Author:
Higbee, Daniel H

Title:
Extra-pulmonary effects of lung function and lung disease

Extra-pulmonary effects of lung function and lung disease

Daniel Harman Higbee

A dissertation submitted to the University of Bristol in accordance with the requirements for the award of the degree of Doctor of Philosophy in the Faculty of Health Sciences

Academic Respiratory Unit, Population Health Sciences, Bristol Medical School

January 2022

35,124 Words

ABSTRACT

Introduction

Traditional observational epidemiology has described an association between reduced lung function and lung disease with extra-pulmonary diseases. However, this method cannot establish causality and is affected by bias from residual confounding and reverse causation. If the observed association is causal, then treatments to maintain or improve lung function could reduce the huge burden of extra-pulmonary co-morbidity. My objective was to utilise novel genetic epidemiological techniques to determine if reduced lung function or lung disease has a *causal* effect on extra-pulmonary disease. I used both traditional and genetic epidemiology to research an understudied pathological lung function state, Preserved Ratio Impaired Spirometry (PRISm), and its extra-pulmonary associations. The COVID pandemic created a need to examine asthma and mental health during lockdown.

Methods

Two Sample Mendelian Randomisation (MR) techniques were used to determine if reduced lung function and COPD have a causal effect on Alzheimer's Disease, cardiovascular disease, and cognitive function. I used the UK Biobank to examine the risk factors, associations, and mortality of PRISm and then performed a Genome-Wide Association Study of PRISm. Traditional epidemiology was used to examine the effect of asthma on mental health outcomes during COVID-19 lockdown in the Avon Longitudinal Parents and Child Cohort.

Results

MR studies showed no strong evidence for an effect of reduced lung function/lung disease on cognitive function or the risk of Alzheimer's disease. However, there was strong evidence that reduced Forced Vital Capacity (FVC), but not Forced Expiratory Volume in one second (FEV_1), causes an increased risk of coronary artery disease. PRISm is common, and despite being a relatively transient state is associated with co-morbidities and an increase in mortality compared to those with normal spirometry. Genetic studies show that there may be shared genetic risk factors for PRISm and its co-morbidities. Mental health outcomes in those with asthma worsened during the lockdown to lower levels than the general population.

Discussion

There is a strong association between lung function, lung disease and extra-pulmonary disease, with genetic epidemiology providing evidence that this relationship is causal, and there may be shared genetic risk factors. This thesis demonstrates the value of triangulating different epidemiological methods and highlights future areas of research that may lead to lung function becoming a screening tool and/or modifiable risk factor for extra-pulmonary diseases.

AUTHOR'S DECLARATION

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:  DATE: 25/01/2022

ACKNOWLEDGEMENTS

First and foremost, I would like to thank James my primary supervisor. The clinical research fellow job he created allowed me to be a part of excellent research teams at the University of Bristol and Southmead Hospital. Through this I have been able to access a wealth of fantastic research and clinical opportunities. His constant drive to improve patient care and produce practice changing research combined with attentive supervision has enabled me to achieve far more than I would have expected three years ago. I am proud to have been involved with the inception of projects that will continue to flourish in the coming years, changing our understanding of airways disease and establishing Bristol as centre for respiratory airways research.

George's work to promote Mendelian Randomization and the improvement of epidemiological research is the foundation of this PhD and would have been impossible without him. I am very fortunate that such an expert was able to give oversight to my projects. Raquel has been a brilliant guide from my challenging early days of negotiating data sets and statistical analysis through to getting papers publication ready and writing this thesis.

My Southmead doctor research buddies, Alex, Dave, Hugh, George, Steve have been fantastic company and always provided help wherever they can. I hope we continue to cross paths many times through our careers. Other team members have made attending work a pleasure; Alice, Emma, Anna M, Caroline, Vicki, S-J, Nick, Millie, Anna B.

I have many people to thank at the IEU, some of whom I never met in person?! I was always so surprised when they took the time to respond to my emails; Eleanor Sanderson, Gibran Hemani, Tom Richardson, Roxana Korologou-Linden, and Wes Spiller.

My parents guiding influence started far before this PhD. Without their unwavering determination and sacrifice to ensure that all their sons fulfilled their maximum potential it's doubtful I would have obtained a place at medical school, let alone everything that has come since.

Lastly, I thank my partner Matt. At the lowest ebbs of my research, he would faithfully collect me after long days at Oakfield House. Lockdowns with him and our cats, Gwen (RIP) and Cosimo di Medici, were a simple happy time, rather than the desolate isolation most faced. His astonishing work ethic has enabled us to make so many dreams come true, with many more to follow.

