View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Helsingin yliopiston digitaalinen arkisto

Helsinki, Finland
and
Clinical Research Unit for Pulmonary Diseases and Division of Pulmonology,
Helsinki University Central Hospital,
Helsinki, Finland

# INDIVIDUAL TRAJECTORIES IN ASTHMA AND COPD: A LONGITUDINAL PERSPECTIVE TO OBSTRUCTIVE LUNG DISEASE

**Jukka Koskela**

**ACADEMIC DISSERTATION**

To be presented with the permission of the Faculty of Medicine,
University of Helsinki, for public examination in Lecture Hall 2,
Haartman Institute, on November 27th 2015, at 12 noon.

Helsinki 2015

**Supervised by:** Professor Tarja Laitinen, M.D., Ph.D.
Division of Medicine, Department of Pulmonary Diseases and Clinical Allergology, Turku University Hospital and University of Turku, Finland

**Reviewed by:** Docent Laura Elo, Ph.D.
Adjunct Professor in Biomathematics, Turku Centre for Biotechnology, University of Turku, Finland

Docent Terttu Harju, M.D., Ph.D.
Department of Internal Medicine, Respiratory Unit, Oulu University Hospital and University of Oulu, Finland
Medical Research Center Oulu, Respiratory Research Group, Oulu University Hospital and University of Oulu, Finland

**Opponent:** Professor Martin Tobin, M.D., Ph.D.
Department of Health Sciences, University of Leicester, Leicester, United Kingdom

*Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis*

ISBN 978-951-51-1707-6 (pbk.)
ISBN 978-951-51-1708-3 (PDF)
ISSN 2342-3161 (print)
ISSN 2342-317X (online)

http://ethesis.helsinki.fi

Layout: Tinde Päivärinta/PSWFolders Oy
Hansaprint
Vantaa 2015

*In memory of my father*

# TABLE OF CONTENTS

# LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications and one unpublished manuscript. Projects are referred to by their Roman numerals in the text.

I Koskela J, Kilpeläinen M, Kupiainen H, Mazur W, Sintonen H, Boezen M, Lindqvist A, Postma D, Laitinen T. Co-morbidities are the key nominators of the health related quality of life in mild and moderate COPD. BMC Pulmonary Medicine 2014 Jun 19;14:102.

II Koskela J, Kupiainen H, Kilpeläinen M, Lindqvist A, Sintonen H, Pitkäniemi J, Laitinen T. Longitudinal HRQoL shows divergent trends and identifies constant decliners in asthma and COPD. Respiratory Medicine 2014 Mar;108(3):463-71.

III Koskela J, Katajisto M, Kallio A, Kilpeläinen M, Lindqvist A, Laitinen T. Individual $FEV_1$ trajectories can be identified from a COPD cohort. Accepted for publication in: COPD: Journal of Chronic Obstructive Pulmonary Disease.

IV Koskela J, Surakka I, Pirinen M, Vasankari T. and HAI-group, Heliövaara M, Salomaa V, Laitinen T, Ripatti S. Single nucleotide polymorphism based heritability of $FEV_1$ development. Unpublished.

# ABBREVIATIONS

| | |
|---|---|
| ACOS | asthma – chronic obstructive pulmonary disease – overlap syndrome |
| AIC | akaike information criterion |
| ANOVA | analysis of variance |
| AUC | area under curve |
| BLUP | best linear unbiased predictors |
| COPD | chronic obstructive pulmonary disease |
| EHRs | electronic health records |
| $FEV_1$ | forced expiratory volume in 1 second |
| FVC | forced vital capacity |
| GWAS | genome-wide association study |
| HRQoL | health related quality of life |
| LD | linkage disequilibrium |
| MAF | minor allele frequency |
| MCID | minimum clinically important difference |
| OR | odds ratio |
| PEF | peak expiratory flow |
| SNP | single nucleotide polymorphism |
| RCT | randomized controlled trial |
| ROC | receiver operating characteristic |
| RPKM | reads per kilobase per million mapped reads |
| QC | quality control |
| VC | vital capacity |

# ABSTRACT

Managing chronic respiratory conditions such as asthma and Chronic Obstructive Pulmonary Disease (COPD) forms a notable burden on the healthcare system while the burden on an individual is equally notable as patients might suffer from a symptomatic disease for decades. However, not all asthma and COPD patients develop a disabling disease with frequent disease exacerbations and highest cost (in Quality Adjusted Life Years lost or healthcare spending). This variation in disease trajectories enables the analytical identification of distinct phenotypes over time. Retrospective data collected from a large number of patients could be used efficiently as the electronic health records are increasingly made available to researchers around the world. The aim of this project is to develop disease models based on longitudinal data to better capture the essential characteristics of obstructive lung disease, mainly focusing on COPD.

Projects I – III in this thesis are based on 2398 asthma and COPD patients retrospectively followed through electronic health records from year 2000 onwards. We aimed to analyse this real-world hospital based data using Hierarchical Models to assess the variation of development between individual patients over time. Unpublished Project IV is based on Health 2000 to 2011 follow-up study consisting of 1113 subjects from random Finnish population. The aim was to estimate Single Nucleotide Polymorphism (SNP) based heritability of Forced Expiratory Volume in 1 s ($FEV_1$) level and development and to perform a Genome-Wide Association Study (GWAS) to identify possible genetic markers associated with $FEV_1$ development over time.

Our results suggest that the major determinants of Health Related Quality of Life (HRQoL) in mild or moderate COPD are the common comorbidities associated with COPD while in severe diseases the accentuated lung function has a major role. Over time, observable individual trajectories of HRQoL are presented in Asthma and COPD. Significant decline of HRQoL in Asthma was found to associate with obesity related diseases and states while the main determinants in COPD were poor lung function and increasing age. Psychiatric conditions were found associated in both Asthma and COPD.

Using an unbalanced data (varying number of measurements and length of follow-up time) of lung function measurements, we were able to observe significant individual trajectories of $FEV_1$ based on the past development. Significant and rapid decline was seen in 30% of the COPD cohort in the study while significant improvement was extremely rare. Rapid decline was associated with numerous exacerbation related markers.

Our unpublished results suggest that development of $FEV_1$ is significantly affected by common variants in DNA as genetic effects were estimated to explain 1/3 of the phenotypic variance in random Finnish population. One locus previously associated with the level of

$FEV_1$ was found associated with the development of $FEV_1$. Suggestive evidence for two novel loci associated with $FEV_1$ development was also identified.

The findings underline the varying trajectories of HRQoL and lung function seen in a homogenous cohort of Asthma and COPD patients.

This thesis aims to provide approaches and aspects to better understand the trajectories of a chosen parameter in asthma and COPD. The variation of e.g. lung function development is abundant, and we should not consider this variation as an obstacle but as a useful source of information as there might be genetic or environmental determinants causing this variation.

# TIIVISTELMÄ

Krooniset ahtauttavat keuhkosairaudet, kuten keuhkoahtaumatauti (COPD) ja astma ovat merkittäviä kansansairauksia Suomessa. Jopa lähes kymmenys maan aikuisväestöstä kärsii astmasta, COPD:n esiintyvyys ikääntyneillä miehillä saattaa olla vielä tätäkin suurempi. Kroonisina tiloina ahtauttavat keuhkosairaudet aiheuttavat huomattavan taakan niitä sairastaville potilaille sekä suuria kustannuksia yhteiskunnalle.

Merkittävä riskitekijä vaikealle keuhkosairaudelle on tupakointi, joka on tunnettu riskitekijä myös monelle muulle krooniselle sairaudelle. Kuitenkin vain osa tupakoivista ahtauttavaa keuhkosairautta sairastavista omaa voimakkaasti etenevän sairauden, joka pahimmillaan johtaa toistuviin intensiivistä hoitoa vaativiin sairauden pahenemisvaiheisiin ja ennenaikaiseen kuolemaan. Nykyisellään suuri osa resursseista kulutetaan intensiivisessä hoidossa, eikä niiden ennaltaehkäisyssä. On mahdollista, että etenevän ja huomattavia kustannuksia aiheuttavan keuhkosairauden taustalla on sille ominaisia perinnöllisiä tai ympäristöllisiä riskitekijöitä. Nämä riskitekijät ovat kuitenkin nykyisellään huonosti tunnettuja.

Keuhkosairaus aiheuttaa oireita vasta keuhkojen toiminnan heikennyttyä jo merkittävästi, jonka jälkeen potilaan toimintakyky heikkenee nopeasti, mikäli sairaus edelleen etenee. Mikäli huonosti kehittyvät potilaat kyettäisiin paremmin tunnistamaan aikaisessa vaiheessa, olisi mahdollista kohdistaa hoito ja interventiot niitä eniten tarvitsevaan ryhmään.

Tässä työssä on tutkittu erityisesti COPD:tä ja astmaa sairastavien potilaiden terveyteen liittyvän elämänlaadun ja keuhkojen toiminnan kehitystä yli ajan. Noin 30 %:lla COPD-potilaista nähtiin tilastollisesti merkitsevää laskua keuhkojen toiminnassa. Heikosti kehittyvät potilaat kärsivät myös muita useammin sairauden pahenemisvaiheista. Huonolle elämänlaadulle COPD:ssä ja elämän laadun kehitykselle astmassa ja COPD:ssä tunnistettiin useita kliinisiä riskitekijöitä. Työssä tarkasteltiin lisäksi väestöaineistossa nähtyä keuhkojen toiminnan muutosta, jolloin noin kolmanneksen vaihtelusta nähtiin selittyvän geneettisillä tekijöillä.

Väitöstyön aineisto on kerätty takautuvasti, mutta kattavasti erilaisista sairaala- ja rekisterilähteistä sekä toistuvin kyselytutkimuksin. Sairaala-aineisto edustaa aitoa potilasmateriaalia, jota ei ole tutkimusta käynnistettäessä merkittävästi rajattu. Ainutlaatuisen materiaalin analysoimiseksi on käytetty kehittyneitä tilastollisia menetelmiä, jotka mahdollistavat yksittäisten potilaiden kehityskulun arvioimisen yli ajan. Nämä menetelmät ovat helposti käyttöönotettavissa myös sairaaloiden tietokannoissa, jolloin hoitava henkilökunta saa paremmin tietoa mahdollisista korkean riskin potilaista entistä aikaisemmin.

# 1 INTRODUCTION

Asthma and Chronic Obstructive Pulmonary Disease (COPD) belong to the category of obstructive lung diseases due to the airflow limitation, which is manifested during expiration. This obstruction of expiratory airflow can be divided based on anatomical structures as in asthma the obstruction is seen in the larger airways (Maddox, Schwartz 2002) and in COPD the most affected airways are the smallest ones (Hogg 2004). Asthmatic symptoms are often varying in time and are caused by airway inflammation and bronchial hyper-responsiveness leading to variably occurring airflow limitation (Global Initiative for Asthma (GINA) 2015). This expiratory airflow limitation is considered reversible after bronchodilation in spirometric testing (Pellegrino et al. 2005). Asthma often first expresses in childhood or in early adulthood and as age increases, the prevalence of COPD becomes more pronounced. COPD is rarely seen in 40-year-old or younger people without alpha-1-antitrypsin deficiency. The most important risk factor for COPD is tobacco smoke whereas especially in low-income countries the exposure to toxic fumes from burning biomass fuels plays a major role (Salvi, Barnes 2009). COPD is characterized by non-reversible airway obstruction in expirium due to resistance in the smallest airways. Elevated resistance in smallest airways is due to chronic bronchitis and collapsing of lung structure after inflammatory response (Hogg 2004). In contrast to asthma, the airflow limitation in COPD is non-reversible and often progressive if exposure to noxious stimuli continues (Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015). Adult populations often constitute diagnostic challenges as incomplete responses to administered bronchodilators in asthma (Kauppi, Kupiainen et al. 2011, Lee et al. 2007) and clinically significant bronchodilation in COPD (Albert et al. 2012) are seen. Patients with characteristics from both asthma and COPD could be seen having Asthma-COPD Overlap Syndrome (ACOS) (Bateman et al. 2015). Asthma is also a known major risk factor of COPD (Silva, Sherrill et al. 2004).

As the diagnosis of most diseases, including obstructive lung diseases, is often made cross-sectionally at a certain period of time, the development of relevant parameters is not assessed at the time of diagnosis. This is well-suited to match the needs of practicing medicine when a patient is very rarely followed to gain a diagnosis, but ill-suited to match the needs of accurate diagnosis. The fundamentals of diagnostics are stagnant in time while it might be extremely useful to follow e.g. response to treatment, quality of life or lung function over time to determine the relevant characteristics of each patient.

Under each obstructive lung disease, many distinct phenotypes exist both in asthma (Wenzel et al. 2012) and in COPD (Turner et al. 2015), each presenting with characteristic underlying disease processes leading to a certain clinical status. Some of the already known or yet unknown phenotypes could be identified based on longitudinal development of a certain parameter, say, lung function over time. As the concept of identifying change on an individual patient-level is not widely known in the field of clinical medicine, the aim

of this study is to provide better understanding and tools for interpreting longitudinal change. As this study shows, there is an abundant variation in the trajectories of each patient even though diagnosis and other relevant criteria are matched. This variation should not be regarded as a nuisance but as valuable information of each patient's own characteristic. The uncertainty related to a specific trajectory might also be considered an advantage in the analysis, in cases where some trajectories could be considered significant in contrast to uncertain ones. Significant trajectories could be statistically weighted more to avoid biased inferences due to insignificant trajectories.

