subjects from hospital environments potentially more inclined to report heightened risk factors. However, the study's findings yielded consistent results across the different study cohorts, effectively alleviating apprehensions related to this plausible bias. Further, in a recent published work by Nurami et al. there was high agreement between self reported and registered age of asthma onset [341].

4.1.4 CAUSAL INFERENCE USING DIRECTED ACYCLIC GRAPHS (DAGS)

In the domain of applied health research, clinical trials are frequently regarded as the benchmark method for establishing causal relationships between variables in real-world contexts. However, ethical limitations and considerations arise when investigating risk factors, as only specific exposures can be ethically explored [342]. This necessitates the emergence of observational studies as an alternative avenue to elucidate intended causal associations. Nevertheless, these studies encounter a host of challenges encompassing biases, data-related intricacies, and methodological complexities [342].

Within the framework of observational studies, a plethora of challenges surfaces during practical implementation [342]. Crafting precise research inquiries becomes intricate, much like the task of selecting pertinent variables that faithfully represent the intended phenomenon. Distinguishing between confounding factors and sources of systematic error presents a conundrum, akin to determining the optimal mathematical and statistical models to address the research query [342]. This intricate landscape rests upon the bedrock of acquiring high-caliber data that genuinely mirrors the phenomenon under investigation.[342]

Against the backdrop of these impediments, the potential outcome framework, complemented by graphical modeling, has emerged as a potent tool. This empowers researchers to visually portray their a priori assumptions concerning hypothesized causal inquiries [342]. This process empowers researchers to render assumed associations into visual constructs prior to embarking on analyses, thereby harnessing insights from the data generation process.

Directed Acyclic Graphs (DAGs) offer researchers engaged with observational data an avenue to transparently convey their comprehension, conceptions, and convictions pertaining to the causal interplay among variables. Scrutinizing a DAG's structure and intricacies allows for the identification of variables that necessitate consideration for the precise calculation of causal effects. This appraisal can be executed through visual scrutiny or by employing specific algorithms. Although the resemblance of the DAG to the genuine data-generation process influences outcome accuracy, the act of crafting and disseminating a DAG accentuates the transparency of assumptions, inviting discerning scrutiny. [342, 343] To construct a DAG that resonates closely with the

natural data-generation process, researchers can invoke visual scrutiny and algorithmic refinements to fine-tune the DAG's architecture. [342]

In **paper 2**, we developed a directed acyclic graph (DAG) represented as a Bayesian network. To achieve this, we utilized data from two studies, namely WSAS 2 and OLIN 2016. The structure of the resulting DAG was a result of capturing the probabilistic connections and interactions between variables within the dataset. These variables were identified by experts as being relevant to our hypothesis regarding the modification of the smoking effect by socioeconomic status (SES). At such step we deviated from deriving a full data driven methods however expert intervention is acknowledged as a mean of improving presumed model of association as some variables on no relevance may introduce noise and distort the intended association/causal path inaccurately [342].

The DAG or Bayesian network structure emerged from the insights gained during the data collection process [342, 343]. This process shed light on how the variables were distributed, how these distributions influenced each other, and how these distributions were impacted by the methods used for collection and definition. This understanding is reflected in the nodes representing variables and the directed arcs representing their statistical or causal interdependencies. As of a study that well designed and with high validity as WSAS we assume the data generation process would add an accurately informing input into the attempt of assessing the presumed effect modification.

To uncover the structure of this network (DAG), we employed a mathematical algorithm known as hill climbing, coupled with the Bayesian Information Criterion (BIC) as a benchmark. The algorithm's approach involved initiating an empty network and then systematically adding or removing edges while assessing the BIC score. At each step, the algorithm selected the edge that contributed the most to enhancing the BIC score. In this endeavor, another computational algorithm, bootstrapping, was harnessed. This algorithm served to ensure that the resultant configuration distinctly embodies the finest representation of both the probability distribution of the data variables and the parameters essential for subsequent outcome estimation. The assessment of the Directed Acyclic Graph (DAG) by subject matter experts unveiled absence of unobserved confounding variables influencing the intended associations under investigation. Notably, it is crucial to acknowledge that arriving at such a conclusion through conventional statistical regression approaches would demand expert input for identifying confounding factors, which might

introduce expert bias. In contrast, the utilization of a data-driven methodology in our approach has yielded this observation, contributing to an elevated level of precision in our estimations. Furthermore, the derived estimations, presented as predicted probabilities, proved to be more readily and easily interpretable.

4.1.5 COMPUTATAIONAL PHENOTYPING OF AIRWAY DISEASES

4.1.5.1 SYSTEMATIC REVIEW

With the mounting volume of evidence in medical literature, keeping abreast of the rapidly expanding information within a specific topic has become a daunting task for researchers, clinicians, and scholars. To address this challenge, comprehensive reviews are conducted to provide an encompassing overview of the current state of knowledge in a particular area. Various types of reviews exist, including expert reviews, scoping reviews, and systematic reviews. Among these, systematic reviews stand apart due to their methodical process of systematically searching the available literature to identify, evaluate, analyze, and synthesize all relevant studies [344, 345]. This approach reduces the potential bias introduced by selectively focusing on specific literature, thereby enhancing the credibility of the conclusions drawn.

In recent decades, the application of unsupervised computational techniques for characterizing airway diseases has shown initial progress. However, a limited understanding exists regarding the phenotypes derived through such methods. As recognition grows regarding the significance of identifying defining characteristics of airway disease heterogeneity, translating these traits into actionable research variables becomes crucial for accurate phenotyping. Moreover, utilizing complex mathematical modeling approaches in computational methods adds a layer of intricacy, necessitating specialized data processing, management, and research design considerations.[69]

In **Paper 3**, an exhaustive systematic search was conducted to compile and analyze the body of literature related to unsupervised computational models for deriving phenotypes of airway diseases. This search was inclusive of studies conducted between 2000 and 2020, spanning various languages and utilizing prominent databases such as PubMed, Web of Science, Scopus, and Google Scholar. The review adhered to the PRISMA guidelines, involving a comprehensive evaluation of 3330 records, leading to the inclusion of 116 relevant studies.

The studies were meticulously analyzed, with a focus on extracting pertinent information. While traditional bias assessment was not applicable to the phenotyping process, a novel approach was taken to evaluate the studies for what we can refer to as good practices in unsupervised phenotyping as indicated in literature [35, 74] [68, 69, 332, 346] . This assessment encompassed several dimensions, including subject selection and inclusion criteria, handling of missing data, outcome definition, variables employed in phenotyping, clinical and scientific relevance of derived phenotypes, and the reproducibility of the phenotyping procedure.

Reproducibility was evaluated based on the provision of detailed methodological information, computational data processing procedures, software and tools employed, and adherence to reproducible research frameworks. Notably, the assessment exposed a significant underreporting of reproducibility aspects in the majority of the reviewed studies. Few studies detailed their handling of data noise and variation, rationale for selecting statistical methods, visualization techniques, or the use of available tools for reproducibility. Conversely, a strong alignment with expert consensus emerged concerning the inclusion of relevant variables in the phenotyping process, with most studies adequately reporting the variables included and their nominations for inclusion.

However, aspects such as reporting random phenotyping sampling and the adoption of gold standard criteria for disease definition were less frequently reported. The in-house developed quality assessment tool effectively highlighted these challenges in the field of computational phenotyping. To further advance this endeavor, refining and enhancing the sensitivity of this tool in the future could contribute to the evolution and growth of computational phenotyping methodologies.

The outcomes of the review were presented in the form of a narrative summary encompassing all the characteristics that were documented as contributors to defining various airway disease phenotypes. In our reporting approach, rather than adhering to the conventional method of describing phenotypes solely from a clinician's interpretable perspective, we chose to deviate from this standard practice. This approach aimed to offer a fresh perspective on the diversity of traits considered for airway disease phenotyping and the specific phenotypes they were linked to within different research contexts. This shift allowed us to explore airway disease heterogeneity without the bias inherent in repeatedly echoing established conventions.

4.1.5.2 MACHINE LEARNING-DERIVED ASTHMA PHENOTYPES IN A REPRESENTATIVE SWEDISH ADULT POPULATION

In **paper 5**, we utilized an advanced machine learning algorithm, specifically deep embedded clustering, to identify distinct phenotypes of asthma. The rationale for applying machine learning techniques in social science research, as opposed to conventional methods, continues to be a topic of debate. The primary argument for incorporating machine learning is the recent availability of substantial and diverse datasets [347]. Machine learning enhances the analytical tools at our disposal for such investigations. Additionally, the nature of research questions in health science has grown increasingly intricate, extending beyond simple direct associations to more complex relationships that require analysis [348].

Given the hypothesis that asthma manifests in clusters, we chose an unsupervised clustering method renowned for its effectiveness in handling large datasets with mixed variable types. Notwithstanding the goal of minimizing bias inherent in human-led analyses, unsupervised machine learning still involves decisions about hyperparameters. Such choices include determining which variables to include for clustering, selecting the model, and deciding on the number of clusters that offer the most meaningful representation of the data. These decisions are made by researchers with the intention of producing clinically relevant results. Models are evaluated based on their sensitivity and specificity, allowing researchers to discern the most suitable model for interpreting the data. Comparisons are then made to determine which model delivers the most clinically valid cluster solution.

While this approach may face criticism for potential bias introduction, it demonstrates machine learning's capability as an adaptive tool that can navigate complex data with pre-existing knowledge, rather than an inflexible, absolute data interpretation method. In **paper 5**, we examined multiple solutions, generating various cluster sets with differing numbers, clinical characteristics, and stabilities. The chosen solution was predicated on the cluster that exhibited the highest scientific validity and enhanced model sensitivity.

4.1.6 ASPECTS OF DATA PROCESSING

4.1.6.1 DEFINITION OF DETERMINANTS, EXPOSURES AND OUTCOMES

Data was gathered through the utilization of multiple standardized questionnaires, which had undergone validation and were previously employed in other research studies. The collection process involved subjects completing postal questionnaires based on their own self-reports. Additionally, during clinical examinations, trained nurses conducted structured interviews and administered another questionnaire, with some questions overlapping between the two questionnaires.

In **paper 1**, the identification of adult-onset asthma was determined by a combination of objective measures from lung function tests and the subjects' self-reported symptoms. Given the ongoing debate regarding the distinction between asthma and COPD, it was believed that while adopting the objective definition approach added validity, there was still a limitation of potentially including COPD patients within the group identified as adult-onset asthmatics. As a result, a sensitivity analysis in the first paper excluded subjects who were likely COPD patients.

Furthermore, in the context of **paper 1**, the definition of outcomes utilized the ACT scoring system for asthma, which is a well-validated approach known for its sensitivity and specificity in capturing control status among individuals with asthma [349, 350]. However, this system lacked consideration for certain aspects such as lung function decline and the frequency of exacerbations, which have more recently been recognized as important factors in defining asthma control. This omission could potentially lead to an underestimation of poor asthma control.

Additionally, in **paper 1**, the evaluation of education level encompassed both a categorical system based on completion of formal schooling years (primary with 9 years, secondary with 12 years, and tertiary), as well as an assessment of education exceeding 12 years. This approach aimed to capture the impact of education in terms of cumulative cognitive development across educational levels, as well as other social indicators associated with education, such as wealth and employment opportunities.

Moving on to **paper 2,** the definition of asthma was solely reliant on subjects' self-reported symptoms and disease as indicated in the postal questionnaires.

Despite the potential for bias in self-reporting, previous studies comparing respiratory symptoms and disease prevalence based on postal and clinical examination revealed no significant differences. This supports the validity of estimates derived from this data. Defining individuals as asthmatics based on their current experience of symptoms ensured that the subjects included in assessing the effect modification between smoking effects and socioeconomic status were not cases of remitting asthma.

Moreover, the definition of COPD or chronic bronchitis outcomes in **paper 2** was based solely on questionnaire data. Due to challenges in distinguishing between these conditions using objective spirometry measures, a definition was adopted that aimed to be as inclusive as possible, considering symptom presentation as indicated by GOLD guidelines, in addition to self-reported use of COPD medications. However, this definition also made it difficult to differentiate these groups from individuals with severe refractory asthma, highlighting a potential limitation.

In **paper 4**, concerning defining rhinitis outcomes, accurate diagnosis of rhinitis based on self-reported symptoms can be difficult due to the similarity between these symptoms and those of conditions like the common cold, which can lead to misidentification. Despite this, self-reported accounts of respiratory symptoms are valued for their contribution to a more comprehensive understanding of the public health impact of these diseases. Despite the potential for bias in self-reporting, our method for classifying rhinitis incorporates individuals' self-reported nasal symptoms such as sneezing, and a runny or blocked nose aligns with the recommended approaches of the International Study of Asthma and Allergies in Childhood (ISAAC) [351] and is endorsed by the Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines [352].

Concerning **paper 1,2,4 and 6** where educational and occupation classes were used as exposures, although education and occupational status are commonly used socioeconomic status (SES) indicators in epidemiological research, their influence on health outcomes, which may be affected by variables such as air pollution, workplace exposures, and domestic environments, can differ. To improve precision, integrating objective indicators of occupational exposure, for instance, through employment exposure matrices, could enhance the reliability and interpretive power of these SES indicators. This method could help to reduce biases that arise from health-related job changes, the selection of employment based on one's health condition, and the concentration of health

issues within certain job categories. In relevance, the studies' cross-sectional structure does not allow us to determine if socioeconomic status (SES) influences the risk of outcomes or if outcomes predispose SES. Thus, the inability to establish temporality in the observed relationships represents a constraint of those studies.

In **paper 5**, defining asthma outcomes based on self-report of either of having asthma or physician diagnosed asthma allows for capturing the timing of asthma diagnosis, regardless of whether asthma status was active or remitting at time of data collection. Our hypothesis motivating this approach is that if remission/activity status of asthma could further be investigated in depth in relevance to those derived phenotypes, allowing for identification of phenotypes that are plausibly related to asthma remission and activity.

Given our adopted unsupervised computational approach, the question of which variables, the extent of their accurate definition and deciding on which to include in the phenotyping process is a controversial question in methodical literature [353, 354]. Previous attempts on phenotyping asthma have found that consideration of different sets of variables results in different sets of resulting phenotypes, with each set presenting different clinical characteristics and degree of stability [332]. Additionally, the best recommended approach presented has been a hybrid of selecting variables based on inputs both from statistical methods and clinical expert insight [332, 354]. In our framework, we utilized a large set of variables that included traditional key clinical defining parameters of asthma like lung function, inflammation and allergic status and severity and symptoms measures. Further, additional aspects that conceivably contributes to asthma like sociodemographic risk factors, sort of triggers and patterns of medication intake. All the variables were further endorsed by clinical experts' view. This provides room for capturing clinically meaningful phenotypes and further exploring whether asthma phenotypes are characterized by features beyond the common reported clinical factors. Yet, the challenge of availability last as a limitation. Some variables on sputum measures and comorbidities were lacking in our data sets.

Another interesting issue related to the aspect of using severity as a feature for defining asthma phenotypes. Severe asthma has long been acknowledged as a separate entity that encompass several phenotypes within. Hence, when including severity as a feature for defining asthma, it is important to bear in mind the probability that two different pathophysiological phenomena being phenotype at once. Hence, the resulting phenotypes may present the iceberg of

deeper phenotypes impeded within. In our framework, severe asthma was defined as per GINA definition using data on self-report of medication usage, which does not present the most precise approach. Aside from the recall bias issue, if we assume that our identification of patients who are uncontrolled despite optimal treatment and management with high dose ICS-LABA treatment was precise. Yet, the data available based on self-report is hard to reflect whether the treatment is optimal without data on adherence, way of using medication, and other contributing factors that plays a role into labelling with severity status. Hence, a most ideal approach would be having severity status defined on hospital setting from register data.

4.1.6.2 MISSING DATA

In cases where data was absent, it was recorded as missing, except for queries related to symptoms or respiratory conditions. For these specific questions, missing data was designated as no to prevent an overestimation of disease or symptom prevalence figures.

In **paper 1, 4**, individuals with incomplete data were excluded from analysis, and the complete analysis was exclusively conducted among subjects who possessed comprehensive data concerning both exposure and outcome variables. Employing a complete case analysis in studies characterized by large samples and well-structured designs is unlikely to yield biased results, as the complete cases generally represent a random subset of the overall study population, and the occurrence of missing data is largely random as well. However, it would have been more accurate to confirm the randomness of the missing data, and imputing missing values plausibly could have enhanced the precision of this approach.

For **paper 2,5 and 6**, missing data was managed using a multiple imputation technique. The randomness of the missing data was evaluated, and subsequently, a systematic algorithm was used to predict missing values based on information derived from the available complete data. This strategy enhances the robustness of the obtained results by strengthening the validity of the data through an effective imputation process [355]. Specifically in **paper 5**, multiple imputation was implemented using random forest approach. Random forest is a sophisticated algorithm that is often used for classifying data or making predictions [356]. It works by building numerous decision trees during the training phase and then taking either the most common outcome (in classification) or the average outcome (in regression) from all the trees. When it comes to imputing missing data, the random forest is particularly effective

because it can navigate complex and nonlinear relationships between variables, allowing for accurate predictions of the missing values [357]. The implementation of such approach was motivated its superiority in handling complexity in the data and to maximize accuracy for the resulting data that was input into further advanced statistical processing for cluster derivation [358].

4.2 DISCUSSION OF RESULTS

4.2.1 EDUCATION LEVEL AND ASTHMA CONTROL AMONG ADULTS WITH ADULT'S ONSET ASTHMA

Our research findings have demonstrated a consistent association between low education levels and inadequate asthma control among individuals with adult-onset asthma. This association held true across the three separate study samples as well as in the pooled analysis. Our observations align with analogous studies investigating the impact of education levels on asthma control in children. These studies have indicated that low parental education levels are linked to suboptimal asthma control among children [359]. Additionally, similar research has highlighted the connection between low education levels and poor asthma control among adults [360], while other investigations have emphasized an independent relationship between low income and inadequate asthma control [361].

Furthermore, our study delved deeper into these findings and unveiled that low education levels were correlated with inadequate asthma control specifically among those who utilize daily inhaled corticosteroids (ICS) and individuals with non-atopic asthma. Interestingly, this correlation was not observed among the group of patients with atopic asthma. This pattern is consistent with a study conducted in the United States, where a low education level was predictive of poor asthma control among adults with moderate to severe asthma who were undergoing treatment [362]. This discrepancy might suggest that the group of daily ICS users forms a more homogeneous subset within the spectrum of asthma patients, displaying distinct characteristics in terms of airway pathophysiology and environmental sensitivities that make them more susceptible to the interplay of low education-related exposures.

When examining the discrepancy in effect observed primarily among those with non-atopic asthma, several factors could be contributing. Despite the

validity of our adopted definition of adult-onset asthma, which was endorsed by physicians in the SAAS study cohort, the cross-sectional nature of our study design and the reliance on subjects' self-reported asthma definitions, with subsequent clinical confirmation, might lead to some inaccuracies. This could result in individuals with remitting asthma showing up in the analysis sample as adult-onset cases, and this group might likely present a more severe, difficult to treat form of asthma with potential atopic predisposition [363]. Consequently, this subgroup could be influencing the distinct pattern of control observed between atopic and non-atopic individuals with adult-onset asthma.

Moreover, with regard to atopy and adult-onset asthma, it is recognized that the manifestation of asthma among sensitized individuals is contingent on various factors, including the extent and duration of allergen exposure and the specific type of allergen to which the patient is sensitized [364]. Therefore, both the duration and intensity of allergen exposure, coupled with the precise type of allergen sensitivity, can contribute to the development of airway hyperresponsiveness [364, 365]. This nuanced interplay might have taken place among the adult-onset subjects, where the exact timing, duration, and types of sensitizations remain insufficiently characterized in our study. It's plausible that among individuals who develop asthma as adults and exhibit atopy, their diverse durations of exposure to specific sensitizers could lead to distinct outcomes. They might develop greater tolerance over time, resulting in better control, or their pattern of sensitization might not be as severe or easily manageable. In contrast, these effects might be less pronounced among those who are non-atopic. As a result, the interaction of low education levels with environmental exposure, triggered by adjuvant exposures, could possibly amplify the non-atopic phenotype compared to the atopic one [366]. Due to the lack of comprehensive data on the timing, duration, and types of allergic sensitization, this mechanism cannot be definitively ruled out.

Additionally, it has been established that the intensity and duration of allergen exposure variably correlate with airway hyperresponsiveness (AHR) and the decline in lung function [365]. Finally, the reflection of multiple facets of social status within low education levels, including lifelong environmental exposures, further underscores the explanation of different control responses among atopic and non-atopic individuals [367].

One frequently proposed explanation for the observed connection between low education levels and poor asthma control is the limited availability of material

resources among individuals with lower education, potentially impacting their access to and adherence to medication use [368, 369]. While this explanation might seem plausible, our questionnaire inquired about the medications used by the subjects, making non-adherence a less likely explanation for the observed effect. Furthermore, a prior Danish study reported an independent link between education and asthma control, irrespective of income levels [368].

Another avenue through which education impacts asthma control is the realm of cognitive capital [367, 370, 371]. Higher education equips patients with the ability to make informed choices about healthy lifestyle behaviors such as exercise, healthy diets, and avoidance of risky habits [372-375]. This extends to patients' awareness of maintaining a healthy body mass index (BMI), which is a crucial determinant of adult-onset asthma [376, 377]. Thus, these factors within our study context could also contribute to the observed association. Although we adjusted for BMI and smoking as potential influencers in this pathway, other routes remain open and could potentially contribute to the observed effect.

A third pathway through which the impact of low education on asthma control can be observed is the aggregation of comorbidities and adverse health outcomes among individuals with lower education levels, influenced by all the aforementioned factors. Comorbidities are directly intertwined with asthma control, particularly given the connection between low education levels and chronic conditions like stroke, hypertension, diabetes, and depression[378, 379].

4.2.2 SMOKING EFFECT MODIFICATION BY SES AND RESPIRATORY OUTCOMES

Smoking had a stronger adverse impact on allergic asthma for lower occupational groups and highly educated individuals, while its negative effect on non-allergic asthma was more notable among higher occupational classes and those with lower education levels.

The SEI and SSYK occupational classification systems categorize individuals based on professional environment, exposures, and to some extent, authority, and income. Research utilizing these systems has highlighted elevated exposure levels to various hazards among specific occupational groups, such as manual workers in industry (SEI) and workers in agriculture, construction, and manufacturing (SSYK) [164]. Additionally, these classifications have revealed respiratory symptoms in manual service and industry workers, higher

asthma rates in healthcare and science workers, and an increased risk of non-allergic asthma among service workers [164].

The finding that former smokers from low socioeconomic groups exhibit a greater likelihood of respiratory diseases compared to both current and never smokers aligns with previous research trends. Studies investigating smoking's impact on respiratory outcomes consistently indicate that former smokers face a higher risk than current smokers [380-382]. Moreover, the heightened probability of respiratory diseases among smokers from lower socioeconomic backgrounds, in contrast to those from higher socioeconomic strata, resonates with previous observations linking combined exposure to occupational dust, fumes, vapors, and smoking to a greater risk of COPD and chronic bronchitis [383-385]. The interaction between vapor, gas, dust, and fumes exposure and smoking has also been linked to the risk of COPD [386]. Though our study didn't specifically address the combined effect of smoking and occupational exposure on COPD or chronic bronchitis, we did observe a combined effect on both allergic and non-allergic asthma across diverse occupations.

Our observed combined detrimental effect of smoking and occupational exposures on the likelihood of allergic asthma in low socioeconomic groups can be attributed to the direct damage inflicted by both factors on airways and through a sensitization effect. Occupational exposures are known to contribute to allergic sensitization, while smoking might increase airway sensitivity to such hazards [387, 388]. Smoking's impact on allergic sensitization remains uncertain, as some studies link it to higher allergic disease risks, while others suggest a potential protective role [389-392]. Smoking's influence on producing airway allergic inflammatory mediators like IgE, IgG antibodies, and histamine might result from airway mucosal damage [393]. This is evident in studies where smokers exhibited higher total serum IgE concentration and allergen-specific skin prick test reactions [393].

Regarding the sensitization-independent harmful effect, research from Sweden, Norway, and Spain reported an elevated risk of adult-onset asthma tied to prolonged exposure to non-sensitizing irritants among workers in industries such as metal, wood, plastics, construction, plumbing, welding, mining, and asphalt roofing [387]. This exacerbates smoking's direct damage to airways, possibly amplifying the likelihood of respiratory diseases [388, 394].

Our finding of higher probabilities of allergic asthma among former smokers in highly educated groups, compared to lower educated, aligns with the idea

that smoking interacts with sensitization pathways that thrive in hygienic settings [395].

The lessened impact of smoking on the probability of non-allergic asthma in low occupational and educational groups, compared to higher professional classes, is intriguing. Understanding non-allergic asthma, its risk factors, and interaction dynamics compared to allergic asthma is still evolving. The impact of smoking on atopic and non-atopic asthma remains conflicting, with studies suggesting gender-specific and smoking habit-related effects [396-398]. The lessened effect of smoking on non-allergic asthma among manual workers in industry and home workers might stem from lower sensitivity to smoking's harm among non-atopic individuals in these settings [399]. Such desensitization could be a result of occupational exposures in these contexts [399]. Smoking's varying effects across socioeconomic groups and by each measure indicate that different occupational exposures trigger airway damage, allergic sensitization, and irritant sensitivity diversely. Socioeconomic settings further interact with host factors, genetics, psychosocial elements, and smoking status [398, 400].

The "healthy smoker effect" explains our finding of higher respiratory disease probabilities among former smokers than among current and never smokers. Respiratory symptoms, severity, and lung function decline might prompt asthma patients who smoked to quit, contributing to increased disease likelihood [401].

Differences in duration and age of smoking initiation between former and never/current smokers could also explain the findings. Former smokers are often older, with longer smoking durations and greater lung function decline [167]. Even among new-onset asthma cases, evidence suggests smoking's negative effects begin before asthma onset [402, 403]. Comorbidities might also drive former smokers to quit, elevating outcome probabilities. While our questionnaire captured age of smoking initiation and cessation, only current cigarette consumption, not past smoking amounts, was queried, limiting our understanding of former smokers' exposure.

4.2.3 COMPUTATIONAL PHENOTYPES OF AIRWAY DISEASES: SYSTEMTIC REVIEW

The review highlights extensive variability in phenotypes derived for obstructive airway diseases and rhinitis in children and adults. Phenotyping methods, study settings, and considered variables also varied widely. Asthma

phenotypes commonly focused on atopic status, disease onset, sex, symptom severity, and lung function. Severe asthma phenotypes emphasized lung function, treatment status, corticosteroid dosing, and GINA steps level. COPD and ACO phenotyping centered on lung function, age, symptoms, comorbidities, and mortality. Rhinitis phenotyping included variables like sex, disease severity, quality of life, comorbidities, and sensitization status. Hierarchical and non-hierarchical clustering were commonly used methods, but from 2015 onwards, alternative methods like mixture-based modeling, structural equation modeling, factor analysis, and latent class modeling gained popularity. The variation in phenotyping approaches underscores the complex nature of airway diseases and the need for comprehensive assessment.

4.2.3.1 VALIDATION OF THE DERIVED PHENOTYPES

Ensuring the reliability and clinical relevance of derived disease clusters is vital, as computational methods can generate clusters without meaningful significance or data quality. Stability involves reproducing similar phenotypes across different datasets or with data modifications. This can be assessed by observing phenotypes' clinical characteristics over time and using statistical testing to examine variations based on important clinical, demographic, and biological factors. For example, studies excluded subgroups of severity to assess their impact on derived phenotypes. Some studies investigated the stability of phenotypes over time and their prediction of outcomes, such as decline in lung function or quality of life. These methods helped identify phenotypes predicting persistent limitation, morbidity status, and adverse respiratory outcomes, adding to the clinical validity of derived clusters.

4.2.3.2 INTERPRETATION OF RESULTS

This study examines the characterization of airway disease phenotypes in both general and clinical settings. Studies often recruited participants from clinical contexts such as primary care centers, tertiary hospitals, and outpatient clinics, while those from the general population reported similar asthma phenotypes, reflecting mild early-onset asthma and mild atopic asthma. Phenotyping within clinical settings enabled the identification of phenotypes defined by easily measured clinic parameters, including high treatment response and adherence, and disease progression trajectories. Similarly, COPD phenotypes in clinical settings encompassed varying degrees of emphysematous and air-entrapment changes, along with comorbidities. Nonetheless, the observed variability in phenotypes between general and clinical settings may stem from differences in available variables and patient presentation.

Phenotyping airway diseases entails considering multiple dimensions, including etiology, disease manifestations, treatment behaviors, and prognostic indicators. Lung function measures were frequently used for asthma and COPD phenotyping, but other significant aspects, such as bronchodilatation (reversibility) and bronchoprovocation tests, were often overlooked. Airway remodeling, a feature of asthma, was considered in only two studies. While certain variables were consistently used to phenotype different airway diseases, the selection process was influenced by expert input, literature, clinical measurement feasibility, and statistical methods like stepwise selection or dimensionality reduction.

The inclusion and exclusion criteria for subjects varied across studies. Asthma was defined through self-report, symptoms, spirometry, and physician diagnosis, while severe asthma often followed established guidelines or physician diagnosis. COPD and rhinitis were similarly diagnosed through physician reports, with some studies focusing on specific subgroups like work-related asthma or NSAID-sensitive asthmatics. Despite outcome definition variations, phenotypes related to key aspects such as atopy, symptomatology, onset age, and comorbidities were generally similar within each outcome category.

4.2.3.3 COMPARISON WITH SIMILAR WORK

Our study aligns with previous research by Cunha et al., finding that common asthma phenotypes are linked to atopic status, age of disease onset, and severity. Hierarchical cluster analysis emerged as the primary method for phenotyping, with subjects mainly recruited from specialized clinical centers. Similarly, Pinto et al.'s work on COPD phenotypes observed that subjects were primarily recruited from clinical settings, mirroring our findings. Both hierarchical and non-hierarchical clustering were commonly used for COPD phenotyping. Unlike Pinto et al., our COPD studies encompassed a wider range of diagnostic criteria, resulting in more diverse subject characteristics.

While Cunha et al.[404] categorized severe asthma as a subtype, our study treated severe asthma as a distinct entity, supported by the inclusion of studies specifically focusing on severe asthma subjects. This distinction underscores the value of separately phenotyping severe asthma and asthma. In rhinitis, we identified allergic phenotypes defined by atopic status, polysensitization, and specific allergen sensitivities. These align with Mullol et al.'s[37] review of rhinitis phenotypes based on clinical, inflammatory, etiological, and demographic traits. Our study further expanded on this by identifying age-

related rhinitis phenotypes and revealing comorbidities like depression, anxiety, diabetes, and fibromyalgia linked to rhinitis.

4.2.4 SOCIOECONOMIC STATUS AND DIFFERENT FORMS OF RHINITIS IN SWEDISH ADULTS

In this representative sample of the adult population, those with secondary or higher education were more likely to report having allergic or chronic rhinitis compared to those with only primary education. Additionally, those working in higher skilled occupations were more likely to report allergic rhinitis than their lower skilled counterparts. There was no significant difference in the reporting of chronic rhinosinusitis across different education levels, nor was there a difference in the risk of either chronic rhinitis or chronic rhinosinusitis across various occupational skill levels. The influence of factors such as sex, age, and body mass index (BMI) on these associations was noted, although these findings were not uniformly observed.

In our study, adults with higher levels of education and those in skilled jobs reported more cases of allergic rhinitis, aligning with previous research that ties socioeconomic status to this condition[175, 207, 405]. Contrastingly, some studies suggested that lower paternal education could increase the risk of allergic rhinitis, while better household conditions were linked to a greater rhinitis risk in children[406]. Individuals with at least secondary education also faced a greater risk of chronic rhinitis, which resonates with findings that associate this risk with exposure to occupational pollutants[177, 407]. Despite no direct link between occupational class and chronic rhinitis in our data, the association with higher education necessitates further exploration. Our results didn't show a correlation between socioeconomic indicators and chronic rhinosinusitis. However, previous studies indicate varying risks based on educational and occupational status, with higher education generally associated with a lower risk and lower-skilled occupations with an increased risk[182, 184, 185, 408, 409].

Our study's alignment with the hygiene hypothesis suggests that individuals from higher social classes, as indicated by education and occupation, are at an increased risk for allergic rhinitis. This hypothesis, supported by the link between hygienic environments and atopy, posits that better living conditions, possibly associated with higher SES, could lead to more allergic conditions like rhinitis [410]. Discrepancies in the link between SES indicators—education and occupation—and chronic rhinitis point to the multifaceted influence of

SES on different rhinitis forms. Education may not always equate to material wealth; therefore, individuals with high education but not necessarily high occupation might face diverse social exposures, impacting their risk of chronic rhinosinusitis differently [156, 411]. Our data did not demonstrate the impact of SES on chronic rhinosinusitis risk, but it doesn't rule out the role of SES-related exposures not captured by our study. Previous research associated occupational exposure to specific agents with an increased risk of chronic rhinosinusitis[412, 413], suggesting that a more detailed examination of occupational environments is necessary.

Sex, age, and BMI could modify the relationship between SES and rhinitis types, hinting at the combined influence of work conditions, biological differences, and environmental factors. Younger individuals in certain occupations may face a unique risk landscape for chronic rhinitis, underscoring the need for a deeper understanding of SES's role in rhinitis, beyond standard indicators. Further studies should dissect the socio-environmental context of rhinitis and how it interacts with pathophysiology and chronic disease management.

4.2.5 MACHINE LEARNING-DERIVED ASTHMA PHENOTYPES IN A REPRESENTATIVE SWEDISH ADULT POPULATION

Cluster 1 in our study, featuring an older demographic with late-onset, severe asthma and a history of smoking, resonates with similar phenotypes identified in adult asthma studies [87, 95, 414]. This group is marked by high symptom prevalence, healthcare usage, and medication intake for severe asthma. Studies like those by Kaneko et al.[95] have described comparable characteristics, while Loureiro et al.[246] observed a related phenotype with obesity and elevated blood eosinophil, neutrophil, and monocyte counts. Other research has linked such phenotypes to systemic inflammation [414], late-onset severe asthma [80], and higher comorbidity and smoking rates [89, 93]. Cluster 2, predominantly female with early adult-onset asthma, exhibits moderate symptoms and nearly normal lung function, echoing findings from studies like those by Dudchenko et al.[242], who noted sensitivity to weather triggers, and Ilmarinen et al.[86], who described a 'female asthma' phenotype with moderate symptoms. This pattern is consistent with young adult-onset, mild female asthma types identified in other works, with stable lung function and a slower progression of the disease[87].