ABBREVIATIONS

2SMR	Two-Sample Mendelian Randomisation
A	Adenine
A&E	Accident and Emergency
ACQ	Asthma Control Questionnaire
ADSP	Alzheimer's Disease Sequencing Project
ALSPAC	Avon Longitudinal Study of Parent and Children
β	Beta
BP	Base pair
BMI	Body Mass Index
C	Cytosine
CAD	Coronary Artery Disease
CHR	Chromosome
CHARGE	Cohort of Heart and Aging Research in Genomic Epidemiology
CI	Confidence Interval
COGENT	Cognitive Genomics Consortium
COPD	Chronic Obstructive Pulmonary Disease
COVID-19	Severe acute respiratory distress syndrome coronavirus 2019
CVD	Cardiovascular Disease
DAG	Directed Acyclic Graph
EPDS-10	Edinburgh Post-natal Depression Score
FEV_1	Forced Expiratory Volume in one second
FVC	Forced Vital Capacity
FEV_1/FVC	Forced Expiratory Volume in one second divided by Forced Vital Capacity
G	Guanine
G0	Parent cohort in ALSPAC
G1	Child cohort in ALSPAC
GAD-7	Global Anxiety Index
GWAS	Genome-Wide Association Study
HR	Hazard ratio

IGAP	International Genomics of Alzheimer's disease
IV	Instrumental variable
IVW	Inverse Variance Weighted
LD	Linkage Disequilibrium
LDSC	Linkage Disequilibrium Score Regression
LLN	Lower Limit of Normal
N	Number
MCI	Mild Cognitive Impairment
MR	Mendelian Randomisation
MVMR	Multivariable Mendelian Randomisation
OR	Odd's ratio
PEFR	Peak Expiratory Flow
PheWAS	Phenome Wide Association Study
PGC	Psychiatric Genomics Consortium
PRISm	Preserved Ratio Impaired Spirometry
QQ	Quantile-Quantile
RCT	Randomised Controlled Trial
RR	Relative Risk
R^2	R-squared
SD	Standard deviation
SE	Standard Error
SMFQ-13	Short Mood and Feeling Questionnaire
SNP	Single Nucleotide Polymorphism
SCR	Symptom Count Ratios
STAI-20	Staite Trait Anxiety Index
T	Thymine
TLC	Total Lung Capacity
WEMWBS-14	Warwick-Edinburgh Mental Wellbeing Scale
X	Exposure
Y	Outcome
Z	Instrumental Variable

CONTENTS

Table of Contents

ABSTRACT 3
Introduction 3
Methods 3
Results 3
Discussion 3
AUTHOR'S DECLARATION 6
ACKNOWLEDGEMENTS 8
ABBREVIATIONS 10
CONTENTS 12
Table of Contents 12
List of Tables 18
List of Figures 18
CHAPTER 1. INTRODUCTION 20
1.1 Lung function 21
1.1.1 Measuring lung function 21
1.1.2 Obstructive lung function 22
1.2 Chronic Obstructive Pulmonary Disease 23
1.3 COPD, reduced lung function and extra-pulmonary disease 24
1.3.1 COPD, reduced lung function, and dementia 24
1.3.2 COPD, reduced lung function and cardiovascular disease 24
1.3.3 COPD, reduced lung function and cognitive function 25
1.3.4 Underlying mechanisms between COPD, reduced lung function and extra-pulmonary disease 26
1.4 Restrictive lung function 26
1.5 Preserved Ratio Impaired Spirometry 27
1.5.1 Underlying mechanism between PRISm and extra-pulmonary disease 27
1.5.2 PRISm – previous research issues and the need for further investigation 28
1.6 Genetic determinants of lung function 28
1.6.1 Current understanding of genetic determinants of lung function, and the need for further research 31
1.7 Limitations of traditional observational epidemiology 31
1.8 The role of genetic epidemiology and Mendelian Randomisation 33

1.9 Thesis overview and research questions 33
CHAPTER 2. GENETIC EPIDEMIOLOGY METHODS 35
2.1 Genetic background 35
2.2 GWAS 35
2.3 MR method 36
2.3.1 Effect Estimate 36
2.3.2 MR assumptions 38
2.3.3 MR Sensitivity analyses 39
2.3.4 One and Two sample MR 42
2.4 Benefits of MR 42
CHAPTER 3. EXAMINING THE POSSIBLE CAUSAL RELATIONSHIP BETWEEN LUNG FUNCTION, COPD AND ALZHEIMER'S DISEASE: A MENDELIAN RANDOMISATION STUDY 43
3.1 Publication and contributions 43
3.2 Introduction 43
3.3 Methods 45
3.3.1 Lung Function 45
3.3.2 Liability to COPD 45
3.3.3 Alzheimer's disease 46
3.3.4 Statistical Analysis 46
3.3.5 Main Mendelian Randomisation Analysis 47
3.4 Results 47
3.4.1 Lung function traits as exposure (Shrine et al GWAS[75]) 47
3.4.2 COPD as exposure (Sakornsakolpat GWAS[109]) 52
3.5 Discussion 54
3.5.1 Evidence before this study 54
3.5.2 Impact of this study 54
3.5.3 Strengths and Limitations 55
3.5.4 Conclusions 56
CHAPTER 4. LUNG FUNCTION AND CARDIOVASCULAR DISEASE: A TWO-SAMPLE MENDELIAN RANDOMISATION STUDY 58
4.1 Publications and contributions 58
4.2 Introduction 58
4.3 Methods 60
4.3.1 Exposure – Lung function traits, Shrine et al. preliminary analysis [75] 60
4.3.2 Outcomes 60
4.3.3 Statistical Analysis 60