# 2 LITERATURE REVIEW

## 2.1 Asthma, COPD and Asthma COPD Overlap Syndrome (ACOS)

The key component in obstructive lung disease is the impaired expiratory airflow. The natural history of these three diseases differs markedly for asthma and COPD but the processes are localized mainly on the conductive airways. Conductive airways deliver the air to the bronchi from pharynx/larynx and do not take part in the gas exchange process, which takes place in the respiratory bronchioles.

Asthmatic airflow limitation is most often initiated by eosinophilic chronic inflammation in the airways leading to hyper-responsiveness in the surrounding smooth muscles finally causing the variable airflow limitation (Anderson et al. 2008). Thickening of basement membrane further obstructs the airways. Airflow limitation in asthma is considered reversible spontaneously or after administration of bronchodilators. The inflammation can be repressed with the use of corticosteroids, usually as administered as an inhalation. The overall response to therapy and prognosis in asthma is good, however, some forms of asthma are related to poor prognosis and frequent disease exacerbations. Numerous phenotypes (subgroups with distinct observable characteristics) in asthma have been suggested, e.g. allergic, eosinophilic and non-eosinophilic, neutrophilic, exercise induced, obesity related, aspirin sensitive and occupational asthma (Turner et al. 2015, Wenzel et al. 2012).

In COPD, noxious stimuli such as tobacco smoke, is the most common cause of inflammatory process, which takes place in the lungs and localizes in the smaller airways (diameter <2mm) compared to asthmatic patients. The inflammatory process (Barnes 2014) stimulates mucous hypersecretion exacerbated by the loss of ciliary function leading to difficulties in expiratory airflow. If persistent, the condition is considered to be chronic bronchitis that very often precedes clinical COPD. Airway remodelling might occur leading to narrowed diameter of bronchioles obstructing the airflow. If noxious stimuli continue or the person is prone to the inflammatory process in question, the balance of proteolysis and antiproteolysis in lungs is disturbed leading to destruction of alveolar structure, normally responsible for gas exchange. Proteolysis leads to loss of elastic recoil in alveolar walls and surrounding structures resulting in air-trapping during expiration. Dilation of alveolar walls leads to fusion of alveolus, also known as emphysema. Emphysema leads to decreased diffusion capacity of the alveolus due to decreased area for gas exchange when small alveoli form larger spaces (Hogg 2004. However, not all these processes are necessarily found in a particular COPD patient. The first phenotypes in advanced COPD were the Blue Bloater and Pink Puffer (Dornhorst 1955) as Pink Puffer represented an older patient with advanced emphysema and severe dyspnea. Blue Bloater in contrast represented a younger patient with chronic bronchitis, recurrent disease exacerbations and commonly suffering from congestive heart failure. These phenotypes are crude generalizations and only offer insight to COPD as an umbrella diagnosis and numerous phenotypes have since been suggested

(Turner et al. 2015). In COPD the mainstay of therapies are muscarinic antagonists and beta-2-adrenergic agonists aimed to decrease air-trapping and mucous hypersecretion. Disease exacerbations often associate with infections, viral or bacterial and occasionally require cures of corticosteroids or antibiotics. The key component to curbing the development of obstruction is avoiding the noxious stimuli causing COPD.

The prevalence of ACOS depends largely on the diagnostic criteria used. Patients with severe or difficult-to-treat asthma have been found with persistent airflow limitation of up to 60% (Lee et al. 2007) while coexistent asthma and COPD diagnosis have been found in less than 20% (Kauppi, Kupiainen et al. 2011, Andersen et al. 2013, Pleasants et al. 2014). Clinically, patients with ACOS are often found with poor lung function compared to COPD patients even though they present less extensive smoking history while comorbidities and disease exacerbations contribute to the condition as well (Nielsen et al. 2015).

## 2.2 Health Related Quality of Life in Asthma and COPD

Self-administered questionnaires are frequently used to assess the Health Related Quality of Life (HRQoL) and disabilities caused by obstructive lung disease. The purpose of the questionnaires is often to provide a summary score (or multiple scores) allowing comparison between patients, patient‘s development over time or to evaluate treatment outcomes. The questionnaires can be divided based on their focus, whether it is on general health or on disease specific HRQoL. Generic questionnaires are used for gaining information about the patient’s general state of health, such as 15D (Saarni, Harkanen et al. 2006), the most commonly used questionnaire in Finland. This is done to estimate the overall burden of separate and overlapping diseases or states affecting the patient‘s HRQoL. Generic questionnaires also allow the comparison of the general population to age or gender matched patients to quantify e.g. Quality Adjusted Life-Years lost due to a disease. When using a generic questionnaire in a study of a specific disease, it should be noted that other diseases could have a confounding effect on the HRQoL if a patient is mainly suffering from the other disease and not the one of interest. However, this effect might be observed with a disease specific questionnaire when the effects of comorbidities provoke pulmonary symptoms as in e.g. congestive heart failure. The ability to observe HRQoL in a wide perspective is a major advantage of the generic questionnaires as diseases might have an impact through yet unknown mechanisms manifesting themselves in different organ systems.

The focus of the disease specific questionnaires such as Airway Questionnaire 20 (AQ20) (Hajiro, Nishimura et al. 1999) is on a specific disease or an organ system and its symptoms or limitations possibly affecting HRQoL. They are frequently used for the obstructive lung disease as the responsiveness is found superior to generic questionnaires (Puhan, Guyatt et al. 2007). HRQoL has been found lowest in patients with ACOS while patients with asthma

have better QoL compared to COPD patients as comorbidities and gender also contribute to HRQoL (Kauppi, Kupiainen et al. 2011, Laitinen, Hodgson et al. 2009).

## 2.3 Co-morbidities and lung function on HRQoL in Asthma and COPD

Although the hallmark of COPD is progressive airflow limitation, many other diseases or states contribute to the HRQoL as well. Airflow limitation has not been found sufficient in predicting the outcome of COPD and correlations between lung function parameters and HRQoL have been modest at most. Asthma is commonly associated with chronic rhino-sinusitis, allergic rhinitis and other atopic diseases, obesity and sleep apnoea (Baiardini, Braido et al. 2006, Braido, Bousquet et al. 2010, Beuther, Sutherland 2007) while COPD has been associated with cardiovascular diseases, diabetes, osteoporosis, kidney and heart failure, musculoskeletal dysfunction, anxiety, depression, addiction diseases and cancer (Corsonello, Antonelli Incalzi et al. 2011, Wijnhoven, Kriegsman et al. 2001, Leidy, Coughlin 1998, van der Molen, Postma et al. 1997, Marks, Dunn et al. 1993, Curtis, Deyo et al. 1994, van Manen, Bindels et al. 2001, Burgel, Escamilla et al. 2013). The background of multimorbidity in COPD remains to be determined, but systemic inflammation has been hypothised as the primus motor, provoked by tobacco or other biomass smoke (Barnes 2010). The shared comorbidities possess common risk factors such as tobacco smoking, but the web of causality is poorly understood. The determinants of the poor development of HRQoL are not well known, but increasing age, smoking, non-compliance to medication, dyspnoea and low Peak Expiratory Flow (PEF) have been found to be associated with the decline in asthma while male gender, low bodyweight, physical inactivity and respiratory symptoms have been found associated in COPD (Hesselink, van der Windt et al. 2006).

## 2.4 Natural history of a disease requires longitudinal assessment

The natural history of a disease is used to describe how a clinical state is achieved in a process with three phases (Natural History of Disease 2008):

- Pathological onset
- Pre-symptomatic change, from the pathological onset to the first symptoms
- Clinical disease

A number of prospective studies have been performed to capture the essential determinants of each disease's natural history, e.g. Framingham Heart Study (Dawber, Meadors et al. 1951) initiated in 1948 to better understand the role of blood pressure as a risk factor of cardiovascular diseases. In the case of COPD, the concept of different susceptibility to tobacco smoke was established by the seminal study (Fletcher, Peto 1977), supportive data was later presented (Burrows, Knudson et al. 1977, Burrows 1990). Even though these studies

confirmed the role of tobacco smoke and differential susceptibility to it, the natural history of COPD remains to be under investigation as it is evident that important determinants of disease progression are still to be found (Mannino, Watt et al. 2006).

The examination of the natural history of a disease requires longitudinal studies to assess the risk factors of different paths e.g. to poor lung function as the follow-up time of 8 years in the seminal study by Fletcher and Peto is arguably short to fully describe the characteristics of COPD (Rennard, Vestbo 2008). Further studies are needed to assess the following basic epidemiological effects of time in COPD: the age, period effects and their possible interactions (Szklo, Nieto 2014a).

## 2.5 Lung function and its development in COPD

While fixed airflow limitation is characteristic of COPD (Mannino, Watt et al. 2006), there are, however, numerous phenotypes in COPD (Han, Agusti et al. 2010, Friedlander, Lynch et al. 2007) such as the frequent exacerbator, emphysema dominant, airway dominant and (lung function) rapid decliners.

Rapid decline of Forced Expiratory Volume in 1 s ($FEV_1$) usually considered as >50ml/year, has been shown to be associated with morbidity and mortality (Beeckman, Wang et al. 2001, Baughman, Marott et al. 2011, Ryan, Knuiman et al. 1999). Rapid decline might also be associated with other phenotypes such as the frequent exacerbator and airway hyper-responsiveness (Han, Agusti et al. 2010, Anzueto 2010). Development of lung function has been shown to be highly variable in COPD (Nishimura, Makita et al. 2012, Vestbo, Edwards et al. 2011, Casanova, de Torres et al. 2011, Tantucci, Modina 2012, Qureshi, Sharafkhaneh et al. 2014). As the variation in $FEV_1$ development is abundant, there is a need to identify the rapid decliners at an individual level (Tashkin 2013) in contrast to mean values of group level decline. The group level decline is used to account for random error of individual trajectories as the grouping is often based on the distribution of individual trajectories, thus discarding the uncertainties related to the individual study subjects. Reliable trajectory estimates require a clinically valid follow-up time with the minimum of 3 measurements of lung function. Assessing the individual trajectories makes use of the uncertainties as an unknown fraction of apparent trajectories with 2 measurements are due to random errors. Trajectories based on 2 measurements do not include the information about the individual level variation and using 2 measurements requires a longer follow-up time and higher validity and reliability to allow solid inferences of the development. Thus trajectories in this thesis are seen as patient specific trends over time, consisting of minimum 2 measurements over clinically valid time.

As the phenotypic variance seen in e.g. lung function can be divided to environmental and to genetic sources, the development of lung function should be thoroughly assessed and understood to allow for further analysis of its determinants.

## 2.6 Heritability of lung function and its development

Heritability estimates can be produced in familial or twin studies or by taking advantage of the genotyped and shared variants of a DNA sequence. Single Nucleotide Polymorphism (SNP) based heritability of lung function level has been assessed in a study enriched with heavy smokers suggesting that 38% of $FEV_1$ variation in non-Hispanic whites and 51% in African Americans is due to genetic effects (Zhou, Cho et al. 2013). Pedigree estimates of lung function level have produced similar but variable estimates depending on the source population and the spirometric parameter (Wilk, Djousse et al. 2000, Ingebrigtsen, Thomsen et al. 2011, Klimentidis, Vazquez et al. 2013). Development of lung function assessed in a cohort of elderly never smoking female twins found that while a third of variation of cross-sectional lung function is due to genetic effects, their contribution in the development of $FEV_1$/FVC is substantially lower (Hukkinen, Kaprio et al. 2011). Even lower estimates were produced in a familial study of elderly population with estimates from 5% ($FEV_1$) to 18% (FVC) but when restricted to concordant for smoking status the estimates were somewhat higher, 18% to 39% respectively (Gottlieb, Wilk et al. 2001). SNP-based and familial estimates have been shown to produce parallel estimates (Klimentidis, Vazquez et al. 2013). Heritability of lung function development has thus been assessed only in familial or twin studies and never using SNP-based analysis.

Also, when the heritability of multiple phenotypes ($FEV_1$/FVC, emphysema, gas trapping) related to COPD was estimated, it was found that the heritability of COPD disease status was 38% (Zhou, Cho et al. 2013).

## 2.7 Genetic susceptibility to poor lung function and COPD

During recent years, a number of loci have been found to associate with the level of lung function (22 associated with $FEV_1$/FVC and 7 with $FEV_1$, Table 1.) (Hancock, Eijgelsheim et al. 2010, Repapi, Sayers et al. 2010, Soler Artigas, Loth et al. 2011) while they only explain some 3% of the variation seen in $FEV_1$/FVC and 1.5% of the variation in $FEV_1$ (Soler Artigas, Loth et al. 2011).