The adult-onset, high-inflammation phenotype (Cluster 3) aligns with phenotypes from other studies reporting elevated eosinophil levels [247, 415]. Although the onset age wasn't specified, Boudir's[84] work underlines a phenotype with marked reversibility and pronounced symptoms, paralleling our findings. Hsaio et al.[92] described a similar phenotype with a smoking history, which complements our observations. Lastly, Cluster 4, which describes early-onset, mild asthma with atopy, matches with many previous findings [86, 95, 242, 248]. Studies have noted physical activity impairment [242] and a younger age at onset, consistent with our young mean age for this group. Additionally, Loza et al.[247] noted an association with low inflammation and a high T2 cell pattern in this phenotype.

Cluster 1, representing late-onset asthma in older adults, is distinguished by a high smoking prevalence and the likelihood of concurrent COPD, suggesting a phenotype where asthma and COPD may overlap [86, 317, 318, 320]. This group, often involving severe asthma in obese women, exhibits high healthcare usage and systemic inflammation, hinting at mixed inflammatory patterns seen in severe late-age asthma [86, 92, 246]. Cluster 2, mainly composed of women with moderate asthma, shows good lung function and lower rates of smoking. Their moderate asthma symptoms and high emergency service use could suggest a proactive approach to managing their condition [416]. Their healthy profile is further characterized by a lower BMI and fewer comorbidities, which contributes to a generally favourable prognosis. The third phenotype (Cluster 3), adult-onset with significant inflammation, is marked by high FeNO and eosinophil levels indicative of a Type 2 immune response [247, 415]. Despite potential responsiveness to inhaled corticosteroids, there may be an undertreatment in this group, which also has a notable history of smoking and allergic comorbidities such as rhinitis. Finally, Cluster 4 aligns with traditional childhood-onset, mild atopic asthma, often leading to good control and a high rate of remission [417]. This phenotype has the lowest proportion of current asthma, suggesting that for some, the condition may have abated over time.

4.2.6 LEVEL OF EDUCATION, BUT NOT OCCUPATION, IS DIFFERENTIALLY ASSOCIATED WITH ASTHMA PHENOTYPES IN ADULTS

In our study, we observed that higher educational levels were linked to particular adult asthma phenotypes. Specifically, individuals with secondary and tertiary education had lower occurrences of mild asthma commonly found

in women, suggesting that socioeconomic status (SES) might have a more significant impact on asthma in women than men [161].

The lower incidence of early-onset, mild, atopic asthma in individuals with higher SES could be due to a higher likelihood of remission in these groups, possibly because of reduced exposure to harmful environmental conditions often associated with lower SES [50]. Additionally, while SES seemed to influence the inflammation associated with the asthma phenotypes we identified, the effect appeared to be smaller compared to other phenotypes. Occupational class did not correlate with any asthma phenotypes in our data, potentially because none of the phenotypes were directly related to occupational exposure—this may be due to the lack of detailed occupational exposure data for our analysis [418]. However, this doesn't rule out the possibility that certain occupational factors or allergens might not be adequately captured by the current occupational classifications used.

5 CONCLUSION

- A lower level of education is identified as a risk factor for inadequate asthma control. Particular attention should be directed toward individuals with moderate to severe asthma and those with non-atopic asthma, as low educational attainment could notably impact their asthma control status.

- The influence of social status as indicated by occupational status and environment and education, marginally modifies the impact of smoking on allergic asthma subtypes. Smoking's effect on allergic asthma is more pronounced within higher occupational classes and lower education groups. Conversely, for non-allergic asthma, the adverse impact is more significant among individuals in lower occupational roles and those possessing higher levels of education.

- Employing computational data-driven methods to derive phenotypes for airway conditions like asthma, COPD, severe asthma, and ACO has led to considerable divergence in derived phenotypes across various studies. This variability can be attributed to dissimilarities in sample selection, outcome metrics, definitions, and the choice of variables for phenotyping. The lack of frequent use of reproducibility measures in computational phenotyping research impedes the ability to investigate the reasons behind such discrepancies. To enhance comprehension and the validity of derived phenotypes, along with their clinical and scientific usefulness, a consistent strategy for defining outcomes and selecting variables, coupled with reproducible techniques for phenotyping airway diseases, is imperative.

- Higher education level seems to be a risk factor for allergic and chronic forms of rhinitis. Similarly, higher occupation skill levels may increase the risk of allergic rhinitis.

- Four distinct asthma phenotypes were identifiable in a study of adults with a history of asthma. These phenotypes were differentiated by factors such as the age when asthma began, the severity of the condition, and various other characteristics including demographic data, risk factors, triggers, respiratory function, symptom patterns, and markers of inflammation.

- Educational level is a more relevant indicator of socioeconomic status (SES) than occupation when it comes to its relationship with different adult asthma phenotypes. The lack of association with occupation may reflect the

complexity of environmental factors affecting asthma, which go beyond the scope of standard job classifications. To deepen our understanding, further investigation into how SES influences asthma phenotypes is needed. Such research could lead to more effective, SES-sensitive prevention and treatment approaches for asthma.

6 FUTURE PERSPECTIVES

In conclusion, this thesis work highlights the importance of considering educational level in asthma management and patient education. Patients with lower education levels should receive increased attention to enhance treatment adherence and lifestyle-related recommendations.

Our findings also demonstrate the usefulness of socio-occupational classifications in defining socioeconomic status (SES) and capturing different respiratory disease patterns across SES levels. These classifications are effective in revealing effect modifications between smoking and respiratory diseases, particularly among different SES groups. Future research could replicate our study using similar SES systems to gain further insights. Regarding effect modification between smoking and SES measures, tailored smoking cessation interventions might be more beneficial for certain high-risk socio-occupational groups. Understanding the combined effects of smoking and socioeconomic exposures on lung health, including potential occupational-induced allergic sensitizations and airways hyperresponsiveness, requires additional investigation.

Further, to advance computational phenotyping of obstructive airway diseases, harmonization of methodological approaches is crucial. Establishing a consensus list of variables for each outcome and enhancing participant selection and phenotyping methods will lead to clearer comparisons between studies. Our quality assessment tool addresses a critical gap in the field, but its wider application will identify areas for improvement. Comprehensive reporting of data processing, statistical methods, and computational processes is essential for ensuring study validity. Future studies should incorporate reproducibility tools to facilitate code and data sharing, version control, and better documentation.

Our finding demonstrated that assessment of rhinitis burden via SES can be one way to develop preventive strategies. Additionally, more nuanced understanding of the underlying social components, such as housing, traffic, childhood, and family factors, is crucial in this respect and when designing effective management strategies.

Our identification of four distinct asthma phenotypes that were differential in clinical, social, and biological characteristics pave the way for in-depth research into such asthma phenotypes, including the detailed immunological

profiling of each phenotype, examination of potential risk factors, exploration of comorbidity patterns, and evaluation of the long-term clinical outcomes and overall societal impact associated with each type of asthma.

Further research into SES variation in various asthma phenotypes is warranted to complement our finding of differential association between education, but not occupation, and presentation of certain asthma phenotypes. These studies can inform the development of tailored SES-related prevention and management strategies of asthma.

7 ACKNOWLEDGEMENT

This thesis work was conducted at the Krefting Research Centre, Department of Internal Medicine and Clinical Nutrition, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg. My profound gratitude extends to every individual who contributed to and assisted in the realization of this endeavor in various ways.

To all the participants of the West Sweden Asthma Study (WSAS), the Study of Obstructive Lung Diseases in Northern Sweden (OLIN), and the Seinäjoki Adult Asthma Study (SAAS) in Finland, I offer my sincerest thanks. Your invaluable contributions have made this work possible, and I am deeply grateful for your participation.

To my main supervisor, **Bright I. Nwaru**, I extend my heartfelt thanks for your dedicated teaching, unwavering attention, invaluable guidance, and unwavering support throughout this journey. Thank you for granting me the opportunity to learn and grow under your mentorship and thanks for your support, valuable advice, and understanding through the hard times. I learned a lot from you as a supervisor, and way more as a human.

To **Hannu Kankaanranta**, my co-supervisor, I am grateful for the rich discussions, insightful comments, and invaluable tips you provided consistently. I deeply appreciate your support and the clinical insights you shared.

To **Linda Ekerljung**, my co-supervisor, thank you for your guidance, attention, and the unending support you provided, especially in dealing with the WSAS data. Your support has been instrumental, and I am thankful for it.

To **Emma Goksor**, my co-supervisor, I appreciate your contribution to this work.

To the memory of **Bo Lundbäck**, my former co-supervisor, I will forever cherish the warmth with which you welcomed me on my first day at Krefting. Our thought-provoking discussions and enjoyable talks about our homelands remain etched in my memory. May your kind soul rest in eternal peace.

To **Rani Basna**, thank you for your support and contributions to this work. I have learned a great deal from our discussions and the valuable resources you consistently provided.

I extend my gratitude to all my co-authors who contributed to and assisted in the completion of this work. Special thanks go to **Pinja Ilmarinen, Caroline Stridsman, Gregorio Paolo Milani, Valentina De Cosmi,** and **Alessandra Mazzocchi** for their valuable contributions. Working with you all has been a pleasure.

A special thank you goes out to the nursing team at Krefting Research Centre for their tremendous assistance, attentive responses, and guidance regarding data matters. I want to express my heartfelt thanks to **Lina Rönnebjerg,** who left Krefting, for the delightful conversations and insightful discussions about the data.

To all my current and former colleagues at Krefting Research Centre, I am immensely grateful for your unwavering help and support. Every shared memory has been a testament to our collective journey. I extend special thanks to **Madeleine Rådinger**, the head of Krefting, who has been an inspiring model of women leadership. Your availability and support during challenging times have been invaluable.

Carina Malmhäll, your supportive and caring presence has been deeply appreciated. Thank you for your assistance and kindness!

To **Emma Boberg**, thank you for your support, guidance, and valuable insights as a senior PhD student.

Mirela Ribic, your consistent presence and support with administrative matters throughout this journey have been invaluable, thank you!

To my coworkers and colleagues, **Saliha Selin** and **Daniil Lisik**, thank you for the assistance and support you provided along the way. Your availability, helpful tips, and shared moments of learning and joy have meant a lot. Especial thanks to **Daniil** for translating the abstract text of this thesis work to Swedish.

To my senior **colleague Abubakr** who has been there with tips and advices, thank you!

To **my family of friends and community** in Gothenburg. To **Hanaa, Esra, Moufida, Roza**, and **Amal**. This journey would not have been bearable without our warm shared moments. Thank you all for your tremendous support, love and the very beautiful memories you have spiced this chapter of my life with!

To my friends**, Marwa, Shahd, Esra, and Hala**, the connection we share has been a constant source of strength and support through the challenging times, flourishing and fortifying despite the miles between us.

To the rock of my life, older brother **Mustafa**, I am everly a grateful and proud sister. To **Hala** and **Moshera**, **Ali**, **Awad**, and **Noof**, my siblings and life gems, your unwavering love has kept fueling my determination, no matter how challenging the path has been.

Last but certainly not least, to **Mom** and **Dad**, my two working-class inspiring heroes, none of this would have been possible without your prayers, love and encouragement. Being your daughter is the one ultimate privilege I was born with and ever had.

To my beloved land, **Sudan**, the cherished cradle of my existence, and its warm-hearted people, who are woven into the very fabric of my being — from this nurturing soil my life's odyssey unfurled, and it is into its embracing arms that all my stories shall return to rest, enveloped in an everlasting embrace of peace and love — Amen!

8 REFERENCES

1. Fabbri L, Peters SP, Pavord I, Wenzel SE, Lazarus SC, MacNee W, et al. Allergic rhinitis, asthma, airway biology, and chronic obstructive pulmonary disease in AJRCCM in 2004. American journal of respiratory and critical care medicine. 2005;**171**(7):686-98.

2. Holgate ST. A brief history of asthma and its mechanisms to modern concepts of disease pathogenesis. Allergy, asthma & immunology research. 2010;**2**(3):165-71.

3. Sakula A. Henry Hyde Salter (1823-71): a biographical sketch. Thorax. 1985;**40**(12):887.

4. Brenner BE. Where have we been?: Marcel Dekker, Inc. New York; 1999.

5. Global initative for asthma. Global Strategy For Asthma Management And Prevention, 2022. 2022 [1/6/2023]. Available from: https://ginasthma.org/.

6. Christopher H Fanta ML-V. Asthma in adolescents and adults: Evaluation and diagnosis. In: UpToDate, editor. UpToDate. Waltham, MA2023.

7. Selroos O, Kupczyk M, Kuna P, Łacwik P, Bousquet J, Brennan D, et al. National and regional asthma programmes in Europe. European Respiratory Review. 2015;**24**(137):474-83.

8. Stern J, Pier J, Litonjua AA. Asthma epidemiology and risk factors. Seminars in Immunopathology. 2020;**42**(1):5-15.

9. Dharmage SC, Perret JL, Custovic A. Epidemiology of asthma in children and adults. Frontiers in pediatrics. 2019;**7**:246.

10. Räisänen P, Backman H, Hedman L, Andersson M, Stridsman C, Kankaanranta H, et al. High but stable incidence of adult-onset asthma in northern Sweden over the last decades. ERJ Open Research. 2021;**7**(3).

11. Ekerljung L, Rönmark E, Larsson K, Sundblad B-M, Bjerg A, Ahlstedt S, et al. No further increase of incidence of asthma: Incidence, remission and relapse of adult asthma in Sweden. Respiratory medicine. 2008;**102**(12):1730-6.

12. Public health agency of Sweden. Diseases and other health problems (self-reported) by employment, sex and year. : Public health agency of Sweden;

2022 [updated 10.2022]. Available from: https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/.

13. Lundbäck B, Vanfleteren LEGW. Letter from Sweden. Respirology. 2021;**26**(8):818-9.

14. Backman H, Räisänen P, Hedman L, Stridsman C, Andersson M, Lindberg A, et al. Increased prevalence of allergic asthma from 1996 to 2006 and further to 2016—results from three population surveys. Clinical & Experimental Allergy. 2017;**47**(11):1426-35.

15. Pakkasela J, Ilmarinen P, Honkamaki J, Tuomisto LE, Andersen H, Piirila P, et al. Age-specific incidence of allergic and non-allergic asthma. BMC Pulm Med. 2020;**20**(1):9.

16. Hansen S, von Bulow A, Sandin P, Ernstsson O, Janson C, Lehtimaki L, et al. Prevalence and management of severe asthma in the Nordic countries: findings from the NORDSTAR cohort. ERJ Open Res. 2023;**9**(2).

17. Larsson K, Stallberg B, Lisspers K, Telg G, Johansson G, Thuresson M, Janson C. Prevalence and management of severe asthma in primary care: an observational cohort study in Sweden (PACEHR). Respir Res. 2018;**19**(1):12.

18. Ronnebjerg L, Axelsson M, Kankaanranta H, Backman H, Radinger M, Lundback B, Ekerljung L. Severe Asthma in a General Population Study: Prevalence and Clinical Characteristics. J Asthma Allergy. 2021;**14**:1105-15.

19. Global initative for chronic obstructive lung disease(GOLD). Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease. 2022 [updated 2022]. Available from: https://goldcopd.org/.

20. Curtis JL. Understanding COPD Etiology, Pathophysiology, and Definition. Respiratory care. 2023;**68**(7):859-70.

21. Agustí A, Celli BR, Criner GJ, Halpin D, Anzueto A, Barnes P, et al. Global initiative for chronic obstructive lung disease 2023 report: GOLD executive summary. American journal of respiratory and critical care medicine. 2023;**207**(7):819-37.

22. World health organization (WHO). Chronic obstructive pulmonary disease (COPD) 2023 [updated 16/3/2023]. Available from: https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd).

23. Adeloye D, Song P, Zhu Y, Campbell H, Sheikh A, Rudan I, Unit NRGRH. Global, regional, and national prevalence of, and risk factors for, chronic obstructive pulmonary disease (COPD) in 2019: a systematic review and modelling analysis. Lancet Respir Med. 2022;**10**(5):447-58.

24. Global initiative for obstrcutive lung diseases. Global Strategy for Prevention, Diagnosis and Management of COPD.; 2022.

25. Gulsvik A, Boman G, Dahl R, Gislason T, Nieminen M. The burden of obstructive lung disease in the Nordic countries. Respiratory medicine. 2006;**100**:S2-S9.

26. Ho T, Cusack RP, Chaudhary N, Satia I, Kurmi OP. Under- and over-diagnosis of COPD: a global perspective. Breathe (Sheff). 2019;**15**(1):24-35.

27. Axelsson M, Backman H, Nwaru BI, Stridsman C, Vanfleteren L, Hedman L, et al. Underdiagnosis and misclassification of COPD in Sweden - A Nordic Epilung study. Respiratory medicine. 2023;**217**:107347.

28. Gulsvik A. The global burden and impact of chronic obstructive pulmonary disease worldwide. Monaldi Archives for Chest Disease= Archivio Monaldi per le Malattie del Torace. 2001;**56**(3):261-4.

29. Lundback B, Lindberg A, Lindstrom M, Ronmark E, Jonsson AC, Jonsson E, et al. Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies. Respiratory medicine. 2003;**97**(2):115-22.

30. Backman H, Eriksson B, Ronmark E, Hedman L, Stridsman C, Jansson SA, et al. Decreased prevalence of moderate to severe COPD over 15 years in northern Sweden. Respiratory medicine. 2016;**114**:103-10.

31. Lindberg A, Jonsson AC, Ronmark E, Lundgren R, Larsson LG, Lundback B. Prevalence of chronic obstructive pulmonary disease according to BTS, ERS, GOLD and ATS criteria in relation to doctor's diagnosis, symptoms, age, gender, and smoking habits. Respiration. 2005;**72**(5):471-9.

32. Danielsson P, Ólafsdóttir IS, Benediktsdóttir B, Gíslason T, Janson C. The prevalence of chronic obstructive pulmonary disease in Uppsala, Sweden - the Burden of Obstructive Lung Disease (BOLD) study: cross-sectional population-based study. The Clinical Respiratory Journal. 2012;**6**(2):120-7.

33. Backman H, Vanfleteren L, Lindberg A, Ekerljung L, Stridsman C, Axelsson M, et al. Decreased COPD prevalence in Sweden after decades of decrease in smoking. Respiratory Research. 2020;**21**(1).

34. Bhatta L, Leivseth L, Mai X-M, Chen Y, Henriksen AH, Langhammer A, Brumpton BM. Prevalence and trend of COPD from 1995–1997 to 2006–2008: the HUNT study, Norway. Respiratory medicine. 2018;**138**:50-6.

35. Weatherall M, Shirtcliffe P, Travers J, Beasley R. Use of cluster analysis to define COPD phenotypes. Eur Respir J. 36. England2010. p. 472-4.

36. Papadopoulos NG, Bernstein J, Demoly P, Dykewicz M, Fokkens W, Hellings P, et al. Phenotypes and endotypes of rhinitis and their impact on management: a PRACTALL report. Allergy. 2015;**70**(5):474-94.

37. Mullol J, Del Cuvillo A, Lockey RF. Rhinitis phenotypes. The Journal of Allergy and Clinical Immunology: In Practice. 2020;**8**(5):1492-503.

38. Guilemany JM, García-Piñero A, Alobid I, Cardelús S, Centellas S, Bartra J, et al. Persistent allergic rhinitis has a moderate impact on the sense of smell, depending on both nasal congestion and inflammation. The Laryngoscope. 2009;**119**(2):233-8.

39. Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, Reitsma S, et al. European Position Paper on Rhinosinusitis and Nasal Polyps 2020. Rhinology. 2020;**58**:I-+.

40. Orlandi RR, Kingdom TT, Hwang PH, Smith TL, Alt JA, Baroody FM, et al. International Consensus Statement on Allergy and Rhinology: Rhinosinusitis. International Forum of Allergy & Rhinology. 2016;**6**:S22-S209.

41. Liva GA, Karatzanis AD, Prokopakis EP. Review of rhinitis: classification, types, pathophysiology. Journal of clinical medicine. 2021;**10**(14):3183.

42. Kakli HA, Riley TD. Allergic Rhinitis. Prim Care. 2016;**43**(3):465-75.

43. Peden D. An overview of rhinitis. In: Post TW, editor. UpToDate Waltham, MA: UpToDate; 2023.

44. Skoner DP. Allergic rhinitis: definition, epidemiology, pathophysiology, detection, and diagnosis. J Allergy Clin Immunol. 2001;**108**(1 Suppl):S2-8.

45. Langley-Hobbs SJ, Demetriou J, Ladlow J. Feline soft tissue and general surgery: Elsevier Health Sciences; 2013.

46. O'Hehir RE, Holgate ST, Hershey GKK, Sheikh A. Allergy essentials: Elsevier Health Sciences; 2021.

47. Bachert C, Marple B, Schlosser RJ, Hopkins C, Schleimer RP, Lambrecht BN, et al. Adult chronic rhinosinusitis. Nature reviews Disease primers. 2020;**6**(1):86.

48. Akhouri S, House SA. Allergic rhinitis: StatPearls; 2019.

49. Goksör E, Wennergren G, Vasileiadou S, Mincheva R, Nwaru B, Lundbäck B, Ekerljung L. Increased prevalence of allergic rhinitis in young men in Western Sweden. Eur Respiratory Soc; 2018.

50. Heldin J, Malinovschi A, Johannessen A, Alving K, Holm M, Franklin KA, et al. Clinical remission of asthma and allergic rhinitis-in a longitudinal population study. Journal of Asthma and Allergy. 2022:1569-78.

51. Cardell L-O, Olsson P, Andersson M, Welin K-O, Svensson J, Tennvall GR, Hellgren J. TOTALL: high cost of allergic rhinitis—a national Swedish population-based questionnaire study. npj Primary Care Respiratory Medicine. 2016;**26**(1):15082.

52. Eriksson J, Ekerljung L, Rönmark E, Dahlén B, Ahlstedt S, Dahlén SE, Lundbäck B. Update of prevalence of self-reported allergic rhinitis and chronic nasal symptoms among adults in Sweden. The clinical respiratory journal. 2012;**6**(3):159-68.

53. Jones N, Smith P, Carney A, Davis A. The prevalence of allergic rhinitis and nasal symptoms in Nottingham. Clinical Otolaryngology & Allied Sciences. 1998;**23**(6):547-54.

54. Hastan D, Fokkens W, Bachert C, Newson R, Bislimovska J, Bockelbrink A, et al. Chronic rhinosinusitis in Europe–an underestimated disease. A GA2LEN study. Allergy. 2011;**66**(9):1216-23.

55. Bengtsson C, Jonsson L, Holmström M, Hellgren J, Franklin K, Gíslason T, et al. Incident chronic rhinosinusitis is associated with impaired sleep quality: results of the RHINE study. Journal of Clinical Sleep Medicine. 2019;**15**(6):899-905.

56. Westman M, Stjärne P, Bergström A, Kull I, Toskala E, Cardell LO, et al. Chronic rhinosinusitis is rare but bothersome in adolescents from a Swedish population-based cohort. Journal of Allergy and Clinical Immunology. 2015;**136**(2):512-4. e6.

57. Snider G. Changes in COPD occurrence; chronic obstructive pulmonary disease. A definition and implication of structural determinants of airflow obstruction for epidemiology. Am Rev Respir Dis. 1989;**140**:s3-s8.

58. Lötvall 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;**127**(2):355-60.

59. Han MK, Agusti A, Calverley PM, Celli BR, Criner G, Curtis JL, et al. Chronic Obstructive Pulmonary Disease Phenotypes. American journal of respiratory and critical care medicine. 2010;**182**(5):598-604.

60. Rice JP, Saccone NL, Rasmussen E. Definition of the phenotype. Adv Genet. 2001;**42**:69-76.

61. Magda S. Susceptibility to exacerbation in chronic obstructive pulmonary disease–Data from the ECLIPSE study. Maedica. 2010;**5**(3):223.

62. Bel EH. Clinical phenotypes of asthma. Current opinion in pulmonary medicine. 2004;**10**(1):44-50.

63. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li HS, Li XN, et al. Identification of Asthma Phenotypes Using Cluster Analysis in the Severe Asthma Research Program. American journal of respiratory and critical care medicine. 2010;**181**(4):315-23.

64. Wenzel SE. Asthma phenotypes: the evolution from clinical to molecular approaches. Nature medicine. 2012;**18**(5):716-25.

65. Corlateanu A, Mendez Y, Wang Y, Garnica RdJA, Botnaru V, Siafakas N. Chronic obstructive pulmonary disease and phenotypes: a state-of-the-art. Pulmonology. 2020;**26**(2):95-100.

66. Loscalzo J, Barabasi AL. Systems biology and the future of medicine. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2011;**3**(6):619-27.

67. Vanfleteren LE, Kocks JW, Stone IS, Breyer-Kohansal R, Greulich T, Lacedonia D, et al. Moving from the Oslerian paradigm to the post-genomic era: are asthma and COPD outdated terms? Thorax. 2014;**69**(1):72-9.

68. Prosperi MC, Marinho S, Simpson A, Custovic A, Buchan IE. Predicting phenotypes of asthma and eczema with machine learning. BMC medical genomics. 2014;**7**(1):S7.

69. Prosperi MC, Sahiner UM, Belgrave D, Sackesen C, Buchan IE, Simpson A, et al. Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. American journal of respiratory and critical care medicine. 2013;**188**(11):1303-12.

70. Pathak J, Kho AN, Denny JC. Electronic health records-driven phenotyping: challenges, recent advances, and perspectives. Journal of the American Medical Informatics Association. 2013;**20**(e2):e206-e11.

71. Bashir MBA, Basna R, Zhang G-Q, Backman H, Lindberg A, Ekerljung L, et al. Computational phenotyping of obstructive airway diseases: protocol for a systematic review. Systematic Reviews. 2022;**11**(1):216.

72. Deliu M, Yavuz S, Sperrin M, Belgrave D, Sackesen C, Sahiner U, et al. P120 Challenges in using hierarchical clustering to identify asthma subtypes: choosing the variables and variable transformation. BMJ Publishing Group Ltd; 2016.

73. Deliu MY, T. S.; Sperrin, M.; Belgrave, D.; Sahiner, U. M.; Sackesen, C.; Kalayci, O.; Custovic, A. Features of asthma which provide meaningful insights for understanding the disease heterogeneity. Clin Exp Allergy. 2018;**48**(1):39-47.

74. Burgel PR. From COPD definitions to COPD phenotypes. Presse Med. 2014;**43**(12 Pt 1):1337-43.

75. 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. Respiratory research. 2016;**17**(1):165.

76. Zhang X, Xia T, Lai Z, Zhang Q, Guan Y, Zhong N. Uncontrolled asthma phenotypes defined from parameters using quantitative CT analysis. Eur Radiol. 2019;**29**(6):2848-58.

77. Kim JH, Chang HS, Shin SW, Baek DG, Son JH, Park CS, Park JS. Lung Function Trajectory Types in Never-Smoking Adults With Asthma: Clinical

Features and Inflammatory Patterns. Allergy Asthma Immunol Res. 2018;**10**(6):614-27.

78. Kim MA, Shin SW, Park JS, Uh ST, Chang HS, Bae DJ, et al. Clinical Characteristics of Exacerbation-Prone Adult Asthmatics Identified by Cluster Analysis. Allergy Asthma Immunol Res. 2017;**9**(6):483-90.

79. Zaihra T, Walsh CJ, Ahmed S, Fugere C, Hamid QA, Olivenstein R, et al. Phenotyping of difficult asthma using longitudinal physiological and biomarker measurements reveals significant differences in stability between clusters. BMC Pulm Med. 2016;**16**(1):74.

80. Nagasaki T, Matsumoto H, Kanemitsu Y, Izuhara K, Tohda Y, Kita H, et al. Integrating longitudinal information on pulmonary function and inflammation using asthma phenotypes. Journal of Allergy and Clinical Immunology. 2014;**133**(5):1474-U406.

81. Folz RJ, Myers J, Jorayeva A, Folz RJ. Phenotyping Older Adults with Asthma Via Clinical Cluster Analysis. American journal of respiratory and critical care medicine. 2018;**197**:2.

82. Sakagami T, Hasegawa T, Koya T, Furukawa T, Kawakami H, Hoshino Y, et al. Identification Of Clinical Asthma Phenotypes By Using Cluster Analysis With Simple Measurable Variables In Japanese Population. American journal of respiratory and critical care medicine. 2011;**183**:1.

83. Siroux V, Basagaña X, Boudier A, Pin I, Garcia-Aymerich J, Vesin A, et al. Identifying adult asthma phenotypes using a clustering approach. Pneumologia. 2011;**60**(3):166-73.

84. 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. American journal of respiratory and critical care medicine. 2013;**188**(5):550-60.

85. 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. European Respiratory Journal. 2013;**41**(6):1308-14.

86. Ilmarinen P, Tuomisto LE, Niemela O, Tommola M, Haanpaa J, Kankaanranta H. Cluster Analysis on Longitudinal Data of Patients with Adult-Onset Asthma. Journal of Allergy and Clinical Immunology-in Practice. 2017;**5**(4):967-78.

87. Kim MA, Shin SW, Park JS, Uh ST, Chang HS, Bae DJ, et al. Clinical characteristics of exacerbation-prone adult asthmatics identified by cluster analysis. Allergy, Asthma and Immunology Research. 2017;**9**(6):483-90.

88. Kim JH, Chang HS, Shin SW, Baek DG, Son JH, Park CS, Park JS. Lung function trajectory types in never-smoking adults with asthma: Clinical features and inflammatory patterns. Allergy, Asthma and Immunology Research. 2018;**10**(6):614-27.

89. Wang L, Liang R, Zhou T, Zheng J, Liang BM, Zhang HP, et al. Identification and validation of asthma phenotypes in Chinese population using cluster analysis. Annals of Allergy Asthma & Immunology. 2017;**119**(4):324-32.

90. Gonem S, Natarajan S, Hartley R, Gupta S, Desai D, Corkill S, et al. Cluster analysis reveals a distinct small airway-predominant phenotype of asthma. Thorax. 2012;**67**:A7-A.

91. Seys SF, Scheers H, Van den Brande P, Marijsse G, Dilissen E, Van den Bergh A, et al. Cluster analysis of sputum cytokine-high profiles reveals diversity in T(h) 2-high asthma patients. Respiratory Research. 2017;**18**:10.

92. Hsiao HP, Lin MC, Wu CC, Wang CC, Wang TN. Sex-Specific Asthma Phenotypes, Inflammatory Patterns, and Asthma Control in a Cluster Analysis. Journal of Allergy and Clinical Immunology-in Practice. 2019;**7**(2):556-67.

93. Tay TR, Choo XN, Yii A, Chung KF, Chan YH, Wong HS, et al. Asthma phenotypes in a multi-ethnic Asian cohort. Respiratory medicine. 2019;**157**:42-8.

94. Jeong A, Imboden M, Hansen S, Zemp E, Bridevaux PO, Lovison G, et al. Heterogeneity of obesity-asthma association disentangled by latent class analysis, the SAPALDIA cohort. Respiratory medicine. 2017;**125**:25-32.

95. Kaneko Y, Masuko H, Sakamoto T, Iijima H, Naito T, Yatagai Y, et al. Asthma phenotypes in japanese adults - their associations with the CCL5 and ADRB2 genotypes. Allergology International. 2013;**62**(1):113-21.

96. Celejewska-Wojcik N, Wojcik K, Ignacak-Popiel M, Cmiel A, Tyrak K, Gielicz A, et al. Subphenotypes of NSAID-exacerbated respiratory disease identified by latent class analysis. Allergy. 2019.

97. Damiens K, Paquet J, Lemiere C. Identification Of Work-Related Asthma Phenotypes By Cluster Analysis. American journal of respiratory and critical care medicine. 2013;**187**:1.

98. Bochenek G, Kuschill-Dziurda J, Szafraniec K, Plutecka H, Szczeklik A, Nizankowska-Mogilnicka E. Certain subphenotypes of aspirin-exacerbated respiratory disease distinguished by latent class analysis. Journal of Allergy and Clinical Immunology. 2014;**133**(1):98-103.

99. Seino Y, Hasegawa T, Koya T, Sakagami T, Mashima I, Shimizu N, et al. A Cluster Analysis of Bronchial Asthma Patients with Depressive Symptoms. Internal Medicine. 2018;**57**(14):1967-75.

100. Wu DW, Bleier BS, Li L, Zhan XJ, Zhang LC, Lv QW, et al. Clinical Phenotypes of Nasal Polyps and Comorbid Asthma Based on Cluster Analysis of Disease History. Journal of Allergy and Clinical Immunology-in Practice. 2018;**6**(4):1297-305.

101. Boudier AC, S.; Accordini, S.; Anto, J. M.; Basagana, X.; Bousquet, J.; Demoly, P.; Garcia-Aymerich, J.; Gormand, F.; Heinrich, J.; Janson, C.; Kunzli, N.; Matran, R.; Pison, C.; Raherison, C.; Sunyer, J.; Varraso, R.; Jarvis, D.; Leynaert, B.; Pin, I.; Siroux, V. Data-driven adult asthma phenotypes based on clinical characteristics are associated with asthma outcomes twenty years later. Allergy. 2019;**74**(5):953-63.

102. Peters SP. Asthma phenotypes: nonallergic (intrinsic) asthma. The journal of allergy and clinical immunology In practice. 2014;**2**(6):650-2.

103. Dixon AE, Shade DM, Cohen RI, Skloot GS, Holbrook JT, Smith LJ, et al. Effect of obesity on clinical presentation and response to treatment in asthma. J Asthma. 2006;**43**(7):553-8.

104. Benton AS, Wang ZY, Lerner J, Foerster M, Teach SJ, Freishtat RJ. Overcoming Heterogeneity in Pediatric Asthma: Tobacco Smoke and Asthma Characteristics Within Phenotypic Clusters in an African American Cohort. Journal of Asthma. 2010;**47**(7):728-34.

105. Amelink M, De Nijs SB, De Groot JC, Van Tilburg PMB, Van Spiegel PI, Krouwels FH, et al. Three phenotypes of adult-onset asthma. Allergy. 2013;**68**(5):674-80.