4.3.4 Main Mendelian Randomisation Analysis 61
4.4 Preliminary Results 62
4.4.1 Shrine et al. preliminary analysis 62
4.5 Collider bias 63
4.5.1 Multivariable Mendelian Randomisation 65
4.6 Exposures – MVMR 66
4.7 MVMR Results 67
4.8 Discussion 71
4.8.1 Limitations 72
4.8.2 Implications 73
4.8.3 Conclusions 74
CHAPTER 5. LUNG FUNCTION, COPD AND COGNITIVE FUNCTION: A MULTIVARIABLE AND TWO SAMPLE MENDELIAN RANDOMISATION STUDY 75
5.1 Publication and contributions 75
5.2 Introduction 75
5.3 Methods 76
5.3.1 Exposure populations 76
5.3.2 Outcome Population 77
5.3.3 Statistical Analysis 78
5.4 Results 79
5.4.1 MVMR Analysis 80
5.4.2 MR Radial 82
5.4.1 Results for 2SMR COPD effect on cognition using MR Radial 83
5.5 Discussion 84
5.5.1 Strengths and Limitations 86
5.5.2 Future research 88
5.5.3 Conclusion 88
CHAPTER 6. PRESERVED RATIO IMPAIRED SPIROMETRY (PRISm): A UKBIOBANK COHORT STUDY 89
6.1 Publication and contributions 89
6.2 Introduction 89
6.3 Methods 91
6.3.1 Baseline 91
6.3.2 Follow-up 92
6.3.3 Survival analysis 93
6.4 Results 93
6.4.1 Prevalence of PRISm 93

6.4.2 Risk factors for PRISm 95
6.4.3 PRISm symptoms and co-morbidities 97
6.4.4 Longitudinal analysis of PRISm 98
6.4.5 PRISm trajectories 101
6.4.6 Persistent PRISm vs PRISm to control trajectories 102
6.4.7 Persistent PRISm vs PRISm progressing to airflow obstruction trajectories 105
6.5 Survival Analysis 107
6.6 Discussion 109
6.6.1 Limitations 111
6.6.2 Future Research 112
6.6.3 Conclusion 112
CHAPTER 7. GENOME WIDE ASSOCIATION STUDY OF PRESERVED RATIO IMPAIRED SPIROMETRY (PRISm) IN UK BIOBANK 114
7.1 Publications and Contributions 114
7.2 Introduction 114
7.3 Method 116
7.3.1 Study Populations 116
7.3.2 GWAS 116
7.3.3 Analysis of GWAS and replicated SNPs 117
7.4 Results 119
7.4.1 Discovery GWAS 119
7.4.2 Replication analysis 120
7.4.3 Analysis of GWAS 120
7.4.4 Analysis of replicated SNPs 121
7.5 Discussion 124
7.5.1 Genomic Inflation 125
7.5.2 Replication method 127
7.6 Conclusion 127
CHAPTER 8. THE IMPACT OF ASTHMA ON MENTAL HEALTH AND WELL-BEING DURING COVID-19 LOCKDOWN 128
8.1 Disclaimer 128
8.2 Publication and contributions 128
8.3 Background 129
8.3.1 Asthma 129
8.3.2 Asthma and co-morbidity 129
8.4 Introduction 132

8.5 Methods 133
8.5.1 Study Samples 134
8.5.2 Self-reported Current Asthma 134
8.5.3 Pre-COVID-19 vs COVID-19 Mental Health 134
8.5.4 Symptoms, Change in Activities, and Worries during lockdown 135
8.5.5 Statistical analyses 135
8.6 Results 137
8.6.1 Demographics 137
8.6.2 Pre-COVID-19 vs COVID-19 Mental Health 139
8.6.3 Symptoms, Change in Activities, and Worries during lockdown 146
8.7 Discussion 151
8.7.1 Strengths & Limitations 153
8.7.2 Future Research 153
8.7.3 Conclusion 154
CHAPTER 9. SUMMARY OF KEY FINDINGS, MY LEARNING, FUTURE RESEARCH, AND CONCLUSIONS 155
9.1 Introduction 155
9.1 Examining the possible relationship between Lung function, COPD and Alzheimer's' Disease: A Mendelian Randomisation Study 155
9.1.1 Summary 155
9.1.2 Future Research 156
9.1.3 My learning 156
9.2 Lung function and cardiovascular disease: A two-sample Mendelian Randomisation study 157
9.2.1 Summary 157
9.2.2 My Learning 157
9.2.3 Future research 157
9.3 Lung function, COPD and cognitive function: A multivariable and two sample Mendelian Randomisation Study 159
9.3.1 Summary 159
9.3.2 Future research 159
9.3.3 My learning 159
9.4 Preserved Ratio Impaired Spirometry (PRISm): A UKBiobank cohort study 162
9.4.1 Summary 162
9.4.2 Future research 162
9.4.3 My learning 163
9.5 Genome-Wide Association Study of Preserved Ratio Impaired Spirometry 165