**Table 1.** *Previously identified loci associated with FEV1 level and development.*

| Locus | Chromosome | Trait |
|---|---|---|
| MECOM | 3 | $FEV_1$ level |
| ZKSCAN3 | 6 | $FEV_1$ level |
| CDC123 | 10 | $FEV_1$ level |
| C10orf11 | 10 | $FEV_1$ level |
| HTR4 | 5 | $FEV_1$ level |
| TNS1 | 2 | $FEV_1$ level |
| GSTCD | 4 | $FEV_1$ level |
| ME3 | 11 | $FEV_1$ development |
| IL16/STARD5/TMC3 | 15 | $FEV_1$ development |
| DLEU7 | 13 | $FEV_1$ development |

The susceptibility for poor development of lung function has been assessed twice (Imboden, Bouzigon et al. 2012, Tang, Kowgier et al. 2014), the first one stratified for asthma status. In the stratified analysis a DLEU7 locus associated with $FEV_1$-development was replicated while the large meta-analysis revealed one locus on ME3 at genome-wide significance in a sub-cohort with >2 measurements and another suggestive locus in IL16/STARD5/TMC3. The known loci, however, explain only a small fraction of variation seen in lung function development considering the heritability estimates for lung function development.

For COPD, seven loci have been identified (Pillai, Ge et al. 2009, Cho, Boutaoui et al. 2010, Cho, McDonald et al. 2014) while again the known loci (Table 2.) explain only a minority of this variation due to genetic effects (Cho, Castaldi et al. 2012).

A known predisposition to emphysematic disease is the alpha-1-antitrypsin deficiency (Laurell, Eriksson 1963) due to mutated SERPINA1 gene is present in 1–3% of COPD cases (Cohen 1980).

**Table 2.** *Loci associated with COPD disease status.*

| Locus | Chromosome | Trait |
|---|---|---|
| CHRNA3/5/IREB | 15 | COPD disease status |
| FAM13A | 4 | COPD disease status |
| HHIP | 4 | COPD disease status |
| RIN3 | 14 | COPD disease status |
| MMP12 | 11 | COPD disease status |
| TGFB2 | 1 | COPD disease status |
| CYP2A6/EGLN2/RAB4B | 19 | COPD disease status |

## 2.8 Questionnaires and tests in studying obstructive lung disease

The validity of a questionnaire or a test is a feature with which the ability to approximate a true but unknown parameter (e.g. HRQoL or $FEV_1$ level) is estimated. Reliability is another feature of a test, and it measures how the estimated parameter would change if the test was repeated. Validity and reliability thus have a relationship affecting the interpretation of the estimates.

The validity of a questionnaire is commonly assessed by comparing the questionnaire in question to the golden standard by means of correlation or other statistical measure. In the field of HRQoL of asthma and COPD, the St. Georges Respiratory Questionnaire (Jones 1992) is considered the golden standard in pulmonary diseases to which both of the questionnaires used in this study have been compared (Kauppinen, Sintonen et al. 1998, Mazur, Kupiainen et al. 2011, Hajiro, Nishimura et al. 1999) while reliability has also been assessed (Stavem 1999, Barley, Quirk et al. 1998). The reliability and repeatability are important measures when measured cross-sectionally, and they become even more vital when used in longitudinal studies as the uncertainty related to an individual trajectory is assessed as more sources of variation are introduced when measured repeatedly.

## 2.9 Bias in epidemiological study

Erroneous interpretation of the results might be due to many sources causing a systematic bias in the results, but they can be roughly divided into selection and measurement biases and confounding.

In the case of selection bias, study participants are selected with a different probability from the target population based on some characteristic. This characteristic might be of interest regarding the study in question and thus distort the inferences. Problems arise if the study aims to depict the whole population, but only a sub-population is included due to e.g. recruitment or follow-up process. The selection probabilities are then affected by exposure or disease status (Dos Santos Silva 1999).

The measurement bias can yet be subdivided into misclassification bias, ecological fallacy and regression towards the mean. The misclassification bias is due to measurement error of exposure or outcome status due to the lack of validity or reliability of questionnaire or test. Sources of misclassification bias are recall, interviewer, reporting and detection biases. Recall bias is a form of misclassification bias relevant in questionnaire-based data when a patient's past answers have a differential impact on the present questionnaire depending on the case-control status. Interviewer bias might appear during interviews, but also lung function testing could be affected by the spirometry staff. Reporting bias might happen unknowingly or knowingly when reporting is affected by the intercourse with the researcher.

Detection bias occurs when a risk factor in question directs to diagnostic procedure, which then leads to excess diagnosis in the exposed. Ecological fallacy takes place when the findings made at group-level are generalized to the individual level. Here ecology refers to a different geographical localization of groups based on which the inferences are made, but broadly the fallacy could take place even within the same region. As in smokers, the decline of lung function is known to be steeper on average, but this is not necessarily true for all individuals as some may not suffer from tobacco smoke. Regression towards the mean is a phenomenon related to extreme values collected at the start of a follow-up period. These extreme values have a tendency to shift towards the mean value of the distribution over time as they have a higher probability to be erroneous (Delgado-Rodriguez, Llorca 2004).

Confounding variables differ from other sources of bias in the sense that they truly exist, in contrast to other sources of bias that are due to erroneous study design. Therefore, confounding variables need to be taken into account when plausible and useful models are developed. Confounding variables are:

- considered to be causally linked to the outcome of interest (a direct risk factor or a proxy),
- considered to be causally or non-causally linked to the exposure in question,
- not considered to be in between of exposure and outcome in the web of causality.

Random error is not a source of bias as it is random by definition and does not affect the estimate but only the uncertainty related to the estimate. Random error can be compensated by increasing the number of samples (Delgado-Rodriguez, Llorca 2004, Szklo, Nieto 2014b).

## 2.10 Retrospective data and Electronic Health Records

In retrospective studies, the outcome and exposures are determined at the initiation phase of the study, and associations of exposures are estimated in a retrospective manner, while in prospective studies the exposure is measured during the follow-up while waiting for the outcome. In prospective studies it is thus possible to plan for data collection systematically to avoid exposure misclassification while this is not possible in retrospective studies. Major challenges epidemiologically and analytically are incomplete data and accuracy of the data due to selection and misclassification bias, which can both cause major misinterpretation of the results. As prospective studies possess qualities superior to retrospective studies, they suffer from higher cost and a time lag from study initiation to the analysis of the results.

Health record data is often recorded only in the case of an event (Hripcsak, Albers 2013) which is usually unfavourable to health and thus the data collection is not systematic as in prospective studies. Missing data could be due to no event or due to event not recorded when it should have been recorded. Data collection is enriched in patients receiving more

intensive care and examination while the non-exposed are exempt from attention from the medical system. Data is often missing not at random, but the missing data is differentially distributed in the exposed and non-exposed, and the use of analytical methods to account for confounding (e.g. multivariate regression) might further aggravate the selection bias. Thus health record data always includes sources of bias and cannot be used as data collected at clinical trials as is, but needs further assessment (Weiskopf, Weng 2013).

## 2.11 Healthcare data in research use

Electronic health records (EHRs) data is abundantly collected for medical decision-making in clinical setting and is used by medical professionals such as medical doctors to enable the diagnosing and treatment of diseases and conditions. After clinical decision-making this data has little use other than its potential for research activities.

EHRs data could, however, be used to define phenotypes in high-throughput manner to refine the clinical data for use in machine learning (Hripcsak, Albers 2013). Possibly the greatest potential of EHRs is in the ability to combine it with other registry and genetic data collected during normal clinical practice (Kohane 2011). Genotype data could be used in conjunction with the whole phenome data at once in contrast to one phenotype-genotype association (Choi, Kim et al. 2013). The use of electronic healthcare data is, however, limited by numerous legal and ethical issues (Taylor 2008).

Systematic bias using EHRs data could, however, persist even after phenotyping and needs to be taken into account whenever using machine learning methods to analyse the EHRs data (Jensen, Jensen et al. 2012).

# 3 AIMS OF THE STUDY

The aim of this thesis was to develop analytical approaches to make use of electronic health care records (EHRs) in identifying trajectories manifesting over time in asthma and COPD. Trajectories are not widely used in the field of clinical medicine, where longitudinal inferences are generally made comparing mean values at the start and at the end of a follow-up time. The use of EHRs necessitated the use of flexible methods for the following reasons: the data includes missing values, measurements are not evenly spaced in time and follow-up times vary depending on patient. The parameters under investigation were Health Related Quality of Life and lung function. The specific aims were:

- to assess the effect of common comorbidities on the HRQoL in asthma and COPD,
- to study whether the development of the chosen parameters would present inter-individual variation in asthma and COPD patients,
- to study whether individual trajectories of HRQoL and $FEV_1$ could be identified and to assess their association to clinical determinants,
- to quantify the variation of $FEV_1$ development due to genetic markers in prospective data and to possibly identify genetic markers associated with the development.

# 4 MATERIAL AND METHODS

## 4.1 Finnish Chronic Airway Disease (FinnCAD) cohort

The cohort used in the Projects I–III is collected from Helsinki and Turku University Central Hospitals and discharged with ICD10 J44 and J45 codes during years 1995–2006. The Coordinating Ethics Committee of the Helsinki and Uusimaa Hospital District (Coordinating Ethics Committee decision 125/E0/04) has approved the study, and the permission to conduct research was granted by the Helsinki and Turku University Hospitals. This cohort consists of 2395 asthmatics, smoking related chronic bronchitis or COPD and Asthma-COPD Overlap Syndrome (ACOS) patients, whose medical records were collected retrospectively from 5–10 years prior to study enrolment during years 2005–2007. Thorough examination of the clinical and diagnostic data was done to determine the main components of the obstructive lung disease as asthma (1329 patients), COPD (739 patients) and ACOS (347 patients), following the GOLD (Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015) and GINA (Global Initiative for Asthma (GINA) 2015) guidelines. All COPD and ACOS patients had a smoking related disease and reversibility in spirometry was required for ACOS diagnosis. Common comorbidities were determined at the time of recruit based on diagnosis, often made by a specialist. Patients will be followed every other year until 10 years has passed from study enrolment while HRQoL, working ability, medicine use and smoking habits are being collected prospectively. During recruitment, blood samples were taken for later analysis. Descriptive data of patients in the cohort are given in the Table 3.

**Table 3.** *Descriptive characteristics of FinnCAD-cohort used in Projects I–III. The data is presented as % of the total or mean (SD) unless stated otherwise. Smoking status is collected in conjunction from medical records and follow-up data.*

| **Total N=2398** | **Asthma, N=1329** | **COPD, N=739** | **ACOS, N=347** |
|---|---|---|---|
| Male % | 26% | 64% | 47% |
| Year of Birth, range | 1951 (12.5), 54 | 1942 (6.8), 46 | 1944 (6.9), 35 |
| $FEV_1$ % | 86.5% (18%) | 57.2% (19%) | 58.4% (19%) |
| $FEV_1$/FVC% | 78.8% (9%) | 63.8% (14%) | 63.6% (14%) |
| Pack Years | 8 (13) | 42 (17) | 16 (39) |
| Current smokers | 123 | 235 | 126 |
| Ex-smokers | 528 | 484 | 221 |
| Non-smokers | 678 | 0 | 0 |
| Age at onset of obstructive lung disease | 43 (16) | 58 (7) | 53 (11) |
| Hypertension | 34% | 41% | 41% |
| Diabetes | 7% | 15% | 14% |
| Alcohol abuse | 4% | 15% | 20% |
| Psychiatric condition | 30% | 33% | 40% |
| Body Mass Index | 27.4 (5.5) | 27.0 (5.5) | 27.6 (6.5) |
| 15D score at inclusion | 0.86 (0.1) | 0.79 (0.1) | 0.79 (0.1) |

## 4.2 GenMets cohort

Cohort used in the Project IV consists of random sample of Finnish adult population for Health 2000 Survey Cohort (Heistaro 2008) with a follow-up on year 2011. The GenMets subcohort consists of 919 metabolic syndrome cases and 1219 matched controls genotyped with Illumina 610K chip. Valid pre-bronchodilatory spirometry at both baseline and follow-up was available for 1113 subjects (Table 4.). Asthma and COPD were determined based on interviews.

**Table 4.** *Descriptive characteristics of the cohort used in Project IV. The data is presented as N, % of total or mean (SD) unless stated otherwise. Study participants with less than 100 cigarettes smoked were considered never-smokers.*

| | **N=1113** |
|---|---|
| Male/Female | 552/561 |
| Age, range | 49 (10), 45 |
| $FEV_1$ | 3.38 (0.83) |
| $FEV_1$/FVC | 80.5% (6.0%) |
| Never-smokers | 52% |
| Ever-smokers | 48% |
| Obstructive lung disease (Asthma/COPD) | 7% |
| Body Mass Index | 27.1 (4.3) |

## 4.3 Statistical analysis

### *4.3.1 Comorbidities in COPD – A cross-sectional analysis*

In Project I linear and logistic regression was used to assess the determinants of HRQoL as continuous and binary (very low HRQoL vs. others) as an outcome. Final models were built using backwards stepwise regression based on Akaike Information Criterion (AIC). Regression estimates from linear regression for HRQoL were standardized to compare the effects of each independent variable on the HRQoL as the scales on independent variables vary. Unadjusted and adjusted coefficients from regression models were given to allow solid assessment of the results as possible confounders are included in adjusted models. Spearman's correlation coefficient was used to allow assessment of nonlinear correlations and non-normal distributions of variables. Differences in the mean values of the estimated parameters between patient groups were determined using ANOVA (Analysis of Variance) followed by Tukey's Honestly Significant Difference as a post-hoc test. Receiver Operator Characteristic (ROC) and Area Under the Curve (AUC) statistic were determined to assess the ability of the selected HRQoL model in predicting mortality during the next 5 years.