106. Schatz M, Hsu JWY, Zeiger RS, Chen WS, Dorenbaum A, Chipps BE, Haselkorn T. Phenotypes determined by cluster analysis in severe or difficult-to-treat asthma. Journal of Allergy and Clinical Immunology. 2014;**133**(6):1549-56.

107. Holguin F, Bleecker ER, Busse WW, Calhoun WJ, Castro M, Erzurum SC, et al. Obesity and asthma: an association modified by age of asthma onset. J Allergy Clin Immunol. 2011;**127**(6):1486-93 e2.

108. Dixon AE, Poynter ME. Mechanisms of Asthma in Obesity Pleiotropic Aspects of Obesity Produce Distinct Asthma Phenotypes. Am J Resp Cell Mol. 2016;**54**(5):601-8.

109. Zoratti EMK, R. Z.; Babineau, D. C.; Pongracic, J. A.; O'Connor, G. T.; Wood, R. A.; Khurana Hershey, G. K.; Kercsmar, C. M.; Gruchalla, R. S.; Kattan, M.; Teach, S. J.; Sigelman, S. M.; Gergen, P. J.; Togias, A.; Visness, C. M.; Busse, W. W.; Liu, A. H. Asthma phenotypes in inner-city children. J Allergy Clin Immunol. 2016;**138**(4):1016-29.

110. Tsukioka K, Koya T, Ueno H, Hayashi M, Sakagami T, Hasegawa T, et al. Phenotypic analysis of asthma in Japanese athletes. Allergology International. 2017;**66**(4):550-6.

111. Gibson P, Simpson J. The overlap syndrome of asthma and COPD: what are its features and how important is it? Thorax. 2009;**64**(8):728-35.

112. 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. European Respiratory Journal. 2016;**48**(3):664-73.

113. Cataldo D, Corhay J-L, Derom E, Louis R, Marchand E, Michils A, et al. A Belgian survey on the diagnosis of asthma–COPD overlap syndrome. International journal of chronic obstructive pulmonary disease. 2017:601-13.

114. Cosio BG, Soriano JB, Lopez-Campos JL, Calle-Rubio M, Soler-Cataluna JJ, de-Torres JP, et al. Defining the Asthma-COPD Overlap Syndrome in a COPD Cohort. Chest. 2016;**149**(1):45-52.

115. Han MK. Asthma and COPD overlap (ACO). UpToDate. Waltham, MA: UpToDate; 2023.

116. McDonald VM, Gibson PG. Treatable traits in asthma: moving beyond diagnostic labels. Medical Journal of Australia. 2022;**216**(7):331-3.

117. Meilan King Han, Mark T Drandfield, Martines FJ. Chronic obstructive pulmonary disease: Diagnosis and staging. UpToDate. Waltham, MA: UpToDate; 2023.

118. Mirza S, Benzo R, editors. Chronic obstructive pulmonary disease phenotypes: implications for care. Mayo Clinic Proceedings; 2017: Elsevier.

119. Marçôa R, Rodrigues DM, Dias M, Ladeira I, Vaz AP, Lima R, Guimarães M. Classification of chronic obstructive pulmonary disease (COPD) according to the new Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2017: comparison with GOLD 2011. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2018;**15**(1):21-6.

120. Nishimura M, Makita H, Nagai K, Konno S, Nasuhara Y, Hasegawa M, et al. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. American journal of respiratory and critical care medicine. 2012;**185**(1):44-52.

121. Schols AM, Broekhuizen R, Weling-Scheepers CA, Wouters EF. Body composition and mortality in chronic obstructive pulmonary disease. The American journal of clinical nutrition. 2005;**82**(1):53-9.

122. Vestbo J, Prescott E, Almdal T, Dahl M, Nordestgaard BG, Andersen T, et al. Body mass, fat-free body mass, and prognosis in patients with chronic obstructive pulmonary disease from a random population sample: findings from the Copenhagen City Heart Study. American journal of respiratory and critical care medicine. 2006;**173**(1):79-83.

123. Camp PG, Ramirez-Venegas A, Sansores RH, Alva LF, McDougall JE, Sin DD, et al. COPD phenotypes in biomass smoke-versus tobacco smoke-exposed Mexican women. European Respiratory Journal. 2014;**43**(3):725-34.

124. Golpe R, Suárez-Valor M, Martín-Robles I, Sanjuán-López P, Cano-Jiménez E, Castro-Añón O, Pérez de Llano LA. Mortality in COPD patients according to clinical phenotypes. International journal of chronic obstructive pulmonary disease. 2018:1433-9.

125. Ji W, Lim MN, Bak SH, Hong SH, Han SS, Lee SJ, et al. Differences in chronic obstructive pulmonary disease phenotypes between non-smokers and smokers. The clinical respiratory journal. 2018;**12**(2):666-73.

126. González-García M, Torres-Duque C. Wood smoke COPD: a new description of a COPD phenotype? European Respiratory Journal. 2014;**44**(1):261-2.

127. Torres-Duque CA, García-Rodriguez MC, González-García M. Is chronic obstructive pulmonary disease caused by wood smoke a different phenotype

or a different entity? Archivos de Bronconeumología (English Edition). 2016;**52**(8):425-31.

128. Perret JL, Abramson MJ. Biomass smoke COPD: A phenotype or a different disease? : Wiley Online Library; 2018. p. 124-5.

129. Bousquet J. World Health Organization; GA2LEN; AllerGen. Allergic Rhinitis and its Impact on Asthma (ARIA) 2008 update (in collaboration with the World Health Organization, GA2LEN and AllerGen. Allergy. 2008;**63**(86):8-160.

130. Group IRMW. International consensus report on the diagnosis and management of rhinitis. Allergy. 1994;**49**(19):5-34.

131. Bousquet J, Van Cauwenberge P, Khaltaev N. Allergic rhinitis and its impact on asthma. Journal of allergy and clinical immunology. 2001;**108**(5):S147-S334.

132. Valero A, Ferrer M, Sastre J, Navarro AM, Monclús L, Martí-Guadaño E, et al. A new criterion by which to discriminate between patients with moderate allergic rhinitis and patients with severe allergic rhinitis based on the Allergic Rhinitis and its Impact on Asthma severity items. Journal of allergy and clinical immunology. 2007;**120**(2):359-65.

133. Rondón C, Campo P, Togias A, Fokkens WJ, Durham SR, Powe DG, et al. Local allergic rhinitis: concept, pathophysiology, and management. Journal of allergy and clinical immunology. 2012;**129**(6):1460-7.

134. Rondón C, Dona I, López S, Campo P, Romero J, Torres M, et al. Seasonal idiopathic rhinitis with local inflammatory response and specific IgE in absence of systemic response. Allergy. 2008;**63**(10):1352-8.

135. Rondón C, Eguiluz-Gracia I, Campo P. Is the evidence of local allergic rhinitis growing? Current Opinion in Allergy and Clinical Immunology. 2018;**18**(4):342-9.

136. Folletti I, Paolocci G, Muzi G, Siracusa A. United airway disease in occupational allergy. Current Treatment Options in Allergy. 2017;**4**:160-70.

137. Mungan D. Occupational allergic rhinitis: what do we know? Current Treatment Options in Allergy. 2015;**2**:10-9.

138. Settipane RA. Other causes of rhinitis: mixed rhinitis, rhinitis medicamentosa, hormonal rhinitis, rhinitis of the elderly, and gustatory rhinitis. Immunology and Allergy Clinics. 2011;**31**(3):457-67.

139. Bernstein JA, Singh U. Neural abnormalities in nonallergic rhinitis. Current allergy and asthma reports. 2015;**15**:1-12.

140. Canonica G, Compalati E. Minimal persistent inflammation in allergic rhinitis: implications for current treatment strategies. Clinical & Experimental Immunology. 2009;**158**(3):260-71.

141. Hellings PW, Klimek L, Cingi C, Agache I, Akdis C, Bachert C, et al. Non-allergic rhinitis: position paper of the European Academy of Allergy and Clinical Immunology. Allergy. 2017;**72**(11):1657-65.

142. Pal I, Babu AS, Halder I, Kumar S. Nasal smear eosinophils and allergic rhinitis. Ear, Nose & Throat Journal. 2017;**96**(10-11):E17-E22.

143. Gelardi M, Landi M, Ciprandi G. The pragmatic role of nasal cytology: a point-of-care testing to implement precision medicine in clinical practice. Revista Alergia México. 2018;**65**(3):259-63.

144. De Corso E, Anzivino R, Galli J, Baroni S, Di Nardo W, De Vita C, et al. Antileukotrienes improve naso-ocular symptoms and biomarkers in patients with NARES and asthma. The Laryngoscope. 2019;**129**(3):551-7.

145. De Corso E, Battista M, Pandolfini M, Liberati L, Baroni S, Romanello M, et al. Role of inflammation in non-allergic rhinitis. Rhinology. 2014;**52**(2):142-9.

146. Cassano M, Cassano P, Mora F, Ciprandi G. Nares patients have more respiratory infections than allergic subjects. International Journal of Immunopathology and Pharmacology. 2011;**24**(3):813-5.

147. Gelardi M, Iannuzzi L, Quaranta N, Landi M, Passalacqua G. NASAL cytology: practical aspects and clinical relevance. Clinical & Experimental Allergy. 2016;**46**(6):785-92.

148. Ibáñez MD, Valero AL, Montoro J, Jauregui I, Ferrer M, Dávila I, et al. Analysis of comorbidities and therapeutic approach for allergic rhinitis in a pediatric population in Spain. Pediatric Allergy and Immunology. 2013;**24**(7):678-84.

149. Bergqvist J, Andersson A, Olin A-C, Murgia N, Schiöler L, Bove M, Hellgren J. New evidence of increased risk of rhinitis in subjects with COPD: a longitudinal population study. International journal of chronic obstructive pulmonary disease. 2016:2617-23.

150. Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, Reitsma S, et al. European position paper on rhinosinusitis and nasal polyps 2020. Rhinology: official organ of the International rhinologic society. 2020.

151. Orlandi RR, Kingdom TT, Hwang PH, Smith TL, Alt JA, Baroody FM, et al., editors. International consensus statement on allergy and rhinology: rhinosinusitis. International forum of allergy & rhinology; 2016: Wiley Online Library.

152. Kowalski ML, Agache I, Bavbek S, Bakirtas A, Blanca M, Bochenek G, et al. Diagnosis and management of NSAID-Exacerbated Respiratory Disease (N-ERD)—a EAACI position paper. Allergy. 2019;**74**(1):28-39.

153. Mullol J, Picado C. Rhinosinusitis and nasal polyps in aspirin-exacerbated respiratory disease. Immunology and Allergy Clinics. 2013;**33**(2):163-76.

154. Organization WH. Social determinants of health. WHO Regional Office for South-East Asia; 2008.

155. Dahlgren G, Whitehead M. European strategies for tackling social inequities in health: Levelling up Part 2. WHO Regional office for Europe Copenhagen; 2006.

156. Cirino PT, Chin CE, Sevcik RA, Wolf M, Lovett M, Morris RD. Measuring socioeconomic status: reliability and preliminary validity for different approaches. Assessment. 2002;**9**(2):145-55.

157. Sullivan K, Thakur N. Structural and Social Determinants of Health in Asthma in Developed Economies: a Scoping Review of Literature Published Between 2014 and 2019. Current allergy and asthma reports. 2020;**20**(2):5-.

158. Link BG, Phelan J. Social Conditions As Fundamental Causes of Disease. Journal of health and social behavior. 1995:80-94.

159. Trinder PM, Croft PR, Lewis M. Social class, smoking and the severity of respiratory symptoms in the general population. Journal of Epidemiology and Community Health. 2000;**54**(5):340.

160. Sullivan K, Thakur N. Structural and social determinants of health in asthma in developed economies: a scoping review of literature published between 2014 and 2019. Current allergy and asthma reports. 2020;**20**(2):1-12.

161. Schyllert C, Lindberg A, Hedman L, Stridsman C, Andersson M, Ilmarinen P, et al. Low socioeconomic status relates to asthma and wheeze, especially in women. ERJ open research. 2020;**6**(3).

162. Li X, Sundquist J, Sundquist K. Socioeconomic and occupational groups and risk of asthma in Sweden. Occupational Medicine. 2008;**58**(3):161-8.

163. Basagaña X, Sunyer J, Kogevinas M, Zock J-P, Duran-Tauleria E, Jarvis D, et al. Socioeconomic status and asthma prevalence in young adults: the European Community Respiratory Health Survey. American journal of epidemiology. 2004;**160**(2):178-88.

164. Schyllert C, Andersson M, Hedman L, Ekström M, Backman H, Lindberg A, Rönmark E. Job titles classified into socioeconomic and occupational groups identify subjects with increased risk for respiratory symptoms independent of occupational exposure to vapour, gas, dust, or fumes. European clinical respiratory journal. 2018;**5**(1):1468715.

165. Thomson NC, Polosa R, Sin DD. Cigarette Smoking and Asthma. The Journal of Allergy and Clinical Immunology: In Practice. 2022.

166. Strachan DP, Cook DG. Parental smoking and childhood asthma: longitudinal and case-control studies. Thorax. 1998;**53**(3):204-12.

167. Tommola M, Ilmarinen P, Tuomisto LE, Haanpää J, Kankaanranta T, Niemelä O, Kankaanranta H. The effect of smoking on lung function: a clinical study of adult-onset asthma. European Respiratory Journal. 2016;**48**(5):1298-306.

168. Hedman L, Bjerg A, Sundberg S, Forsberg B, Rönmark E. Both environmental tobacco smoke and personal smoking is related to asthma and wheeze in teenagers. Thorax. 2011;**66**(1):20-5.

169. Bakakos P, Kostikas K, Loukides S. Smoking asthma phenotype: diagnostic and management challenges. Current opinion in pulmonary medicine. 2016;**22**(1):53-8.

170. Hekking P-PW, Bel EH. Developing and Emerging Clinical Asthma Phenotypes. The Journal of Allergy and Clinical Immunology: In Practice. 2014;**2**(6):671-80.

171. Gilliland FD, Islam T, Berhane K, Gauderman WJ, McConnell R, Avol E, Peters JM. Regular Smoking and Asthma Incidence in Adolescents. American journal of respiratory and critical care medicine. 2006;**174**(10):1094-100.

172. Chen Y, Dales R, Jiang Y. The Association Between Obesity and Asthma Is Stronger in Nonallergic Than Allergic Adults. Chest. 2006;**130**(3):890-5.

173. Emanuel M. Hay fever, a post industrial revolution epidemic: a history of its growth during the 19th century. Clinical & Experimental Allergy. 1988;**18**(3):295-304.

174. Heinrich J, Popescu MA, Wjst M, Goldstein IF, Wichmann H-E. Atopy in children and parental social class. American Journal of Public Health. 1998;**88**(9):1319-24.

175. Mercer M, Joubert G, Ehrlich R, Nelson H, Poyser M, Puterman A, Weinberg E. Socioeconomic status and prevalence of allergic rhinitis and atopic eczema symptoms in young adolescents. Pediatric allergy and immunology. 2004;**15**(3):234-41.

176. Lee K-S, Rha Y-H, Oh I-H, Choi Y-S, Choi S-H. Socioeconomic and sociodemographic factors related to allergic diseases in Korean adolescents based on the Seventh Korea Youth Risk Behavior Web-based Survey: a cross-sectional study. BMC pediatrics. 2016;**16**(1):1-9.

177. Hisinger-Mölkänen H, Piirilä P, Haahtela T, Sovijärvi A, Pallasaho P. Smoking, environmental tobacco smoke and occupational irritants increase the risk of chronic rhinitis. World Allergy Organization Journal. 2018;**11**(1):1-7.

178. Mady LJ, Schwarzbach HL, Moore JA, Boudreau RM, Willson TJ, Lee SE, editors. Air pollutants may be environmental risk factors in chronic rhinosinusitis disease progression. International Forum of Allergy & Rhinology; 2018: Wiley Online Library.

179. Fukuoka A, Yoshimoto T. Barrier dysfunction in the nasal allergy. Allergology International. 2018;**67**(1):18-23.

180. van den Brekel L, Lenters V, Mackenbach JD, Hoek G, Wagtendonk A, Lakerveld J, et al. Ethnic and socioeconomic inequalities in air pollution exposure: a cross-sectional analysis of nationwide individual-level data from the Netherlands. The Lancet Planetary Health. 2024;**8**(1):e18-e29.

181. Geramas I, Terzakis D, Hatzimanolis E, Georgalas C. Social factors in the development of chronic rhinosinusitis: a systematic review. Current allergy and asthma reports. 2018;**18**:1-18.

182. Fu Q-L, Ma J-X, Ou C-Q, Guo C, Shen S-Q, Xu G, Shi J. Influence of self-reported chronic rhinosinusitis on health-related quality of life: a population-based survey. PloS one. 2015;**10**(5):e0126881.

183. Gao W-X, Ou C-Q, Fang S-B, Sun Y-Q, Zhang H, Cheng L, et al. Occupational and environmental risk factors for chronic rhinosinusitis in China: a multicentre cross-sectional study. Respiratory research. 2016;**17**:1-7.

184. Thilsing T, Rasmussen J, Lange B, Kjeldsen AD, Al-Kalemji A, Baelum J. Chronic rhinosinusitis and occupational risk factors among 20-to 75-year-old Danes—A GA2LEN-based study. American journal of industrial medicine. 2012;**55**(11):1037-43.

185. Koh DH, Kim HR, Han SS. The relationship between chronic rhinosinusitis and occupation: the 1998, 2001, and 2005 Korea National health and nutrition examination survey (KNHANES). American journal of industrial medicine. 2009;**52**(3):179-84.

186. Ahlroth Pind C, Gunnbjörnsdottír M, Bjerg A, Järvholm B, Lundbäck B, Malinovschi A, et al. Patient-reported signs of dampness at home may be a risk factor for chronic rhinosinusitis: A cross-sectional study. Clinical & Experimental Allergy. 2017;**47**(11):1383-9.

187. Bergmark RW, Hoehle LP, Chyou D, Phillips KM, Caradonna DS, Gray ST, Sedaghat AR. Association of socioeconomic status, race and insurance status with chronic rhinosinusitis patient-reported outcome measures. Otolaryngology–Head and Neck Surgery. 2018;**158**(3):571-9.

188. Forey BA, Thornton AJ, Lee PN. Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. BMC Pulm Med. 2011;**11**(1):1-61.

189. Kamga A, Rochefort-Morel C, Le Guen Y, Ouksel H, Pipet A, Leroyer C. Asthma and smoking: A review. Respiratory Medicine and Research. 2022;**82**:100916.

190. Saulyte J, Regueira C, Montes-Martínez A, Khudyakov P, Takkouche B. Active or passive exposure to tobacco smoking and allergic rhinitis, allergic

dermatitis, and food allergy in adults and children: a systematic review and meta-analysis. PLoS medicine. 2014;**11**(3):e1001611.

191. Starobin D, Bargutin M, Rosenberg I, Yamolovsky A, Levi T, Fink G. Asthma control and compliance in a cohort of adult asthmatics: first survey in Israel. IMAJ-RAMAT GAN-. 2007;**9**(5):358.

192. Ulrik C, Frederiksen J. Mortality and fatal risk factors in 1075 outpatients treated for asthma. Ugeskrift for Laeger. 1996;**158**(25):3608-11.

193. Vignoud L, Pin I, Boudier A, Pison C, Nadif R, Le Moual N, et al. Smoking and asthma: disentangling their mutual influences using a longitudinal approach. Respiratory medicine. 2011;**105**(12):1805-14.

194. Bellou V, Gogali A, Kostikas K. Asthma and tobacco smoking. Journal of Personalized Medicine. 2022;**12**(8):1231.

195. Graff S, Bricmont N, Moermans C, Henket M, Paulus V, Guissard F, et al. Clinical and biological factors associated with irreversible airway obstruction in adult asthma. Respiratory medicine. 2020;**175**:106202.

196. Van De Ven MO, van Zundert RM, Engels RC. Effects of asthma on nicotine dependence development and smoking cessation attempts in adolescence. Journal of Asthma. 2013;**50**(3):250-9.

197. Bakakos A, Vogli S, Dimakou K, Hillas G. Asthma with fixed airflow obstruction: from fixed to personalized approach. Journal of Personalized Medicine. 2022;**12**(3):333.

198. Lanckacker EA, Tournoy KG, Hammad H, Holtappels G, Lambrecht BN, Joos GF, Maes T. Short cigarette smoke exposure facilitates sensitisation and asthma development in mice. European Respiratory Journal. 2013;**41**(5):1189-99.

199. Muddaluru V, Valenta R, Vrtala S, Schlederer T, Hindley J, Hickey P, et al. Comparison of house dust mite sensitization profiles in allergic adults from Canada, Europe, South Africa and USA. Allergy. 2021;**76**(7):2177-88.

200. McLeish AC, Zvolensky MJ. Asthma and cigarette smoking: a review of the empirical literature. Journal of Asthma. 2010;**47**(4):345-61.

201. Hiscock R, Bauld L, Amos A, Fidler JA, Munafò M. Socioeconomic status and smoking: a review. Annals of the New York Academy of Sciences. 2012;**1248**(1):107-23.

202. Grant T, Croce E, Matsui EC. Asthma and the social determinants of health. Annals of Allergy, Asthma & Immunology. 2022;**128**(1):5-11.

203. Schikowski T, Sugiri D, Reimann V, Pesch B, Ranft U, Krämer U. Contribution of smoking and air pollution exposure in urban areas to social differences in respiratory health. BMC Public Health. 2008;**8**:1-10.

204. Trinder PM, Croft PR, Lewis M. Social class, smoking and the severity of respiratory symptoms in the general population. Journal of Epidemiology & Community Health. 2000;**54**(5):340-3.

205. Herr M, Just J, Nikasinovic L, Foucault C, Le Marec A-M, Giordanella J-P, Momas I. Risk factors and characteristics of respiratory and allergic phenotypes in early childhood. Journal of Allergy and Clinical Immunology. 2012;**130**(2):389-96. e4.

206. Sullivan K, Thakur N. Structural and social determinants of health in asthma in developed economies: a scoping review of literature published between 2014 and 2019. Current allergy and asthma reports. 2020;**20**:1-12.

207. Bråbäck L, Hjern A, Rasmussen F. Social class in asthma and allergic rhinitis: a national cohort study over three decades. European Respiratory Journal. 2005;**26**(6):1064-8.

208. Chittleborough CR, Taylor AW, Dal Grande E, Gill TK, Grant JF, Adams RJ, et al. Gender differences in asthma prevalence: variations with socioeconomic disadvantage. Respirology. 2010;**15**(1):107-14.

209. Masoompour SM, Mahdaviazad H, Ghayumi SMA. Asthma and its related socioeconomic factors: the Shiraz Adult Respiratory Disease Study 2015. The Clinical Respiratory Journal. 2018;**12**(6):2110-6.

210. Statistics Sweden. Population by region, marital status, age and sex. Year 1968 - 2022 2023. Available from: https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101A/BefolkningNy/.

211. West Sweden. 2023. Available from: https://www.vastsverige.com/.

212. Statistics Sweden. Total earned income, mean income for persons registered in the national population register during the whole year by region, sex, age, income bracket and year 2023. Available from:

https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__HE__HE0110__HE0110A/SamForvInk1/table/tableViewLayout1/.

213. Statistics Sweden. Living Conditions Surveys, employment (gainfully employed persons aged 16-74 years). Proportion of persons in percent by study domain, sex, period and indicator 2023. Available from: https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__LE__LE0101__LE0101S/LE01012021S10/table/tableViewLayout1/.

214. Regionfakta. Norrbotten county, facts & prespectives. 2023.

215. Nwaru BI, Ekerljung L, Rådinger M, Bjerg A, Mincheva R, Malmhäll C, et al. Cohort profile: the West Sweden Asthma Study (WSAS): a multidisciplinary population-based longitudinal study of asthma, allergy and respiratory conditions in adults. BMJ open. 2019;**9**(6):e027808-e.

216. Barón NB, Kankaanranta H, Hedman L, Andersson M, Stridsman C, Lindberg A, et al. Body mass index increase: a risk factor for forced expiratory volume in 1 s decline for overweight and obese adults with asthma. ERJ Open Research. 2022;**8**(4).

217. Schyllert C, Rönmark E, Andersson M, Hedlund U, Lundbäck B, Hedman L, Lindberg A. Occupational exposure to chemicals drives the increased risk of asthma and rhinitis observed for exposure to vapours, gas, dust and fumes: a cross-sectional population-based study. Occupational and environmental medicine. 2016;**73**(10):663-9.

218. Kankaanranta H, Ilmarinen P, Kankaanranta T, Tuomisto LE. Seinäjoki Adult Asthma Study (SAAS): a protocol for a 12-year real-life follow-up study of new-onset asthma diagnosed at adult age and treated in primary and specialised care. npj Primary Care Respiratory Medicine. 2015;**25**(1):15042.

219. International Labour Organization. International Classification of Occupations (ISCO) 2023. Available from: https://ilostat.ilo.org/resources/concepts-and-definitions/classification-occupation/.

220. Statistics Sweden. The swedish socioeconomic classifcation system (SEI) 1982 [cited 2023 15.1]. Available from: https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-yrkesklassificering-ssyk/.

221. Statistics Sweden. The Swedish standard classification of occupations 2022 [cited 2021 19.2]. Available from:

https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-yrkesklassificering-ssyk/.

222. Ekerljung L, Bossios A, Lötvall J, Olin A-C, Rönmark E, Wennergren G, et al. Multi-symptom asthma as an indication of disease severity in epidemiology. Clinical and Translational Allergy. 2013;**3**:1-.

223. Reddel HK, Bacharier LB, Bateman ED, Brightling CE, Brusselle GG, Buhl R, et al. Global Initiative for Asthma Strategy 2021: executive summary and rationale for key changes. American journal of respiratory and critical care medicine. 2022;**205**(1):17-35.

224. Bashir MBA, Basna R, Zhang G-Q, Backman H, Lindberg A, Ekerljung L, et al. Computational phenotyping of obstructive airway diseases: protocol for a systematic review. Systematic reviews. 2022;**11**(1):1-5.

225. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic reviews. 2015;**4**(1):1.

226. Scutari M. Learning Bayesian networks with the bnlearn R package 2009 [cited 2021 14.5].

227. Rothman KJ, Greenland S, Lash TL. Modern epidemiology: Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia; 2008.

228. VanderWeele TJ, Knol MJ. A tutorial on interaction. Epidemiologic Methods. 2014;**3**(1):33-72.

229. Richardson DB, Kaufman JS. Estimation of the Relative Excess Risk Due to Interaction and Associated Confidence Bounds. American Journal of Epidemiology. 2009;**169**(6):756-60.

230. Armstrong RA. When to use the Bonferroni correction. Ophthalmic and Physiological Optics. 2014;**34**(5):502-8.

231. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? International journal of methods in psychiatric research. 2011;**20**(1):40-9.

232. Wilson S. miceRanger: Multiple Imputation by Chained Equations with Random Forests 2021. Available from: https://CRAN.R-project.org/package=miceRanger.

233. Xie J, Girshick R, Farhadi A, editors. Unsupervised deep embedding for clustering analysis. International conference on machine learning; 2016: PMLR.

234. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. Journal of statistical software. 2014;**61**:1-36.

235. John CR, Watson D, Russ D, Goldmann K, Ehrenstein M, Pitzalis C, et al. M3C: Monte Carlo reference-based consensus clustering. Sci Rep. 2020;**10**(1).

236. John CR, Watson D, Russ D, Goldmann K, Ehrenstein M, Pitzalis C, et al. M3C: Monte Carlo reference-based consensus clustering. Sci Rep. 2020;**10**(1):1816.

237. Amelink M, de Nijs SB, De Groot JC, Van Tilburg PMB, Van Spiegel P, Krouwels FH, et al. Phenotypes of adult-onset asthma by cluster analysis. European Respiratory Journal. 2012;**40**:2.

238. Bhargava S, Mahesh PA, Holla AD, Jayaraj BS, Praveena AS, Ravi S, Khurana S. Cluster analysis identifies distinct clinical phenotypes with poor treatment responsiveness in asthma. European Respiratory Journal. 2018;**52**:2.

239. Chanoine S, Pin I, Sanchez M, Temam S, Pison C, Le Moual N, et al. Asthma Medication Ratio Phenotypes in Elderly Women. Journal of Allergy and Clinical Immunology-in Practice. 2018;**6**(3):897-906.

240. Couto M, Stang J, Horta L, Stensrud T, Severo M, Mowinckel P, et al. Two distinct phenotypes of asthma in elite athletes identified by latent class analysis. Journal of Asthma. 2015;**52**(9):897-904.

241. Cruz AA, Bansal AT, Ponte EV, Souza-Machado A, Almeida PC, Biao-Lima V, et al. Unbiased clinical cluster analysis of patients of ProAR Asthma Cohort. European Respiratory Journal. 2018;**52**:2.

242. Dudchenko LS, Savchenko VM. Cluster analysis classification of asthmatic pathologic manifestations during stay at the resort. Tuberculosis and Lung Diseases. 2018;**96**(2):16-21.

243. Hilvering B, Jansen J, Lammers JWJ, Koenderman L. Phenotyping Asthma Using An Unsupervised Prediction Model Based On Blood Granulocyte Responsiveness. American journal of respiratory and critical care medicine. 2015;**191**:1.

244. Kim S, Lim MN, Hong Y, Han SS, Lee SJ, Kim WJ. A cluster analysis of chronic obstructive pulmonary disease in dusty areas cohort identified three subgroups. BMC Pulm Med. 2017;**17**:9.

245. Liang ZY, Liu LY, Zhao HJ, Xia Y, Zhang WZ, Ye YM, et al. A Systemic Inflammatory Endotype of Asthma With More Severe Disease Identified by Unbiased Clustering of the Serum Cytokine Profile. Medicine. 2016;**95**(25):7.

246. Loureiro CC, Sa-Couto P, Todo-Bom A, Bousquet J. Cluster analysis in phenotyping a Portuguese population. Revista Portuguesa de Pneumologia (English Edition). 2015;**21**(6):299-306.

247. Loza MJ, Djukanovic R, Chung KF, Horowitz D, Ma KY, Branigan P, et al. Validated and longitudinally stable asthma phenotypes based on cluster analysis of the ADEPT study. Respiratory Research. 2016;**17**:21.

248. Makikyro EMS, Jaakkola MS, Jaakkola JJK. Subtypes of asthma based on asthma control and severity: a latent class analysis. Respiratory Research. 2017;**18**:11.

249. Mason PF, A. C.; Scarpa, M. C.; Maestrelli, P.; Guarnieri, G. Cluster analysis of occupational asthma caused by isocyanates. Journal of Allergy and Clinical Immunology. 2018;**142**(6):2011-2.

250. Mastalerz L, Celejewska-Wojcik N, Wojcik K, Gielicz A, Cmiel A, Ignacak M, et al. Induced sputum supernatant bioactive lipid mediators can identify subtypes of asthma. Clinical and Experimental Allergy. 2015;**45**(12):1779-89.

251. Qiu RH, Xie JX, Chung KF, Li NJ, Yang ZW, He MZ, et al. Asthma Phenotypes Defined From Parameters Obtained During Recovery From a Hospital-Treated Exacerbation. Journal of Allergy and Clinical Immunology-in Practice. 2018;**6**(6):1960-7.

252. Sendin-Hernandez MP, Avila-Zarza C, Sanz C, Garcia-Sanchez A, Marcos-Vadillo E, Munoz-Bellido FJ, et al. Cluster Analysis Identifies 3 Phenotypes within Allergic Asthma. Journal of Allergy and Clinical Immunology-in Practice. 2018;**6**(3):955-61.

253. Watanabe S, Koya T, Hasegawa T, Hayashi M, Sakagami T, Kimura Y, et al. Cluster Analysis Of Uncontrolled Asthma In Japanese Population. American journal of respiratory and critical care medicine. 2016;**193**:1.

254. Zaihra T, Walsh CJ, Ahmed S, Fugere C, Hamid QA, Olivenstein R, et al. Phenotyping of difficult asthma using longitudinal physiological and biomarker measurements reveals significant differences in stability between clusters. BMC Pulm Med. 2016;**16**:8.

255. Zhang XX, Xia TT, Lai ZD, Zhang QL, Guan YB, Zhong NS. Uncontrolled asthma phenotypes defined from parameters using quantitative CT analysis. European Radiology. 2019;**29**(6):2848-58.

256. Brew B, Chiesa F, Lundholm C, Almqvist C. Combining supervised and unsupervised models to characterize asthma phenotypes in children. European Respiratory Journal. 2018;**52**:2.

257. Cabral ALB, Sousa AW, Mendes FAR, De Carvalho CRF. Phenotypes of asthma in low income children and adolescents: cluster analysis. Jornal Brasileiro De Pneumologia. 2017;**43**(1):44-50.

258. Deliu M, Yavuz ST, Sperrin M, Sahiner U, Sackesen C, Custovic A, Kalayci O. Hierarchical clustering identifies novel subgroups of childhood asthma. Allergy. 2016;**71**:98-.

259. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvarinen A, Kaulek V, et al. Clinical and Epidemiologic Phenotypes of Childhood Asthma. American journal of respiratory and critical care medicine. 2014;**189**(2):129-38.

260. Gower WA, Li X, Li H, Ampleford EJ, Bleecker ER, Meyers DA, Consortium EVE. Cluster Analysis Of Asthma Phenotypes In Children Ages 6 To 18 Years From The Eve Consortium. American journal of respiratory and critical care medicine. 2013;**187**:1.

261. Just JG-E, R.; Rouve, S.; Wanin, S.; Moreau, D.; Annesi-Maesano, I. Two novel, severe asthma phenotypes identified during childhood using a clustering approach. Eur Respir J. 2012;**40**(1):55-60.

262. Just JS-P, P.; Gouvis-Echraghi, R.; Laoudi, Y.; Roufai, L.; Momas, I.; Annesi Maesano, I. Childhood allergic asthma is not a single phenotype. J Pediatr. 2014;**164**(4):815-20.

263. Kwon J, Seo J, Kim H, Kim B, Kim H, Lee S, et al. Asthma Phenotypes in School-aged Children from the Population Study: Cluster Analysis. Journal of Allergy and Clinical Immunology. 2012;**129**(2):AB7-AB.