9.5.1 Summary 165
9.5.2 Future research 165
9.5.3 My learning 165
9.6 The impact of asthma on mental health and wellbeing during COVID-19 lockdown 166
9.6.1 Summary 166
9.6.2 Future research 167
9.6.3 My learning 167
9.7 Traditional observational epidemiology and the benefits of MR 169
9.8 Limitations of genetic epidemiology and MR 169
9.8.1 Demography, assortative mating, and dynastic effects 169
9.8.2 Within family MR 171
9.8.3 Limited genetic data for different ancestral populations 171
9.8.4 Mendelian Randomisation and RCTs 172
9.9 Final summary 174
REFERENCES 175

List of Tables

Table 1. Two-sample MR results of lung function traits[75] effect on risk of Alzheimer's disease[112] 50
Table 2. Two-sample MR results of COPD[194] effect on risk of Alzheimer's disease[206] 53
Table 3. Results of decreasing lung function traits on risk of coronary artery disease 62
Table 4. Results of decreasing lung function traits on risk of ischaemic stroke 63
Table 5. Multivariable MR results of FEV1 and FVC on Coronary Artery Disease and Ischaemic Stroke using UKBiobank lung function GWAS 68
Table 6. Multivariable MR results of and FEV1/FVC <0.7 on Coronary Artery Disease and Ischaemic Stroke using UKBiobank lung function GWAS 69
Table 7. 2SMR, decreasing lung function effect on cognitive function 80
Table 8. 2SMR, COPD effect on cognitive function 80
Table 9. MVMR Analysis results. Effect of decreasing Lung function trait on cognitive function with and without conditioning for covariates 82
Table 10. Results of 2 Sample MR all lung function SNPs (Shrine et al GWAS) effect on cognition using MR Radial after outlier exclusion 82
Table 11. Results of 2SMR COPD effect on cognition using MR Radial after outlier exclusion 84
Table 12. Baseline demographics of participants with PRISm, Control and Airflow Obstruction 94
Table 13. Baseline demographics of participants that only had baseline data vs participants included in follow up 99
Table 14. Demographics of participants by PRISm trajectory at follow up 100
Table 15. Baseline demographics of participants by PRISm trajectory 104
Table 16. PRISm trajectory in different subgroups 106
Table 17. Multivariable multinomial logistic regression of baseline demographics association with control to prism and control to airflow obstruction vs. persistent control trajectories 107
Table 18. Results of LD-correlation study between PRISm and both pulmonary and extra-pulmonary GWAS 121
Table 19. Replicated SNPs results from discovery and replication GWAS 123
Table 20. Demographics in G1-Index children and G0 parents' cohorts 138
Table 21. Lockdown and pre-existing mental health: asthma vs. no asthma 140
Table 22. Effect of current asthma on outcomes specific to lockdown adjusted for age, gender, smoking, overweight and pre-existing mental health in G1-Index Children 144
Table 23. Effect of current asthma on outcomes specific to lockdown adjusted for age, gender, smoking, overweight and pre-existing mental health in G0-parents 145
Table 24. Lockdown characteristics: asthma vs. no asthma in G1-Index Children 147
Table 25. Lockdown characteristics: asthma vs. no asthma in G0 Parents 148
Table 26. Comparison of asthma-effects in ALSPAC-G0 vs. ALSPAC-G1 150

List of Figures

Figure 1. A patient performing spirometry 22
Figure 2. Flow volume loops produced by spirometry 22
Figure 3. Adaption of Fletcher & Peto 1977.[70] The (previously believed) natural history of FEV1 % predicted leading to COPD. 29
Figure 4. Lung function trajectories[71] 30
Figure 5. A Directed Acyclic Graph (DAG) demonstrating a traditional observational epidemiological study 32