### 4.3.2 HRQoL development in Asthma and COPD – focus on significant individual trajectories

The individual trajectories of HRQoL over the 5-year follow-up time were assessed in the Project II when patients had a variable number of HRQoL measurements distributed unevenly in time. A linear mixed effects model (Robinson 1991) was used to model the trajectories consisting of multiple measurements. The Best Linear Unbiased Predictors (BLUPs) for trajectory and intercept were let to vary at random from patient to another as it is assumed that patients present variation in their baseline and trajectories of HRQoL. Subsequently Markov Chain Monte Carlo simulations using Bayesian inference (Martin, Quinn et al. 2011) were run to create a sample of the posterior distribution of the trajectories to identify individual patients with significant decline. The decline was considered significant when 85% of the simulated samples were negative (85% probability level) for a particular patient. ROC and AUC statistic were used to estimate the value of cross-sectional HRQoL measurement in predicting future development. Optimal cut points were determined for cross-sectional HRQoL-measurement to estimate the Odds Ratios related to lower baseline HRQoL in future HRQoL development. Bayesian Models Averaging (Wintle, McCarthy et al. 2003) was used to determine the important determinants of the development of HRQoL by means of averaging over the competing models to estimate the posterior effect probabilities for each variable in the input (Hoeting 1999). Clinical determinants possibly affecting the development were included in a generalized linear model when trajectories of HRQoL were treated as a continuous outcome. Missing values in the 15D questionnaire were imputed up to three dimensions using a regression method based on other dimensions with an algorithm provided with 15D instruments (Sintonen 1994).

### 4.3.3 Lung function development in COPD – assessing the variability and identifying individual trends in unbalanced data set

The focus of the Project III was on individual $FEV_1$ trajectories over time, again distributed unevenly between patients and over time. The number of available measurements varied greatly between patients. Thus a Hierarchical Bayesian Model (Gelman, Hill 2007) was used to allow the flexible estimation of the model parameters using a non-informative prior distribution. Linear fit was assumed as the aim of this study was not to assess the age or time effects in trajectories. Linear fit is also less prone to overfitting. Logistic regression was used to determine the clinical determinants associated with significant lung function decline.

### 4.3.4 Genetic background of lung function development

Lung function development in project IV was calculated by subtracting the $FEV_1$ of the year 2000 from the $FEV_1$ of the year 2011. Estimating the heritability of the lung function parameters was based only on genotyped (HumanHap610-Quad Genotyping BeadChip) common variants (SNP Minor Allele Frequency, MAF >5%). Quality Control (QC) (Turner, Armstrong et al. 2011) was performed to exclude individuals and markers with call

rate <95%. Individuals were also excluded due to unexpected relatedness, unusually high heterozygocity and gender checks while SNPs were excluded due to deviations from the Hardy-Weinberg equilibrium $<1\times10^{-6}$. A suggested method (Yang, Benyamin et al. 2010) aims to estimate the narrow sense heritability of unrelated subjects by fitting the SNP effects (Identity by State matrix) in a mixed effects model. SNP effects were included as random effects to estimate the variance explained while assessing the effects simultaneously.

Prior to the screening of the genetic association, SNPs with call rate <99% were further excluded to allow robust genotype imputation using the 1000 Genomes reference panel with added subjects from the Finnish population (Sequencing Initiative Suomi). Genotype imputation was done to achieve a more comprehensive set of SNPs while it was based on predicting the variants by using the genotyped variants and a dense reference panel of haplotypes. The two haplotypes were then compared and aligned and missing genotypes imputed according to the reference panel (Marchini, Howie 2010).

Screening for variants associated with the $FEV_1$ development during the follow-up was done on residuals adjusted for possible confounding variables and population structure by using multidimensional scaling (Purcell, Neale et al. 2007) to avoid spurious associations. The residuals were transformed to ranks and subsequently to Z-scores for screening in SNPTEST (Marchini, Howie et al. 2007, Marchini, Howie 2010).

# 5 RESULTS

## 5.1 Poor HRQoL in COPD is associated with characteristic determinants depending on severity of disease

Compared to age and gender matched Finnish population the COPD patients in FinnCad-cohort were found to often suffer from major comorbidities previously known to be associated with COPD. The prevalence of the comorbidities was, however, lower compared to some previous reports in COPD cohorts.

Comorbidities were found to contribute to the generic HRQoL significantly as standardized regression estimates were found comparable for psychiatric conditions and $FEV_1$ while alcohol abuse, cardiovascular diseases, diabetes and arterial hypertension contributed as well. Respiratory specific AQ20 was found mostly affected by $FEV_1$ followed by psychiatric conditions, female gender, alcohol abuse and hypertension.

Risk factors for very low HRQoL in 15D were psychiatric conditions (Odds Ratio, OR = 4.7, $p < 0.001$), $FEV_1$ <40% of predicted (OR = 3.1, p = 0.03), alcohol abuse (OR = 2.3, p = 0.007) and diabetes (OR = 2.1, p = 0.03).

AQ20 was mostly affected by lung function as $FEV_1$ <40 % of predicted was associated with OR=5.2 (p=0.001), but also $FEV_1$ 40-64% (OR=2.5, p=0.05) had an effect. Alcohol abuse (OR=3.0, p=0.001), female gender (OR=2.1, p=0.004) and psychiatric conditions (OR=2.0, p=0.007) were also associated with poor airway specific HRQoL.

The risk factors identified for poor HRQoL were also associated with 5-year mortality while $FEV_1$ <40% was found to have a considerable effect (OR=6.9, p0.001). Alcohol abuse and cardiovascular diseases were found to contribute a well. For both instruments the overall HRQoL score was clinically and significantly lower at baseline in patients who died during the 5-year period.

## 5.2 Individual HRQoL trajectories are identifiable in Asthma and COPD

During the 5-year follow-up period most of the COPD patients (60–80% depending on the questionnaire used) showed decline in their HRQoL, whereas in asthma the majority (46–71%) presented no decline. Both asthma and COPD showed decline more often with the generic 15D.

With the use of the 15D, 6% of both asthma and COPD patients were shown to present significant decline. Quantified by Minimum Clinically Important Difference (MCID), the significant decliners reached MCID in two years in both asthma and COPD. In the residual

patients the mean time to MCID in asthma would take 75 years and 7 years in COPD. The number of valid measurements over time was suggested to be important as most of the significant decliners were identified with 4 (against 3) measurements over the 5-year follow-up (77–96% depending on diagnosis and instrument).
Significant decline could be predicted with the low baseline HRQoL alone while the optimal cut-points were associated with significant decline: OR = 4–6 depending on the instrument in asthma and OR = 5–11 in COPD respectively.

Clinical determinants for significant decline in asthma were obesity linked, such as hypertension, diabetes mellitus and gastroesophageal reflux disease. Decline in COPD was associated with increasing age and lung function level. Psychiatric conditions were associated with decline in both asthma and COPD.

## 5.3 Longitudinal $FEV_1$ presents variation in COPD, but individual trajectories are identifiable

Development of pre-bronchodilatory $FEV_1$ in COPD presented variation at the cohort level while positive trajectories were detected for 10% of the cohort population. Thus a great majority, 90% of the cohort population, presented decline of $FEV_1$ during the variable follow-up period (mean 5 years) with highly variable number of measurements for each patient (mean 8 measurements).

However, when the individual trajectories were assessed, we found that there was significant improvement in 2% and decline in 30% of the population. 68% of the population did not present any significant development at 95% probability level suggesting significant diversity in the development of $FEV_1$ in COPD. Significant decliners could not be identified accurately from the distribution of trajectories as there was considerable overlap in the development in decliners and non-decliners (Figure 1.).

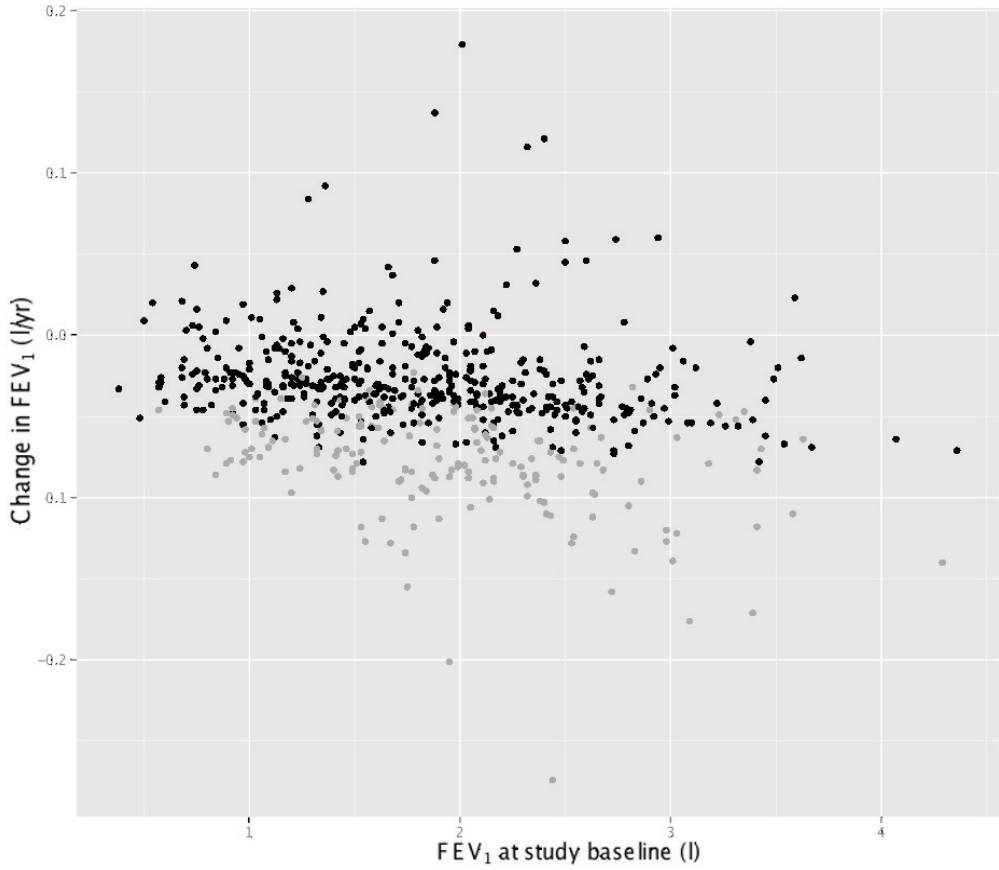


**Figure 1.** *Development of FEV1 in the cohort. Significant decliners (grey) and non-decliners (black) present considerable overlap in the development of FEV1.*

The uncertainty related to each trajectory is not observed unless the patients are assessed individually (Figure 2.). Two patients might have identical trajectories (a line fitted on measurements of $FEV_1$ in time), but the other might not present statistically significant decline and is thus not labelled as a significant decliner.

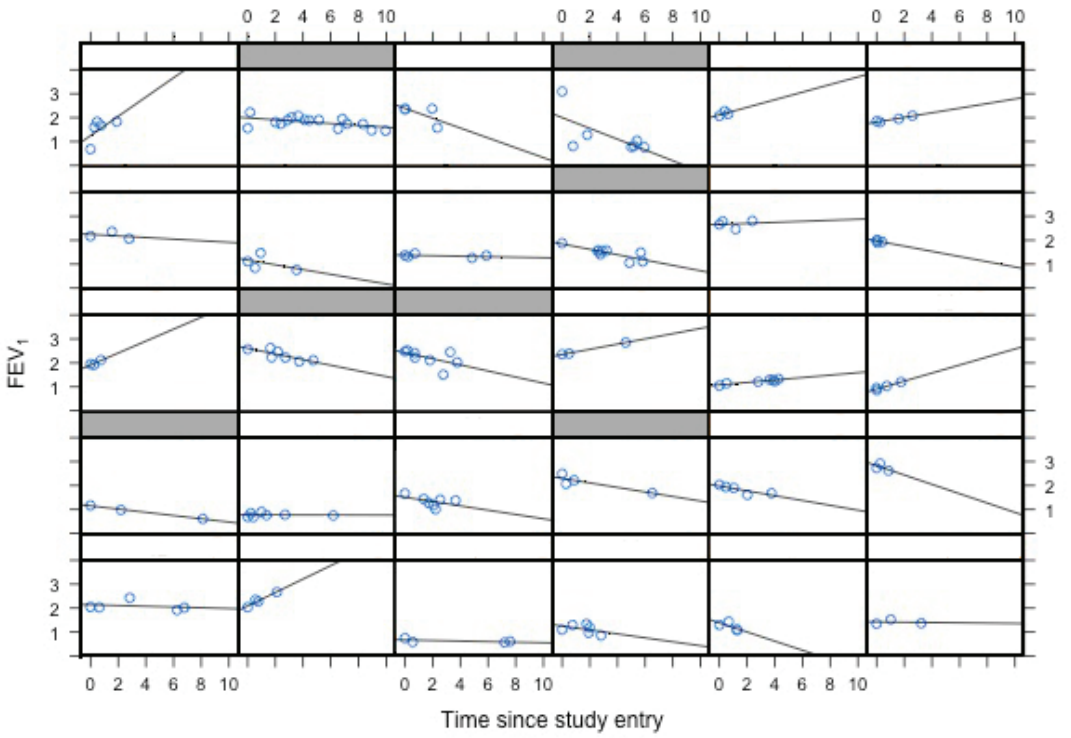


**Figure 2.** *Random sample of trajectories of $FEV_1$ (litres) over time. Significant decliners are indicated in grey.*

Significant decline was found associated with several disease exacerbation related markers such as emergency visits, hospital admissions, pneumonias and purchases of oral corticosteroids or antibiotics. Common comorbidities were not found associated with the decline while continuous smoking was found to be borderline significant.