264. Lee E, Lee SH, Kwon JW, Kim YH, Yoon J, Cho HJ, et al. Persistent asthma phenotype related with late-onset, high atopy, and low socioeconomic status in school-aged Korean children. BMC Pulm Med. 2017;**17**:11.

265. Mahut BP, S.; Delclaux, C. Exhaled nitric oxide and clinical phenotypes of childhood asthma. Respir Res. 2011;**12**:65.

266. Nasreen SW, P.; Mullowney, T.; Karp, I. Asthma exacerbation trajectories and their predictors in children with incident asthma. Annals of Allergy Asthma & Immunology. 2019;**123**(3):293-300.

267. Schimdlin KA, Brokamp C, LeMasters GK, Bernstein DI, Lockey JE, Hershey GKK, Ryan P. Distinct Phenotypes of Childhood Asthma: Cluster Analysis in a Longitudinal Birth Cohort. Journal of Allergy and Clinical Immunology. 2015;**135**(2):AB85-AB.

268. Yu J, Yoon J, Eom EJ, Kim JT, Lim DH, Kim WK, et al. Heterogeneity of childhood asthma in Korea: Cluster analysis of children with asthma from Korean childhood asthma study (KAS). Allergy. 2019;**74**:717-8.

269. Augustin IML, Spruit MA, Houben-Wilke S, Franssen FME, Vanfleteren L, Gaffron S, et al. The respiratory physiome: Clustering based on a comprehensive lung function assessment in patients with COPD. Plos One. 2018;**13**(9):14.

270. Bafadhel M, McKenna S, Terry S, Mistry V, Reid C, Haldar P, et al. Acute Exacerbations of Chronic Obstructive Pulmonary Disease Identification of Biologic Clusters and Their Biomarkers. American journal of respiratory and critical care medicine. 2011;**184**(6):662-71.

271. Bertini I, Luchinat C, Miniati M, Monti S, Tenori L. Phenotyping COPD by 1H NMR metabolomics of exhaled breath condensate. Metabolomics. 2014;**10**(2):302-11.

272. Burgel PR, Paillasseur JL, Caillaud D, Tillie-Leblond I, Chanez P, Escamilla R, et al. Clinical COPD phenotypes: a novel approach using principal component and cluster analyses. European Respiratory Journal. 2010;**36**(3):531-9.

273. Burgel PR, Paillasseur JL, Peene B, Dusser D, Roche N, Coolen J, et al. Two Distinct Chronic Obstructive Pulmonary Disease (COPD) Phenotypes Are Associated with High Risk of Mortality. Plos One. 2012;**7**(12):9.

274. Burgel PR, Paillasseur JL, Janssens W, Piquet J, ter Riet G, Garcia-Aymerich J, et al. A simple algorithm for the identification of clinical COPD phenotypes. European Respiratory Journal. 2017;**50**(5):11.

275. Chen CZ, Wang LY, Ou CY, Lee CH, Lin CC, Hsiue TR. Using Cluster Analysis to Identify Phenotypes and Validation of Mortality in Men with COPD. Lung. 2014;**192**(6):889-96.

276. Chubachi S, Tsutsumi A, Kameyama N, Sato M, Sasaki M, Miyazaki M, et al. Cluster Analysis Based On Comorbidities For Japanese COPD Patients. American journal of respiratory and critical care medicine. 2016;**193**:1.

277. De Torres JP, Marin JM, Martinez-Gonzalez C, de Lucas-Ramos P, Cosio B, Casanova C. The importance of symptoms in the longitudinal variability of clusters in COPD patients: A validation study. Respirology. 2018;**23**(5):485-91.

278. Divo M, Casanova C, Marin JM, Celli B, de Torres JP, Polverino F, et al. Identification of clinical phenotypes in patients with and without COPD using cluster analysis. European Respiratory Journal. 2016;**48**:3.

279. Fens N, van Rossum AGJ, Zanen P, van Ginneken B, van Klaveren RJ, Zwinderman AH, Sterk PJ. Subphenotypes of Mild-to-Moderate COPD by Factor and Cluster Analysis of Pulmonary Function, CT Imaging and Breathomics in a Population-Based Survey. Copd-Journal of Chronic Obstructive Pulmonary Disease. 2013;**10**(3):277-85.

280. Guillamet RV, Ursu O, Iwamoto G, Moseley PL, Oprea T. Chronic obstructive pulmonary disease phenotypes using cluster analysis of electronic medical records. Health informatics journal. 2018;**24**(4):394-409.

281. Harrison SL, Robertson N, Graham CD, Williams J, Steiner MC, Morgan MDL, Singh SJ. Can we identify patients with different illness schema following an acute exacerbation of COPD: A cluster analysis. Respiratory medicine. 2014;**108**(2):319-28.

282. Kilk K, Aug A, Ottas A, Soomets U, Altraja S, Altraja A. Phenotyping of Chronic Obstructive Pulmonary Disease Based on the Integration of Metabolomes and Clinical Characteristics. International Journal of Molecular Sciences. 2018;**19**(3):17.

283. Kim WJG, V.; Nishimura, M.; Makita, H.; Idolor, L.; Roa, C.; Loh, L. C.; Ong, C. K.; Wang, J. S.; Boonsawat, W.; Gunasekera, K. D.; Madegedara, D.; Kuo, H. P.; Wang, C. H.; Wang, C.; Yang, T.; Lin, Y. X.; Ko, F. W. S.; Hui, D. S.

C.; Lan, L. T. T.; Vu, Q. T. T.; Bhome, A. B.; Ng, A.; Seo, J. B.; Lee, B. Y.; Lee, J. S.; Oh, Y. M.; Lee, S. D. Identification of chronic obstructive pulmonary disease subgroups in 13 Asian cities. International Journal of Tuberculosis and Lung Disease. 2018;**22**(7):820-6.

284. Li X, Han MK, Ortega VE, Dransfield MT, Li H, Barr RG, et al. Cluster Analysis Of Chronic Obstructive Pulmonary Disease (COPD) Related Phenotypes In The Subpopulations And Intermediate Outcome Measures In COPD Study (spiromics). American journal of respiratory and critical care medicine. 2016;**193**:1.

285. Lopes AC, Xavier RF, Pereira A, Stelmach R, Fernandes FLA, Harrison SL, Carvalho CRF. Identifying COPD patients at risk for worse symptoms, HRQoL, and self-efficacy: A cluster analysis. Chronic Illness. 2019;**15**(2):138-48.

286. Peters JB, Boer LM, Molema J, Heijdra YF, Prins JB, Vercoulen JH. Integral Health Status-Based Cluster Analysis in Moderate-Severe COPD Patients Identifies Three Clinical Phenotypes: Relevant for Treatment As Usual and Pulmonary Rehabilitation. International Journal of Behavioral Medicine. 2017;**24**(4):571-83.

287. Pikoula M, Quint JK, Nissen F, Hemingway H, Smeeth L, Denaxas S. Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. BMC medical informatics and decision making. 2019;**19**:14.

288. Rodrigues A, Camillo CA, Furlanetto KC, Paes T, Morita AA, Spositon T, et al. Cluster analysis identifying patients with COPD at high risk of 2-year all-cause mortality. Chronic respiratory disease. 2018;**16**:8.

289. Scarlata S, Finamore P, Santangelo S, Giannunzio G, Pennazza G, Grasso S, et al. Cluster analysis on breath print of newly diagnosed COPD patients: effects of therapy. Journal of Breath Research. 2018;**12**(3):10.

290. Xavier RFP, Acac; Lopes, A. C.; Cavalheri, V.; Pinto, R. M. C.; Cukier, A.; Ramos, E. M. C.; Carvalho, C. R. F. Identification of Phenotypes in People with COPD: Influence of Physical Activity, Sedentary Behaviour, Body Composition and Skeletal Muscle Strength. Lung. 2019;**197**(1):37-45.

291. Yoon HY, Park SY, Lee CH, Byun MK, Na JO, Lee JS, et al. Prediction of first acute exacerbation using COPD subtypes identified by cluster analysis. International journal of chronic obstructive pulmonary disease. 2019;**14**:1389-97.

292. Brinkman P, Wagener AH, Bansal AT, Knobel HH, Vink TJ, Rattray N, et al. Electronic Noses Capture Severe Asthma Phenotypes By Unbiased Cluster Analysis. American journal of respiratory and critical care medicine. 2014;**189**:2.

293. Desai D, May RD, Haldar P, Shah S, Gupta S, Bafadhel M, et al. Cytokine Profiling In Severe Asthma Subphenotypes Using Factor And Cluster Analysis. American journal of respiratory and critical care medicine. 2011;**183**:1.

294. Diver S, Richardson M, Desai D, Mistry V, Haldar K, Barer M, Brightling CE. Hierarchical cluster analysis of subjects with severe asthma according to microbiological profile. Thorax. 2018;**73**:A20-A.

295. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, et al. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. J Allergy Clin Immunol. 2011;**127**(2):382-9.

296. Gomez JL, Yan XT, Holm CT, Grant N, Liu Q, Cohn L, et al. Characterisation of asthma subgroups associated with circulating YKL-40 levels. European Respiratory Journal. 2017;**50**(4):12.

297. Jang ASK, H. S.; Cho, Y. S.; Bae, Y. J.; Kim, T. B.; Park, J. S.; Park, S. W.; Uh, S. T.; Choi, J. S.; Kim, Y. H.; Hwang, H. K.; Moon, H. B.; Park, C. S. Identification of Subtypes of Refractory Asthma in Korean Patients by Cluster Analysis. Lung. 2013;**191**(1):87-93.

298. Konstantellou E, Papaioannou AI, Loukides S, Patentalakis G, Papaporfyriou A, Hillas G, et al. Persistent airflow obstruction in patients with asthma: Characteristics of a distinct clinical phenotype. Respiratory medicine. 2015;**109**(11):1404-9.

299. Newby C, Heaney LG, Menzies-Gow A, Niven RM, Mansur A, Bucknall C, et al. Statistical cluster analysis of the british thoracic society severe refractory asthma registry: Clinical outcomes and phenotype stability. PLoS ONE. 2014;**9**(7).

300. Raherison-Semjen C, Prudhomme A, Nocent-Eijnani C, Oster JP, Maurer C, Coetmoeur D, et al. FASE-CPHG Study: Identification of asthma phenotypes in the French Severe Asthma Study using cluster analysis. European Respiratory Journal. 2018;**52**:2.

301. Sekiya K, Nakatani E, Fukutomi Y, Kaneda H, Iikura M, Yoshida M, et al. Severe or life-threatening asthma exacerbation: patient heterogeneity identified by cluster analysis. Clinical and Experimental Allergy. 2016;**46**(8):1043-55.

302. Serrano-Pariente J, Rodrigo G, Fiz JA, Crespo A, Plaza V. Identification and characterization of near-fatal asthma phenotypes by cluster analysis. Allergy. 2015;**70**(9):1139-47.

303. Ye WJ, Xu WG, Guo XJ, Han FF, Peng J, Li XM, et al. Differences in airway remodeling and airway inflammation among moderate-severe asthma clinical phenotypes. Journal of Thoracic Disease. 2017;**9**(9):2904-14.

304. Youroukova VMD, D. G.; Valerieva, A. D.; Lesichkova, S. S.; Velikova, T. V.; Ivanova-Todorova, E. I.; Tumangelova-Yuzeir, K. D. Phenotypes Determined by Cluster Analysis in Moderate to Severe Bronchial Asthma. Folia Med (Plovdiv). 2017;**59**(2):165-73.

305. Moore WC, Meyers DA, Wenzel SE, Teague WG, Li HS, Li XN, et al. Identification of Asthma Phenotypes Using Cluster Analysis in the Severe Asthma Research Program. American journal of respiratory and critical care medicine. 2010;**181**(4):315-23.

306. Freitas PD, Xavier RF, Da Silva STC, Carvalho-Pinto RM, Stelmach R, Cukier A, et al. Identification of asthma phenotypes using cluster analysis. European Respiratory Journal. 2018;**52**:2.

307. Adnane C, Adouly T, Khallouk A, Rouadi S, Abada R, Roubal M, Mahtar M. Using preoperative unsupervised cluster analysis of chronic rhinosinusitis to inform patient decision and endoscopic sinus surgery outcome. European Archives of Oto-Rhino-Laryngology. 2017;**274**(2):879-85.

308. Agache I, Ciobanu C. Risk Factors and Asthma Phenotypes in Children and Adults with Seasonal Allergic Rhinitis. Physician and Sportsmedicine. 2010;**38**(4):81-6.

309. Bousquet PJ, Devillier P, Tadmouri A, Mesbah K, Demoly P, Bousquet J. Clinical Relevance of Cluster Analysis in Phenotyping Allergic Rhinitis in a Real-Life Study. International Archives of Allergy and Immunology. 2015;**166**(3):231-40.

310. Burte E, Bousquet J, Varraso R, Gormand F, Just J, Matran R, et al. Characterization of Rhinitis According to the Asthma Status in Adults Using an Unsupervised Approach in the EGEA Study. Plos One. 2015;**10**(8):18.

311. Herr M, Just J, Nikasinovic L, Foucault C, Le Marec AM, Giordanella JP, Momas I. Risk factors and characteristics of respiratory and allergic phenotypes in early childhood. Journal of Allergy and Clinical Immunology. 2012;**130**(2):389-96.e4.

312. Kurukulaaratchy RJ, Zhang HM, Patil V, Raza A, Karmaus W, Ewart S, Arshad SH. Identifying the heterogeneity of young adult rhinitis through cluster analysis in the Isle of Wight birth cohort. Journal of Allergy and Clinical Immunology. 2015;**135**(1):143-U225.

313. Lee E, Lee SH, Kwon JW, Kim Y, Cho HJ, Yang SI, et al. A rhinitis phenotype associated with increased development of bronchial hyperresponsiveness and asthma in children. Ann Allergy Asthma Immunol. 2016;**117**(1):21-8.e1.

314. Soler ZM, Hyer JM, Ramakrishnan V, Smith TL, Mace J, Rudmik L, Schlosser RJ. Identification of chronic rhinosinusitis phenotypes using cluster analysis. International Forum of Allergy & Rhinology. 2015;**5**(5):399-407.

315. T. A. Nakayama D, Yoshikawa M, Okushi T, Matsuwaki Y, Moriyama H, Otori N. Identification of chronic rhinosinusitis phenotypes using cluster analysis. Am J Rhinol Allergy. 2012;**26**(3):172-6.

316. Z. M. H. Soler JM, Ramakrishnan V, Smith TL, Mace J, Rudmik L, Schlosser RJ. Identification of chronic rhinosinusitis phenotypes using cluster analysis. International Forum of Allergy & Rhinology. 2015.

317. De Vries R, Dagelet YWF, Spoor P, Snoey E, Jak PMC, Brinkman P, et al. Clinical and inflammatory phenotyping by breathomics in chronic airway diseases irrespective of the diagnostic label. European Respiratory Journal. 2018;**51**(1):10.

318. Fingleton J, Huang KW, Weatherall M, Guo YF, Ivanov S, Bruijnzeel P, et al. Phenotypes of symptomatic airways disease in China and New Zealand. European Respiratory Journal. 2017;**50**(6):10.

319. Gorska K, Paplinska-Goryca M, Nejman-Gryz P, Goryca K, Krenke R. Eosinophilic and Neutrophilic Airway Inflammation in the Phenotyping of Mild-to-Moderate Asthma and Chronic Obstructive Pulmonary Disease. Copd-Journal of Chronic Obstructive Pulmonary Disease. 2017;**14**(2):181-9.

320. Rootmensen G, van Keimpema A, Zwinderman A, Sterk P. Clinical phenotypes of obstructive airway diseases in an outpatient population. Journal of Asthma. 2016;**53**(10):1026-32.

321. Shaw DE, Sousa AR, Fowler SJ, Fleming LJ, Roberts G, Corfield J, et al. Clinical and inflammatory characteristics of the European U-BIOPRED adult severe asthma cohort. Eur Respir J. 2015;**46**(5):1308-21.

322. Mahut B, Peyrard S, Delclaux C. Exhaled nitric oxide and clinical phenotypes of childhood asthma. Respir Res. 2011;**12**:65.

323. Mason P, Frigo AC, Scarpa MC, Maestrelli P, Guarnieri G. Cluster analysis of occupational asthma caused by isocyanates. Journal of Allergy and Clinical Immunology. 2018;**142**(6):2011-+.

324. Jang AS, Kwon HS, Cho YS, Bae YJ, Kim TB, Park JS, et al. Identification of subtypes of refractory asthma in Korean patients by cluster analysis. Lung. 2013;**191**(1):87-93.

325. Hsiao HPL, M. C.; Wu, C. C.; Wang, C. C.; Wang, T. N. Sex-Specific Asthma Phenotypes, Inflammatory Patterns, and Asthma Control in a Cluster Analysis. Journal of Allergy and Clinical Immunology-in Practice. 2019;**7**(2):556-+.

326. Just J, Gouvis-Echraghi R, Rouve S, Wanin S, Moreau D, Annesi-Maesano I. Two novel, severe asthma phenotypes identified during childhood using a clustering approach. Eur Respir J. 2012;**40**(1):55-60.

327. Just J, Saint-Pierre P, Gouvis-Echraghi R, Laoudi Y, Roufai L, Momas I, Maesano IA. Childhood Allergic Asthma Is Not a Single Phenotype. Journal of Pediatrics. 2014;**164**(4):815-20.

328. Burgel PR, Paillasseur JL, Peene B, Dusser D, Roche N, Coolen J, et al. Two distinct chronic obstructive pulmonary disease (COPD) phenotypes are associated with high risk of mortality. PLoS One. 2012;**7**(12):e51048.

329. Kim WJ, Gupta V, Nishimura M, Makita H, Idolor L, Roa C, et al. Identification of chronic obstructive pulmonary disease subgroups in 13 Asian cities. International Journal of Tuberculosis and Lung Disease. 2018;**22**(7):820-+.

330. Youroukova VM, Dimitrova DG, Valerieva AD, Lesichkova SS, Velikova TV, Ivanova-Todorova EI, Tumangelova-Yuzeir KD. Phenotypes Determined by Cluster Analysis in Moderate to Severe Bronchial Asthma. Folia Med (Plovdiv). 2017;**59**(2):165-73.

331. Nasreen S, Wilk P, Mullowney T, Karp I. Asthma exacerbation trajectories and their predictors in children with incident asthma. Annals of Allergy Asthma & Immunology. 2019;**123**(3):293-+.

332. Deliu M, Yavuz TS, Sperrin M, Belgrave D, Sahiner UM, Sackesen C, et al. Features of asthma which provide meaningful insights for understanding the disease heterogeneity. Clin Exp Allergy. 2018;**48**(1):39-47.

333. . !!! INVALID CITATION !!! 230, 233, 234, 236, 238-240, 244, 245, 247, 249, 250.

334. Rothman KJ. Epidemiology: an introduction: Oxford university press; 2012.

335. Toren K, Brisman J, Järvholm B. Asthma and asthma-like symptoms in adults assessed by questionnaires: a literature review. Chest. 1993;**104**(2):600-8.

336. Larsson L. Incidence and prevalence of asthma: Relation to differences in utilisation of asthma drugs between two neighbouring Swedish provinces. 1997.

337. Samet JM. Epidemiologic approaches for the identification of asthma. Chest. 1987;**91**(6):74S-8S.

338. Rönmark EP, Ekerljung L, Lötvall J, Torén K, Rönmark E, Lundbäck B. Large scale questionnaire survey on respiratory health in Sweden: effects of late- and non-response. Respiratory medicine. 2009;**103**(12):1807-15.

339. Mannetje At, Eng A, Douwes J, Ellison-Loschmann L, McLean D, Pearce N. Determinants of non-response in an occupational exposure and health survey in New Zealand. Australian and New Zealand journal of public health. 2011;**35**(3):256-63.

340. Groves RM. Nonresponse rates and nonresponse bias in household surveys. International Journal of Public Opinion Quarterly. 2006;**70**(5):646-75.

341. Nurmi E, Vähätalo I, Ilmarinen P, Andersén H, Tuomisto L, Sovijärvi A, et al. Agreement between self-reported and registered age at asthma diagnosis in Finland. 2023.

342. Hernán MA, Robins JM. Causal Inference: What If: CRC Boca Raton, FL; 2010.

343. Ben-Gal I. Bayesian networks. Encyclopedia of statistics in quality and reliability. 2008.

344. Deeks JJ, Bossuyt P, Leeflang MM, Takwoingi Y. Cochrane handbook for systematic reviews of diagnostic test accuracy: John Wiley & Sons; 2023.

345. Uman LS. Systematic reviews and meta-analyses. Journal of the Canadian Academy of Child and Adolescent Psychiatry. 2011;**20**(1):57.

346. Deliu M, Sperrin M, Belgrave D, Custovic A. Identification of asthma subtypes using clustering methodologies. Pulmonary therapy. 2016;**2**(1):19-41.

347. Moshontz H, Campbell L, Ebersole CR, IJzerman H, Urry HL, Forscher PS, et al. The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. Advances in Methods and Practices in Psychological Science. 2018;**1**(4):501-15.

348. Jacobucci R, Grimm KJ, Zhang Z. Machine Learning for Social and Behavioral Research: Guilford Publications; 2023.

349. Nathan RA, Sorkness CA, Kosinski M, Schatz M, Li JT, Marcus P, et al. Development of the asthma control test: a survey for assessing asthma control. Journal of Allergy and Clinical Immunology. 2004;**113**(1):59-65.

350. Schatz M, Sorkness CA, Li JT, Marcus P, Murray JJ, Nathan RA, et al. Asthma Control Test: reliability, validity, and responsiveness in patients not previously followed by asthma specialists. Journal of Allergy and Clinical Immunology. 2006;**117**(3):549-56.

351. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International Study Of Asthma And Allergies In Childhood (Isaac) - Rationale And Methods. European Respiratory Journal. 1995;**8**(3):483-91.

352. Bousquet J, van Cauwenberge P, Khaltaev N, Ait-Khaled N, Annesi-Maesano I, Bachert C, et al. Allergic rhinitis and its impact on asthma. Journal of Allergy and Clinical Immunology. 2001;**108**(5):S147-S334.

353. Howard R, Rattray M, Prosperi M, Custovic A. Distinguishing Asthma Phenotypes Using Machine Learning Approaches. Current Allergy and Asthma Reports. 2015;**15**(7):10.

354. Prosperi MCF, Sahiner UM, Belgrave D, Sackesen C, Buchan IE, Simpson A, et al. Challenges in Identifying Asthma Subgroups Using Unsupervised

Statistical Learning Techniques. American journal of respiratory and critical care medicine. 2013;**188**(11):1303-12.

355. Pereira RC, Rodrigues PP, Figueiredo MA, Abreu PH, editors. Automatic Delta-Adjustment Method Applied to Missing Not At Random Imputation. International Conference on Computational Science; 2023: Springer.

356. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;**28**(1):112-8.

357. Hong S, Lynn HS. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. BMC Medical Research Methodology. 2020;**20**(1).

358. Hennig C, Meila M, Murtagh F, Rocci R. Handbook of cluster analysis: CRC press; 2015.

359. Strömberg Celind F, Wennergren G, Vasileiadou S, Alm B, Åberg N, Goksör E. Higher parental education was associated with better asthma control. Acta Paediatrica. 2019;**108**(5):920-6.

360. Stanford RH, Gilsenan AW, Ziemiecki R, Zhou X, Lincourt WR, Ortega H. Predictors of uncontrolled asthma in adult and pediatric patients: analysis of the Asthma Control Characteristics and Prevalence Survey Studies (ACCESS). Journal of Asthma. 2010;**47**(3):257-62.

361. Schatz M, Mosen DM, Kosinski M, Vollmer WM, Magid DJ, O'Connor E, Zeiger RS. Predictors of asthma control in a random sample of asthmatic patients. Journal of Asthma. 2007;**44**(4):341-5.

362. Peters SP, Jones CA, Haselkorn T, Mink DR, Valacer DJ, Weiss ST. Real-world Evaluation of Asthma Control and Treatment (REACT): findings from a national Web-based survey. Journal of Allergy and Clinical Immunology. 2007;**119**(6):1454-61.

363. To M, Tsuzuki R, Katsube O, Yamawaki S, Soeda S, Kono Y, et al. Persistent asthma from childhood to adulthood presents a distinct phenotype of adult asthma. The Journal of Allergy and Clinical Immunology: In Practice. 2020;**8**(6):1921-7. e2.

364. Palmans E, Kips JC, Pauwels RA. Prolonged allergen exposure induces structural airway changes in sensitized rats. American journal of respiratory and critical care medicine. 2000;**161**(2):627-35.

365. Gianniou N, Katsaounou P, Dima E, Giannakopoulou C-E, Kardara M, Saltagianni V, et al. Prolonged occupational exposure leads to allergic airway sensitization and chronic airway and systemic inflammation in professional firefighters. Respiratory medicine. 2016;**118**:7-14.

366. Cook DN. Role of environmental adjuvants in asthma development. Current allergy and asthma reports. 2020;**20**:1-11.

367. Krieger N. A glossary for social epidemiology. Journal of Epidemiology & Community Health. 2001;**55**(10):693-700.

368. Davidsen JR, Søndergaard J, Hallas J, Siersted HC, Knudsen TB, Lykkegaard J, Andersen M. Impact of socioeconomic status on the use of inhaled corticosteroids in young adult asthmatics. Respiratory medicine. 2011;**105**(5):683-90.

369. Apter AJ, Boston RC, George M, Norfleet AL, Tenhave T, Coyne JC, et al. Modifiable barriers to adherence to inhaled steroids among adults with asthma: it's not just black and white. Journal of Allergy and Clinical Immunology. 2003;**111**(6):1219-26.

370. Parks CA, Blaser C, Smith TM, Calloway EE, Oh AY, Dwyer LA, et al. Correlates of fruit and vegetable intake among parents and adolescents: findings from the Family Life, Activity, Sun, Health, and Eating (FLASHE) study. Public health nutrition. 2018;**21**(11):2079-87.

371. Vaara JP, Vasankari T, Koski HJ, Kyröläinen H. Awareness and knowledge of physical activity recommendations in young adult men. Frontiers in public health. 2019;**7**:310.

372. Huovinen E, Kaprio J, Laitinen LA, Koskenvuo M. Social predictors of adult asthma: a co-twin case-control study. Thorax. 2001;**56**(3):234-6.

373. Ilmarinen P, Tuomisto LE, Kankaanranta H. Phenotypes, risk factors, and mechanisms of adult-onset asthma. Mediators of inflammation. 2015;**2015**.

374. Rönmark E, Andersson C, Nyström L, Forsberg B, Järvholm B, Lundbäck B. Obesity increases the risk of incident asthma among adults. European Respiratory Journal. 2005;**25**(2):282-8.

375. Iikura M, Yi S, Ichimura Y, Hori A, Izumi S, Sugiyama H, et al. Effect of lifestyle on asthma control in Japanese patients: importance of periodical exercise and raw vegetable diet. PLoS One. 2013;**8**(7):e68290.

376. Barros R, Moreira A, Fonseca J, Ferraz de Oliveira J, Delgado L, Castel-Branco M, et al. Adherence to the Mediterranean diet and fresh fruit intake are associated with improved asthma control. Allergy. 2008;**63**(7):917-23.

377. Stoodley I, Williams L, Thompson C, Scott H, Wood L. Evidence for lifestyle interventions in asthma. Breathe. 2019;**15**(2):e50-e61.

378. Dalstra JA, Kunst AE, Borrell C, Breeze E, Cambois E, Costa G, et al. Socioeconomic differences in the prevalence of common chronic diseases: an overview of eight European countries. International journal of epidemiology. 2005;**34**(2):316-26.

379. Lotfaliany M, Bowe SJ, Kowal P, Orellana L, Berk M, Mohebbi M. Depression and chronic diseases: Co-occurrence and communality of risk factors. Journal of affective disorders. 2018;**241**:461-8.

380. Backman H, Hedman L, Jansson S-A, Lindberg A, Lundbäck B, Rönmark E. Prevalence trends in respiratory symptoms and asthma in relation to smoking - two cross-sectional studies ten years apart among adults in northern Sweden. The World Allergy Organization journal. 2014;**7**(1):1-.

381. Lötvall J, Ekerljung L, Rönmark EP, Wennergren G, Lindén A, Rönmark E, et al. West Sweden Asthma Study: prevalence trends over the last 18 years argues no recent increase in asthma. Respiratory research. 2009;**10**(1):94-.

382. Viegi G, Pedreschi M, Baldacci S, Chiaffi L, Pistelli F, Modena P, et al. Prevalence rates of respiratory symptoms and diseases in general population samples of North and Central Italy. The International Journal of Tuberculosis and Lung Disease. 1999;**3**(11):1034-42.

383. Boggia B, Farinaro E, Grieco L, Lucariello A, Carbone U. Burden of smoking and occupational exposure on etiology of chronic obstructive pulmonary disease in workers of southern sItaly. Journal of occupational and environmental medicine. 2008:366-70.

384. de Meer G, Kerkhof M, Kromhout H, Schouten JP, Heederik D. Interaction of atopy and smoking on respiratory effects of occupational dust exposure: a general population-based study. Environ Health. 2004;**3**(1):6.

385. Meer Gd, Kerkhof M, Kromhout H, Schouten JP, Heederik D. Interaction of atopy and smoking on respiratory effects of occupational dust exposure: a general population-based study. Environmental health. 2004;**3**(1):1-7.

386. van der Plaat DA, De Matteis S, Steven S, Debbie J, Paul C, Cosetta M. Interaction between occupational exposures and antioxidant genes on chronic obstructive pulmonary disease in UK biobank. EUROPEAN MEDICAL JOURNAL RESPIRATORY. 2020;**8**(1):77-8.

387. Graff P, Fredrikson M, Jönsson P, Flodin U. Non-sensitising air pollution at workplaces and adult-onset asthma in the beginning of this millennium. International archives of occupational and environmental health. 2011;**84**:797-804.

388. Shirakawa T, Kusaka Y, Morimoto K. Combined effect of smoking habits and occupational exposure to hard metal on total IgE antibodies. Chest. 1992;**101**(6):1569-76.

389. Bottema RW, Reijmerink NE, Kerkhof M, Koppelman GH, Stelma FF, Gerritsen J, et al. Interleukin 13, CD14, pet and tobacco smoke influence atopy in three Dutch cohorts: the allergenic study. European Respiratory Journal. 2008;**32**(3):593-602.

390. Kim SY, Sim S, Choi HG. Atopic dermatitis is associated with active and passive cigarette smoking in adolescents. PloS one. 2017;**12**(11):e0187453.

391. Noakes PS, Hale J, Thomas R, Lane C, Devadason SG, Prescott SL. Maternal smoking is associated with impaired neonatal toll-like-receptor-mediated immune responses. European Respiratory Journal. 2006;**28**(4):721-9.

392. Shargorodsky J, Garcia-Esquinas E, Galán I, Navas-Acien A, Lin SY. Allergic sensitization, rhinitis and tobacco smoke exposure in US adults. PloS one. 2015;**10**(7):e0131957.

393. Zetterström O, Osterman K, Machado L, Johansson S. Another smoking hazard: raised serum IgE concentration and increased risk of occupational allergy. Br Med J (Clin Res Ed). 1981;**283**(6301):1215-7.

394. De Matteis S. Occupational causes of chronic obstructive pulmonary disease: an update. Current Opinion in Allergy and Clinical Immunology. 2022;**22**(2):73-9.

395. Liu AH. Hygiene theory and allergy and asthma prevention. Paediatric and Perinatal Epidemiology. 2007;**21**:2-7.

396. Flodin U, Ponsson P, Ziegler J, Axelson O. An epidemiologic study of bronchial asthma and smoking. Epidemiology. 1995;**6**(5):503-5.

397. Godtfredsen N, Lange P, Prescott E, Osler M, Vestbo J. Changes in smoking habits and risk of asthma: a longitudinal population based study. European Respiratory Journal. 2001;**18**(3):549-54.

398. Rönmark E, Lundbäck B, Jonsson E, Jonsson AC, Lindström M, Sandström T. Incidence of asthma in adults–report from the Obstructive Lung Disease in Northern Sweden Study. Allergy. 1997;**52**(11):1071-8.

399. Lajunen TK, Jaakkola JJ, Jaakkola MS. Different effects of smoking on atopic and non-atopic adult-onset asthma. Clinical and Translational Allergy. 2021;**11**(8):e12072.

400. Lindström M, Jönsson E, Kotaniemi J, Lundbäck B. Smoking, respiratory symptoms, and diseases: a comparative study between northern Sweden and northern Finland: report from the FinEsS study. Chest. 2001;**119**(3):852-61.

401. Lundbäck B, Rönmark E, Jönsson E, Larsson K, Sandström T. Incidence of physician-diagnosed asthma in adults—a real incidence or a result of increased awareness? Report from the Obstructive Lung Disease in Northern Sweden Studies. Respiratory medicine. 2001;**95**(8):685-92.

402. Jaakkola JJ, Hernberg S, Lajunen TK, Sripaijboonkij P, Malmberg LP, Jaakkola MS. Smoking and lung function among adults with newly onset asthma. BMJ open respiratory research. 2019;**6**(1):e000377.

403. Liu C, Cheng W, Zeng Y, Zhou Z, Zhao Y, Duan J, et al. Different characteristics of ex-smokers and current smokers with COPD: a cross-sectional study in China. International journal of chronic obstructive pulmonary disease. 2020:1613-9.

404. Cunha F, Amaral R, Jacinto T, Sousa-Pinto B, Fonseca JA. A Systematic Review of Asthma Phenotypes Derived by Data-Driven Methods. Diagnostics (Basel). 2021;**11**(4).

405. Torfi Y, Bitarafan N, Rajabi M. Impact of socioeconomic and environmental factors on atopic eczema and allergic rhinitis: a cross sectional study. EXCLI journal. 2015;**14**:1040.

406. Biagini JM, LeMasters GK, Ryan PH, Levin L, Reponen T, Bernstein DI, et al. Environmental risk factors of rhinitis in early infancy. Pediatric Allergy and Immunology. 2006;**17**(4):278-84.

407. Lu MX, Yang HJ, Wang JY, An Z, Li J, Wu ZN, et al. Acute effects of ambient air pollution on outpatients with chronic rhinitis in Xinxiang, China. Environ Sci Pollut Res. 2021;**28**(8):9889-97.

408. Kilty SJ, McDonald JT, Johnson S, Al-Mutairi D. Socioeconomic status: a disease modifier of chronic rhinosinusitis? Rhinology. 2011;**49**(5):533.