Figure 6. A DAG showing how to calculate the effect of an exposure on an outcome 37
Figure 7. DAG to demonstrate the assumptions of IV analysis 38
Figure 8. Single-SNP analysis plot with obvious outlier..... 40
Figure 9. Scatter plots showing each SNPs effect 41
Figure 10. Scatter plot of the SNP-effect on lung function trait and SNP-effect on Alzheimer's disease 49
Figure 11. Single SNP analysis of lung function traits on Alzheimer's disease 51
Figure 12. Leave-one out analysis of lung function on risk of Alzheimer's' Disease..... 51
Figure 13. Funnel plot. SNPs affecting lung function have their effect plotted against the inverse of their standard error 52
Figure 14. Directed acyclic graph demonstrating possible pathways leading to bias 64
Figure 15. DAG demonstrating how MVMR can be used to estimate the direct effect (βMR) of two exposures by conditioning their effects..... 66
Figure 16. MR Radial plots for 2 sample MR all lung function SNPs (Shrine et al GWAS) effect on cognition 83
Figure 17. MR Radial plot for 2SMR COPD effect on cognition 84
Figure 18. Forest plot showing factors associated with PRISm vs control spirometry 96
Figure 19. CONSORT diagram..... 98
Figure 20. Lung function trajectories from Baseline to Follow-up 102
Figure 21. Kaplan-Meier survival estimates based on spirometric group at baseline 108
Figure 22. Manhattan plot of discovery PRISm GWAS 119
Figure 23. Flow chart of analysis..... 120
Figure 24. QQ plot of PRISm GWAS..... 126
Figure 25. WEMWBS-14 Wellbeing score during lockdown in G0-ALSPAC children with asthma vs G1-ALSPAC parents with asthma 142
Figure 26. SMFQ-13 Depression score during lockdown in G0-ALSPAC children with asthma vs G1-ALSPAC parents with asthma 142
Figure 27. GAD-7 Anxiety score during lockdown in G0-ALSPAC children with asthma vs G1-ALSPAC parents with asthma 143
Figure 28. Effect of current asthma on factors specific to lockdown adjusted for age, gender, smoking status, overweight and pre-existing mental health. 146
Figure 29. Showing the global frequency of allele B for blood type [269] 170
Figure 30. Schematic demonstrating how it was thought that MR and RCT's were analogous[274] 172

CHAPTER 1. INTRODUCTION

Diseases of the lungs are common and are leading causes of morbidity and death worldwide.[1] In 2017, 544.9 million people worldwide had a chronic respiratory disease and 3.91 million people died from this.[2] Many lung diseases were previously thought of as single organ diseases, however, there is increasing evidence they are associated with diseases outside the lungs, known as extra-pulmonary disease.[3] For example, in the UK >80% of patients with Chronic Obstructive Pulmonary Disease (COPD) will have at least one other medical condition.[4] This can range from cognitive issues such as dementia, to physical health problems including cardiovascular disease. When lung and extra-pulmonary diseases co-exist, there is an increased treatment burden, healthcare utilisation, and worse health outcomes including death.[5] Evidence shows that even in those with severe COPD, two-thirds of deaths are not from respiratory failure, but from extra-pulmonary diseases.[6] Therefore, successful management of patients with lung disease, will require understanding of the relationship between lung disease and extra-pulmonary disease.

However, determining the causality of the relationship between lung and extra-pulmonary disease is challenging. The common causes of both these disease groups are shared e.g. smoking. Traditional epidemiological methods are only able to describe associations between lung diseases and extra-pulmonary diseases rather than determine causality, as this method suffers from bias including residual confounding and reverse causation. Genetic epidemiological techniques may be able to determine causality, if they are used in robust studies.[7]

Determining causality is essential to justify ongoing research and trial clinical interventions. If reduced lung function and lung diseases are causing these extra-pulmonary diseases, then measures to improve lung function and reduce lung disease could reduce the burden of extra-pulmonary disease, improving patients' quality of life and survival.

The overall aim of this work was to use different epidemiological techniques to examine associations between reduced lung function and lung disease on extra-pulmonary diseases, with a focus on using genetic epidemiology.

1.1 Lung function

The measurement of lung function is essential in clinical medicine. The diagnosis and severity of some lung diseases are defined by the pattern of lung function abnormality they produce. Treatments for lung diseases are sometimes restricted to those with certain degrees of lung function deficit.[8] Although it is possible to have significant respiratory symptoms with normal lung function, the measurement of lung function traits is widely available, inexpensive and one of the most powerful predictors of clinically relevant outcomes, including symptoms, exacerbations, and mortailty.[9] Reduced lung function, even in the absence of a specific respiratory diagnosis, has been associated with extra-pulmonary disease and all-cause mortality.[10-12] Therefore I thought it was important to investigate the effect of reduced lung function, as well as lung disease, on the risk of extra-pulmonary disease.

1.1.1 Measuring lung function

Lung function is commonly tested with spirometry, as shown in **Figure 1**, which measures the ability to move air in and out of the lungs, the "bellows" function. Spirometry can be performed easily in hospital, at GP practices, and even by patients independently at home. Multiple types of measurements can be obtained, but the most clinically useful measures for assessing the airways are recorded when the patient is exhaling maximally.[13]

Figure 1. A patient performing spirometry

As depicted in **Figure 2**, Forced Expiratory Volume in 1 second (FEV_1) is the volume of air exhaled within the first second of a maximal exhalation. The total volume exhaled during a forced manoeuvre is the Forced Vital Capacity (FVC). A ratio is produced by dividing the FEV_1 by the FVC, known as the FEV_1/FVC ratio. Normal values of FEV_1 and FVC vary dependent on factors including ethnicity, sex, age, and height. Reference equations can be used to determine what the patient's results are as a percentage of predicted results after taking these factors into account.[14] This produces the FEV_1 and FVC percent predicted (FEV_1 % predicted, and FVC % predicted).