## 5.4 Heritability and genetics of lung function trajectories

In the Project IV it was suggested that the variation in $FEV_1$ development in a randomly chosen population cohort is notably affected by genetic markers. The variation explained was estimated at 31% (Likelihood Ratio Test p-value = 0.07) for the level of $FEV_1$ at the start of the follow-up and 32% (p = 0.02) for the development of $FEV_1$.

Of the previously known loci for $FEV_1$ level or development or COPD status, only one locus was seen associated. Most significant SNP in the C10orf11 locus was rs7476758 ($p=8.2\times10^{-4}$). C10orf11 was not found to be notably expressed in human lung (GTEx Consortium 2015).

The GWA screening was performed to identify the loci associated with $FEV_1$ development as two suggestive associations were found. In 1p31.1 the most significant SNP was an intronic rs10874272 (p-value=$4.11\times10^{-7}$), which was in Linkage Disequilibrium in regard to neighbouring SNPs, and Minor Allele Frequency was found to be 19%. The SNPs are located on the LPHN2 gene, which was found to be the most notably expressed in human lung of all tissues with Reads Per Kilobase per Million mapped reads (RPKM) of 29.4 (GTEx Consortium 2015).

The other novel association for rs9599484 (p=$3.99\times10^{-7}$) was found in 13q13.3 with a MAF of 6% and in strong Linkage Disequilibrium (LD) to neighbouring SNPs. The nearest genes in the region were LINC00457 and NBEA, which were not found notably expressed in human lung (GTEx Consortium 2015).

# 6 DISCUSSION

In this thesis we have identified comorbidities with major effect on the Health Related Quality of Life in mild or moderate COPD as the comorbidities were found to play a minor role also in severe COPD. We identified significantly poor development of HRQoL in individual patients both in asthma and COPD during the 5-year follow-up when the major determinants of poor development were characteristic of asthma and COPD. We found that while the variation in the development of lung function in COPD is abundant, individual trajectories are identifiable. Significant decline in lung function was associated with numerous exacerbation related markers. In the unpublished Project IV results suggest that the decline of lung function in the Finnish population is significantly affected by genetic variation. A previously reported locus (c10orf11) for $FEV_1$ level was also found associated with the development of lung function. GWA screening suggests two novel loci (LPHN2 and LINC00457/NBEA) associated to lung function development.

## 6.1 HRQoL in mild or moderate COPD is greatly affected by comorbidities

Although the prevalence of comorbidities has been studied in COPD, their relative effect on HRQoL has not been assessed. In addition, HRQoL in COPD has not been widely studied and the role of comorbidities, when recognized, has remained under speculation (Curtis, Deyo et al. 1994, van Manen, Bindels et al. 2001, Burgel, Escamilla et al. 2013, Putcha, Puhan et al. 2013). Our results suggest that the common comorbidities play an integral part in mild or moderate COPD whereas in severe COPD the HRQoL is mostly affected by poor lung function. The number of comorbidities was not found significant, a congruent finding with previous findings (Putcha, Puhan et al. 2013). The model built to describe poor HRQoL also showed predictive ability of mortality during the 5-year follow-up.

HRQoL is often a secondary endpoint in studies with multiple exclusion criteria (Calverley, Anderson et al. 2007, Decramer, Celli et al. 2009), and thus a considerable number of patients are excluded from the analysis for justified reasons. This affects the prevalence of comorbidities in trials as some are deliberately excluded to improve the internal validity of a trial. Also, as a diagnosis found in medical records was required in contrast to self-reported comorbidity, the prevalence of comorbidities was lower compared to some previous studies. However, the external validity suffers from the extent of exclusion criteria as in random population of COPD patients only 5% were found to fulfill the inclusion criteria for the major Randomized Controlled Trials (RCTs) published in COPD (Travers, Marsh et al. 2007).

## 6.2 HRQoL in asthma and COPD during the 5-year follow-up

To our knowledge, individual trajectories of HRQoL have not been previously assessed in clinical pulmonology and the first findings of patient-level trajectories are presented in the Project II. Our results suggest that asthmatic patients often avoid the decline of HRQol especially when generic HRQoL is assessed. COPD patients, however, often presented progressive loss of HRQoL and also reached the Minimum Clinically Important Difference in a shorter period of time compared to asthmatics. Significant individual decline was seen for both asthma and COPD using both generic and airway specific instruments. Clinical determinants of decline in asthma were obesity related diseases and states while decline in COPD was affected by lung function and increasing age. Psychiatric conditions were suggested to be an important determinant for both asthma and COPD.

## 6.3 Individuals with rapid decline of lung functions

The development of lung function has only recently been assessed more thoroughly (Nishimura, Makita et al. 2012, Vestbo, Edwards et al. 2011, Casanova, de Torres et al. 2011, Tantucci, Modina 2012, Tashkin 2013, Tashkin, Li et al. 2013) and individual (patient-level) trajectories are still poorly known (Casanova, de Torres et al. 2011, Tashkin 2013). The use of retrospective and un-balanced data was shown to produce results comparable to previous prospective studies when the cohort-level development was assessed. We were able to identify a notable fraction of patients (30% of cohort) as significant decliners when compared to the only previous report (Casanova, de Torres et al. 2011). Previous studies have found the prevalence of improving trajectories unexpectedly common considering the progressive nature of COPD up to 15% of cohort population (Vestbo, Edwards et al. 2011). Our results underline the importance of assessing the individual trajectories and the uncertainty related to them as significantly improving trends were found in only 2% of the cohort population.

Significant decline was associated with exacerbation related markers further strengthening the conception of overlap of the frequent exacerbator and rapid decliner phenotype (Anzueto 2010).

As the cohort represented hospital based COPD patients, even the patients with the most severe disease were not lost from follow-up, which is common in prospective studies and might have a notable effect on inferences of randomized controlled trials (Akl, Briel et al. 2012).

As spirometry is a non-invasive and often used examination in clinical medicine, analysis of lung function trajectories might be used as an endpoint in genetic analysis.

## 6.4 Heritability and genetic susceptibility of lung function development

So far, only estimates from familial or twin studies exist to assess the amount of genetic markers explaining the phenotypic variation of lung function development (Hukkinen, Kaprio et al. 2011, Gottlieb, Wilk et al. 2001). This study suggests the first SNP-based estimate for heritability of $FEV_1$ development, which was found substantial and significantly different from zero. This estimate is also the first to assess the $FEV_1$ level heritability in the general population as previous SNP-based estimate is based on cohort of heavy smokers.

As the level of lung function, or as in this study, especially the development of lung function is of interest, it should be noted that estimates also include remnants of determinants essential in developing lung function in utero and during childhood. More studies are needed to clarify age-specific effects also in adulthood.

Of the previously known loci associated with the $FEV_1$ level, development of $FEV_1$ or COPD disease status, only one locus was found associated with the development of $FEV_1$ in the present study. This suggests that different loci are causal to lung function level and development. Our results suggest novel association in two loci: LPHN2 (p-value $4.11\times10^{-7}$) and LINC00457/NBEA ($p=3.99\times10^{-7}$). LPHN2 is mostly expressed in human lung (GTEx Consortium 2015) and has been previously found mutated in non-small cell lung carcinomas (NSCLC) (Zheng et al. 2013).

# 7 CONCLUSIONS

Asthma and COPD are major causes of morbidity, and according to the World Health Organization COPD will become the third leading cause of death worldwide by 2030 (World Health Organization). In Finland, the prevalence of COPD in males aged 65–74 years has been found to be 13% (Vasankari, Impivaara et al. 2010) as even 50% of smokers are estimated to develop COPD (Lundback, Lindberg et al. 2003). COPD is considered a progressive disease leading to disabilities unless smoking is discontinued and thorough medical and lifestyle changes are introduced (Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2015). So far though, the progression of lung function obstruction has not been assessed extensively, and the amount of heritable markers affecting the development is poorly known. This study provides perspectives on the development of HRQoL and $FEV_1$ to better understand the entities of lung function development and significant individual trajectories.

The results of this study suggest that Health Related Quality of Life (HRQoL) is notably affected by common comorbidities in COPD. Obesity-related comorbidities and characteristics were found significantly associated with the development of HRQoL in asthma, whereas in COPD, the main determinants were increasing age and the severity of lung function obstruction. Severe psychiatric conditions were seen as determinants of development of HRQoL both in asthma and COPD.

The development of lung function was seen variable between individuals and significant individual trajectories were identified. The development of lung function was shown to be substantially affected by genetic markers and should continue to be analysed in the future.

This study is mostly based on Electronic Health Records (EHRs) and thus is more prone to bias compared to prospective studies as the use of retrospective data needs to be shown valid for decision-making in medical research. However, the use of EHRs could prove fruitful as it offers considerably lower costs and does not require a period of time after study initiation before the analysis can be performed as the data is already available. In future, clinical data will be increasingly released from hospitals to researchers, but the potential can only be utilized if the data can be treated correctly.

Nowadays, medicine is designed for the "average patient", and not enough attention is paid to individuals presenting deviation from the average. The trajectories in clinical medicine are determined at the group level to account for random error related to individual trajectories. As moving towards P4 medicine (preventive, predictive, personalized and participatory medicine), assessing individual trajectories has become more momentous (Hood 2013). However, it is challenging as biological processes/physiological measurements include

random variation due to measurement errors and natural (e.g. daily) oscillation in the state. The within-subject variation needs to be assessed as trajectories present between-subject variation, which could be taken advantage of in future studies. Significant trajectories could be accentuated in the analysis while insignificant trajectories would be given less weight. Thus, the analysis would not be based on the distribution of trajectories but on the combination of trajectories and their credibility.

In clinical medicine, the level of a parameter or measurement has the highest value at the time of diagnosis, after which the future values are compared to the first ones. This underlines the importance of the trajectories as development within a parameter is always relevant and calls for their assessment in the future. As the level of $FEV_1$ always contains some of the decline after reaching its full volume in early adulthood, it would be necessary to assess whether the same determinants have an effect on the decline of $FEV_1$ throughout the human lifespan. Diagnosis of COPD is made cross-sectionally, which has been shown to lead to overestimation of COPD in the elderly and underdiagnosis in the young (Hardie, Buist et al. 2002, Swanney, Ruppel et al. 2008, Cerveri, Corsico et al. 2008). The diagnosis could be more accurate when based on the development of lung function in uncertain cases. This study suggests that rapid decliners are identifiable from real-world clinical data, which enables the targeting of medical interventions effectively to those in risk of disability and mortality. The decline of $FEV_1$, especially if more thoroughly investigated for age and period effects, could prove useful as an endpoint in genetic analysis. As typical effect sizes (Odds Ratio) in case-control GWA studies are in the range of 1.1–1.3, the accuracy of diagnosis is of critical importance as 70% sensitivity and specificity can deflate ORs of 1.3 due to non-differential misclassification (Szklo, Nieto 2014b). Therefore, the phenotypes used in genetic analysis should be relevant and well-defined. The phenotypes in question could represent clinically relevant traits not necessarily corresponding to the classical diagnosis of e.g. COPD.

# ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Tarja Laitinen for introducing me to the field of medical science. You have enabled me exceptional views over this field. Thank you for your never-ending ideas and for your confidence in me.

I am also grateful to late Prof. Vuokko Kinnula for her energy and support, and for contacting me with Prof. Tarja Laitinen.

I am very grateful to Prof. Martin Tobin for accepting the invitation to be my opponent.

Also, I would like to thank Docent Janne Pitkäniemi for encouraging me to open-mindedly study and use biostatistics.

I am deeply grateful to Prof. Samuli Ripatti and Dr. Ida Surakka for introducing me to genetics and welcoming me to Institute for Molecular Medicine Finland (FIMM).

I would like to thank my thesis revisers Docent Laura Elo and Docent Terttu Harju for your constructive comments and thorough evaluation of my thesis.