409. Lu-Myers Y, Deal AM, Miller JD, Thorp BD, Sreenath SB, McClurg SM, et al. Comparison of socioeconomic and demographic factors in patients with chronic rhinosinusitis and allergic fungal rhinosinusitis. Otolaryngology–Head and Neck Surgery. 2015;**153**(1):137-43.

410. Liu AH. Hygiene theory and allergy and asthma prevention. Paediatr Perinat Epidemiol. 2007;**21 Suppl 3**:2-7.

411. Haghdoost AA. Complexity of the Socioeconomic Status and its Disparity as a Determinant of Health. International journal of preventive medicine. 2012;**3**(2):75.

412. Hox V, Delrue S, Scheers H, Adams E, Keirsbilck S, Jorissen M, et al. Negative impact of occupational exposure on surgical outcome in patients with rhinosinusitis. Allergy. 2012;**67**(4):560-5.

413. Jarvis D, Newson R, Lotvall J, Hastan D, Tomassen P, Keil T, et al. Asthma in adults and its association with chronic rhinosinusitis: the GA2LEN survey in Europe. Allergy. 2012;**67**(1):91-8.

414. Nadif R, Febrissy M, Andrianjafimasy M, Le Moual N, Gormand F, Just J, et al. Adult asthma phenotypes identified by a cluster analysis on clinical and biological characteristics. European Respiratory Journal. 2018;**52**:2.

415. Koike F, Otani Y, Oyama S, Furuie W, Endo J, Nakamura Y, et al. Cluster analysis of cough variant asthma using exhaled value of forced oscillation technique. European Respiratory Journal. 2018;**52**:3.

416. Zein JG, Erzurum SC. Asthma is Different in Women. Current Allergy and Asthma Reports. 2015;**15**(6).

417. Honkamäki J, Piirilä P, Hisinger-Mölkänen H, Tuomisto LE, Andersén H, Huhtala H, et al. Asthma Remission by Age at Diagnosis and Gender in a Population-Based Study. The Journal of Allergy and Clinical Immunology: In Practice. 2021;**9**(5):1950-9.e4.

418. Quirce S, Sastre J. Occupational asthma: clinical phenotypes, biomarkers, and management. Current opinion in pulmonary medicine. 2019;**25**(1):59-63.

419. Baptist APH, W.; Karamched, K. R.; Kaur, B.; Carpenter, L.; Song, P. X. K. Distinct Asthma Phenotypes Among Older Adults with Asthma. Journal of Allergy and Clinical Immunology-in Practice. 2018;**6**(1):244-+.

420. Bafadhel MM, S.; Terry, S.; Mistry, V.; Reid, C.; Haldar, P.; McCormick, M.; Haldar, K.; Kebadze, T.; Duvoix, A.; Lindblad, K.; Patel, H.; Rugman, P.; Dodson, P.; Jenkins, M.; Saunders, M.; Newbold, P.; Green, R. H.; Venge, P.; Lomas, D. A.; Barer, M. R.; Johnston, S. L.; Pavord, I. D.; Brightling, C. E. Acute Exacerbations of Chronic Obstructive Pulmonary Disease Identification of Biologic Clusters and Their Biomarkers. American Journal of Respiratory and Critical Care Medicine. 2011;**184**(6):662-71.

421. Bhargava SH, A. D.; Jayaraj, B. S.; Praveena, A.; Ravi, S.; Khurana, S.; Mahesh, P. A. Distinct asthma phenotypes with low maximal attainment of lung function on cluster analysis. Journal of Asthma. 2019:12.

422. Brew BC, F.; Lundholm, C.; Almqvist, C. Combining supervised and unsupervised models to characterize asthma phenotypes in children. European Respiratory Journal. 2018;**52**:2.

423. Brew BKC, F.; Lundholm, C.; Ortqvist, A.; Almqvist, C. A modern approach to identifying and characterizing child asthma and wheeze phenotypes based on clinical data. PLoS One. 2019;**14**(12):e0227091.

424. Castaldi PJB, M.; Petersen, H.; Rafaels, N.; Finigan, J.; Paoletti, M.; Boezen, H. M.; Vonk, J. M.; Bowler, R.; Pistolesi, M.; Puhan, M. A.; Anto, J.; Wauters, E.; Lambrechts, D.; Janssens, W.; Bigazzi, F.; Camiciottoli, G.; Cho, M. H.; Hersh, C. P.; Barnes, K.; Rennard, S.; Boorgula, M. P.; Dy, J.; Hansel, N. N.; Crapo, J. D.; Tesfaigzi, Y.; Agusti, A.; Silverman, E. K.; Garcia-Aymerich, J. Do COPD subtypes really exist? COPD heterogeneity and clustering in 10 independent cohorts. Thorax. 2017;**72**(11):998-1006.

425. Chang TSL, R. F.; Mauger, D. T.; Fitzpatrick, A. M.; Sorkness, C. A.; Szefler, S. J.; Gangnon, R. E.; Page, C. D.; Jackson, D. J.; Childhood Asthma Res Educ, Care. Childhood asthma clusters and response to therapy in clinical trials. Journal of Allergy and Clinical Immunology. 2014;**133**(2):363-+.

426. Chupp GLP, M. F.; Gomez-Arroyo, J. G.; He, S.; Holm, C.; Cook, D.; Zhao, H.; Cohn, L. E.; Yan, X. Determination Of Asthma Phenotypes By

Unsupervised Transcriptomic Cluster Analysis Of Induced Sputum. American journal of respiratory and critical care medicine. 2013;**187**:1.

427. Delgado-Eckert EF, O.; Kumar, N.; Pekkanen, J.; Dalphin, J. C.; Riedler, J.; Lauener, R.; Kabesch, M.; Kupczyk, M.; Dahlen, S. E.; Mutius, E. V.; Frey, U. Functional phenotypes determined by fluctuation-based clustering of lung function measurements in healthy and asthmatic cohort participants. Thorax. 2018;**73**(2):107-15.

428. Ding LLL, D.; Wathen, M.; Altaye, M.; Mersha, T. B. African ancestry is associated with cluster-based childhood asthma subphenotypes. Bmc Medical Genomics. 2018;**11**:11.

429. Diver SR, M.; Haldar, K.; Ghebre, M. A.; Ramsheh, M. Y.; Bafadhel, M.; Desai, D.; Cohen, E. S.; Newbold, P.; Rapley, L.; Rugman, P.; Pavord, I. D.; May, R. D.; Barer, M.; Brightling, C. E. Sputum microbiomic clustering in asthma and COPD reveals a Haemophilus-predominant subgroup. Allergy. 2020:10.

430. Eom EJY, J. S.; Kim, Y. K.; Jung, S. S.; Suh, D. I.; Kim, H. Y.; Lim, D. H.; Kim, H. S.; Kim, J. T.; Shin, M. Y.; Baek, H. S.; Woo, S. I.; Shin, Y. H.; Yoo, Y.; Kim, J. K.; Park, J. S.; Kwon, J. W.; Lee, E.; Yang, H. J.; Lee, J. S.; Seo, J. H.; Jang, G. C.; Kim, W. K.; Song, D. J.; Yu, J. H. Heterogeneity of Childhood Asthma in Korea: Cluster Analysis of Children with Asthma from the Korean Childhood Asthma Study (KAS). Pediatric Pulmonology. 2018;**53**:S107-S8.

431. Fens NvR, A. G. J.; Zanen, P.; van Ginneken, B.; van Klaveren, R. J.; Zwinderman, A. H.; Sterk, P. J. COPD subphenotypes in a population-based survey by factor and cluster analysis. European Respiratory Journal. 2012;**40**:1.

432. Fernandez DS, R. J.; Dostal, M.; Pastorkova, A.; Gmuender, H.; Choi, H. Modeling Unobserved Heterogeneity in Susceptibility to Ambient Benzo[a]pyrene Concentration among Children with Allergic Asthma Using an Unsupervised Learning Algorithm. Int J Environ Res Public Health. 2018;**15**(1).

433. Guo YFN, P.; Sun, T. Y. LATE-BREAKING ABSTRACT: Exploratory study of clinical phenotypes of airways disease with wheezing by cluster analysis. European Respiratory Journal. 2015;**46**:2.

434. He ZAD, K.; Nava-Guerra, L.; Khoo, M. C. K.; DiFrancesco, M.; Xu, Y.; McConnell, K.; Hossain, M. M.; Amin, R. Phenotype of Ventilatory Control in Children With Moderate to Severe Persistent Asthma and Obstructive Sleep Apnea. Sleep. 2019.

435. Howrylak JAF, A. L.; Strunk, R. C.; Zeiger, R. S.; Weiss, S. T.; Raby, B. A.; Childhood Asthma Management, Progra. Classification of childhood asthma phenotypes and long-term clinical responses to inhaled anti-inflammatory medications. Journal of Allergy and Clinical Immunology. 2014;**133**(5):1289-U494.

436. Humphries SMN, A. M.; Centeno, J. P.; Strand, M. J.; Crapo, J. D.; Silverman, E. K.; Lynch, D. A. Deep Learning Enables Automatic Classification of Emphysema Pattern at CT. Radiology. 2019:191022.

437. Kharevich OL, I.; Lapteva, E. Two phenothypes of severe asthma identified by cluster analysis. European Respiratory Journal. 2018;**52**:2.

438. Khoo SKR, J.; Franks, K.; Zhang, G.; Bizzintino, J.; Coleman, L.; McCrae, C.; Oberg, L.; Troy, N. M.; Prastanti, F.; Everard, J.; Oo, S.; Borland, M. L.; Maciewicz, R. A.; Le Souef, P. N.; Laing, I. A.; Bosco, A. Upper Airway Cell Transcriptomics Identify a Major New Immunological Phenotype with Strong Clinical Correlates in Young Children with Acute Wheezing. J Immunol. 2019;**202**(6):1845-58.

439. Khusial RJS, J. K.; Loijmans, R. J. B.; Snoeck-Stroband, J. B.; Assendelft, P. J. J.; Schermer, T. R. J.; Honkoop, P. J.; Grp, Accurate Study. Longitudinal outcomes of different asthma phenotypes in primary care, an observational study. Npj Primary Care Respiratory Medicine. 2017;**27**:6.

440. Kilk KA, A.; Ottas, A.; Soomets, U.; Altraja, S.; Altraja, A. Phenotyping of Chronic Obstructive Pulmonary Disease Based on the Integration of Metabolomes and Clinical Characteristics. International Journal of Molecular Sciences. 2018;**19**(3):17.

441. Kuhlen JLW, A. E.; Nietert, P. J.; Bains, S. N. Identification of Asthma Phenotypes in a Tertiary Care Medical Center. American Journal of the Medical Sciences. 2014;**348**(6):480-5.

442. Kuo CHSP, S.; Loza, M.; Baribaud, F.; Rowe, A.; Pandis, I.; Rossios, C.; Wilson, S.; Djukanovic, R.; Sterk, P.; Chung, K. F.; Adcock, I. M.; Guo, Y.; U, Biopred. Asthma Phenotypes From Semi-Supervised Machine-Learning Approach Of Bronchial Biopsy And Brush Transcriptomics In U-Biopred. American journal of respiratory and critical care medicine. 2015;**191**:2.

443. Kupczyk MD, B.; Sterk, P. J.; Nizankowska-Mogilnicka, E.; Papi, A.; Bel, E. H.; Chanez, P.; Howarth, P. H.; Holgate, S. T.; Brusselle, G.; Siafakas, N. M.; Gjomarkaj, M.; Dahlen, S. E.; Investigators, Bioair. Stability of phenotypes

defined by physiological variables and biomarkers in adults with asthma. Allergy. 2014;**69**(9):1198-204.

444. Mason PS, M. C.; Liviero, F.; Guarnieri, G.; Maestrelli, P. Distinct Clinical Phenotypes of Occupational Asthma due to Diisocyanates. Journal of occupational and environmental medicine. 2017;**59**(6):539-42.

445. Panico LS, B.; Bartley, M.; Kelly, Y. Asthma trajectories in early childhood: identifying modifiable factors. PLoS One. 2014;**9**(11):e111922.

446. Patrawalla PK, A.; Rogers, L.; Shao, Y.; Liu, M.; Cheng, Q.; Fernandez-Beros, M. E.; Reibman, J. Replication Of The Severe Asthma Research Program Cluster Analysis In An Urban Population. American journal of respiratory and critical care medicine. 2011;**183**:1.

447. Prosperi MCS, U. M.; Belgrave, D.; Sackesen, C.; Buchan, I. E.; Simpson, A.; Yavuz, T. S.; Kalayci, O.; Custovic, A. Challenges in identifying asthma subgroups using unsupervised statistical learning techniques. American journal of respiratory and critical care medicine. 2013;**188**(11):1303-12.

448. Riley JB, A.; Ponte, E.; Souza-Machado, A.; Almeida, P.; Biao-Lima, V.; Bates, S.; F-Chung, K.; Sterk, P.; Cruz, A. A. Unbiased clinical cluster analysis across a combined European and a Brazilian severe asthma (SA) population. European Respiratory Journal. 2018;**52**:3.

449. Rennard SIL, N.; Delafont, B.; Tal-Singer, R.; Silverman, E. K.; Vestbo, J.; Miller, B. E.; Bakke, P.; Celli, B. P.; Calverley, P. M. A.; Coxson, H.; Crim, C.; Edwards, L. D.; Lomas, D. A.; MacNee, W.; Wouters, E. F. M.; Yates, J. C.; Coca, I.; Agusti, A.; Evaluation, Copd Longitudinally. Identification of Five Chronic Obstructive Pulmonary Disease Subgroups with Different Prognoses in the ECLIPSE Cohort Using Cluster Analysis. Annals of the American Thoracic Society. 2015;**12**(3):303-12.

450. Sato ST, Y.; Misa, K.; Fukuhara, N.; Nikaido, T.; Uematsu, M.; Fukuhara, A.; Wang, X. T.; Ishida, T.; Munakata, M. Identification of Clinical Phenotypes in Idiopathic Interstitial Pneumonia with Pulmonary Emphysema. Internal Medicine. 2016;**55**(12):1529-35.

451. Sharma SM, D. P.; Emmett, A.; Li, H. Cluster Analysis For The Identification And Replication Of Distinct Subject Clusters From COPD Clinical Trials. American journal of respiratory and critical care medicine. 2010;**181**:1.

452. Sohn SW, C. I.; Wu, S. T.; Liu, H. F.; Ryu, E.; Krusemark, E.; Seabright, A.; Voge, G. A.; Juhn, Y. J. Ascertainment of asthma prognosis using natural language processing from electronic medical records. Journal of Allergy and Clinical Immunology. 2018;**141**(6):2292-4.

453. Sukhan VS. Cluster analysis of the phenotype of asthma and obesity. Wiad Lek. 2018;**71**(2):319-21.

454. Wang YZ, Y.; Therneau, T. M.; Atkinson, E. J.; Tafti, A. P.; Zhang, N.; Amin, S.; Limper, A. H.; Khosla, S.; Liu, H. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. J Biomed Inform. 2019;**102**:103364.

455. Yang JA, E.; Balte, P.; Hoffman, E.; Austin, J.; Ben, Smith; Bleecker, E.; Bowler, R.; Cooper, C.; Couper, D.; Dransfield, M.; Han, M.; Hansel, N.; Martinez, F.; Paine, R.; Woodruff, P.; Ortega, V.; Laine, A.; Barr, R. G. Late Breaking Abstract - Unsupervised machine learning of novel emphysema subtypes: SPIROMICS. European Respiratory Journal. 2018;**52**:2.

456. Li P, Stuart EA, Allison DB. Multiple Imputation. JAMA. 2015;**314**(18):1966.

457. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials – a practical guide with flowcharts. BMC Medical Research Methodology. 2017;**17**(1).

458. Pedersen A, Mikkelsen E, Cronin-Fenton D, Kristensen N, Pham TM, Pedersen L, Petersen I. Missing data and multiple imputation in clinical epidemiological research. Clinical Epidemiology. 2017;**Volume 9**:157-66.

459. Tibshirani R, Walther G. Cluster validation by prediction strength. Journal of Computational and graphical Statistics. 2005;**14**(3):511-28.

9 APPENDIX 1: PAPER 3

9.1 MAIN TEXT TABLES

Table S 1 Number of derived phenotypes and their descriptions for studies on asthma

Reference, country	Population age	Number of phenotypes	Phenotype description
Amaral et al 2019. USA	Adults	3	High symptoms with poor lung function and more emergency visits among those > 40 years. Low symptoms with better lung function and less emergency visits among those < 40 years Low symptoms with better lung function and less emergency visits among those > 40 years
Amelink et al. 2013. Netherlands	Adults	3	Severe eosinophilic inflammation-predominant asthma and persistent airflow limitation despite high-intensity anti-inflammatory treatment, with relatively low symptom scores
			Obese female, frequent symptoms, high HCU and low eosinophilia.
			Males with mild to moderate, well controlled asthma and low HCU, low symptoms and normal function.
Bhargava et al. 2018. India	Adults	4	Treatment responsive, normal weight, childhood asthma.
			Poor treatment response, obese, males with late onset disease.
			Poor treatment response, old, male, with late-onset, obese.
			Good treatment response, obese females with late onset disease.
Benton et al. 2010. USA	Children and adolescents	4	Male neutrophilic asthma
			Female obese, late onset asthma
			Atopic, uncontrolled, obese eosinophilic asthma
			Mild asthma
Bochenek et al.2014. Poland	Adults	4	Asthma with a moderate course, intensive upper airway symptoms, and blood eosinophilia.
			Asthma with a mild course, relatively well controlled, and with low health care use.
			Asthma with a severe course, poorly controlled, and with severe exacerbations and airway obstruction.
			Poorly controlled asthma with frequent and severe exacerbations in female subjects.
Boudier et al. 2013.	Adults	7	Mild atopic with no treatment needed.

Multicounty			Non atopic mild with no treatment need
			Non atopic sever with treatment needed.
			Atopic, severe with high reactivity and treatment need.
			Atopica moderate asthma with high reactivity
			Atopic moderate asthma with normal function.
			Non atopic moderate with no need to treatment.
Cabral et al. 2016. Brazil	Children and adolescents	3	Normal lung function, mild eosinophile, late onset, mild atopy and few exacerbations.
			Normal lung function, moderate atopy, more severe eosinophilia and early onset disease.
			Poor lung function, frequent exacerbations, severe eosinophilia, and severe atopy.
Celejewska-Wójcik et al. 2020. Poland	Adults	3	Younger females with mild paucigranulocytic, low eicosanoids and low pro inflammatory markers.
			Older Females with severe, eosinophilic asthma and high proinflammatory markers.
			Old women, obese, mild eosinophilic asthma with high pro and anti-inflammatory markers
Chanoine et al. 2017. France	Adults	5	Never regular treatment
			Persistence high ratio of maintenance therapy
			Increasing ration of maintenance therapy from low to high.
			Initiating maintenance therapy at high level
			Cessation of maintenance therapy during study period.
Couto et al. 2015. Portugal	Adults	2	Atopic asthma with rhinitis, allergic comorbidities and high FeNO.
			Sport asthma: non-allergic exercise induced asthma with high BHR.
Cruz et al. 2018 (abstract). Brazil	Adults	4	Normal baseline function with less treatment needed.
			Low baseline function with obesity.
			Low baseline function and high medication and reversibility.
			Moderate baseline impairment and low revisability.
Damiens K et al.2013. Canada	Adults	6	Old sever atopic WRA with rhinitis. (Clusters were identified in subjects with WRA)
			Young mild atopic, eosinophilic OA with rhinitis with exposure to HMW agents.
			Non atopic, men, with WEA with exposure to LMW.
			Atopic, LMW agents' exposure and delayed asthma reaction and low occupational rhinitis. (Clusters were identified in the OA subgroup)
			HMW exposure, immediate long duration reaction.
			Nonsmoking women with normal function and immediate eosinophilic reaction.
Deliu et al. 2016.	Children	6	Early onset mild asthma with impaired function and high medication use.

Turkey			Non atopic children with normal function and controlled asthma.
			Late onset eosinophilic asthma with impaired function, high medication uses and sensitization.
			Elderly with late onset severe, poor controlled asthma.
			Elderly with mild asthma and impaired function
			Mild atopic asthma with normal function
	Children	5	High symptom and medication use. (HC after dimensionality reduction)
Deliu et al. 2018.			Males with late onset severe disease with normal function.
Turkey			Late onset, mild asthma with impaired function and multiple sensitizations.
			Female with early onset, mild, atopic asthma.
			Mild atopic asthma
			Female, early onset severe asthma. (HC using all available variables)
			late onset mild atopic asthma
			Moderate, highly atopic asthma,
			Male mild non atopic asthma.
			Late onset, atopic, severe asthma.
			Difficult asthma. (HC using the informative set of features)
			Early-onset mild atopic asthma.
			Early-onset mild non-atopic asthma.
			Late-onset asthma.
			Exacerbation-prone asthma.
Depner et al. 2013.	Children	5	Persistent wheeze.
Multi-country			Late-onset wheeze.
			Intermediate wheeze.
			Transient wheeze.
			No/infrequent wheeze.
Dudchenko et al. 2018.	Adults	7	Mild symptoms, infrequent attacks relieved by SABA, moderate S of B. and limited physical activity and sensitivity to weather changes.
Russia			Uncontrolled asthma, severe inflammation, daily attacks and severely impaired physical activity and night sleep
			Uncontrolled disease, high catarrhal manifestation, daily frequent attacks, moderate symptoms with high sleep and activity impairment.
			Mild, well controlled, minimal inflammation, low symptoms severity and frequency, low need for SABA and minimal impairment of PA and night sleep.

			Controlled to partially controlled, moderate catarrhal manifestation, within normal lung function, with excretion related limitation of PA and moderate symptoms.
			Uncontrolled asthma, moderate catarrhal symptoms, occasional symptoms, and usage of low dose BD and mild to moderate lung function impairment.
			Variable control, moderate inflammatory features, frequent breathlessness and diverse pattern of nocturnal disturbance, high dependence on weather changes and severely impaired lung function.
Folz et al.2018. **USA**	Adults	3	Three clusters that shows significant difference in BMI, FEV1%, FVC, FEV1, exercises induced asthma, asthma control, age at asthma onset and Feno.
Fontanella et al. 2018. **UK**	Children	4	Multiple sensitizations with sIgE to multiple components
			Predominantly dust mite sensitization (sIgE responses mainly to components from C. sIgE-1)
			Predominantly grass and tree sensitisation (sIgE to multiple components across C. sIgE-4–7)
			Lower-grade sensitisation
Gonem S et al. 2012 (abstract). **UK**	Adults	2	High air entrapment, abnormal lung mechanics, limited function, poor control and QOL.
			Normal lung physiology compared to 1.
Gower WA et al. 2013 (abstract). **USA**	Children	5	Non-Hispanic white, early onset and normal baseline function.
			Hispanic, early onset and low baseline function.
			Obese, late onset, low baseline function.
			Males with low baseline function.
			African American, obese, atopic, high disease duration and hospitalization.
Hilvering et al. 2015 (abstract). **Netherland**	Adults	Non indicated	Eosinophilic and non-eosinophilic asthma phenotypes that are differentiated by peripheral blood eosinophil count, FeNO (Fraction of Exhaled Nitric Oxide), ACQ (Asthma Control Questionnaire), medication use, nasal polyposis, aspirin sensitivity and neutrophil/eosinophil responsiveness upon stimulation.
Hsiao et al. 2018. **Taiwan**	Adults	5	Females, late onset, non-atopic asthma with normal function and low inflammation.
			Females, young, atopic, eosinophilic, low neutrophilic asthma.
			Females, late onset, non-atopic, obese, low eosinophilic, high neutrophilic asthma.
			Males, late onset, non-atopic, low eosinophilic and normal function.
			Males, young, atopic, smoking, neutrophilic asthma.
			Males who are former smokers with late onset high eosinophilic disease.
Ilmarinen et al. 2017.	Adults	5	Non-rhinitis asthma.

Finland			Smoking asthma.
			Female asthma.
			Obesity-related asthma.
			Early-onset atopic adult asthma.
Jeong, A. et al. 2016. Switzerland	Adults	4	Women, persistent multiple symptoms, late onset asthma with high percent body fat.
			Symptom presenting asthma with obesity.
			Symptom-free atopic asthma, atopic, high rhinitis,
			Symptom-free non-atopic asthma.
Just, J. et al. 2012. France	Children	3	Atopic, severe, uncontrolled, eosinophilic, and basophilic, with high treatment asthma.
			Obesity, non-atopic, low function, neutrophilic asthma.
			Mild asthma
Just, J. et al. 2014. France	Children	4	Multiple Allergies and Severe Asthma.
			Pollen Sensitization with Severe Exacerbations.
			Multiple Allergic Sensitizations and mild Asthma.
			House dust mite (HDM) Sensitization and Mild Asthma.
Kaneko, Y. et al. 2013; Japan	Adults	5	Late-Onset, mild, less-Atopic.
			Early-Onset, mild, atopic.
			Early-Onset, moderate-to-Severe Atopic.
			Late-Onset, severe.
			Middle-Age onset, female-Dominant.
			Late-Onset, moderate Less Atopic
Kim JH. et al. 2018; Korea	Adults	5	Persistent normal lung functions, female, middle/old age, obese, low treatment, and low symptoms.
			Persistent normal lung functions, young, female, atopic, low dose treatment, and low symptoms.
			Mild baseline impairment with slight improvement, low function, early onset, long duration, high dosage treatment and higher symptoms.
			Baseline marked impairment with fast improvement, atopic, high symptoms, low reversibility, and high dosage treatment.
			Marked baseline impairment with slow improvement, non-atopic, low reversibility, long duration, high symptoms, and high dosage treatment usage.
Kim MA. et al. 2017; Korea	Adults	4	Young, early onset, atopic asthma with normal function.
			Elderly, late onset, non-atopic asthma with low function.
			High atopic with severely impaired function.
			Elderly, late onset, slight atopy, and normal function asthma.

Kim TB. et al. 2013; Korea	Adults	4	Male, late onset, smokers, with preserved lung function.
			Atopic, late onset, high HCU, low function, low reversibility, and high severity.
			Young, early onset, and atopic asthma.
			Older, late onset, mild asthma with high function.
Koike et al. 2018 Not indicated.	Adults	4	High FeNo asthma
			Low value of exhaled R5-R20
			Abnormal exhaled reactance.
			Abnormal resistance and reactance.
Kwon et al. 2012 (abstract). Korea	Children	3	Atopic asthma
			Male, eosinophilic asthma.
			Non-atopic asthma
Lee E et al. 2017. South Korea	Children	4	Early onset, atopic asthma with mild symptoms and life impairment.
			Mild asthma with infrequent symptoms attacks and high SES, normal function and low BHR.
			Atopic, frequent symptoms attacks, daily impairment, eosinophilic, low SES and high BHR.
			Non atopic, infrequent symptoms attacks with minimal daily impairment, low eosinophilia and BHR and normal lung function.
Liang et al. 2016. China	Adults	4	Males, high basophilic, uncontrolled, high anti (IL- 10, TGF-b, and sRAGE) and pro inflammatory markers (IFN-g, IL- 4, IL-5, IL-6, IL-9, IL-17, IL-23, EGF, GM-CSF, and TNF-a) and high baseline function.
			Low pro (INF-g, IL-4, IL-5, IL-6, IL-8, IL-9, IL-13, IL-17, IL- 23, EGF, GM-CSF, TNF-a, and VEGF) and anti-inflammatory markers (IL-10 and sRAGE) uncontrolled, neutrophilic, and basophilic with high baseline function.
			Females, controlled, neutrophilic, low basophilic, low baseline function and high leptin and VFGF levels but low sRAGE.
Loureiro et al. 2015. Portugal	Adults	5	Young males, normal weight, well controlled, early onset, mild asthma, high function and low HCU.
			Old females, obese, long duration, good control and less severity, neutrophilic asthma.
			Young females, early onset atopic asthma, severe and frequent symptoms, no eosinophilia, normal function
			Females, obese, late onset, non-atopic, mixed low inflammation, uncontrolled, high impairment on daily life, high medication and HCU and normal lung function.
			Male, smoking, atopic, obese, late onset, aspirin sensitive, comorbid, long duration, eosinophilic, severe asthma with low function and high medication and HCU.

Loza et al. 2016. Multi-country	Adults	4	Early onset, mild disease, low BDR, low AHR, and low inflammation and predominant T2 high inflammation.
			Atopic, moderate to severe, mild reversable, eosinophilic, high BHR asthma with high T2 inflammation.
			Non atopic, controlled, low function, neutrophilic, with mixed symptoms severity, with T2 low inflammation
			Non atopic, severe, uncontrolled, high BDR and AHR with mixed inflammation with moderate T2 inflammation.
Mahut et al. 2011. France	Children	4	Nonsmoking exposed male with well controlled asthma.
			Nonsmoking exposed females with well controlled asthma.
			Nonsmoking exposed, well controlled disease with high airways tone.
			Parental smoking, well controlled asthma with small airways and lung size ratio.
Mäkikyrö et al. 2017. Finland	Adults	5	Well controlled mild asthma
			Partially controlled mild asthma
			Partially controlled moderate severity asthma.
			Uncontrolled with mixed severity.
			Uncontrolled severe asthma.
Mason et al. 2018. Italy	Adults	3	Male, nonsmoking, overweight, toluene diisocyanate (TDI) sensitized, late specific inhalation challenge (SIC) response with long exposure duration.
			Sensitization to TD1 and late SIC response.
			Atopic, sensitized to methylene diisocyanate (MDI) with early SIC response.
Mastalerz et al. 2015. Poland	Adults	4	High CRS, high atopy, mixed inflammation, mild to moderate intermittent, mostly controlled, low dose of ICS and higher levels of PGE2.
			Mostly AERD, females, High CRS, severe uncontrolled, eosinophilic, Aspirin sensitive, highest PGD2, LTE4, LTD 4 in ISS.
			Atopic, controlled, paucigranulocytic inflammation, no CRS, and low levels of proinflammatory mediators.
			Atopic, female, obese, early onset, severe, uncontrolled asthma, with low proinflammatory mediators, cycLTs and PGD2, and high levels of PGE2.
Nadif et al 2018. France	Adults	3	Adult-onset asthma, poor lung function, treatments, cough and phlegm, exacerbations, high neutrophil count, and high fluorescence's oxidative products (FlOPs) level.
			Paucigranulocytic asthma, normal lung function, rhinitis and low IgE level.
			Predominantly men with childhood-onset asthma, eosinophilic asthma, allergic comorbidities, and high IgE level.
Nagasaki et al. 2014.	Adults	4	Late onset, paucigranuolcytic asthma with high function.

Japan			Early onset, mild eosinophilic asthma.
			Late onset, severe, eosinophilic with serum Periostin.
			Uncontrolled, severe, serum neutrophilic, high IL6 and high comorbidities.
Nasreen et al. 2019. Canada	Children	3	Low initial attack rates that increase to high.
			Medium initial attack rates that decrease to none.
			High initial attack rates that decrease to medium.
Qui et al. 2018. China	Adults	4	Young, early onset, sputum neutrophilia and low eosinophilia, with moderate function impairment.
			Female, severe disease, eosinophilic, hypoxemic asthma with impaired function.
			Females, elderly, neutrophilic asthma with moderate to severe function impairment.
			Male, smoking, mixed inflammation asthma with moderate to severe impairment.
Sakagami et al. 2011. (abstract) Japan	Adults	5	Female, early onset, controlled long term asthma.
			Female, uncontrolled, and high depression.
			Female, atopic, late onset well controlled asthma.
			Elderly, female, non-atopic late onset asthma.
			Chinese, elderly, female, late onset well controlled asthma.
Schatz et al. 2013. USA	Adults and children	3	Children
			Atopic white male patients with no smoking exposure, normal function
			Mostly females, normal function
			Non atopic asthma
			Passive smoking exposure
			White race children, with higher BMI.
			Adolescents and adults
			White female, adult onset, non-aspirin sensitive with lower total IgE levels
			High atopy and atopic dermatitis.
			Mostly male patients
			Nonwhite race patients
			Aspirin sensitive asthma
Schimdlin et al. 2015. (Abstract). USA	Children	4	Atopic, males with late onset wheezing.
			Non-atopic, males, early onset, and low function.
			Atopic asthma with persistent wheeze.
			Female, infrequent wheeze and normal function.
Sendín-Hernández et al. 2018. Spain	Adults	3	Non atopic, familial atopy, mild, persistent asthma.
			Mild or intermittent asthma.
			Atopic, severe asthma that needs high treatment

Seino et al. 2018. Japan	Adults	3	Elderly, sever, uncontrolled, high treatment and high adherence barrier.
			Elderly, normal weight, sever, uncontrolled asthma, with no adherence barrier.
			Young, obese, controlled, persistent asthma with no adherence barrier.
Seys et al. 2017. Belgium	Adults	5	Sustained low function, eosinophilic neutrophilic sputum, and high IL-5-, IL-17/A/F and IL-25.
			High IL-5 and/or IL-10 and normal IL-17 F levels.
			High IL-6 profile.
			High IL-6 and IL-1β profile.
			Normal levels of all above cytokines.
Siroux V et al. 2011. Multicounty	Adults	5	EGEA
			Inactive allergic childhood onset asthma
			Active allergic childhood onset asthma
			Inactive adult-onset asthma
			Active adult-onset asthma
			ECRHS
			Inactive asthma
			Active allergic asthma
			Severe asthma
			Adult-onset non-allergic asthma
Tay et al. 2019. Singapore	Adults	3	Chinese, female, old, late onset obese, with controlled low symptom asthma.
			Non-Chinese, females, obese, uncontrolled, high symptoms, high GINA step and highest comorbidity.
			Male, multiethnic, atopic, early onset, smokers with moderate control level.
Tsukioka et al. 2017. Japan	Adults	3	Athletes with moderate levels of FeNO and IgE.
			Female athletes, lowest FeNO and IgE levels, worst function despite low symptoms.
			Male athletes' childhood onset asthma, atopic, higher FeNO and IgE and higher response to methacholine.
Wang LL et al. 2017. (abstract) China	Adults	5	Men, former or current smoker with high depression and anxiety, well controlled asthma with high lung function.
			Young, non-smoking women, atopic, with high psychological morbidity.
			Smoking, high HCU and sustained low function.
			Smoking men, high psychological morbidity and poor control, low SES with high HCU
			Women, non-smoking, non-allergic, slight obstruction, worst control and high anxiety and depression.
Watanabe et al. 2016	Adults	3	Older men with high morbidity of COPD and hypertension

(abstract).			Middle age me n with higher treatment adherence behavior.
Japan			Women, middle age, low treatment levels.
Wisnivesky et al;2019 USA	Adults	3	Decreased lung function, poor control, high hospitalization, and high sinusitis and depression rates.
			Normal lung function, good control, moderate hospitalization rate, high depression, and post-traumatic stress disorders.
			Men, near normal lung function, good control, low rate of emergency hospitalization and medication use, low mental health diseases and GERD.
Wu et al. 2017. China	Adults	3	Smoking, nonatopic, late onset, eosinophilic, short duration, NSAID sensitive, low function and impaired CT scan asthma with prior sinus surgery.
			Elderly, non-atopic, late onset, non-eosinophilic, long duration asthma, with prior sinus surgery.
			Male atopic asthma.
Zaihra T et al. 2016. Canada	Adults	4	Severe asthma, late onset, high dose ICS, least cluster stability
			Females, obese, severe asthma with highest cluster stability.
			Mild air entrapment with proximal airway remodeling, early onset, reduced lung functions.
			Moderate asthma with good lung functions.
Zhang X et al. 2019 (abstract). China	Adults	5	Moderate air entrapment with/without proximal airway remodeling
			Sever air entrapment with proximal airway remodeling
			Non atopic, infrequent symptoms and rhinitis asthma with normal function.
Zoratti E et al. 2018. USA	Adults	5	Non-atopic, low inflammation, frequent symptoms and rhinitis and high treatment.
			Moderate atopy, mild symptoms and rhinitis, low medication, and minimal function impairment.
			Atopic, high inflammation, infrequent symptoms, and intermediate treatment.
			Atopic, frequent symptoms, high treatment, and highly impaired function.