Figure 2. Flow volume loops produced by spirometry

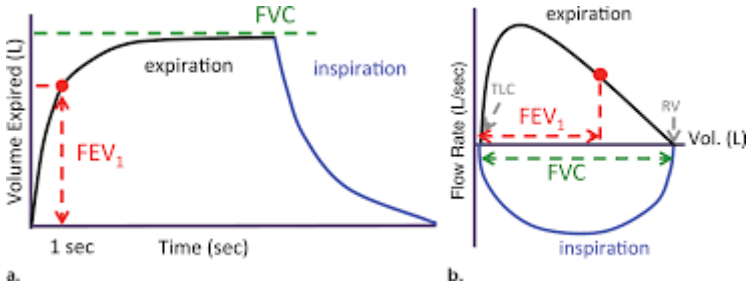


Figure 2 shows the flow volume loops produced when performing spirometry, and is labelled with the volumes that are measured for FEV_1 and FVC

1.1.2 Obstructive lung function

Many lung diseases can be categorised into two types based on spirometry. When the FEV_1/FVC ratio is reduced below 70%, the lung function results can be defined as obstructive.[13] Other defining

criteria can be used, such as when the FEV_1/FVC ratio is below the lower limit of normal (LLN).[15] The lower limit of normal value is calculated as the 5th percentile of the standard deviation for a lung function trait in a healthy non-smoking population. Obstructive lung diseases denote an inability to exhale air as quickly as would be expected. Common underlying mechanisms are inflammation and thickening of the airway wall, reducing the width of the airways, thus increasing resistance to the flow of air. Mucus produced by inflamed airways can worsen this obstruction.[16,17] One of the most common causes of obstructive lung disease is COPD.[13]

1.2 Chronic Obstructive Pulmonary Disease

COPD is characterised by persistent respiratory symptoms and airflow obstruction that is due to bronchitis affecting the airways and/or emphysema causing alveolar abnormalities.[18] Spirometry showing airflow obstruction is essential for the diagnosis of COPD, and the severity of the disease has been defined by different thresholds of FEV_1 % predicted e.g. FEV_1 >80% is mild, and FEV_1 <30% is very severe. Therefore, reduced lung function and COPD are inextricably linked. Historically COPD has been thought to be due to smoking. Although smoking remains the predominant risk factor, biomass fuel exposure, air pollution as well as host factors such as genetic abnormalities are involved.[18] COPD was considered a single organ lung disease.[19] However, COPD is associated with multiple extra-pulmonary co-morbidities, such as cardiovascular disease, cognitive dysfunction, dementia, diabetes, depression, and osteoporosis.[3,20] Understanding the association between lung function, COPD and these extra-pulmonary diseases is essential to managing complex patients. If the association is causal, lung function and COPD diagnoses may be a useful screening tool to estimate the future risk of extra-pulmonary disease. Treatments for reduce lung function and COPD may be used as modifiable risk factor to reduce the risk or severity of extra-pulmonary disease.

1.3 COPD, reduced lung function and extra-pulmonary disease

1.3.1 I chose to examine the possible causal effect of COPD and reduced lung function on Alzheimer's Disease, cardiovascular disease, and cognitive function. These three extra-pulmonary associations were chosen for a variety of reasons. Cardiovascular diseases are an extremely common cause of morbidity and death globally, and particularly in those with reduced lung function and COPD, and so deserve research attention.[6,21,22] The lungs may have a different effect on cognitive function & dementia compared to cardiovascular diseases as they comprise different disease groups with different pathological processes. There had not been previous adequate research utilising genetic techniques exploring the relationship between lung function or COPD and these extra-pulmonary outcomes. To conduct an MR study genetic data with exposure and outcome measures are required. This was available for Alzheimer's disease, cardiovascular disease, and cognitive function. COPD, reduced lung function, and dementia

Dementia is a syndrome in which there is deterioration in memory, cognition, behaviour, and the ability to perform everyday activities.[23] It is not part of normal aging. Its prevalence is rapidly increasing, and it is a significant co-morbidity. In England there were 232,000 people with dementia in 2008 which increased to 850,000 in 2014.[24] Globally there are an estimated 50 million people with dementia and this is set to reach 152 million in 2050.[23] The global cost of care provision is high and is estimated that to amount to 1.09% of aggregated global gross domestic product.[25] Alzheimer's disease (AD) is the most common type of dementia,[26] and reports of a potential link between COPD and AD was first described nearly 30 years ago.[27] Large retrospective observational case-control cohorts have reported increased risk of AD in patients with both COPD and reduced lung function.[28,29] A large cohort study reported an OR of 1.24 for AD-type dementia or mild cognitive impairment (MCI) in patients with COPD and OR 1.79 for those with a restrictive impairment compared to controls.[29] Chapter 3 uses Mendelian Randomization to determine if there is evidence of a causal effect of lung function and COPD on the risk of AD.