This work would have not been possible without support of the Clinical Research Unit for Pulmonary Diseases and Dr. Ari Lindqvist. Thank you Kirsi Sariola, Kerstin Ahlskog, Sari Nummijoki and Tinja Kanerva for your endless support, care and coffee. I also thank my collaborators Henna Kupiainen, Milla Katajisto, Dr. Witold Mazur, Prof. Harri Sintonen, Dr. Maritta Kilpeläinen and Aleksi Kallio.

This study was financially supported by The Organisation for Respiratory Health in Finland, Ida Montin Foundation, Väinö and Laina Kivi Foundation and the University of Helsinki Funds.

My sister Outi is thanked for language editing of my thesis, and my mom Päivi for always offering to help. Thank you, my little Vuokko and Valma, for always giving me a reason to stop working and for all the joy in my life. Lastly, I would like to thank my wife Sanna, for working so hard for my happiness. You make every day feel like a celebration.

*Jukka Koskela*
*Helsinki, November 2015*

# REFERENCES


Akl, E.A., Briel, M., You, J.J., Sun, X., Johnston, B.C., Busse, J.W., Mulla, S., Lamontagne, F., Bassler, D., Vera, C., Alshurafa, M., Katsios, C.M., Zhou, Q., Cukierman-Yaffe, T., Gangji, A., Mills, E.J., Walter, S.D., Cook, D.J., Schunemann, H.J., Altman, D.G. and Guyatt, G.H., 2012. Potential impact on estimated treatment effects of information lost to follow-up in randomised controlled trials (LOST-IT): systematic review. *BMJ (Clinical research ed.),* **344**, pp. e2809.

Albert, P., Agusti, A., Edwards, L., Tal-Singer, R., Yates, J., Bakke, P., Celli, B.R., Coxson, H.O., Crim, C., Lomas, D.A., Macnee, W., Miller, B., Rennard, S., Silverman, E.K., Vestbo, J., Wouters, E. and Calverley, P., 2012. Bronchodilator responsiveness as a phenotypic characteristic of established chronic obstructive pulmonary disease. *Thorax,* **67**(8), pp. 701-708.

Andersen, H., Lampela, P., Nevanlinna, A., Saynajakangas, O. and Keistinen, T., 2013. High hospital burden in overlap syndrome of asthma and COPD. *The clinical respiratory journal,* **7**(4), pp. 342-346.

Anzueto, A., 2010. Impact of exacerbations on COPD. *European respiratory review : an official journal of the European Respiratory Society,* **19**(116), pp. 113-118.

Anderson, G.P., 2008. Endotyping asthma: new insights into key pathogenic mechanisms in a complex, heterogeneous disease. *Lancet (London, England),* **372**(9643), pp. 1107-1119.

Baiardini, I., Braido, F., Brandi, S. and Canonica, G.W., 2006. Allergic diseases and their impact on quality of life. *Annals of Allergy, Asthma & Immunology : Official Publication of the American College of Allergy, Asthma, & Immunology,* **97**(4), pp. 419-28; quiz 429-30, 476.

Barley, E.A., Quirk, F.H. and Jones, P.W., 1998. Asthma health status measurement in clinical practice: validity of a new short and simple instrument. *Respiratory medicine,* **92**(10), pp. 1207-1214.

Barnes, P.J., 2010. Chronic obstructive pulmonary disease: effects beyond the lungs. *PLoS medicine,* **7**(3), pp. e1000220.

Barnes, P.J., 2014. Cellular and molecular mechanisms of chronic obstructive pulmonary disease. *Clinics in chest medicine,* **35**(1), pp. 71-86.

Bateman, E.D., Reddel, H.K., Van Zyl-Smit, R.N. and Agusti, A., The asthma–COPD overlap syndrome: towards a revised taxonomy of chronic airways diseases? *The Lancet Respiratory Medicine,* 2015, [Epub ahead of print].

Baughman, P., Marott, J.L., Lange, P., Andrew, M. and Hnizdo, E., 2011. Health outcomes associated with lung function decline and respiratory symptoms and disease in a community cohort. *Copd,* **8**(2), pp. 103-113.

Beeckman, L.A., Wang, M.L., Petsonk, E.L. and Wagner, G.R., 2001. Rapid declines in FEV1 and subsequent respiratory symptoms, illnesses, and mortality in coal miners in the United States. *American journal of respiratory and critical care medicine,* **163**(3 Pt 1), pp. 633-639.

Beuther, D.A. and Sutherland, E.R., 2007. Overweight, obesity, and incident asthma: a meta-analysis of prospective epidemiologic studies. *American journal of respiratory and critical care medicine,* **175**(7), pp. 661-666.

Braido, F., Bousquet, P.J., Brzoza, Z., Canonica, G.W., Compalati, E., Fiocchi, A., Fokkens, W., Gerth Van Wijk, R., La Grutta, S., Lombardi, C., Maurer, M., Pinto, A.M., Ridolo, E., Senna, G.E., Terreehorst, I., Todo Bom, A., Bousquet, J., Zuberbier, T. and Baiardini, I., 2010. Specific recommendations for PROs and HRQoL assessment in allergic rhinitis and/or asthma: a GA(2) LEN taskforce position paper. *Allergy,* **65**(8), pp. 959-968.

Burgel, P.R., Escamilla, R., Perez, T., Carre, P., Caillaud, D., Chanez, P., Pinet, C., Jebrak, G., Brinchault, G., Court-Fortune, I., Paillasseur, J.L., Roche, N. and Initiatives BPCO Scientific Committee, 2013. Impact of comorbidities on COPD-specific health-related quality of life. *Respiratory medicine,* **107**(2), pp. 233-241.

Burrows, B., 1990. Airways obstructive diseases: pathogenetic mechanisms and natural histories of the disorders. *The Medical clinics of North America,* **74**(3), pp. 547-559.

Burrows, B., Knudson, R.J., Cline, M.G. and Lebowitz, M.D., 1977. Quantitative relationships between cigarette smoking and ventilatory function. *The American Review of Respiratory Disease,* **115**(2), pp. 195-205.

Calverley, P.M., Anderson, J.A., Celli, B., Ferguson, G.T., Jenkins, C., Jones, P.W., Yates, J.C., Vestbo, J. and TORCH investigators, 2007. Salmeterol and fluticasone propionate and survival in chronic obstructive pulmonary disease. *The New England journal of medicine,* **356**(8), pp. 775-789.

Casanova, C., De Torres, J.P., Aguirre-Jaime, A., Pinto-Plata, V., Marin, J.M., Cordoba, E., Baz, R., Cote, C. and Celli, B.R., 2011. The progression of chronic obstructive pulmonary disease is heterogeneous: the experience of the BODE cohort. *American journal of respiratory and critical care medicine,* **184**(9), pp. 1015-1021.

Cerveri, I., Corsico, A.G., Accordini, S., Niniano, R., Ansaldo, E., Anto, J.M., Kunzli, N., Janson, C., Sunyer, J., Jarvis, D., Svanes, C., Gislason, T., Heinrich, J., Schouten, J.P., Wjst, M., Burney, P. and De Marco, R., 2008. Underestimation of airflow obstruction among young adults using FEV1/FVC <70% as a fixed cut-off: a longitudinal evaluation of clinical and functional outcomes. *Thorax,* **63**(12), pp. 1040-1045.

Cho, M.H., Boutaoui, N., Klanderman, B.J., Sylvia, J.S., Ziniti, J.P., Hersh, C.P., Demeo, D.L., Hunninghake, G.M., Litonjua, A.A., Sparrow, D., Lange, C., Won, S., Murphy, J.R., Beaty, T.H., Regan, E.A., Make, B.J., Hokanson, J.E., Crapo, J.D., Kong, X., Anderson, W.H., Tal-Singer, R., Lomas, D.A., Bakke, P., Gulsvik, A., Pillai, S.G. and Silverman, E.K., 2010. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nature genetics,* **42**(3), pp. 200-202.

Cho, M.H., Castaldi, P.J., Wan, E.S., Siedlinski, M., Hersh, C.P., Demeo, D.L., Himes, B.E., Sylvia, J.S., Klanderman, B.J., Ziniti, J.P., Lange, C., Litonjua, A.A., Sparrow, D., Regan, E.A., Make, B.J., Hokanson, J.E., Murray, T., Hetmanski, J.B., Pillai, S.G., Kong, X., Anderson, W.H., Tal-Singer, R., Lomas, D.A., Coxson, H.O., Edwards, L.D., Macnee, W., Vestbo, J., Yates, J.C., Agusti, A., Calverley, P.M., Celli, B., Crim, C., Rennard, S., Wouters, E., Bakke, P., Gulsvik, A., Crapo, J.D., Beaty, T.H., Silverman, E.K., ICGN investigators, ECLIPSE investigators and COPDGene investigators, 2012. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Human molecular genetics,* **21**(4), pp. 947-957.

Cho, M.H., Mcdonald, M.L., Zhou, X., Mattheisen, M., Castaldi, P.J., Hersh, C.P., Demeo, D.L., Sylvia, J.S., Ziniti, J., Laird, N.M., Lange, C., Litonjua, A.A., Sparrow, D., Casaburi, R., Barr, R.G., Regan, E.A., Make, B.J., Hokanson, J.E., Lutz, S., Dudenkov, T.M., Farzadegan, H., Hetmanski, J.B., Tal-Singer, R., Lomas, D.A., Bakke, P., Gulsvik, A., Crapo, J.D., Silverman, E.K., Beaty, T.H. and NETT GENETICS, ICGN, ECLIPSE and COPDGene Investigators, 2014. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *The Lancet.Respiratory medicine,* **2**(3), pp. 214-225.

Choi, I.Y., Kim, T.M., Kim, M.S., Mun, S.K. and CHUNG, Y.J., 2013. Perspectives on clinical informatics: integrating large-scale clinical, genomic, and health information for clinical care. *Genomics & informatics,* **11**(4), pp. 186-190.

Cohen, B.H., 1980. Chronic obstructive pulmonary disease: a challenge in genetic epidemiology. *American Journal of Epidemiology,* **112**(2), pp. 274-288.

Corsonello, A., Antonelli Incalzi, R., Pistelli, R., Pedone, C., Bustacchini, S. and Lattanzio, F., 2011. Comorbidities of chronic obstructive pulmonary disease. *Current opinion in pulmonary medicine,* **17 Suppl 1**, pp. S21-8.

Curtis, J.R., Deyo, R.A. and Hudson, L.D., 1994. Pulmonary rehabilitation in chronic respiratory insufficiency. 7. Health-related quality of life among patients with chronic obstructive pulmonary disease. *Thorax,* **49**(2), pp. 162-170.

Dawber, T.R., Meadors, G.F. and Moore, F.E.,Jr, 1951. Epidemiological approaches to heart disease: the Framingham Study. *American Journal of Public Health and the Nation's Health,* **41**(3), pp. 279-281.

Decramer, M., Celli, B., Kesten, S., Lystig, T., Mehra, S., Tashkin, D.P. and UPLIFT Investigators, 2009. Effect of tiotropium on outcomes in patients with moderate chronic obstructive pulmonary disease (UPLIFT): a prespecified subgroup analysis of a randomised controlled trial. *Lancet,* **374**(9696), pp. 1171-1178.

Delgado-Rodriguez, M. and Llorca, J., 2004. Bias. *Journal of epidemiology and community health,* **58**(8), pp. 635-641.

Dornhorst, A.C., 1955. Respiratory insufficiency. *Lancet,* **268**(6876), pp. 1185-1187.

Dos Santos Silva, I., 1999. Cancer Epidemiology: Principles and Methods. *International Agency for Research on Cancer*, pp. 277-295.

Fletcher, C. and Peto, R., 1977. The natural history of chronic airflow obstruction. *British medical journal,* **1**(6077), pp. 1645-1648.

Friedlander, A.L., Lynch, D., Dyar, L.A. and Bowler, R.P., 2007. Phenotypes of chronic obstructive pulmonary disease. *Copd,* **4**(4), pp. 355-384.

World Health Organization., 2000-last update, World Health Report. Geneva: World Health Organization. Available from http://www.who.int/whr/2000/en/ statistics.htm.

Gelman, A. and Hill, J., 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Global strategy for asthma management and prevention. GINA. http://www.ginasthma.org (2014)

Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease, GOLD. http://www.goldcopd.org/uploads/users/files/GOLD_Report_2015.pdf (2015)

Gottlieb, D.J., Wilk, J.B., Harmon, M., Evans, J.C., Joost, O., Levy, D., O'connor, G.T. and Myers, R.H., 2001. Heritability of longitudinal change in lung function. The Framingham study. *American journal of respiratory and critical care medicine,* **164**(9), pp. 1655-1659.

GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (New York, N.Y.),* **348**(6235), pp. 648-660.

Hajiro, T., Nishimura, K., Jones, P.W., Tsukino, M., Ikeda, A., Koyama, H. and Izumi, T., 1999. A novel, short, and simple questionnaire to measure health-related quality of life in patients with chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine,* **159**(6), pp. 1874-1878.