Table S.2

Table S 3. Number of derived phenotypes and their descriptions for studies on severe asthma

Reference ID and country	Numbers	Phenotypes description
Brinkman et al, 2011. (abstract). Muti-countries	3	Moderate disease with moderate use of OCS and low eosinophilia.
		Low eosinophilia and mild disease with low OCS usage.
		Obstructed eosinophilic with high OCS usage.
Desai et al. 2011. (abstract). UK	4	Obesity with discordant high symptoms/ low eosinophilia.
		Late onset disease with concordant high symptom/eosinophilia
		Early onset, discordant low symptoms/eosinophilia, low FEV1
		Early onset concordant symptom and eosinophilia.
Diver et al. 2018. UK	2	Microbial predominance with Hemophilus and Moraxella, high Gammaproteobacterial (G) to Firmicutes (F) ratio
		Low microbial prevalence and low G:F ratio
Fitzpatrick et al, 2018. USA	4	Late-onset, non-atopic, non- Hispanic white, normal function and symptomatic asthma
		Early-onset atopic asthma, lower lung function and high symptoms and medication usage.
		Early-onset atopic asthma with high comorbidities, high bronchial responsiveness, and low lung function.
		Early-onset atopic asthma with sever obstruction, highest symptoms and medication usage.
Freitas PD et al. 2018 (abstract). Brazil	3	Physically unactive, obese, with depression and high treatment.
		Physically active, young, less comorbidities and good control and QOL.
		Intermediate physical activity.
Gomez et al, 2017. USA	4	Young females' mild atopic asthma with low YKL levels.
		Males with non-atopic asthma of good treatment response and low YKL level.
		Severe asthma, elderly with high treatment and high YKL levels.
		Obese elderly with sever short term disease and high YKL levels.
Jang et al, 2017; Korea	4	Mild, atopic, low rhinitis, eosinophilic asthma among young
		Non atopic neutrophilic severe asthma.
		Female highly reactive asthma with mixed inflammation
		Male smoking severely obstructed with high rhinitis.
	3	Non atopic mild obstruction with low treatment and non-SRA criteria.
		Atopic, severe obstructed, with treatment and positive SRA criteria.

Konstantellou et al. 2015. Greece		Atopic, mild disease with low treatment and non-SRA criteria.
Lau et al, 2017: Singapore	2	High Absolut eosinophilic asthma, young, males and requires more hospital admission
		Low eosinophilic asthma
Moore et al 2010; USA	6	Young females, childhood onset mild atopic asthma with low treatment and health care utilization.
		Old females' childhood onset mild disease with high treatment use.
		Elderly, late onset, obesity, non-atopic sever disease with high treatment usage and HCU.
		Early onset, atopic severe asthma, with treatment requirement and HCU.
Newby et al, 2018; UK	5	Atopic, early onset, low function, high HCU, high treatment, and high BDR.
		Late onset, obese, frequent symptoms, mild function decline, high treatment and high depression.
		Non atopic, normal function, infrequent symptoms and low treatment.
		Marked function decline, high treatment and infrequent exacerbations.
Raherson et al, 2018 (abstract). France	3	Atopic asthma.
		Male dominant eosinophilic asthma.
		Non atopic asthma.
Sekiya et al. 2015; Japan	5	Early onset, sever symptoms, activity limitation, high treatment and high HCU.
		Elderly, female, long disease duration and high CRS and nasal polyps.
		Atopic low treatment and better hospitalization prognosis.
		Elderly, male, with COPD.
		Mild disease with previous hospitalization.
Simpson et al; 2017. Multi-country		Classic asthma: high airway reversibility and high eosinophiles.
		Pulmonary treatable traits asthma (high prevalence of multiple pulmonary traits.
		Steroid insensitive: high eosinophiles despite good medication adherence.
		Reflux and cough: cluster highly identified by traits of reflux and cough.
		High treatable traits: high prevalence of treatable traits across subjects.
		Low prevalence of treatable traits across subjects.
Taniguchi et al; 2014;(abstract) Japan	5	Young onset atopic severe asthma.
		Older onset, female, obesity severe asthma
		High smoking pack years and high eosinophilia.
		High smoking pack years, low DLCO and low eosinophile
		Non indicated

Serrano Pariente et al. 2015. Spain	3	Elderly, with severe asthma.
		Respiratory arrest, impaired consciousness and need for mechanical ventilation
		Young, no sufficient treatment and sensitization to alternata and soybean.
Wu et al; 2014. USA		Healthy controls with normal lung function and no symptoms.
		Mild asthma, less symptoms, better quality of life, early onset, less atopic, more allergen skin test reaction, better clinical outcome, and high BAL eosinophile and neutrophile than cluster 1.
		Hispanic women, frequent symptoms, low QOL, high allergic sensitization, low inflammation, and near normal FEV1 value with low hospitalization.
		Female, non- Caucasian, high BMI, high symptoms, and early onset asthma; high familial asthma, low lung function high BHR and high inflammation.
		Elderly, late onset asthma, non-allergic, high nasal polyposis and sinusitis and high CS dosage; low function, high eosinophile and neutrophiles.
		Early onset, high symptoms, lowest lung function, high HCU, high sinusitis, high OCS, high FeNO, eosinophile and neutrophiles and high osteoporosis.
Weng-Jing Ye et al. 2017. China	4	Early onset atopic asthma.
		small airway obstruction and atopic asthma.
		late-onset and non-atopic asthma.
		Severe airflow obstruction and obvious airway remodeling.
Yourouko va et al. 2017. Bulgaria	4	Late-onset, non-atopic asthma with impaired lung function.
		Late-onset, atopic asthma.
		Late-onset, aspirin sensitivity, eosinophilic asthma.
		Early-onset, atopic asthma.

Table S 4. Number of derived phenotypes and their descriptions for studies on COPD AND ACO

Study ID; Year	Number of phenotypes	Phenotypes description
Augustin et al; 2018. Netherland	7	Male, mild obstruction with mildly impaired diffusion.
		Elderly, males, moderate obstruction with moderately impaired diffusion.
		Sex non differential, moderate to severe obstruction and diffusion impairment.
		Sex non differential, moderate to severe obstruction with mild diffusion impairment.
		Elderly, males, severe obstruction, severe hyperinflation, moderate diffusion impairment and respiratory muscle weakness.
		Elderly, female, severe obstruction, sever hyperinflation and moderate diffusion impairment.
		Elderly, males, severe obstruction, severe hyperinflation, sever impaired diffusion, respiratory muscle weakness and alveolar hypoventilation.
Bafadhel et al.; 2011 UK	4	Bacteria-predominant.
		Eosinophil-predominant.
		Virus- predominant.
		Pauci-inflammatory reaction. Clusters further varied by Sputum IL-1b, serum CXCL10 as biomarker.
Bertini et al.; 2013. Italy	3	High Formate, Serine, Valine, Lysine, Acetate, Alanine, Isoleucine and Leucine.
		High radiologic emphysema.
		Low Proline.
Burgel et al.; 2017 France	5	Elderly, high mortality, high CVD and diabetes and less severe disease.
		Intermediate mortality, low comorbidities and severe to moderate disease.
		Elderly, obese, intermediate mortality and high comorbidities.
		High mortality, severe disease with low comorbidities.
		Low mortality, mild disease, and low comorbidities.
Burgel et al.; 2010 France	4	Young, severe airflow limitation (GOLD stage 3 and 4), low BMI, frequent exacerbations, high levels of anxiety and depression; cardiovascular comorbidities were infrequent.
		Older individuals, mild airflow limitation, low dyspnea, mild overweight, low anxiety and depression levels, almost no exacerbations, and mild impairment in HRQoL; higher prevalence of cardiovascular morbidity.
		Moderate to severe obstruction, young with low prevalence of cardiovascular and depression comorbidity.
		Moderate to severe obstruction, older with high prevalence of cardiovascular and depression comorbidity.
Burgel et al.; 2012 Belgium	3	Mild to moderate obstruction, mild emphysema, mild dyspnea, normal nutrition status and low comorbidities.
		Younge, males, underweight, Severe obstruction, marked emphysema, sever dyspnea, impaired QOL, high musculoskeletal disease
		Olde, obese, moderate to severe obstruction with bronchial thickening, mild emphysema than C2, high rates of CVD and diabetes.

Chen et al; 2014. **Taiwan and China**	5	Young, mild obstruction, mild symptoms, and infrequent exacerbations.
		Elderly, mild obstruction, mild and infrequent symptoms, and high CIS usage.
		Elderly, underweight, moderate obstruction with sever exacerbation and dyspnea symptoms.
		Severe obstruction, high symptoms, low BOS and mild exacerbations.
		Severe obstruction, low BOS, severe frequent exacerbations, high ICS and high mortality.
Chubachi et al (abstract); 2016 **Japan**	5	Low comorbidity
		Lung and other cancers.
		Metabolic and CVD.
		Psychological and GERD.
		Cachectic, anemia and Osteoporosis.
De Torres et al; 2017. **Spain**	3	Young, mild obstruction, low BMI, and low CVD.
		Characteristics between C1 and C3.
		Elderly, high BMI, marked obstruction, high symptoms and high HTN, diabetes, OSA and CVD.
Divo et al. abstract; 2016. **Not indicated**	4	Young, male, obese, moderate impairment on QOL, physical ability and health, moderate diffusion capacity
		Elder, obese, sex undeferential, severe obstruction, moderate diffusion capacity with marked health, physical impairment, and high mortality.
		Elderly, female, obese, mild obstruction, high diffusion capacity, least health impairment and low mortality.
		Reference: Young, female, no obstruction, high diffusion capacity, normal to overweight, with the least overall impairment and mortality.
Fens et al.; 2013. **Netherland**	4	Females, mild obstruction, low symptoms, good QOL with high lung density and little emphysema.
		Combined CB and emphysema, moderate to severe obstruction, GOLD stage 3, impaired diffusion capacity and emphysema.
		Mild obstruction, GOLD stage 1, Hypercholesterolemia and low lung density and high emphysema.
		Smoking, high symptoms, preserved function, low QOL, with moderate emphysema and lung density.
Guillamet RV et al.; 2018. **USA**	9	Elderly, high depression and mild comorbidities.
		Low comorbidities and low remission rates.
		Elderly, with high CHD and CHF.
		Young, low comorbidities, high medication, and readmission rates.
		Advanced disease and frequent readmissions.
		Young, with high CVA.
		Young, atopic asthma with high readmission rates.
		Young, high CKD and diabetes with few readmissions.
		Advanced disease with frequent readmissions.
Harrison SL et al.; 2014 **UK**	3	Controlled, infrequent symptoms, low emotional sensitivity, and short illness duration.
		Uncontrolled, frequent symptoms, high emotional sensitivity, and short illness duration.
		Coherent illness, frequent symptoms, emotional sensitivity, and cyclical illness timeline.

Haghighi et al, 2019. USA	4	Asymptomatic, normal airway structure, normal lung function and airway wall thickening and moderate emphysema.
		Obese, female, increase tissue fraction at inspiration and minimal emphysema and lowest emphysema progression rate.
		Elderly, male, small airway narrowing, decreased tissue fraction ay expiration and high air entrapment.
		Lean, male, severe COPD and high emphysema progression rate.
Kim W.J et al; 2017. Asia	3	Elderly, male, low function, and few symptoms,
		Very Low function, high symptoms of dyspnea and health impairment.
		Nonsmoking, obese, mild disease and normal function.
Kim S et al 2017. Korea		Young, mild, low symptoms, low disease impairment on life and low inflammation.
		Male, old, heavy smokers, severe obstruction, underweight, high inflammation, high symptoms and disease impairment of daily life, high emphysematous changes.
		Females, non-smokers, moderate dyspnea and disease impairment on life.
Kukol et al. 2019;(abstract). Russia		Normal body weight, with disease duration more than 5 years; with a frequency of exacerbations less than 2 times a year among women
		Elderly, younger than cluster 1, overweight, disease duration <5years and exacerbation less than twice a time among women.
		Overweight, less than 6 years of disease duration and exacerbation less than twice a year among men.
		Under weight, more 7 years of disease duration, and exacerbation less than twice yearly among men.
		Overweight more than 8 years disease duration and less than 2 exacerbations yearly.
Lee et al, 2019, Korea	4	ACO, second best lung function to mild COPD, second highest age, highest BMI, least smoking, longest walking distance and lowest CAT score and highest rate of asthma.
		Mild COPD
		Moderate COPD
		Severe COPD
Li et al, 2016, USA	5	Resistant smokers with normal function and early emphysema.
		Resistant smokers with mild function decline and no emphysema.
		Heavy smokers, mild COPD minimal obstruction and emphysema.
		Less smokers, moderate COPD high HCU and ICS usage.
		Sever COPD, sever obstruction, sever emphysema and high ICS and HCU usage.
Liang et al, 2019: Korea	4	Seven trajectories of association between clinical and autoantigens parameters, retrospective exacerbation (AE) increased with CAT score.
		Five trajectories of association between clinical and immunological parameters, retrospective exacerbation associated negatively with age, lung function and sputum antibodies (P0, Scl70, Sm, U1-SnRNP, PR3 and Ro/SSA) and serum globulin (Glb) and positively related to white blood cells, sputum anti-PR3, sputum anti-Ro/SSA, and sputum anti-U1-SnRNP were significantly negatively correlated with AE
		Five trajectories of association between clinical and immunological parameters, AE was positively associated with the CAT score and sputum autoantibodies (U1-SnRNP, PR3, MPO and Ro/SSA).

		Six trajectories of association between clinical and immunological parameters, AE was negatively associated with serum uric acid and blood neutrophil count.
Lopes et al; 2019. Brazil	2	Frequent symptoms, high consequences, and cyclical timeline of illness, less coherence and high emotional sensitivity. (destressed)
		Infrequent symptoms, less cyclical timeline of illness, low emotional sensitivity, and high illness coherence. (coping)
Ning et al; 2016, (abstract). China	4	Chronic bronchitis in smokers with normal pulmonary function.
		Chronic bronchitis or mild chronic obstructive pulmonary disease (COPD) patients with mild airflow limitation.
		Heavy smoking, poor quality of life and severe airflow limitation.
		Atopic patients with mild airflow limitation, elevated serum IgE and clinical features of asthma
Peters et al; 2016 Netherlands	3	Moderate COPD, normal weight, high exercise performance and mild impact on quality of life, functionality, and symptoms.
		Moderate COPD, overweight, high exercise performance and high impact on symptoms, functionality, and quality of life.
		Severe COPD, overweight, moderate exercise performance and mild impact on symptoms, quality of life and functionality.
Pikoula et al; 2019. UK	5	Young, female, smoking, high psychological illness, highly deprived group with high IMD.
		Male, late diagnosis, severe airflow obstruction, frailty and the lowest comorbidities.
		Elderly, male, former smokers, high IHD, CVD and diabetes.
		Under weight, females, eosinophilic, sever disease with low CVD high obesity and atopy.
		Females, smoking, obese, atopic mild asthma high CRS and GERD.
Rodrigues et al.; 2018. Brazil	2	Under to normal weight, severe disease, low function, impaired daily activity, and low muscle strength, compared to cluster 2.
Scarlata et al; 2018 Italy	3	Mild central and peripheral obstruction and high KCO.
		Intermediate peripheral obstruction with no comorbidities
		Sever obstruction, low KCO and short estimated survival.
Xavier F et al; 2019. Brazil	3	Young, physically active, low obstruction, low dyspnea, and good body composition.
		Elderly, physical inactive with low mortality.
		Physical inactive, poor QOL, and low mortality
Yoon et al.; 2019. Korea	4	Asthma and COPD overlap
		Mild COPD
		Moderate COPD
		Severe COPD
Asthma COPD overlap		
Rootmensen et al.2016. Netherlands	4	Cluster1: excessive smoking history COPD without sign of emphysema
		Cluster 2: emphysematous COPD

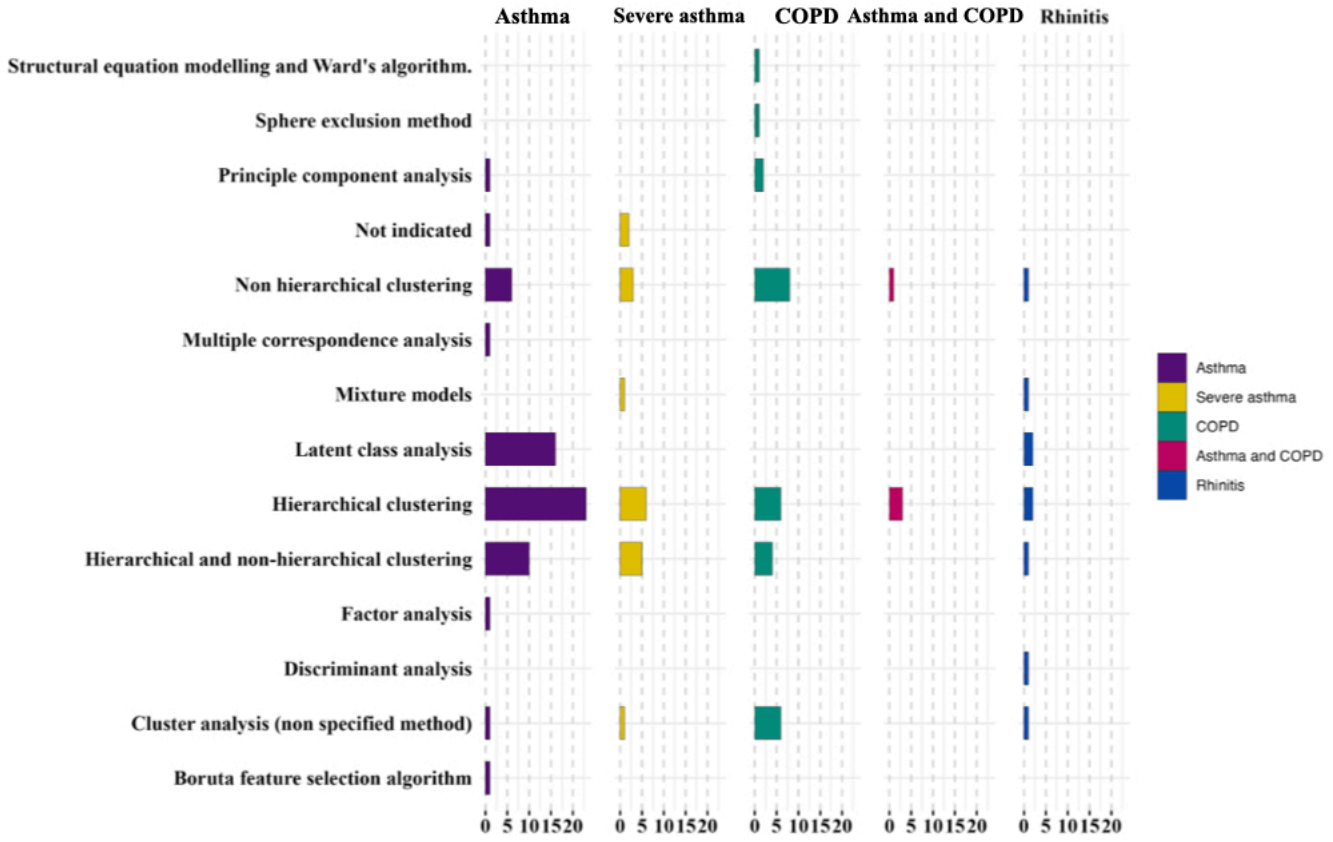
		Cluster3: patients with allergic asthma characteristics
		Cluster4 features suggesting asthma and COPD overlap
De Vries et al. 2018. Netherlands	5	Cluster 1: female, obese, high symptoms, combined asthma and COPD, and low inflammation.
		Cluster 2: males, combined asthma, and COPD, eosinophilic, high FENO and low treatment step.
		Cluster3: non-Caucasian, combined asthma, and COPD, highly obstructed, eosinophilia, low exacerbations, and low treatment steps.
		Cluster4: atopic, combined asthma and COPD, neutrophilia, and high exacerbations,
		Cluster5: mostly asthmatics, high lung function, and low exacerbations.
Fingleton et al. 2017. New eland and China	5	Cluster1: smokers, atopic, severe, late onset, asthma/ chronic bronchitis/ emphysema overlaps with systemic inflammation.
		Cluster2: smokers, moderately severe, early onset asthma combined asthma and COPD, type two dominant inflammation.
		Cluster3: minimal smoking, atopic, eosinophilia and type 2 inflammation with minimal airflow obstruction.
		Cluster4: late onset non atopic with minimal airflow obstruction.
		Cluster 5: atopic, early onset and mild/ intermittent phenotype.
Górska K t al.2017; Poland	3	Cluster1: atopic asthmatics.
		Cluster2: non atopic asthmatics,
		Cluster 3: smokers with highly obstructed COPD
		Cluster 3:1: smokers, highly obstructed COPD with high eosinophilia.
		Cluster 3:2: smokers, highly obstructed COPD with low eosinophilia.
		Cluster 1: asthmatics, atopic with eosinophilia.
		Cluster 2: combined asthma and COPD that has further 5 subgroups:
		2:1: mainly COPD with high MM6.
		2:2: smokers, mainly COPD, highly obstructed,
		2:3: mainly combined asthma and COPD
		2:4: atopic, asthma with high eosinophilia.
		2:5: low smoking, mainly COPD, with minimal obstruction.

9.1.1.1 TABLE S 5. NUMBER OF DERIVED PHENOTYPES AND THEIR DESCRIPTIONS FOR STUDIES ON RHINITIS

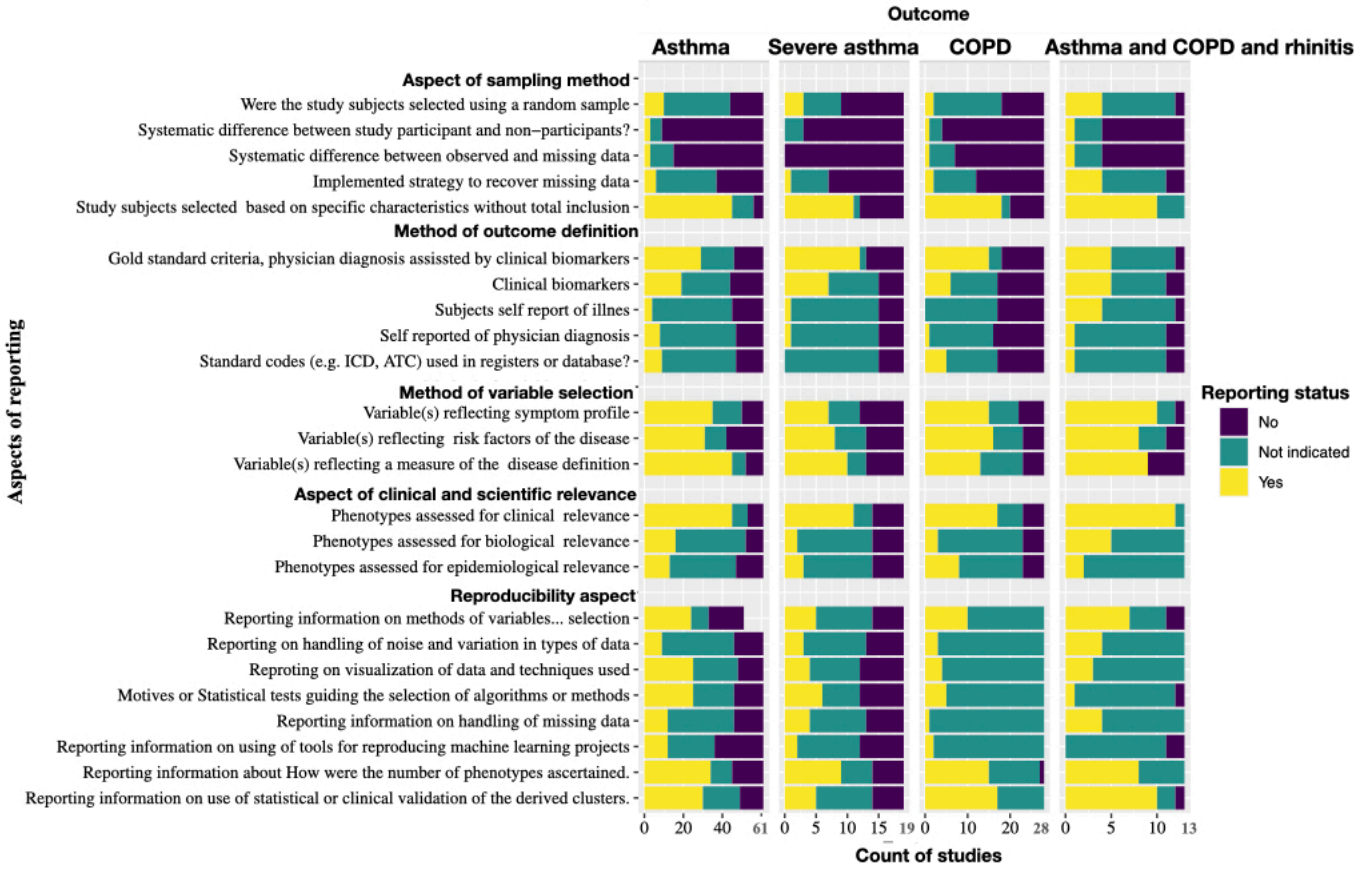
Refence ID and country	Number of phenotypes	Phenotypes description
Adnane et al.2017. Morocco	3	Female, eosinophilic, high CRS with nasal polyps, low endoscopy score, low CT and sino-nasal outcome test
		Male, non-eosinophilic rhinitis.
		Female, high CRSsNP, high endoscopy score, CT and sino-nasal outcome score.
Agache et al. 2010; France	5	Short breast-feeding duration and severe rhinitis- children.
		Male, polysensitization and sever rhinitis - children.
		Sever rhinitis with polysensitization – adults.
		Male, high pets' exposure and sever rhinitis- adults.
		High atopy and polysensitization- adults.
Bousquet et al. 2015. France	4	Moderate to severe rhinitis, low QOL, high symptoms, high disease burden than C3.
		Elderly, Female, early onset, mild intermittent rhinitis, high QOL, low symptoms and low comorbidities.
		Males, moderate to severe rhinitis, less symptoms and better QOL than C1.
		Young, female, atopic, later onset, severe to moderate persistent rhinitis, high symptoms, low QOL and high disease burden.
Burte et al. 2015. France	3	No nasal symptoms,
		High nasal symptoms through year, high sinusitis, and low sensitization.
		Nasal symptoms at spring, high sensitization, high hay fever, high allergic rhinitis and conjunctivitis and high polysensitization.
Herr M et al.2012. France	3	Highly atopic, sever occasional wheeze, low HCU, high sensitivity to cowmilk, egg white, nuts, cat and house mites' dust.
		Mild atopic, severe wheeze, impaired activity with high nocturnal cough and respiratory infections.
		Reference: mild occasional wheeze, with low respiratory and allergic outcomes.
Kurukulaaratchy et al. 2015. UK	4	Atopic, early onset, moderate severity, normal function, low BHR, BDR and prevalence of asthma.
		Female, non-atopic, late onset, mild severity, normal function with low BHR and BDR, low asthma and low inflammation
		Severe earliest-onset rhinitis with asthma, the youngest rhinitis onset, the highest comorbid asthma (of simultaneous onset) and atopy.
		Male, atopic, early onset, seasonal disease of mild severity, low function with moderate BHR and BDR, moderate asthma and high inflammation.
Lee E.L. et al 2016.	4	Non atopic rhinitis with low SES.
		Atopic rhinitis with normal lung function.

Korea		Atopic, low function, high food allergy, high atopic dermatitis, high symptoms with high treatment need,
		Non atopic rhinitis with high SES.
T.A Nakayama et al. 2012. Japan	4	Young, low peripheral eosinophil, basophile and mucosal eosinophile count, low CT score and polyps score and low symptoms.
		Low peripheral eosinophil, basophile and mucosal eosinophile count, higher CT and polyp score than cluster 1, and high symptoms score.
		highest mucosal eosinophile count and peripheral basophile and eosinophile counts and low polyp and symptoms score.
		High CT and polyps score, the highest incidence of ATA as comorbidity, high peripheral eosinophile, basophile and mucosal eosinophile count and high symptoms.
Soler et al.2015. USA and Canada	5	Elderly, males, high alcohol intake, high diabetes, low depression, moderate severity with low QOL.
		Elderly, male, high depression, moderate QOL and low endoscopy and control score.
		Female, high depression, low endoscopy score and intermediate control score.
		Young, males, high depression, mild disease and low QOL.
		Females, high depression and fibromyalgia, low control, and endoscopy score and low QOL.

9.2 MAIN TEXT FIGURES


Asthma
Severe asthma
COPD
Asthma and COPD
Rhinitis
Structural equation modelling and Ward's algorithm.
Sphere exclusion method
Principle component analysis
Not indicated
Non hierarchical clustering
Multiple correspondence analysis
Mixture models
Latent class analysis
Hierarchical clustering
Hierarchical and non-hierarchical clustering
Factor analysis
Discriminant analysis
Cluster analysis (non specified method)
Boruta feature selection algorithm
0 5 10 15 20
Asthma
Severe asthma
COPD
Asthma and COPD
Rhinitis

9.2.1.1 FIGURE 22. NUMBER OF STUDIES USING EACH UNSUPERVISED PHENOTYPING METHOD FOR EACH RESPIRATORY OUTCOME


Outcome
Asthma
Severe asthma
COPD
Asthma and COPD and rhinitis
Aspects of reporting
Aspect of sampling method
Were the study subjects selected using a random sample
Systematic difference between study participant and non−participants?
Systematic difference between observed and missing data
Implemented strategy to recover missing data
Study subjects selected based on specific characteristics without total inclusion
Method of outcome definition
Gold standard criteria, physician diagnosis assissted by clinical biomarkers
Clinical biomarkers
Subjects self report of illnes
Self reported of physician diagnosis
Standard codes (e.g. ICD, ATC) used in registers or database?
Method of variable selection
Variable(s) reflecting symptom profile
Variable(s) reflecting risk factors of the disease
Variable(s) reflecting a measure of the disease definition
Aspect of clinical and scientific relevance
Phenotypes assessed for clinical relevance
Phenotypes assessed for biological relevance
Phenotypes assessed for epidemiological relevance
Reproducibility aspect
Reporting information on methods of variables... selection
Reporting on handling of noise and variation in types of data
Reproting on visualization of data and techniques used
Motives or Statistical tests guiding the selection of algorithms or methods
Reporting information on handling of missing data
Reporting information on using of tools for reproducing machine learning projects
Reporting information about How were the number of phenotypes ascertained.
Reporting information on use of statistical or clinical validation of the derived clusters.
0 20 40 61
0 5 10 15 19
0 10 20 28
0 5 10 13
Count of studies
Reporting status
No
Not indicated
Yes

9.2.1.2 FIGURE 23. QUALITY ASSESSMENT ITEMS REPORTING FOR STUDIES ON ASTHMA, SEVERE ASTHMA, COPD, ASTHMA AND COPD AND RHINITIS.