1.3.2 COPD, reduced lung function and cardiovascular disease

Impaired FEV_1 and FVC are strongly associated with multi-morbidity and are reported as independent predictors of cardiovascular disease.[30] Although research has often focused on the

contribution of FEV_1 and obstructive airways disease to cardiovascular risk, FVC is a stronger predictor of overall survival and appears to add value when predicting of mortality in those with intermediate risk of cardiovascular disease.[31,32]

Observational studies have reported that COPD, decreased FEV_1, FVC and FEV_1/FVC ratio are all associated with an increased risk of coronary artery disease.[33,34] However results are inconsistent, with some studies reporting no association,[35] or that the association is limited to those with abnormally high blood pressure.[36] There is evidence suggesting that COPD and impaired lung function are associated with an increased risk of both haemorrhagic and ischaemic stroke.[37,38] Chapter 4 uses Mendelian randomisation to examine for a causal effect of reduced lung function and COPD on the risk of cardiovascular disease.

1.3.3 COPD, reduced lung function and cognitive function

Cognitive function impacts important physical and mental health outcomes including mortality and educational attainment.[39] It exists on a continuum from normal cognitive function, to the potentially reversible state of MCI, which can lead to irreversible dementia.[40,41] There are very limited therapeutic options that effectively increase cognitive function or treat MCI, so finding modifiable risk factors is important.

Co-morbid lung disease and cognitive impairment is associated with worse health outcomes, quality of life, and increased healthcare utilisation. [42] The association between lung function and cognition may be due to shared risk factors seen more commonly in those with lung disease e.g. smoking.[43,44] The pattern of cognitive impairment has also been studied, with evidence of a global impairment in some groups, whilst others show a COPD specific pattern of deficits in attention, memory, learning, and motor functions.[45]. Neuroimaging provides further evidence of a relationship between lung function and cognition. After adjustment for smoking, reduced lung function remains associated with white matter brain lesions, [46] and a 'dose response' relationship is seen between severity of lung function deficit and risk of cognitive impairment.[47] Chapter 5 uses Mendelian Randomisation to

determine if there is evidence of lung function and COPD having a causal effect on cognitive function.

1.3.4 Underlying mechanisms between COPD, reduced lung function and extra-pulmonary disease

The described association between COPD, reduced lung function and extra-pulmonary disease may be related to confounding factors such as smoking and ageing. However, plausible independent mechanisms have been proposed. There may be a "spill-over" of inflammatory mediators from diseased lungs into the circulation increasing oxidative stress.[3,20,48] Plasma fibrinogen, interlukin-6, and interlukin-7 have been found in higher levels in those with COPD and heart disease.[49] Other possible inflammatory mediators include tumour necrosis factor related apoptosis-inducing ligand which is reduced in smokers with comorbid emphysema and Coronary Artery Disease (CAD), and is associated with reduced survival.[50] Inflammation may be driven by airway bacterial colonisation, which is increased in patients with COPD and higher levels of systemic inflammation.[51]

Hypoxia mediated neuronal damage or reduction in neurotransmitters that require oxygen-dependent enzymes for synthesis has been suggested as causing impairment to cognitive function. A meta-analysis of nine studies found a negative correlation between partial pressure of oxygen in arterial blood and cognitive function, but cognitive deficits have been found in both hypoxic and non-hypoxaemic patients with COPD.[52,53] The relationship with carbon dioxide is even less clear with some studies showing a relationship with executive function, attention, and verbal memory with others showing no relationship.[45]

1.4 Restrictive lung function

Restrictive lung diseases are less common than obstructive lung diseases. They are a heterogenous group of conditions that can be divided into conditions which are intrinsic and extrinsic to the lungs. Intrinsic causes occur when inflammatory changes damage the alveolar Interstitium. Extrinsic causes are due to restrictions of chest wall movements such as obesity, chest wall abnormalities, or

neuromuscular disease.[54] This can reduce the FVC and FEV_1, so the FEV_1/FVC ratio remains above 70%, and is often increased as the reduction in FVC compared to FEV_1 is proportionally higher. Restrictive lung function requires the measurement of Total Lung Capacity (TLC) to confirm that the volume of the lungs has reduced. The TLC cannot be measured by spirometry, but requires more in-depth assessment, generally only possible in the hospital setting. Not all those with restrictive spirometry will have true restriction, therefore it cannot be diagnosed based on spirometry alone.[55,56]

1.5 Preserved Ratio Impaired Spirometry

Preserved Ratio Impaired Spirometry (PRISm), also referred to as 'restrictive pattern' spirometry, is defined as an FEV_1 <80% predicted, with a FEV_1/FVC ratio ≥0.70. The diagnosis is purely based on spirometry, with no requirement to measure TLC, nor are there specific characteristic symptoms or confirmed underlying pathological mechanism. PRISm is currently being used as a research term, rather than a clinical diagnosis. PRISm has received considerably less research attention compared to obstructive and restrictive lung function. This is probably because the two most common respiratory diseases, asthma and COPD, both cause obstruction. Additionally there are likely to be many different pathological subgroups causing PRISm.[57] The true population prevalence of PRISm is unknown with estimates from 4% to 48% depending on gender, ancestry, geographical location, and smoking history.[58-61] Analyses from cohort studies show an association between PRISm and respiratory symptoms, increased healthcare utilisation, co-morbidities such as obesity, diabetes, cardiac disease, and increased overall mortality.[57,60,62-64] I believe PRISm deserves research attention, due to its prevalence and association with symptoms and extra-pulmonary disease.