Han, M.K., Agusti, A., Calverley, P.M., Celli, B.R., Criner, G., Curtis, J.L., Fabbri, L.M., Goldin, J.G., Jones, P.W., Macnee, W., Make, B.J., Rabe, K.F., Rennard, S.I., Sciurba, F.C., Silverman, E.K., Vestbo, J., Washko, G.R., Wouters, E.F. and Martinez, F.J., 2010. Chronic obstructive pulmonary disease phenotypes: the future of COPD. *American journal of respiratory and critical care medicine,* **182**(5), pp. 598-604.

Hancock, D.B., Eijgelsheim, M., Wilk, J.B., Gharib, S.A., Loehr, L.R., Marciante, K.D., Franceschini, N., Van Durme, Y.M., Chen, T.H., Barr, R.G., Schabath, M.B., Couper, D.J., Brusselle, G.G., Psaty, B.M., Van Duijn, C.M., Rotter, J.I., Uitterlinden, A.G., Hofman, A., Punjabi, N.M., Rivadeneira, F., Morrison, A.C., Enright, P.L., North, K.E., Heckbert, S.R., Lumley, T., Stricker, B.H., O'connor, G.T. and London, S.J., 2010. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nature genetics,* **42**(1), pp. 45-52.

Hardie, J.A., Buist, A.S., Vollmer, W.M., Ellingsen, I., Bakke, P.S. and Morkve, O., 2002. Risk of over-diagnosis of COPD in asymptomatic elderly never-smokers. *The European respiratory journal,* **20**(5), pp. 1117-1122.

Heistaro, S., 2008. *METHODOLOGY REPORT Health 2000 Survey* B : 26/2008. Helsinki, Finland: Kansanterveyslaitos.

Hesselink, A.E., Van Der Windt, D.A., Penninx, B.W., Wijnhoven, H.A., Twisk, J.W., Bouter, L.M. and Van Eijk, J.T., 2006. What predicts change in pulmonary function and quality of life in asthma or COPD? *The Journal of asthma : official journal of the Association for the Care of Asthma,* **43**(7), pp. 513-519.

Hoeting, J.A., 1999. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statist.Sci.14, no.4 (1999), 382-417,* **14**(4), pp. 382-417.

Hogg JC. Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *Lancet*. 2004; 364:709-721.

Hood, L., 2013. Systems biology and p4 medicine: past, present, and future. *Rambam Maimonides medical journal,* **4**(2), pp. e0012.

Hripcsak, G. and Albers, D.J., 2013. Next-generation phenotyping of electronic health records. *Journal of the American Medical Informatics Association : JAMIA,* **20**(1), pp. 117-121.

Hukkinen, M., Kaprio, J., Broms, U., Viljanen, A., Kotz, D., Rantanen, T. and Korhonen, T., 2011. Heritability of lung function: a twin study among never-smoking elderly women. *Twin research and human genetics : the official journal of the International Society for Twin Studies,* **14**(5), pp. 401-407.

Imboden, M., Bouzigon, E., Curjuric, I., Ramasamy, A., Kumar, A., Hancock, D.B., Wilk, J.B., Vonk, J.M., Thun, G.A., Siroux, V., Nadif, R., Monier, F., Gonzalez, J.R., Wjst, M., Heinrich, J., Loehr, L.R., Franceschini, N., North, K.E., Altmuller, J., Koppelman, G.H., Guerra, S., Kronenberg, F., Lathrop, M., Moffatt, M.F., O'connor, G.T., Strachan, D.P., Postma, D.S., London, S.J., Schindler, C., Kogevinas, M., Kauffmann, F., Jarvis, D.L., Demenais, F. and Probst-Hensch, N.M., 2012. Genome-wide association study of lung function decline in adults with and without asthma. *The Journal of allergy and clinical immunology,* **129**(5), pp. 1218-1228.

Ingebrigtsen, T.S., Thomsen, S.F., Van Der Sluis, S., Miller, M., Christensen, K., Sigsgaard, T. and Backer, V., 2011. Genetic influences on pulmonary function: a large sample twin study. *Lung,* **189**(4), pp. 323-330.

Jensen, P.B., Jensen, L.J. and Brunak, S., 2012. Mining electronic health records: towards better research applications and clinical care. *Nature reviews.Genetics,* **13**(6), pp. 395-405.

Jones, P.W., 1992. A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. *The American Review of Respiratory Disease,* **145**(6), pp. 1321-1327.

Kauppi, P., Kupiainen, H., Lindqvist, A., Tammilehto, L., Kilpelainen, M., Kinnula, V.L., Haahtela, T. and Laitinen, T., 2011. Overlap syndrome of asthma and COPD predicts low quality of life. *The Journal of asthma : official journal of the Association for the Care of Asthma,* **48**(3), pp. 279-285.

Kauppinen, R., Sintonen, H., Vilkka, V., Pekurinen, M. and Tukiainen, H., 1998. Quality-of-life measures and clinical parameters in asthmatics during three year follow-up. *Monaldi archives for chest disease = Archivio Monaldi per le malattie del torace / Fondazione clinica del lavoro, IRCCS [and] Istituto di clinica tisiologica e malattie apparato respiratorio, Universita di Napoli, Secondo ateneo,* **53**(4), pp. 400-404.

Klimentidis, Y.C., Vazquez, A.I., De Los Campos, G., Allison, D.B., Dransfield, M.T. and Thannickal, V.J., 2013. Heritability of pulmonary function estimated from pedigree and whole-genome markers. *Frontiers in genetics,* **4**, pp. 174.

Kohane, I.S., 2011. Using electronic health records to drive discovery in disease genomics. *Nature reviews.Genetics,* **12**(6), pp. 417-428.

Laitinen, T., Hodgson, U., Kupiainen, H., Tammilehto, L., Haahtela, T., Kilpelainen, M., Lindqvist, A. and Kinnula, V.L., 2009. Real-world clinical data identifies gender-related profiles in chronic obstructive pulmonary disease. *Copd,* **6**(4), pp. 256-262.

Laurell, C.B. and Eriksson, S., 1963. The electrophoretic alpha1-globulin pattern of serum in alpha1-antitrypsin deficiency. *Scandinavian Journal of Clinical and Laboratory Investigation*, **15** (2), pp. 132–140.

Lee, J.H., Haselkorn, T., Borish, L., Rasouliyan, L., Chipps, B.E. and Wenzel, S.E., 2007. Risk factors associated with persistent airflow limitation in severe or difficult-to-treat asthma: insights from the TENOR study. *Chest,* **132**(6), pp. 1882-1889.

Leidy, N.K. and Coughlin, C., 1998. Psychometric performance of the Asthma Quality of Life Questionnaire in a US sample. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation,* **7**(2), pp. 127-134.

Lundback, B., Lindberg, A., Lindstrom, M., Ronmark, E., Jonsson, A.C., Jonsson, E., Larsson, L.G., Andersson, S., Sandstrom, T., Larsson, K. and Obstructive Lung Disease in Northern Sweden Studies, 2003. Not 15 but 50% of smokers develop COPD?--Report from the Obstructive Lung Disease in Northern Sweden Studies. *Respiratory medicine,* **97**(2), pp. 115-122.

Maddox, L. and Schwartz, D.A., 2002. The pathophysiology of asthma. *Annual Review of Medicine,* **53**, pp. 477-498.

Mannino, D.M., Watt, G., Hole, D., Gillis, C., Hart, C., Mcconnachie, A., Davey Smith, G., Upton, M., Hawthorne, V., Sin, D.D., Man, S.F., Van Eeden, S., Mapel, D.W. and Vestbo, J., 2006. The natural history of chronic obstructive pulmonary disease. *The European respiratory journal,* **27**(3), pp. 627-643.

Marchini, J. and Howie, B., 2010. Genotype imputation for genome-wide association studies. *Nature reviews.Genetics,* **11**(7), pp. 499-511.

Marchini, J., Howie, B., Myers, S., Mcvean, G. and Donnelly, P., 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics,* **39**(7), pp. 906-913.

Marks, G.B., Dunn, S.M. and Woolcock, A.J., 1993. An evaluation of an asthma quality of life questionnaire as a measure of change in adults with asthma. *Journal of clinical epidemiology,* **46**(10), pp. 1103-1111.

Martin, A.D., Quinn, K.M. and Park, J.H., 2011. MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software,* **42**(9), pp. 1-21.

Mazur, W., Kupiainen, H., Pitkaniemi, J., Kilpelainen, M., Sintonen, H., Lindqvist, A., Kinnula, V.L. and Laitinen, T., 2011. Comparison between the disease-specific Airways Questionnaire 20 and the generic 15D instruments in COPD. *Health and quality of life outcomes,* **9**, pp. 4.

Natural History of Disease. 2008. In: M. PORTA, ed, *A Dictionary of Epidemiology.* 5 edn. Oxford University Press, pp. 193-194.

Nielsen, M., Barnes, C.B. and Ulrik, C.S., 2015. Clinical characteristics of the asthma-COPD overlap syndrome--a systematic review. *International journal of chronic obstructive pulmonary disease,* **10**, pp. 1443-1454.

Nishimura, M., Makita, H., Nagai, K., Konno, S., Nasuhara, Y., Hasegawa, M., Shimizu, K., Betsuyaku, T., Ito, Y.M., Fuke, S., Igarashi, T., Akiyama, Y., Ogura, S. and Hokkaido Copd Cohort Study Investigators, 2012. Annual change in pulmonary function and clinical phenotype in chronic obstructive pulmonary disease. *American journal of respiratory and critical care medicine,* **185**(1), pp. 44-52.

Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *European Respiratory Journal*. 2005; **26**(5), pp. 948-968.

Pillai, S.G., Ge, D., Zhu, G., Kong, X., Shianna, K.V., Need, A.C., Feng, S., Hersh, C.P., Bakke, P., Gulsvik, A., Ruppert, A., Lodrup Carlsen, K.C., Roses, A., Anderson, W., Rennard, S.I., Lomas, D.A., Silverman, E.K., Goldstein, D.B. and ICGN Investigators, 2009. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS genetics,* **5**(3), pp. e1000421.

Pleasants, R.A., Ohar, J.A., Croft, J.B., Liu, Y., Kraft, M., Mannino, D.M., Donohue, J.F. and Herrick, H.L., 2014. Chronic obstructive pulmonary disease and asthma-patient characteristics and health impairment. *Copd,* **11**(3), pp. 256-266.

Puhan, M.A., Guyatt, G.H., Goldstein, R., Mador, J., Mckim, D., Stahl, E., Griffith, L. and Schunemann, H.J., 2007. Relative responsiveness of the Chronic Respiratory Questionnaire, St. Georges Respiratory Questionnaire and four other health-related quality of life instruments for patients with chronic lung disease. *Respiratory medicine,* **101**(2), pp. 308-316.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* **81**(3), pp. 559-575.

Putcha, N., Puhan, M.A., Hansel, N.N., Drummond, M.B. and Boyd, C.M., 2013. Impact of co-morbidities on self-rated health in self-reported COPD: an analysis of NHANES 2001-2008. *Copd,* **10**(3), pp. 324-332.

Qureshi, H., Sharafkhaneh, A. and Hanania, N.A., 2014. Chronic obstructive pulmonary disease exacerbations: latest evidence and clinical implications. *Therapeutic advances in chronic disease,* **5**(5), pp. 212-227.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rennard, S.I. and Vestbo, J., 2008. Natural histories of chronic obstructive pulmonary disease. *Proceedings of the American Thoracic Society,* **5**(9), pp. 878-883.

Repapi, E., Sayers, I., Wain, L.V., Burton, P.R., Johnson, T., Obeidat, M., Zhao, J.H., Ramasamy, A., Zhai, G., Vitart, V., Huffman, J.E., Igl, W., Albrecht, E., Deloukas, P., Henderson, J., Granell, R., Mcardle, W.L., Rudnicka, A.R., Wellcome Trust Case Control Consortium, Barroso, I., Loos, R.J., Wareham, N.J., Mustelin, L., Rantanen, T., Surakka, I., Imboden, M., Wichmann, H.E., Grkovic, I., Jankovic, S., Zgaga, L., Hartikainen, A.L., Peltonen, L., Gyllensten, U., Johansson, A., Zaboli, G., Campbell, H., Wild, S.H., Wilson, J.F., Glaser, S., Homuth, G., Volzke, H., Mangino, M., Soranzo, N., Spector, T.D., Polasek, O., Rudan, I., Wright, A.F., Heliovaara, M., Ripatti, S., Pouta, A., Naluai, A.T., Olin, A.C., Toren, K., Cooper, M.N., James, A.L., Palmer, L.J., Hingorani, A.D., Wannamethee, S.G., Whincup, P.H., Smith, G.D., Ebrahim, S., Mckeever, T.M., Pavord, I.D., Macleod, A.K., Morris, A.D., Porteous, D.J., Cooper, C., Dennison, E., Shaheen, S., Karrasch, S., Schnabel, E., Schulz, H., Grallert, H., Bouatia-Naji, N., Delplanque, J., Froguel, P., Blakey, J.D., Nshd Respiratory Study Team, Britton, J.R., Morris, R.W., Holloway, J.W., Lawlor, D.A., Hui, J., Nyberg, F., Jarvelin, M.R., Jackson,

C., Kahonen, M., Kaprio, J., Probst-Hensch, N.M., Koch, B., Hayward, C., Evans, D.M., Elliott, P., Strachan, D.P., Hall, I.P. and Tobin, M.D., 2010. Genome-wide association study identifies five loci associated with lung function. *Nature genetics,* **42**(1), pp. 36-44.