9.3 SUPPLEMENTARY TABLES

9.3.1.1 TABLE S 6. ARTICLES EXCLUDED AND BASIS OF EXCLUSION

Author	Year	Outcome	Reason for exclusion
Baptist et al.[419]	2018	Asthma	Study data was retrieved from a clinical trial
Bafadhel et al.[420]	2011	COPD	Duplicate
Bhargava et al .[421]	2019	Asthma	Duplicate
Boudir et al.[101]	2019	Asthma	Comparison of outcome of phenotypes included in previous record.
Brew et al.[422]	2018	Asthma	Combined supervised and non-supervised methods
Brew et al.[423]	2019	Asthma	Duplicate
Castaldi et al.[424]	2017	COPD	The study was aiming at comparing stability metrics for COPD phenotypes reproducibility
Chang et al.[425]	2014	Asthma	Study data was retrieved from a clinical trial
Chupp et al.[426]	2013	Asthma	Phenotyping based on genetic data (genotyping)
Delgado Eckert et al.[427]	2017	Asthma	Part of the data is derived from clinical trial
Ding et al.[428]	2018	Asthma	Data from clinical trial
Diver et al.[429]	2018	Severe asthma and COP	Duplicate abstract of fully included study

Eom et al. [430]	2018	Asthma	Full data could not be retrieved
Fens et al. [431]	2012	COPD	Full data could not be retrieved
Fernandez et al.[432]	2018	Asthma	Phenotyping based on genetic data (genotyping)
Guo Y et al. [433]	2015	Asthma	Full data could not be retrieved
He z et al.[434]		Obstructive sleep apnea	Non relevant outcome and method
Howrylaket al. [435]		Asthma	Data from clinical trial
Humphroies et al.[436]	2019	Emphysema	Non relevant methods: comparing patterns of CT emphysematous scan to mortality
Khavarich et al. [437]	2018	Asthma	Full data could not be retrieved
Khoo et al.[438]	2020	Wheezing	Outcome is wheezing and phenotyping based on data including genetic data (genotyping)
Khusial et al.[439]	2017	Asthma	Data from clinical trial
Klick et al.[440]	2018	COPD	Comparative review of COPD characterization using different methodological approaches
Kuhlen et al [441]	2015	Asthma	Supervised phenotyping method
Kuo et al.[442]	2011	Asthma	Non relevant method: using semi supervised method for phenotypes derivation

Kupczyk et al.[443]	2018	Asthma	Supervised phenotyping method
Mason et al.[444]	2017	Occupational asthma	Supervised phenotyping method
Panico et al.[445]	2019	Wheeze	Non relevant outcome: wheeze
Patrawalla et al. [446]	2017	Asthma	Use of supervised methods based on variables applied in the SARP analysis
Prosperi et al [447]	2014	Airway diseases	Review article
Reiley et al[448]	2018	severe asthma	Using supervised methods
Rennard et al [449]	2015	COPD	Supervised phenotyping method
Sato Set al[450]	2016	Idiopathic intestinal pneumonia with emphysema	Non relevant outcome
Sharma et al[451]	2010	COPD	Data from clinical trial
Sohn et al.[452]		Asthma	Supervised phenotyping method
Sukhan et al.[453]	2018		Full data could not be retrieved
Wang et al.[454]	2019		Review of comparison of methods of unsupervised phenotyping of diseases

Yang et al.[455]	2018	Emphysema	Supervised phenotyping method

9.3.1.2 TABLE S 7. QUALITY ASSESSMENT RESULTS, INDICATING THE NUMBER OF STUDIES REPORTING ON THE DIFFERENT QUALITY ITEMS FOR EACH OUTCOME.

Outcome (number of studies)	Asthma n = 61			COPD n =28			Severe asthma n = 19			Asthma and COPD n =4			Rhinitis n =9		
Quality items	Yes	No	Not indicated	Yes	No	Not indicated	Yes	No	Not indicated	Yes	No	Not indicated	Yes	No	Not indicated
Aspect of sampling method															
Study subjects selected based on specific characteristics without total inclusion	42	11	7	19	3	2	11	1	3	2	2	0	7	2	1
Were the study subjects selected using a random sample	10	34	16	1	16	7	2	5	8	2	2	0	1	7	2
Systematic difference between study participant and non-participants?	4	6	50	0	3	21	0	3	12	0	1	3	1	3	6
Systematic difference between observed and missing data	5	12	43	1	5	18	0	0	15	0	0	4	1	2	7
Implemented strategy to recover missing data	7	32	16	1	10	17	1	6	12	2	2	0	2	5	2
Method of outcome definition															
Gold standard criteria, physician diagnosis assisted by clinical biomarkers	38	8	14	17	2	5	12	0	3	2	2	0	4	4	2
Clinical biomarkers	19	24	17	9	9	6	6	5	4	3	1	0	3	5	2
Subjects self-report of illness	6	39	15	1	17	6	1	10	4	1	3	0	3	5	2
Self-reported of physician diagnosis	7	35	16	2	14	8	1	10	4	0	4	0	1	6	3
Standard codes (e.g., ICD, ATC) used in registers or database?	6	38	11	5	11	12	0	10	9	0	4	0	2	5	2
Method of variables selection															
Variable(s) reflecting symptom profile	34	11	15	16	3	5	7	5	3	3	1	0	7	1	2
Variable(s) reflecting risk factors of the disease	32	12	16	17	2	5	9	3	3	3	0	1	5	3	2
Variable(s) reflecting a measure of the disease definition	36	9	10	14	5	9	8	4	7	3	0	1	6	1	3
Aspect of clinical and scientific relevance															
Phenotypes assessed for clinical relevance	45	9	6	18	4	2	11	2	2	3	1	0	9	0	1
Phenotypes assessed for biological relevance	16	36	8	5	17	2	1	12	2	2	2	0	4	5	1

Phenotypes assessed for epidemiological relevance	18	35	2	9	13	6	3	10	6	0	4	0	2	7	0
Reproducibility aspect															
Reporting information on methods of variables' selection	24	10	26	11	4	9	5	2	8	2	0	2	7	1	2
Reporting on handling of noise and variation in types of data	9	39	12	3	16	5	3	7	5	1	3	0	4	5	1
Reporting on visualization of data and techniques used	25	25	10	6	14	4	4	6	5	2	2	0	2	7	1
Motives or Statistical tests guiding the selection of algorithms or methods	25	23	12	5	13	6	5	4	6	1	2	1	0	8	2
Reporting information on handling of missing data	12	35	13	1	18	5	4	6	5	3	1	0	1	8	0
Reporting information on using of tools for reproducing machine learning projects	20	25	10	5	15	8	3	6	10	0	2	2	2	8	0
Reporting information about How were the number of phenotypes ascertained.	34	11	15	16	3	5	9	0	6	2	2	0	6	3	1
Reporting information on use of statistical or clinical validation of the derived clusters.	30	19	11	18	3	3	5	6	4	2	1	1	8	1	1

9.3.1.3 TABLE S 8. NUMBER AND LIST OF VARIABLES USED FOR ASTHMA PHENOTYPING.

Reference ID and country	Population age	Number of variables	Variables considered for analysis
Amaral et al. 2019, US	Adults	15	Limit activity by wheezing; Wheezing with exercise; Sleep disturbance by wheezing; Wheezing attack; FEV1/FVC; Absenteeism by wheezing; Emergency department visit by asthma; Female; FEV1; FeNO ≥ 35 ppb; B-Eos ≥ 300/mm3; Ever-smoked; Hay fever; BMI ≥ 30 kg/m2 and Early asthma onset.
Amelink et al. 2013. Amsterdam	Adults	35	Reversibility in FEV1; pbFEV1 in percentage predicted; pbFVC in percentage predicted; pbFEV1/FVC in percentage predicted; KCO in percentage predicted; RV%/TLC in percentage predicted; gender; race; medication use and number of exacerbations
Bhargava et al. 2018. India	Adults		Demographic characteristics, age at disease onset, disease duration, smoking and indoor air pollution, BMI, serial lung functions and allergen sensitization.
Benton et al. 2010. USA	Children and adolescents	11	ITG Daytime Symptoms Score; ITG Nighttime Symptoms Score; ITG Functional Limitations Score; Asthma Control TestTM; Post-bronchodilator FEV1, % predicted; Post-bronchodilator FEF25-75, % predicted; Post-bronchodilator FEV1/FVC, % predicted; Nasal eosinophils, %; Nasal neutrophils, %; Blood eosinophils, % and Total serum IgE, IU/mL

Bochenek et al.2014. Poland	Adults	68	Asthma age of onset, body mass index (BMI), asthma control, asthma related emergency department (ED) visits, hospitalizations, and stays in the intensive care unit (ICU), upper airway symptoms, occurrence of nasal polyps, and history of polypectomies as indicators of chronic rhinosinusitis, FEV1 percent predicted, DFEV1 after bronchodilator, skin prick test responses, total IgE levels, and blood eosinophilia, sex and asthma duration.
Boudier et al. 2013. Multicounty	Adults	9	Asthma symptom score, waking up by attack of cough, 12-month productive cough, asthma attack within 12 months, asthma treatment, allergic sensitization, total IgE and bronchial hyperresponsiveness.
Cabral et al. 2016. Brazil	Children and adolescents	20	Gender; obesity; race; asthma severity; age at the onset of asthma; asthma triggers; blood eosinophils; number of previous asthma hospitalizations; tendency toward exacerbation; history of ICU admission; specific serum IgE levels; gastroesophageal reflux; sinus infection; baseline FEV1 (% predicted); FEV1/FVC ratio; labile FEV1; presence of fixed airway obstruction; best response to bronchodilator
Celejewska-Wójcik et al. 2020. Poland	Adults	16	Sex; BMI; Age at asthma onset; Asthma control and severity; Use of inhaled and/or oral steroids; History of asthma exacerbations; FEV1; Blood eosinophil count; Total serum IgE; Atopy status determined by skin prick tests; Induced sputum inflammatory cells; Selected eicosanoids – PGD2, PGE2, LTE4.
Chanoine et al.2017. France	Adults	1	The controller-to-total asthma medication ratio was described among elderly women with ever asthma using a 0.5-threshold.
Couto et al. 2015. Portugal	Adults	9	Presence of respiratory symptoms; current use of asthma medication; presence of rhinitis or other allergic diseases; conjunctivitis; urticaria; eczema; anaphylaxis; drug, food, and venom allergies.

Cruz et al. 2018 (abstract). Brazil	Adults	8	Not indicated
Damiens K et al.2013. Canada	Adults	12	Age, sex, body mass index, smoking, atopy, occupational rhinitis, inhaled corticosteroids, FEV, PC20, sputum eosinophils and neutrophils, occupational agent (high [HMW] or low [LMW] molecular weight agents), duration of exposure and work status (at or off work) at the time of investigation.
Deliu et al. 2016. Turkey	Children	68	Not indicated
Deliu et al. 2018. Turkey	Children	45	Sex; maternal allergic disease; Father with allergic disease; Exposure to tobacco smoke; Pet ownership; Atopic; Sensitized to house dust mite; Sensitized to grass; Sensitized to trees; Sensitized to weeds; Sensitized to molds; Sensitized to cat; Sensitized to dog; Sensitized to cockroach; Blood Eosinophil % 0.15-0.3; Blood Eosinophil % 0.3-0.5; Blood Eosinophil % >0.50; IgE total; Use of long-acting beta2 agonist; Use of Montelukast; Use of regular controller medication; No use of inhaled corticosteroids.
Depner et al. 2013. Multicountry	Children	Not indicated	Specific immunoglobulin levels, Feno measurement and lung functions asthma medication.
Dudchenko et al. 2018. Russia	Adults	23	Disease control according to Asthma Control Test (AST), the frequency of asthma attacks per day, the severity of asthma attacks, the use of short-acting bronchodilators, the severity of shortness of breath, the frequency of coughing per day, sleep disturbances, physical activity during the daytime, constancy, prevalence and number of dry wheezing, the presence and shortness of breath in the previous year, choking in the previous year, reaction to irritants and weather changes, vital lung capacity (VC), forced VC (FVC), FEV1, peak expiratory flow rate (PEFV), average volumetric expiratory flow rate at the level of 25-75 % FVC (SOS25-75), instantaneous volumetric expiratory flow rates at the level of 25, 50 and 75% FVC (respectively MOS25, MOS50 and MOS75).

Folz et al.2018. USA	Adults	Not indicated	Not indicated
Fontanella et al. 2018. UK	Children	Not indicated	Serum IgE to 112 allergenic molecules using ImmunoCAP ISAC.
Gonem S et al. 2012 (abstract). UK	Adults	Not indicated	Not indicated
Gower WA et al. 2013 (abstract). USA	Children	9	Age of asthma onset, asthma duration, gender, race, BMI, lung function- FEV1, FEV1/FVC, presence of allergy and Hospitalization.
Hilvering et al. 2015 (abstract). Netherland	Adults	Not indicated	Not indicated
Hsiao HP et al.2018. Taiwan	Adults	29	Age, Age of asthma onset, Early-onset asthma, Asthma duration, BMI, Obesity, Body fat, Waistline, Waist circumference/hip circumference, Smoking history, FEV1, % predicted, FVC, % predicted, FEV1/FVC ratio, Total IgE, Allergic rhinitis, Atopic dermatitis, WBC count, Neutrophil %, Eosinophil %, Monocyte %, Lymphocyte %, Basophil %, Breathlessness (wk), Breathlessness, during sleeping, Sneezing and nasal congestion, inhaled corticosteroid (ICS), oral corticosteroid (OCS), Number of controllers, Antihistamine.
Ilmarinen P et al.2017. Finland	Adults	Not indicated	Duration of respiratory symptoms before diagnosis; Feno at follow-up; Rhinitis at follow-up (yes/no); "Did you have respiratory symptoms suggesting asthma during childhood (<16 y) even though asthma had not been diagnosed?" (yes/no); Blood neutrophils at follow-up; "Does your asthma aggravate during working day?" at follow-up (yes/no); Maximal response in FEV1 (from diagnosis to Max0-2.5); BMI at follow-up; FEV1 reversibility at diagnosis; FEV1 reversibility at follow-up; Number of other than asthma/allergy-related drugs in

			use at follow-up; ACT score at follow-up; Smoking status at diagnosis (current/ex/never); Post-BD FEV1/FVC at diagnosis; Pre-BD FVC at follow-up.
Jeong A et al.2016. Switzerland	Adults	7	Asthma attack in the last 12 months, current asthma medication, number of asthma symptoms in the last 12 months, five typical respiratory symptoms were considered: breathless while wheezing, chest tightness, shortness of breath at rest, shortness of breath after exercise, and woken by shortness of breath at night, number of asthma symptoms repeatedly reported from baseline to the second follow-up, atopy defined by positive skin prick test at baseline for common allergens (cat fur, dog epithelia, house dust mite, dermatophagoides, pteronyssinus , timothy grass pollen, birch pollen, Parietaria pollen, and the moulds, Alternaria and Cladosporium), nasal allergy as hay fever, age of asthma onset.
Just J et al. 2012. France	Children	10	Age; Body mass index (BMI); Asthma duration (,<5 or ≥5 years); Maternal or paternal asthma; Allergic sensitization to aeroallergens and trophoallergens (positive skin-prick test and confirmed positive specific immunoglobulin E); Total IgE level; Inflammatory markers (lymphocyte, neutrophil, eosinophil, basophil, monocyte counts, serum IgG, IgA, IgM levels); Lung function (FEV1 expressed as a percentage of predicted value); Severity (mild persistent, moderate persistent, severe persistent); Control (controlled, partially controlled, uncontrolled) according to Global Initiative for Asthma (GINA) guidelines.
Just J et al. 2014. France	Children	18	Sex, age, eczema, food allergy, asthma duration, mild or moderate to severe asthma, and hospitalization for exacerbation. Four classes were defined for asthma control as previously defined: controlled or uncontrolled (or partially controlled), with or without high doses of ICS. paraclinical parameters (total IgE level, type and number of allergens to which the child was sensitized, FeNO, FVC, FEV1, and FEF25-75) and environmental parameters (tobacco smoke, mold, cockroaches, and furred pet exposure).
Kaneko Y et al.2013. Japan	Adults	8	Age, sex, smoking status (current, past or never), age at onset, total IgE levels, atopic status (specific IgE responsiveness to common inhaled allergens), baseline FEV1, and FEV1/FVC.

Kim HJ. et al.2018. Korea	Adults	9	Age at enrollment, sex, body mass index (BMI), age at asthma onset, and asthma duration, was collected at the baseline visit; all patients underwent standardized assessments, which encompassed peripheral blood total and differential cell counts, serum total immunoglobulin E (IgE), chest radiography, and skin prick tests with 24 common inhalant allergens
Kim MA. et al.2017. Korea	Adults	12	Age at enrollment, sex, BMI, age at asthma onset, asthma duration, smoking amount, atopy, peripheral blood eosinophils (%), total IgE level, FEV1, ΔFEV1, and PC20.
Kim TB. et al. 2013. Korea	Adults	10	Demographic parameters; Body mass index (BMI); Complete blood cell count; Total serum IgE levels; Pulmonary function tests (FEV1, FVC, FEV1/FVC ratio); Airway hyperresponsiveness (AHR); Methacholine bronchial provocation test; Skin prick tests with 12 common allergens; Smoking status; History of hospital visits due to asthma exacerbation in the previous year.
Koike et al. 2018. Not indicated	Adult	2	Fraction exhaled Nitric oxide and forced oscillation technique (FOT).
Kwon et al. 2012 (abstract).	Children	Not indicated	Blood tests for total IgE, eosinophil fraction, skin prick test, spirometry, and methacholine bronchial challenge test.
Lee E et al. 2017. South Korea	Children	14	Sex; body mass index (BMI); monthly house hold income; maternal education level; environmental tobacco smoke (ETS), parental history of allergic diseases (AD, AR, and/or asthma); onset age of the first wheezing, number of absence days from school and nocturnal awakening days owing to asthma attacks, number of asthma attacks in the previous 12 months, number of days with treatment for asthma in the previous 12 months at the time of enrollment, and SPT, blood tests, and lung function tests at the time of enrollment.

Liang et al. 2016. China	Adults	39	The levels of EGF; GM-CSF; IFN-g; IL-1b; IL-4; IL-5; IL-6; IL-8; IL-9; IL-10; IL-13; IL-17; IL-23; IL-33; Leptin; sRAGE; TGF-b1; TNF-a; TSLP; VEGF; (Age; Sex; BMI; Atopy; Smoking Duration; Serum total IgE; Blood WBC; Blood eosinophil; Blood neutrophil; Blood basophil; Sputum eosinophil; Sputum neutrophil; Sputum macrophage; Sputum lymphocyte; Baseline FVC; Baseline FVC% pred; Baseline FEV; Baseline FEV1% pred; Daytime symptom score; Night time symptom score; ACQ-5)
Loureiro et al. 2015. Portugal	Adults	22	Gender ; age ; BMI ; duration of illness ; age at onset of diseases ; IgE sensitization ; exposure to any smoke ; Pneumonia history ; Rhinitis ; Anxiety ; depression ; No of comorbidities.; Disease control ; Asthma control test (ACT) ; CARAT (Rhinitis) ; CARAT (Asthma) ; Asthma life quality (ALQ) ; Severe asthma score (SOA) ;Therapeutics (GINA as Until step3) ; High-dose ICS ; frequency of using OCS (last year) ; Severe exacerbation (previous year) ; FEV1 (%) besline; FEV1/CVF (%) at baseline ; Fixed obstruction ; WHO severity status ; FeNO (ppb) ; Serum IgE (log (mmol/L)) ; Blood eosinophils (%):; Sputum eosinophils (%) ; Sputum neutrophils (%) .
Loza et al. 2016. Multicountry	Adults	10	FEV1; FVC expressed as % predicted; FEV1/FVC ratio; bronchodilator reversibility (BDR); airway hyper-responsiveness (AHR), i.e., log-transformed provocative concentration of methacholine resulting in a 20% decline in FEV1 from baseline (PC20); Asthma Control Questionnaire (ACQ-7); Asthma Quality of Life Questionnaire (AQLQ); log-transformed FENO; and blood eosinophil counts expressed as absolute counts per μL.

Mahut et al. 2011. France	Children	28	Sex; Age; Height; Weight; BMI; Atopic status (negative, 1 positive, > 1 positive); Tobacco exposure (maternal, paternal, paternal and maternal); Clinical events within past 3 months (uncontrolled, partially or uncontrolled, number of days with symptoms, severe exacerbation, number of days with systemic steroid); Treatment (beta-agonist on demand, low ICS dose with mean ± SD dose, medium ICS dose with mean ± SD dose, high ICS dose with mean ± SD dose, LABA); Pulmonary function tests (sRaw % predicted, FEV1 % predicted, FEV1/FVC %, FVC % predicted, FEF75-25% predicted, FEF50% predicted, TLC % predicted, FRC % predicted, RV % predicted, RV/TLC, FEF50 %/TLC, FENO 0.05 ppb median [IQ])
Mäkikyrö et al. 2017. Finland	Adults	7	Use of controller asthma medication, bronchodilators, oral corticosteroids, and/or antibiotics during asthma exacerbations, and use of various healthcare services. All of these were inquired for the past 12 months. In addition, St. George's Respiratory Questionnaire score in the past 4 weeks was applied.
Mason et al. 2018. Italy	Adults	13	Age; sex; BMI; smoking status; Atopy; FEVQ/FVC % predicted at baseline; FEV1 predicted baseline; pre=SIC pd20 methacholine test; response to agent (TDI/HDI/MSI); eosinophils in peripheral blood %; neutrophiles in peripheral blood %; type of reaction (early/late/dual); age at onset.
Mastalerz et al. 2015. Poland	Adults	19	Sex, age at asthma onset, BMI, asthma control and severity, use of inhaled and/or oral steroids, presence of chronic rhinosinusitis, history of sinus surgery, FEV1, blood eosinophil count, total serum IgE, skin prick tests, induced sputum inflammatory cells and selected eicosanoids – PGD2, PGA2, PGE2, LTE4, 5-HETE; was applied on collected data set.
Nadif et al, 2018. France	Adults	25	Not inclusive (respiratory symptoms, asthma exacerbation and treatment, lung function, allergic sensitization, and added blood eosinophil and neutrophil counts, and fluorescent oxidation products (FlOPs) level.)
Nagasaki et al. 2014. Japan	Adults	7	Asthma Control Test (ACT) scores; serum IgE levels; blood eosinophil counts; prebronchodilator percent predicted FEV1; age; gender; BMI.

Nasreen et al. 2019. Canada	Children	5	Time since asthma diagnosis; sex of the child, child having health professional diagnosed allergy as a proxy for allergen sensitization or atopy of the child, at least 1 biological parent with a history of health professional diagnosed asthma or allergy, smoking habit of parent or spouse as a proxy for exposure to environmental tobacco smoke at home, number of siblings at home, and age at asthma diagnosis.
Qui et al. 2018. Korea	Adults	13	Gender, BMI, smoking, age at onset, age at enrollment, FEV1 predicted and FEV1/FVC (%), fef 25,74, blood eosinophilia and blood neutrophile, serum IgA, PACO2, PAO2. sputum neutrophil and sputum eosinophiles.
Sakagami et al. 2011. (abstract) Japan	Adults	7	Gender, body Mass Index, asthma Control Test Score, Patient Health Questionnaire -9 (PHQ-9) Score, disease duration, age of onset and type of asthma.
Schatz et al, 2013. USA	Adults and children	8	Sex, race, atopy, age of asthma onset, smoking (adolescents and adults), passive smoke exposure (children), obesity, and aspirin sensitivity
Schimdlin et al. 2015. (Abstract). USA	Children	8	Sex, race, peak flow, FEV1%, FVC%, FEF25-75%, early aeroallergen and food sensitization, and wheezing at ages 1, 2, 3, 4, and 7.
Sendín-Hernández et al. 2018. Spain	Adults	20	Age; sex; body mass index (BMI); family history of atopy (defined as a first-degree antecedent with atopic dermatitis, asthma, or allergic rhinitis); family history of asthma; family history of rhino conjunctivitis; age of onset of asthma; age of onset of rhinitis; atopic dermatitis; allergic rhino conjunctivitis; smoking status; lung function; fractional exhaled nitric oxide (FENO); eosinophil peripheral blood count; serum total IgE levels; severity of asthma according to GEMA 2009,15; use of inhaled steroids; use of long-acting bronchodilators; use of oral corticosteroids
Seino et al. 2018. Japan	Adults	10	BMI, smoking status, OCS burst episode, frequency of asthma attacks in the previous year, comorbidities (heart diseases), the drug used [long-acting muscarinic antagonists (LAMA), oral sustained-released theophylline (OSRT), oral corticosteroids (OCS)], and ACT-J and ASK-12 scores.

Seys et al. 2017. Belgium	Adults	12	Sputum mRNA expression level of the following cytokines: IFN-γ, IL-4, IL-5, IL-13, IL-10, IL-17A, IL-17F, IL-22, IL-25, TNF, IL-6 and IL-1β.
Siroux V et al. 2011 Multicountry	Adults	14	Age; sex; asthma symptoms over the past 12 months; asthma attacks over the past 12 months; age at asthma onset; asthma exacerbation; asthma treatment; chronic cough or phlegm; allergic rhinitis; atopic dermatitis; specific sensitization; total IgE; FEV1 % predicted.
Tay et al. 2019. Singapore	Adults	9	Age of asthma onset, sex, ethnicity, smoking status, body mass index, lung function, blood eosinophil count, asthma control test score, and exacerbation frequency.
Tuskiko et al. 2017. Japan	Adults	6	Age at onset, eosinophilic inflammation, airflow variability, and IgE levels, sex and age.
Wang LL et al. 2017. China	Adults	18	Sex; age; body mass index (BMI); atopy; history of smoking; socioeconomic status (SES; including educational level, health insurance, monthly income status, and employment status); asthma duration; age at asthma onset; asthma severity; FEV1 (liters); FEV1/FVC ratio; FEV1 percentage predicted (FEV1%predicted); ACT score; AQLQ score; and psychological symptoms.
Watanabe et al. 2016 (abstract). Japan	Adults	Not indicated	Not indicated
Wisnivesky et al;2019 USA	Adults	20	Sociodemographic, asthma history, severity, medication, and control; resource utilization as emergency and drug utilization and comorbidities as mental health GERD, serum IgE to sensitization to indoor allergen and lung function spirometry measures.

Wu et al.2017. China	Adults	26	Age, atopy, FEV1, FVC, and LM score, sex, age at first symptoms of rhinitis, age at first symptoms of asthma, disease duration, age at asthma diagnosis, adult-onset, smoker, prior sinus surgery, history of NSAID sensitivity, history of family asthma, Dmin value, total IgE, FeNO, blood eosinophil percentage, tissue eosinophil, blood neutrophil percentage, eosinophilia phenotypes of upper airway, eosinophilia phenotypes of lower airway severe asthma, FEV1/FVC ratio, and PEF.
Yoon et al. 2019. Korea	Children	12	Sex, age, current diagnosis of allergic rhinitis, current diagnosis of atopic dermatitis, history of acute bronchiolitis, puberty stage, age at asthma onset, PC20 from the methacholine challenge test, atopy defined as a positive response to at least 1 allergen on skin prick tests, baseline predicted FEV, and frequency of asthma symptoms.
Zaihra T et al.2016. Canada	Adults	13	FEV1 < 80 %, Beclomethasone or equivalent dose, AOO, year ACQ, Female, Severe asthmatic, Atopic, Non-Smoker, Age, BMI, Sputum Neutrophils, Sputum Eosinophils, FENO ppb.
Zhang X et al. 2019 (abstract). China	Adults	11	Inspiratory: mean of LA/BSA, mean of T/BSA, Mean of WA/BSA, mean of TA/BSA mean of WA, mean of Pi10WA, expiratory: MLD air trapping indexes: expiratory VI−856, MLD E/I, VI−856/−950 E-I, VI−856 E-I.
Zoratti E et al. 2018; USA	Adults	8	Sex, race, markers of atopy, including the number of self-reported allergic triggers (pollen, animals, mold, dust), history of allergic rhinitis or atopic dermatitis, and serum total IgE level; age at asthma onset; smoking status (adults) or passive smoke exposure (children); body mass index (BMI); self-reported history of aspirin sensitivity; and (8) ratio of FEV1 to forced vital capacity (FVC) (FEV1:FVC) from spirometry performed according to American Thoracic Society guidelines.

9.3.1.4 TABLE S 9. NUMBER AND LIST OF VARIABLES USED FOR COPD AND ACO PHENOTYPING.

Reference ID and country	Number of variables	Variables considered for analysis
Augustin et al; 2018. Netherland	27	FEV1; % predicted; FEV1/FVC; %; FVC; % predicted; PEF; % predicted; ITGV; % predicted; RV; % predicted; TLC; % predicted; TLCO; % predicted; KCO; % predicted; MIP; % predicted; MEP; % predicted; arterial blood gases (PaO2; PaCO2); SaO2; %; FEV1; FVC; PEF; ITGV; RV; TLC; TLCO; KCO; TLCH; VIN; TA; MIP; MEP.
Bafadhel et al.; 2011. UK	Not indicated	Not indicated
Bertini et al.; 2013. Italy	35	Not indicated
Burgel et al.; 2017 France	7	Age; body mass index (BMI); FEV1 (% predicted); modified Medical Research Council (mMRC) dyspnea scale; number of exacerbations in the previous 12 months; and presence/absence of cardiovascular comorbidities (hypertension; coronary artery disease and/or left heart failure) and/or diabetes.
Burgel et al.; 2010 France	8	Age, tobacco smoking, severity of airflow obstruction (assessed by FEV1 % pred); exacerbations (number per patient per year); nutritional status (assessed by BMI kg*m-2); dyspnea (assessed by the MMRC scale); HRQoL (assessed by the SGRQ total score); and anxiety and depression (assessed by the HAD total score).
Burgel et al.; 2012 Belgium	7	Age, body mass index (BMI), FEV1 (% predicted), mMRC scale, CCQ total score, thoracic gas volume (TGV, % predicted) and DLCO (% predicted).
Chen et al; 2014	7	Age; postbronchodilator FEV1 (% predicted); BMI; the number of severe exacerbations in the preceding 3 years; mMRC dyspnea scale; SpO2 %; and Charlson index.

Taiwan and China		
Chubachi et al (abstract); 2016 Japan	19	Anemia; underweight; obesity; chronic renal failure; chronic heart failure; osteoporosis; anxiety; depression; lung cancer; other cancer; diabetes mellitus; dyslipidemia; hypertension; hyperuricemia; cerebrovascular disease; ischemic heart disease; gastric ulcer; gastro-esophageal reflex disease (GERD) and benign prostatic hypertrophy.
De Torres et al; 2017 Spain	8	Age; pack-year history; BMI; FEV1%; FVC%; 6MWD; MMRC and CAT.
Divo et al. abstract; 2016 Not indicated	8	Not indicated
Fens et al.; 2013. Netherland	12	Postbronchodilator FEV1 %pred; CT emphysema score (perc15); Chronic bronchitis; Use of long-acting bronchodilators; Dyspnea at rest; cardiovascular disease; Body Mass Index; Diabetes; Reversibility in FEV1; eNose PC1; eNose PC4; Pack-years
Guillamet RV et al.; 2018. USA	14	Demographics, comorbidities included in Charlson's comorbidity index, presence of atopy, obesity, number of admissions, prescriptions for inhalers grouped as short acting beta-agonist, long-acting beta-agonist, anticholinergics, steroids and their combinations, prescriptions for oral steroids, beta-blockers and statins.
Harrison SL et al.; 2014 UK	Not indicated	Not indicated
Haghighi et al, 2019. USA	75	The segmental variables included bifurcation angle (θ), air- way circularity (Cr), wall thickness (WT) and hydraulic diameter (Dh), fractional air volume change (ΔVairF), the determinant of Jacobian (Jacobian), and aniso- tropic deformation index (ADI), functional small airway disease percentage (fSAD%) and emphysema percentage (Emph%), TLC and RV as (βtissueTLC and βtissueRV), ratio of apical-basal distance.

		over ventral-dorsal distance at TLC (lung shape). Overall, there were 32 local/segmental structural variables, 35 lobar structural variables and 8 global structural variables.
Kim W.J et al; 2017 Asia	8	Age, body mass index (BMI), cigarette smoking, mMRC dyspnoea scale, the St George's Respiratory Questionnaire (SGRQ) score, Charlson comorbidity index (CCI),15 FEV1, FVC and FEV1/ FVC.
Kim S et al; 2017. Asia	15	Body mass index (BMI), cigarette smoking amount, mMRC score, CAT score, white blood cells (WBCs) with polymorphonuclear neutrophils (PMNs), eosinophils, interleukin (IL)-6, C-reactive protein (CRP), uric acid, EI, FEV1, FVC, and FEV1/FVC.
Kukol et al. 2019;(abstract). Russia	Not indicated	Not indicated
Lee et al, 2019, Korea	7	Age, BMI, post BD FEV1%, CAT score, history of asthma, DLCO % and smoking status.
Li et al. abstract; 2016	13	Post-bronchodilator FEV1/FVC, % predicted FEV1, % predicted FVC, FEV1 reversibility, pack-years cigarette smoking, TLC -950 HU, RV -856 HU, Apex/Base ratio -950 HU, Apex/Base slope -950 HU, FACIT physical wellbeing score, pulse following six-minute walk, gender, and BMI.
Liang et al, 2019 Korea	14	Smith antigen (Sm), Ribosomal phosphoprotein P0 (P0), Ro/Sjögren syndrome type A antigen (SS-A), La/Sjögren syndrome type B antigen (SS-B), DNA topoisomerase I (Scl70), Histidyl-tRNA synthetase (Jo1), U1 small nuclear ribonucleoprotein (U1-SnRNP), Thyroid peroxidase (TPO), Proteinase-3 (PR3), Myeloperoxidase (MPO), demographics, lung function, CAT score, dyspnea score.
Lopes et al; 2019 Brazil	13	Age, gender, education level, marital status, smoking status and lung function, dyspnea, illness perception, social support, health efficacy, clinical control, health related quality of life and physical activity.
Ning et al; 2016, (abstract).	Not indicated	Not indicated

China		
Peters et al; 2016. Netherlands	11	FEV1% of predicted, body composition (BMI kg m−2), exercise capacity (VO2 max% of predicted), subjective symptoms, dyspnea emotions, fatigue, subjective impairment, behavioral impairment, general quality of life, health-related quality of life and satisfaction with relationships.
Pikoula et al; 2019. UK	15	Body mass index (BMI), smoking status (current or ex), atopy, airflow obstruction as defined by Global Initiative for Chronic Obstructive Lung Disease (GOLD) stage [21]: 1 (FEV1% predicted > = 80%), 2 (50% < = FEV1% predicted < 80%), 3 (30% < = FEV1% predicted <50%) and 4 (<= FEV1% predicted < 30%)– and eosinophil % of white blood cell counts, gastro-esophageal reflux disease (GERD), chronic rhinosinusitis (CRS), diabetes, anxiety, depression, ischemic heart disease (IHD), hypertension, heart failure, sex, height. treatment as COPD therapy type with regards to different combinations of inhaled corticosteroids (ICS), Long Acting Muscarinic Antagonists (LAMA) and/or Long Acting Beta-2 Antagonists (LABA) as: a) no therapy (none of LAMA, LABA prescribed), b) mono-therapy (prescription of LABA or LAMA only), c) dual therapy (prescription of either LABA&LAMA or LABA&ICS or LAMA&ICS), and c) triple therapy: prescription of all LABA, LAMA and ICS.
Rodrigues et al.; 2018 Brazil	19	Age; Gender; BMI; FVC (l); FVC (%pred); FEV1 (l); FEV1 (%pred); FEV1/FVC; GOLD, I/II/III/IV; FFMI; 6MWT (m); 6MWT (%pred); IRMQF (kg); m MRC (0–4 pts); MIP (cmH2O); MIP (%pred); BODE index, I/II/III/IV; Comorbidities (n); Smoking history (pack-years); Follow-up (months).
Scarlata et al; 2018. Italy	Not indicated	Not indicated
Xavier F et al; 2019. Brazil	Not indicated	Not indicated

Yoon et al.; 2019 Korea	7	Age, BMI, smoking status, history of asthma, COPD assessment test (CAT) score, FEV1, DLCO.