Robinson, G.K., 1991. That BLUP is a Good Thing: The Estimation of Random Effects. , pp. 15-32.

Ryan, G., Knuiman, M.W., Divitini, M.L., James, A., Musk, A.W. and Bartholomew, H.C., 1999. Decline in lung function and mortality: the Busselton Health Study. *Journal of epidemiology and community health,* **53**(4), pp. 230-234.

Saarni, S.I., Harkanen, T., Sintonen, H., Suvisaari, J., Koskinen, S., Aromaa, A. and Lonnqvist, J., 2006. The impact of 29 chronic conditions on health-related quality of life: a general population survey in Finland using 15D and EQ-5D. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation,* **15**(8), pp. 1403-1414.

Salvi SS, Barnes PJ. 2009. Chronic obstructive pulmonary disease in non-smokers. *The Lancet*. **374**(9691), pp. 733-743.

Silva, G.E., Sherrill, D.L., Guerra, S. and Barbee, R.A., 2004. Asthma as a risk factor for COPD in a longitudinal study. *Chest,* **126**(1), pp. 59-65.

Sintonen, H., 1994. The 15D-measure of health-related quality of life. I. Reliability, validity and sensitivity of its health state descriptive system. **Working Paper 41**.

Sequencing Initiative Suomi, http://www.sisuproject.fi.

Soler Artigas, M., Loth, D.W., Wain, L.V., Gharib, S.A., Obeidat, M., Tang, W., Zhai, G., Zhao, J.H., Smith, A.V., Huffman, J.E., Albrecht, E., Jackson, C.M., Evans, D.M., Cadby, G., Fornage, M., Manichaikul, A., Lopez, L.M., Johnson, T., Aldrich, M.C., Aspelund, T., Barroso, I., Campbell, H., Cassano, P.A., Couper, D.J., Eiriksdottir, G., Franceschini, N., Garcia, M., Gieger, C., Gislason, G.K., Grkovic, I., Hammond, C.J., Hancock, D.B., Harris, T.B., Ramasamy, A., Heckbert, S.R., Heliovaara, M., Homuth, G., Hysi, P.G., James, A.L., Jankovic, S., Joubert, B.R., Karrasch, S., Klopp, N., Koch, B., Kritchevsky, S.B., Launer, L.J., Liu, Y., Loehr, L.R., Lohman, K., Loos, R.J., Lumley, T., Al Balushi, K.A., Ang, W.Q., Barr, R.G., Beilby, J., Blakey, J.D., Boban, M., Boraska, V., Brisman, J., Britton, J.R., Brusselle, G.G., Cooper, C., Curjuric, I., Dahgam, S., Deary, I.J., Ebrahim, S., Eijgelsheim, M., Francks, C., Gaysina, D., Granell, R., Gu, X., Hankinson, J.L., Hardy, R., Harris, S.E., Henderson, J., Henry, A., Hingorani, A.D., Hofman, A., Holt, P.G., Hui, J., Hunter, M.L., Imboden, M., Jameson, K.A., Kerr, S.M., Kolcic, I., Kronenberg, F., Liu, J.Z., Marchini, J., Mckeever, T., Morris, A.D., Olin, A.C., Porteous, D.J., Postma, D.S., Rich, S.S., Ring, S.M., Rivadeneira, F., Rochat, T., Sayer, A.A., Sayers, I., Sly, P.D., Smith, G.D., Sood, A., Starr, J.M., Uitterlinden, A.G., Vonk, J.M., Wannamethee, S.G., Whincup, P.H., Wijmenga, C., Williams, O.D., Wong, A., Mangino, M., Marciante, K.D., Mcardle, W.L., Meibohm, B., Morrison, A.C., North, K.E., Omenaas, E., Palmer, L.J., Pietilainen, K.H., Pin, I., Pola Sbreve Ek, O., Pouta, A., Psaty, B.M., Hartikainen, A.L., Rantanen, T., Ripatti, S., Rotter, J.I., Rudan, I., Rudnicka, A.R., Schulz, H., Shin, S.Y., Spector, T.D., Surakka, I., Vitart, V., Volzke, H., Wareham, N.J., Warrington, N.M., Wichmann, H.E., Wild, S.H., Wilk, J.B., Wjst, M., Wright, A.F., Zgaga, L., Zemunik, T., Pennell, C.E., Nyberg, F., Kuh, D., Holloway, J.W., Boezen, H.M., Lawlor, D.A., Morris, R.W., Probst-Hensch, N., International Lung Cancer Consortium, Giant Consortium, Kaprio, J., Wilson, J.F., Hayward, C., Kahonen, M., Heinrich, J., Musk, A.W., Jarvis, D.L., Glaser, S., Jarvelin, M.R., Ch Stricker, B.H., Elliott, P., O'connor, G.T., Strachan, D.P., London, S.J., Hall, I.P., Gudnason, V. and Tobin, M.D., 2011. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nature genetics,* **43**(11), pp. 1082-1090.

Stavem, K., 1999. Reliability, validity and responsiveness of two multiattribute utility measures in patients with chronic obstructive pulmonary disease. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation,* **8**(1-2), pp. 45-54.

Swanney, M.P., Ruppel, G., Enright, P.L., Pedersen, O.F., Crapo, R.O., Miller, M.R., Jensen, R.L., Falaschetti, E., Schouten, J.P., Hankinson, J.L., Stocks, J. and Quanjer, P.H., 2008. Using the lower limit of normal for the FEV1/FVC ratio reduces the misclassification of airway obstruction. *Thorax,* **63**(12), pp. 1046-1051.

Szklo, M. and Nieto, F.J., 2014a. Analysis of Age, Birth Cohort and Period Effects. *Epidemiology: Beyond the Basics.* 3 edn. Maryland, United States of America: Aspen Publishers, Inc, Gaithersburg, pp. 4-14.

Szklo, M. and Nieto, F.J., 2014b. Understanding Lack of Validity: Bias. *Epidemiology: Beyond the Basics.* 3 edn. Maryland, United States of America: Aspen Publishers, Inc, Gaithersburg, pp. 131-132.

Tang, W., Kowgier, M., Loth, D.W., Soler Artigas, M., Joubert, B.R., Hodge, E., Gharib, S.A., Smith, A.V., Ruczinski, I., Gudnason, V., Mathias, R.A., Harris, T.B., Hansel, N.N., Launer, L.J., Barnes, K.C., Hansen, J.G., Albrecht, E., Aldrich, M.C., Allerhand, M., Barr, R.G., Brusselle, G.G., Couper, D.J., Curjuric, I., Davies, G., Deary, I.J., Dupuis, J., Fall, T., Foy, M., Franceschini, N., Gao, W., Glaser, S., Gu, X., Hancock, D.B., Heinrich, J., Hofman, A., Imboden, M., Ingelsson, E., James, A., Karrasch, S., Koch, B., Kritchevsky, S.B., Kumar, A., Lahousse, L., Li, G., Lind, L., Lindgren, C., Liu, Y., Lohman, K., Lumley, T., Mcardle, W.L., Meibohm, B., Morris, A.P., Morrison, A.C., Musk, B., North, K.E., Palmer, L.J., Probst-Hensch, N.M., Psaty, B.M., Rivadeneira, F., Rotter, J.I., Schulz, H., Smith, L.J., Sood, A., Starr, J.M., Strachan, D.P., Teumer, A., Uitterlinden, A.G., Volzke, H., Voorman, A., Wain, L.V., Wells, M.T., Wilk, J.B., Williams, O.D., Heckbert, S.R., Stricker, B.H., London, S.J., Fornage, M., Tobin, M.D., O'connor, G.T., Hall, I.P. and Cassano, P.A., 2014. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PloS one,* **9**(7), pp. e100776.

Tantucci, C. and Modina, D., 2012. Lung function decline in COPD. *International journal of chronic obstructive pulmonary disease,* **7**, pp. 95-99.

Tashkin, D.P., 2013. Variations in FEV(1) decline over time in chronic obstructive pulmonary disease and its implications. *Current opinion in pulmonary medicine,* **19**(2), pp. 116-124.

Tashkin, D.P., Li, N., Halpin, D., Kleerup, E., Decramer, M., Celli, B. and Elashoff, R., 2013. Annual rates of change in pre- vs. post-bronchodilator FEV1 and FVC over 4 years in moderate to very severe COPD. *Respiratory medicine,* **107**(12), pp. 1904-1911.

Taylor, P., 2008. Personal genomes: when consent gets in the way. *Nature,* **456**(7218), pp. 32-33.

Travers, J., Marsh, S., Caldwell, B., Williams, M., Aldington, S., Weatherall, M., Shirtcliffe, P. and Beasley, R., 2007. External validity of randomized controlled trials in COPD. *Respiratory medicine,* **101**(6), pp. 1313-1320.

Turner, S., Armstrong, L.L., Bradford, Y., Carlson, C.S., Crawford, D.C., Crenshaw, A.T., De andrade, M., Doheny, K.F., Haines, J.L., Hayes, G., Jarvik, G., Jiang, L., Kullo, I.J., Li, R., Ling, H., Manolio, T.A., Matsumoto, M., Mccarty, C.A., Mcdavid, A.N., Mirel, D.B., Paschall, J.E., Pugh, E.W., Rasmussen, L.V., Wilke, R.A., Zuvich, R.L. and Ritchie, M.D., 2011. Quality control procedures for genome-wide association studies. *Current protocols in human genetics / editorial board, Jonathan L.Haines ...[et al.],* **Chapter 1**, pp. Unit1.19.

Turner, A.M., Tamasi, L., Schleich, F., Hoxha, M., Horvath, I., Louis, R. and Barnes, N., 2015. Clinically relevant subgroups in COPD and asthma. *European respiratory review : an official journal of the European Respiratory Society,* **24**(136), pp. 283-298.

Van Der Molen, T., Postma, D.S., Schreurs, A.J., Bosveld, H.E., Sears, M.R. and Meyboom De Jong, B., 1997. Discriminative aspects of two generic and two asthma-specific instruments: relation with symptoms, bronchodilator use and lung function in patients with mild asthma. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation,* **6**(4), pp. 353-361.

Van Manen, J.G., Bindels, P.J., Dekker, E.W., Ijzermans, C.J., Bottema, B.J., Van Der Zee, J.S. and Schade, E., 2001. Added value of co-morbidity in predicting health-related quality of life in COPD patients. *Respiratory medicine,* **95**(6), pp. 496-504.

Vasankari, T.M., Impivaara, O., Heliovaara, M., Heistaro, S., Liippo, K., Puukka, P., Saarelainen, S., Kanervisto, M. and Jousilahti, P., 2010. No increase in the prevalence of COPD in two decades. *The European respiratory journal,* **36**(4), pp. 766-773.

Vestbo, J., Edwards, L.D., Scanlon, P.D., Yates, J.C., Agusti, A., Bakke, P., Calverley, P.M., Celli, B., Coxson, H.O., Crim, C., Lomas, D.A., Macnee, W., Miller, B.E., Silverman, E.K., Tal-Singer, R., Wouters, E., Rennard, S.I. and Eclipse Investigators, 2011. Changes in forced expiratory volume in 1 second over time in COPD. *The New England journal of medicine,* **365**(13), pp. 1184-1192.

Weiskopf, N.G. and Weng, C., 2013. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association : JAMIA,* **20**(1), pp. 144-151.

Wenzel, S.E., 2012. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nature medicine,* **18**(5), pp. 716-725.

Wijnhoven, H.A., Kriegsman, D.M., Hesselink, A.E., Penninx, B.W. and De Haan, M., 2001. Determinants of different dimensions of disease severity in asthma and COPD : pulmonary function and health-related quality of life. *Chest,* **119**(4), pp. 1034-1042.

Wilk, J.B., Djousse, L., Arnett, D.K., Rich, S.S., Province, M.A., Hunt, S.C., Crapo, R.O., Higgins, M. and Myers, R.H., 2000. Evidence for major genes influencing pulmonary function in the NHLBI family heart study. *Genetic epidemiology,* **19**(1), pp. 81-94.

Wintle, B.A., Mccarthy, M.A., Volinsky, C.T. and Kavanagh, R.P., 2003. The Use of Bayesian Model Averaging to Better Represent Uncertainty in Ecological Models; El Uso de Promedios de Modelo Bayesiano para Mejorar la Representación de la Incertidumbre en Modelos Ecológicos. *Conservation Biology,* **17**(6), pp. 1579-1590.

Yang, J., Benyamin, B., Mcevoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E. and Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics,* **42**(7), pp. 565-569.

Zhou, J.J., Cho, M.H., Castaldi, P.J., Hersh, C.P., Silverman, E.K. and Laird, N.M., 2013. Heritability of Chronic Obstructive Pulmonary Disease and Related Phenotypes in Smokers. *Am J Respir Crit Care Med,* **188**(8), pp. 941-947.