9.3.1.5 TABLE S 10. NUMBER AND LIST OF VARIABLES USED FOR SEVERE ASTHMA PHENOTYPING.

Reference ID and country	Population age	Number of variables	Variables considered for analysis
Brinkman et al, 2011 (abstract). Multicountries	Adult	Not indicated	Not indicated
Desai et al. 2011. UK	Adults	Not indicated	Not indicated
Diver et al. 2018. UK	Adults	1	Microbial profiles were obtained through 16S rRNA gene sequencing.
Fitzpatrick et al, 2018. USA	Children	12	Duration of asthma in months, baseline FEV1 percent predicted, and the best postbronchodilator FEV1 percent predicted, sex, race (white, black, or other) and ICS group (none, low-dose, or high-dose), Beta agonist use over the previous 3 months, the frequency of symptoms, the magnitude of atopic sensitization, and exhaled nitric oxide quartile.
Freitas PD et al. 2018 (abstract); Brazil	Adults	Not indicated	Not indicated
Gomez et al, 2017. USA	Adults	12	Baseline FEV1% predicted, baseline FVC% predicted, baseline FEV1/FVC ratio, FEV1% predicted and FVC% predicted after maximal bronchodilation with albuterol, % of FEV1 bronchodilator response, age of asthma onset, asthma duration, gender, Treatment, and serum YKL-40 (ng/mL).
Jang et al, 2017; Korea	Adults	5	Age at the onset of asthma, BMI, FEV_1/FVC ratio, amount of cigarette smoking, and PC_{20}.

Konstantellou et al. 2015; Greece	Adults	5	Patients' demographics, pulmonary function tests, inflammatory cells in induced sputum, bronchial hyper- responsiveness (BHR, PD15 to methacholine) and treatment regimens were recorded.
Lau et al, 2017: Singapore	Adults	Not indicated	Not indicated
Moore et al 2010; USA	Adults	35	Gender; race; BMI; age at enrollment; age of asthma onset; asthma duration ; pre-bronchodilator >6 hours withholding of bronchodilator: fev1 % predicted; FFVC % predicted; FEV1/FVC; post-bronchodilator as best values after 6-8 puffs of albuterol: fev1 % predicted; FFVC % predicted; maximal% change in FEV1; number of positive SPT; composite medications used as: corticosteroids used; total number of controllers ;type of controllers; beta-agonist frequency score; composite health care utilization as : frequency/severity past year; intensity/ICU ever in lifetime; composite asthma symptoms as : general symptoms score; symptoms with activities ; composites pattern of skin prick test as :cats/dogs; dust mites/cockroach; molds and pollens; composites triggers as : severity of allergy symptoms ; aspirin sensitivity/nasal polyps; sinusitis and sinus surgery; bronchitis/ pneumonia; composite co-morbidities as : GERD and HTN; composite family history as : parental asthma ;siblings with asthma; composite tobacco exposure: passive and remote ; composite of women hormone exposure.
Newby et al, 2018; UK	Adults	23	Gender, body mass index (BMI), smoking status, atopy, unscheduled visits to GP in last 12 months, intensive care unit (ITU) visits in last 12 months, hospital admissions in last 12 months, perennial rhinitis, seasonal rhinitis , eczema, nasal Polyps, reflux history, pre-bronchodilatorFEV1% predicted, pre-bronchodilator FVC% predicted, FEV1/FVC ratio pre-bronchodilator, blood eosinophils, total blood IgE, rescue steroid courses taken in last 12 months, beclomethasone dipropionate (BDP) equivalent inhaled corticosteroid dose, age at first assessment, age at onset of symptoms, oral steroid dose, and center.
Raherson et al, 2018	Adults	Not indicated	Not indicated

(abstract); France			
Sekiya et al. 2015; Japan	Adults	24	Gender, n (%), Age, BMI, smoking status, pack years, Pack year, Hypersensitivity symptoms after exposure to furred pets, Allergic rhinitis, Chronic hyperplastic rhinosinusitis/nasal polyposis, Diabetes, Chronic obstructive pulmonary disease, Inactive pulmonary tuberculosis, Any physiological diseases, Asthma duration, History of hospitalization for asthma, Hospitalizations for asthma in the past year, History of NFA requiring mechanical ventilation, Daytime symptoms, Limitation of activities, Nocturnal symptoms/awakening, ICS, SABA, OCS, Viral infection/common cold, Bacterial infection.
Serrano Pariente et al. 2015; Spain	Adults	44 (reduced form 77)	Sex; Age; Level of education; Smoking; GHQ score; Psychiatric comorbidity; TAS score; Presence of alexithymia; Asthma severity (GINA); FEV1 (post NFA episode); Age at onset of asthma; Presumptive NSAID intolerance; History of atopy; History of rhinosinusitis; Medical care for asthma; Number of prior hospitalizations for asthma; Past mechanical ventilation for asthma; Hospitalization for asthma in the past 12 months; Past ED visits for asthma in the past 12 months; Past NFA attacks; Past hospitalization for asthma; Treatment for asthma; Inhaled corticosteroids dose (mcg/day of budesonide) Long-acting beta agonist (formoterol; salmeterol); Oral corticosteroids; Theophylline; Ipratropium bromide; Trigger of the NFA episode; Sudden onset of NFA episode (<2 h); Impaired consciousness level on hospital admission; Tracheal intubation required; Respiratory arrest; Hours under mechanical ventilation; pH on arterial blood gas analysis on hospital admission PaCO2 on arterial blood gas analysis; Written asthma action plan followed; Season of the NFA episode; Days in hospital; Days in ICU; Total dose of systemic steroids in first 5 days in hospital (mg prednisone); Medical complications during hospitalization; Eosinophil count cells on hospital admission; Number of positive skin prick-tests per patient; Sensitization to Alternaria alternata; Sensitization to fungi.

Simpson et al (abstract); 2017. Multi-country	Adults	Not indicated	Not indicated
Taniguchi et al; 2014;(abstract) Japan	Adults	13	Not indicated
Wu et al; 2014. USA	Adults	45	16 symptom-related, 10 medication-related, 10 asthma history/comorbidities–related, 7 allergy-related, 5 lung function–related, and 3 health care use (HCU)–related variables.
Weng-Jing Ye et al. 2017; China	Adults	22	Sex, age, age at onset, asthma duration, BMI, smoking index, family history, Disease severity and risk Respiratory failure, COPD Lab tests IgE, EOS, allergen detection Lung function, FVC% pred, FEV1% pred, FEV1/FVC, post FVC% pred, post FEV1% pred, post FEV1/FVC, change in FEV1, Medicine use ICS, Health care utilization in the past year, Health care utilization in the past.
Youroukova et al. 2017; Bulgaria	Adults	8	Age of asthma onset, duration of diseases, atopy, smoking, blood eosinophils, nonsteroidal anti-inflammatory drugs hyper- sensitivity, baseline FEV1/FVC and symptoms severity.

9.3.1.6 TABLE S 11.NUMBER AND LIST OF VARIABLES USED FOR RHINITIS PHENOTYPING.

Reference ID and country	Number of variables	Variables considered for analysis
Adnane et al.2017. Morocco	9	Not indicated
Agache et al. 2010; France	11	Male sex; family history of asthma; breastfeeding < 2 months; passive/active smoking; obesity; pets/molds exposure; high total serum immunoglobulin E (IgE); polysensitization (sensitized to 3 seasonal pollens with different structure); mixed sensitization (seasonal and perennial allergens); severe rhinitis (according to the Allergic Rhinitis and its Impact on Asthma guidelines); and lack of allergen-specific immunotherapy (SIT) for rhinitis preceding asthma diagnosis.
Bousquet et al. 2015; France	18	Not indicated
Burte et al. 2015; France	21	Report of nasal symptoms; current/ever symptoms; persistence and disturbance of these symptoms; seasonal pattern; sensitivity to seven triggers; report of allergic rhinitis; hay fever; conjunctivitis; sinusitis and eczema; report of diagnostic of allergy by a physician; SPT; report of spray; report of drug except spray; and allergic immunotherapy since the last survey.
Herr M et al.2012. France	27	Wheeze occasional, wheeze recurrent, limitations of daily activities, sleep disorders, Nocturnal dry cough apart from a cold, Allergic rhinitis–like symptoms, Atopic dermatitis, 1-3 Upper respiratory tract infections (apart from colds), 4 or more Upper respiratory tract infections (apart from colds), 1-3 Lower respiratory tract infections, 4 or

		more lower respiratory tract infections, Inhaled bronchodilators, Inhaled corticosteroids, Hospital-based care because of respiratory disorders, Emergency department visit, Admission to hospital, Total IgE (U/mL), Total IgE >_45 U/mL, Blood eosinophilia (eosinophils/mm3, Eosinophilia >_470 eosinophils/mm3, Allergenic sensitization, Food .allergens, Inhalant allergens, Allergenic sensitization >_0.70 U/mL, Sensitization to >_2 allergens.
Kurukulaaratchy et al. 2015. UK	13	Atopic status at age 18 years, asthma at age 18 years, eczema at age 18 years, age at which rhinitis appeared, seasonality of rhinitis symptoms, total IgE level (log10), BDR, BHR DRS, mean Feno value (log10), FEV1, forced vital capacity (FVC), FEV1/FVC ratio, and forced expiratory flow at 25% to 75% (FEF25-75).
Lee E.L. et al 2016. Korea	15	Sex; body mass index; socioeconomic status, including income and maternal educational level; environmental factors, including prior exposure to environmental tobacco smoke (ETS); parental history of allergic diseases (AD, AR, and/or asthma); presence of allergic disease; comorbidities at the time of enrollment; treatment of rhinitis within the previous 12 months; SPT, total serum IgE, blood eosinophil percentages, and pulmonary function test results at the time of enrollment; and the presence of rhinitis at the time of enrollment.
T.A Nakayama et al. 2012. Japan	16 reduced to 5	Anterior nasal discharge, posterior nasal discharge, facial pain, nasal obstruction, specific IgE to house-dust mites, specific Ige to dermatophagoides pteronyssinus, total IgE (IU/ml), polyp score, CT score, staphylococcus aureus, decreased sense of smell ,ATA ,mucin ,Peripheral basophil count (/µl),mucosal eosinophil count (/HPF)peripheral eosinophil count (/µl). Eigenvalue

ZM Soler et al.2015. USA and Canada	32	This included 3 demographic variables, 14 comorbidity/exposure variables, 3 objective measures of CRS, and 83 patient-reported outcome measure (PROM) variables, including those from 5 validated questionnaires. Prior to performing a cluster analysis, PROM variables were reduced to meaningful factors that contained questions with a high degree of correlation. Corresponding composite scores were determined. For instance, the SNOT-22 questionnaire, which consisted of 22 individual questions related to Sino-nasal QOL, was reduced to 3 factors. Similarly, the RSDI was reduced to 4 factors and the SF-12 to 2 factors (Table 1). The PSQI and PHQ did not reduce to meaningful factors, thus total scores were used.

9.3.1.7 TABLE S 12. NUMBER AND LIST OF VARIABLES USED FOR ACO PHENOTYPING.

Reference ID and country	Number of variables	Variables considered for analysis
De Vries et al. 2018. Netherlands	Not indicated	Not indicated
Fingleton et al. 2017. New Zealand and China	12	(FEV1) (% predicted), FEV1/ (FVC) ratio (%), bronchodilator reversibility (% change in FEV1 from baseline after salbutamol administration), peak expiratory flow (PEF) variability (expressed as % of mean), KCO (% predicted), log FENO (ppb), log IgE (IU·L−1), log hsCRP (mg·L−1), age of onset of respiratory symptoms, body mass index, total SGRQ score and cigarette smoke exposure (pack-years). Functional residual capacity was used as an additional variable in the NZRHS.

Górska K t al. 2017. Poland	Not indicated	Not indicated
Rootmensen et al. 2016. Netherlands	Not indicated	Not indicated

10 APPENDIX 2: PAPER 4

10.1 ADDITIONAL SUPPLEMENTARY MATERIAL

10.1.1 SUPPLEMENTARY TABLES

10.1.1.1 TABLE S1. PREVALENCE OF FORMS OF RHINITIS BY EDUCATION AND OCCUPATION LEVELS WITHIN MALE AND FEMALES

	Male				Female			
	Primary	Secondary	Tertiary	P value	Primary	Secondary	Tertiary	P value
n	2880	6928	6457		2746	7281	9821	
Allergic rhinitis	522 (18.1)	1901 (27.4)	2070 (32.1)	<0.001	579 (21.1)	2081 (28.6)	3107 (31.6)	<0.001
Chronic rhinitis	543 (18.9)	1409 (20.3)	1203 (18.6)	0.032	486 (17.7)	1513 (20.8)	1819 (18.5)	<0.001
Chronic rhinosinusitis	83 (2.9)	178 (2.6)	145 (2.2)	0.167	88 (3.2)	235 (3.2)	252 (2.6)	0.023

	Male					Female				
	Lowest	Upper low	lower high	Highest	P value	Lowest	Upper low	lower high	Highest	P value
n	345	6576	2921	3308		721	7629	2265	5323	
Allergic rhinitis	101 (29.3)	1782 (27.1)	852 (29.2)	1041 (31.5)	<0.001	214 (29.7)	2281 (29.9)	699 (30.9)	1630 (30.6)	0.728
Chronic rhinitis	94 (27.)	1348 (20.5)	566 (19.4)	570 (17.2)	<0.001	171 (23.7)	1470 (19.3)	437 (19.3)	940 (17.7)	0.001
Chronic rhinosinusitis	11 (3.2)	187 (2.8)	53 (1.8)	60 (1.8)	0.001	36 (5.0)	237 (3.1)	54 (2.4)	124 (2.3)	<0.001

10.1.1.2 TABLE S2. PREVALENCE OF FORMS OF RHINITIS BY EDUCATION AND OCCUPATION GROUPS WITHIN STRATA OF AGE GROUPS AND THEIR P VALUES

	Less than 30				30- 45			
	Primary	**Secondary**	**Tertiary**	**P value**	**Primary**	**Secondary**	**Tertiary**	**P value**
n	434	2571	2810		468	3265	5358	
Allergic rhinitis	108 (24.9)	867 (33.7)	979 (34.8)	<0.001	128 (27.4)	1078 (33.0)	1896 (35.4)	0.001
Chronic rhinitis	105 (24.2)	689 (26.8)	639 (22.7)	0.003	94 (20.1)	698 (21.4)	948 (17.7)	<0.001
Chronic rhinosinusitis	12 (2.8)	86 (3.3)	82 (2.9)	0.611	27 (5.8)	132 (4.0)	140 (2.6)	<0.001
	46 - 60				**>=61**			
	Primary	**Secondary**	**Tertiary**	**P value**	**Primary**	**Secondary**	**Tertiary**	**P value**
n	1520	4536	4598		3204	3837	3512	
Allergic rhinitis	368 (24.2)	1266 (27.9)	1472 (32.0)	<0.001	497 (15.5)	771 (20.1)	830 (23.6)	<0.001
Chronic rhinitis	294 (19.3)	877 (19.3)	784 (17.1)	0.011	536 (16.7)	658 (17.1)	651 (18.5)	0.119
Chronic rhinosinusitis	67 (4.4)	136 (3.0)	106 (2.3)	<0.001	65 (2.0)	59 (1.5)	69 (2.0)	0.236

	Less than 30					30- 45				
	Lowest	**Upper low**	**Lower high**	**Highest**	**P value**	**Lowest**	**Upper low**	**Lower high**	**Highest**	**P value**
n	237	1830	624	1086		311	3613	1609	2842	
Allergic rhinitis	87 (36.7)	646 (35.3)	210 (33.7)	368 (33.9)	0.722	101 (32.5)	1257 (34.8)	569 (35.4)	981 (34.5)	0.791
Chronic rhinitis	72 (30.4)	471 (25.7)	145 (23.2)	233 (21.5)	0.008	72 (23.2)	756 (20.9)	281 (17.5)	480 (16.9)	<0.001
Chronic rhinosinusitis	13 (5.5)	67 (3.7)	18 (2.9)	36 (3.3)	0.303	15 (4.8)	137 (3.8)	36 (2.2)	69 (2.4)	0.001
	46 - 60					**>=61**				
	Lowest	**Upper low**	**Lower high**	**Highest**	**P value**	**Lowest**	**Upper lowe**	**Lower high**	**Highest**	**P value**
n	308	4992	1685	2593		210	3770	1268	2110	
Allergic rhinitis	89 (28.9)	1431 (28.7)	509 (30.2)	823 (31.7)	0.046	38 (18.1)	729 (19.3)	263 (20.7)	499 (23.6)	0.001
Chronic rhinitis	79 (25.6)	911 (18.2)	329 (19.5)	419 (16.2)	<0.001	42 (20.0)	680 (18.0)	248 (19.6)	378 (17.9)	0.546
Chronic rhinosinusitis	16 (5.2)	144 (2.9)	36 (2.1)	49 (1.9)	0.001	3 (1.4)	76 (2.0)	17 (1.3)	30 (1.4)	0.234

10.1.1.3 TABLE S3. PREVALENCE OF FORMS OF RHINITIS BY EDUCATION AND OCCUPATION LEVELS WITHIN STRATA OF THOSE WHO WERE AND WERE NOT RAISED ON A FARM DURING CHILDHOOD.

	Raised on a farm				**Not raised on a farm**			
	Primary	**Secondary**	**Tertiary**	**P value**	**Primary**	**Secondary**	**Tertiary**	**P value**
n	1358	1724	1424		4173	12278	14592	
Allergic rhinitis	220 (16.2)	360 (20.9)	343 (24.1)	<0.001	856 (20.5)	3567 (29.1)	4762 (32.6)	<0.001
Chronic rhinitis	222 (16.3)	342 (19.8)	252 (17.7)	0.039	788 (18.9)	2542 (20.7)	2738 (18.8)	<0.001
Chronic rhinosinusitis	31 (2.3)	44 (2.6)	30 (2.1)	0.705	135 (3.2)	362 (2.9)	363 (2.5)	0.01

	Raised on a farm					**Not Raised on a farm**				
	Lowest	**Upper Low**	**Lower high**	**Highest**	**P value**	**Lowest**	**Upper low**	**Lower high**	**Highest**	**P value**
n	169	2148	469	768		875	11903	4658	7748	
Allergic rhinitis	44 (26.0)	443 (20.6)	109 (23.2)	174 (22.7)	0.221	264 (30.2)	3578 (30.1)	1427 (30.6)	2462 (31.8)	0.085
Chronic rhinitis	43 (25.4)	386 (18.0)	107 (22.8)	126 (16.4)	0.003	219 (25.0)	2399 (20.2)	890 (19.1)	1368 (17.7)	<0.001
Chronic rhinosinusitis	9 (5.3)	48 (2.2)	13 (2.8)	9 (1.2)	0.007	381 (4.3)	368 (3.1)	93 (2.0)	175 (2.3)	<0.001

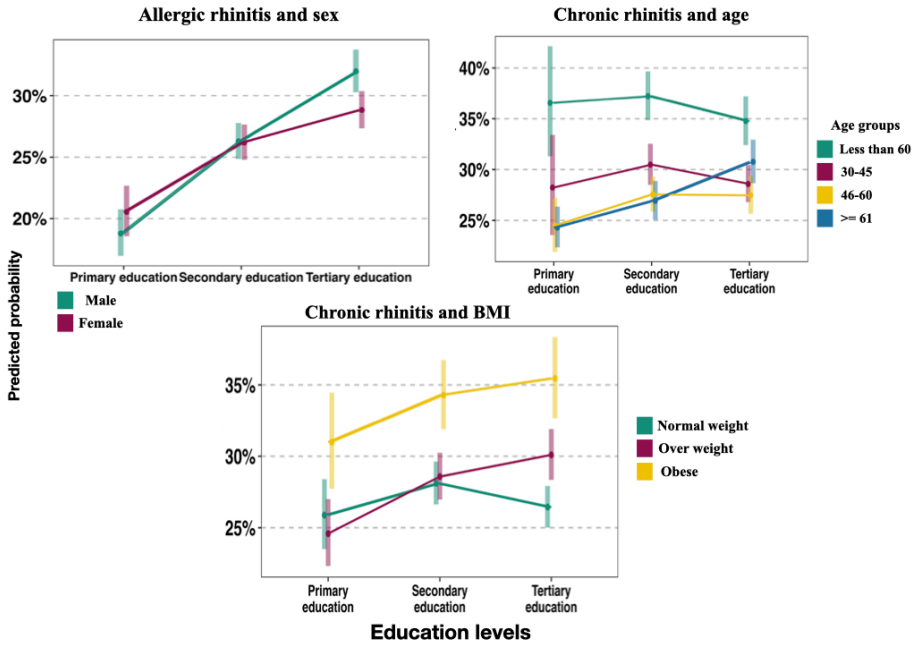
10.1.1.4 TABLE S4. RESULTS OF GENDER- AND AGE-MODIFIED ASSOCIATION BETWEEN EDUCATION AND ALLERGIC RHINITIS.

	Sex	Odds. ratio	Lower 95%	Upper 95%
Allergic rhinitis				
Secondary education / Primary education	Male	1.53	1.30	1.81
Tertiary education / Primary education	Male	2.02	1.71	2.38
Tertiary education / Secondary education	Male	1.32	1.18	1.47
Secondary education / Primary education	Female	1.36	1.16	1.61
Tertiary education / Primary education	Female	1.56	1.33	1.83
Tertiary education / Secondary education	Female	1.14	1.04	1.26
	Age groups	**Odds. ratio**	**Lower 95%**	**Upper 95%**
Chronic rhinitis				
Secondary education / Primary education	less than 30	1.02	0.75	1.39
Tertiary education / Primary education	less than 30	0.92	0.68	1.25
Tertiary education / Secondary education	less than 30	0.90	0.77	1.06
Secondary education / Primary education	30-45	1.12	0.82	1.54
Tertiary education / Primary education	30-45	1.03	0.75	1.40
Tertiary education / Secondary education	30-45	0.92	0.79	1.05
Secondary education / Primary education	46-60	1.18	0.97	1.44
Tertiary education / Primary education	46-60	1.18	0.96	1.44
Tertiary education / Secondary education	46-60	1.00	0.87	1.15
Secondary education / Primary education	>=61	1.14	0.96	1.35
Tertiary education / Primary education	>=61	1.39	1.18	1.65
Tertiary education / Secondary education	>=61	1.22	1.05	1.43

10.1.1.5 TABLE S5. RESULTS OF BODY MASS INDEX -MODIFIED ASSOCIATION BETWEEN EDUCATION AND CHRONIC RHINITIS.

	BMI	odds. Ratio	Lower 95%	Upper 95%
Secondary education / Primary education	Normal weight	1.11	0.93	1.31
Tertiary education / Primary education	Normal weight	1.03	0.87	1.22
Tertiary education / Secondary education	Normal weight	0.93	0.83	1.03
Secondary education / Primary education	Overweigh t	1.23	1.04	1.47
Tertiary education / Primary education	Overweigh t	1.33	1.11	1.58
Tertiary education / Secondary education	Overweigh t	1.08	0.95	1.22
Secondary education / Primary education	Obese	1.17	0.93	1.47
Tertiary education / Primary education	Obese	1.24	0.98	1.58
Tertiary education / Secondary education	Obese	1.06	0.88	1.28

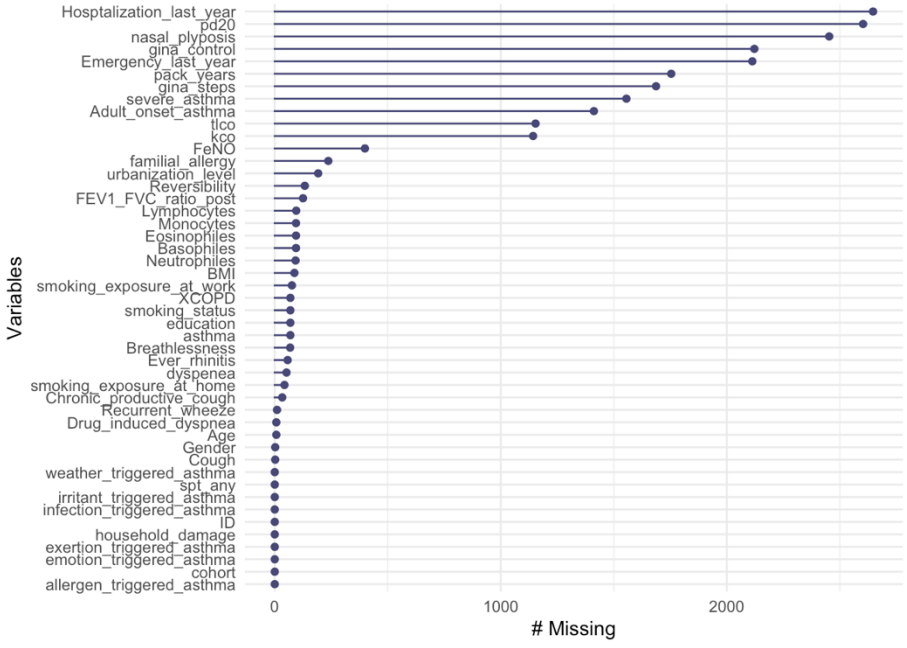
10.1.2 SUPPLEMENTARY FIGURES



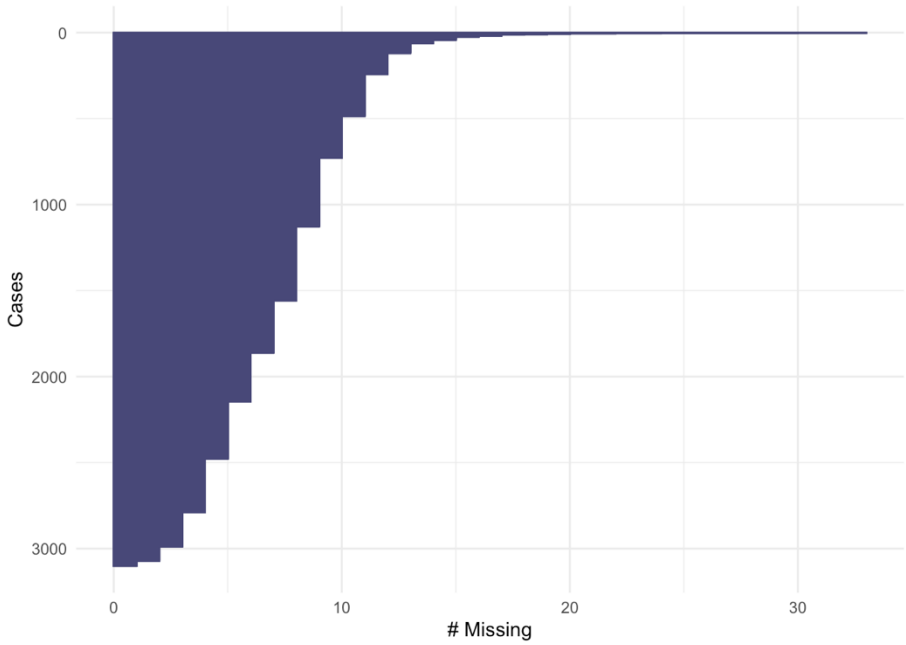
10.1.2.1 FIGURE 24. INTERACTION BETWEEN EDUCATION AND BASELINE CHARACTERISTICS TOWARDS RISK OF FORMS OF RHINITIS

11 APPENDIX 3: PAPER 5

11.1 SUPPLEMENTARY FIGURES



11.1.1.1 FIGURE 25. THE COUNT OF MISSINGNESS ACROSS STUDY VARIABLES.



11.1.1.2 FIGURE 26. THE COUNT OF MISSING DATA BY STUDY SUBJECTS

11.2 ADDITIONAL SUPPLEMENTARY DATA

11.2.1 METHODS

11.2.1.1 HANDLING OF MISSING DATA AND IMPUTATION

The dataset for this study on imputation is a mix of multivariate types, comprising 19 continuous and 57 categorical variables, amounting to 76 in total. We find that 42,209 out of 193,467 observations are incomplete, representing 21.8% of the dataset. The objective of this research is to identify clusters of asthma from a dataset of mixed categorical and continuous variables. Multiple imputation is a well-regarded technique for addressing missing data, provided the data adheres to certain assumptions. When the missing data is assumed to be Missing At Random (MAR) or Missing Completely At Random (MCAR), the unobserved data is believed to depend on the observed data within the dataset, or to be entirely random, respectively [456]. Imputation under these conditions is considered to reduce bias [457]. It is also

advised to consider and assess the nonignorability assumption, where the characteristics of the missing data differ from those of the complete dataset, during the imputation process. In epidemiological studies employing regression models, it has been established that the least bias in population parameter estimates occurs when the missing data is not simultaneously related to both the exposure and the outcome [458].

A suitable solution for this study, which focuses on asthma as the primary outcome for cluster identification, is to investigate whether the missing data correlates with the asthma status or with variables that are believed to significantly influence the formation of asthma clusters.

The imputation was performed using the random forest method for multiple imputation. This approach, using random forests for imputation, has shown computational advantages in terms of robustness, particularly with complex and multidimensional data [458].

Missing data framework included:

- Assessment of the missing data mechanism and assumption, plus confirmation of weather missingness in the data was associated with outcome and exposures at once. Determination of exposure variables was based on expert assessment of which variables are likely to contribute the construct of the derived cluster. Although some variables were missing not at random yet using MICE approach with auxiliary variables deemed helping in reducing any presumed bias.
- We impute the data using MICE approach, generating 100 imputation sets (in this case the prediction was based on random forest model).
- We validate the results visually and insure the convergence of the MICE model.
- A random set of the imputed was selected to run the cluster analysis on.

11.2.1.2 CLUSTERING PROCESS AND VALIDATIONS

This study employs a comprehensive computational approach for the clustering of asthma-related datasets through unsupervised deep-learning

methodologies. Initially, the dataset undergoes a preprocessing phase to ensure its readiness for subsequent analysis. The determination of the optimal number of clusters, a critical step in the clustering process, is achieved through a dual-method approach. Primarily, we utilize the R package NbClust, which leverages voting consensus methods to ascertain the most suitable number of clusters [459].

11.2.1.3 ASSESSMENT OF OPTIMAL CLUSTERING SOLUTION

The NbClust package in R is designed for determining the optimum number of clusters within a dataset. It incorporates over 30 internal validation metrics that consider intra-cluster compactness and inter-cluster separation, among other factors. These indices allow for a comprehensive evaluation of clustering results, providing insights into the most suitable number of clusters. Key methods include the Silhouette method, the Gap statistic, and the Dunn index, each offering a unique perspective on clustering validity. Euclidean and Manhattan distances are both used as the main distance metrics. For a detailed exploration of these methods, the original paper should be consulted [234].

The motivation for using Euclidean and Manhattan distances to determine the optimal number of clusters in the context of Deep Embedded Clustering (DEC) arises from DEC's utilization of the K-means algorithm for initial cluster formation. These distances help in grouping data points based on their spatial relationship, a key step in K-means, which DEC builds upon in the initial phase of training. This strategy aims to align the initialization phase of DEC clustering with the intrinsic data structure captured by Euclidean and Manhattan distances, enhancing the overall clustering outcome facilitated by DEC.

To complement and validate this approach, we further employ the M3C method—Monte Carlo reference-based consensus clustering. The M3C method employs a Monte Carlo reference-based consensus clustering technique to rigorously determine the optimal number of clusters within a dataset. It leverages statistical significance testing against reference distributions to ensure the reliability and validity of the clustering solution. This method, detailed in [235] provides a robust framework for cluster validation, enhancing confidence in the selected number of clusters for analysis.

Following the determination of the optimal number of clusters using both NbClust and M3C, this study advances by inputting the identified cluster count into the Deep Embedded Clustering (DEC) algorithm. The DEC algorithm

then executes the clustering, grouping the datasets into distinct clusters based on the unsupervised deep learning model's analysis.

11.2.1.4 CLUSTERS VALIDATION

To ensure the integrity and applicability of our clustering results, we conducted a validation process using the prediction strength validation method, a technique used in evaluating the coherence and sanity of clustering outcomes. This validation step confirms that the derived clusters are meaningful and accurately reflect the underlying patterns within the asthma data.

11.2.2 RESULT

11.2.2.1 VARIATION BETWEEN SEVERAL CLUSTER NUMBER SOLUTION DERIVED FROM THE CLUSTERING PROCESS.

The clustering analysis yielded several potential solutions, each proposing a different optimal number of clusters. This analysis was meticulously reviewed by a team of experts, including senior clinicians specializing in respiratory diseases, as well as professors with expertise in asthma and allergy epidemiology. Two noteworthy solutions emerged from the use of an enhanced deep clustering algorithm, indicating either four or three distinct clusters. Similarly, the deep embedding clustering algorithm suggested two configurations: one with five clusters and another with four.

Upon closer examination, we found that the five-cluster configurations proposed by both the deep embedding and improved methods did not exhibit significant clinical differences among various asthma patient subgroups. These configurations lacked distinct variations in critical characteristics such as lung function, inflammation levels, symptom presentation, and severity.

However, one of the solutions, indicating three clusters and derived from the refined embedding approach, demonstrated clinically meaningful distinctions. This solution identified a cluster characterized by early-onset mild asthma with minimal symptom presentation, optimal lung function, and the least severity, representing the best control status. The second cluster encompassed cases of late adolescent-onset asthma associated with high inflammation and elevated blood eosinophil counts. The third cluster included late adult-onset cases, marked by the most severe symptoms, frequent hospitalizations, and the worst control status.

Despite recognizing the clinical validity of this three-cluster solution, we concluded that the four-cluster solution obtained through the deep embedding clustering technique provided a more nuanced and differentiated perspective. This four-cluster solution revealed a distinct phenotype characterized by female dominance, good lung function, but frequent symptom occurrence, typically developing in late adolescence. Consequently, we acknowledge the four-cluster solution as the superior model, offering a clinically valid and meaningful delineation of asthma phenotypes. Additionally, the four-cluster solution that was adopted clinically was the second-best guess nominated by the validation metrics of our computational approach.