1.5.1 Underlying mechanism between PRISm and extra-pulmonary disease

PRISm has received considerably less research attention compared to asthma or COPD, so underlying mechanisms of co-morbidity haven't yet been studied. However, as it is a construct of

reduced lung function, mechanisms leading to extra-pulmonary effects of reduced lung function described above (1.3.4) would be relevant.

1.5.2 PRISm – previous research issues and the need for further investigation

Definitive epidemiological understanding of PRISm has been limited to cohorts with relatively small patient numbers, rarely containing >1000 cases.[57,60] Some cohorts have used selected populations e.g. only smokers.[57,59,65] A finding of previous PRISm research had been that up to 50% of those with PRISm would transition to COPD on follow up. This finding may not be generalizable to non-smokers, which make up the majority of the general population and the majority of those with PRISm. Duration of follow-up has often been limited to ≤ 5 years, which limits insight into possible longer term associations such as mortality.[57,60,66] These issues of small samples, selection bias, and confounding limit conclusions.[67] Therefore there was a clear need to study PRISm in a large generalizable cohort, that has not restricted recruitment to smokers, with a prolonged follow up period. The UK Biobank fulfilled these requirements. It is an ongoing cohort study that started recruiting participants in 2006 aged between 40- and 69-years old living in the UK. It has collected an unprecedented amount of biological and medical data on ~500,000 participants that includes lung function measures, genetic data, lifestyle information, and symptoms with linkage to health records.[68] It is now the largest and most in depth cohort study in the world. I applied to use its data to produce the largest and most generalisable study of PRISm published to date, with novel findings that differed from previous research which had been impaired by bias. This is described in Chapter 6.

1.6 Genetic determinants of lung function

It was believed that people attained predicted maximal lung function for their age/sex/height as young adults, and then the rate of decline was uniform, except in those who smoked and experienced a rapid decline.[69] This hypothesis suggested that those who smoked developed obstruction and COPD due to the rapidly declining FEV1 % predicted as shown in **Figure 3**.

Figure 3. Adaption of Fletcher & Peto 1977.[70] The (previously believed) natural history of FEV1 % predicted leading to COPD

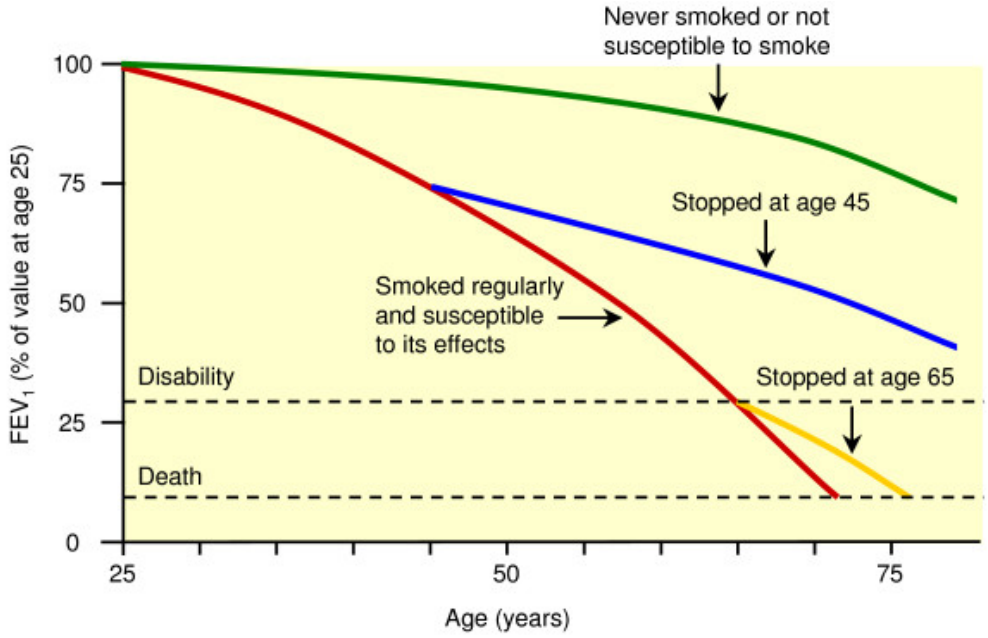


Figure 3 shows what was previously considered to be lung function trajectory of FEV1 % predicted. Note that the X-axis shows the graph starting at age 25 (rather than birth) and that everyone starts from 100% FEV1 % predicted on the y-axis.

However, it is now known that there are a number of lung function trajectories that start from different baselines at birth, before again diverging during childhood, puberty and adulthood. Some people will never achieve their predicted maximal lung function. The rate of decline from maximal lung function varies, with some declining quicker than others, not just due to smoking. These trajectories can be seen in **Figure 4**.

Figure 4. Lung function trajectories[71]

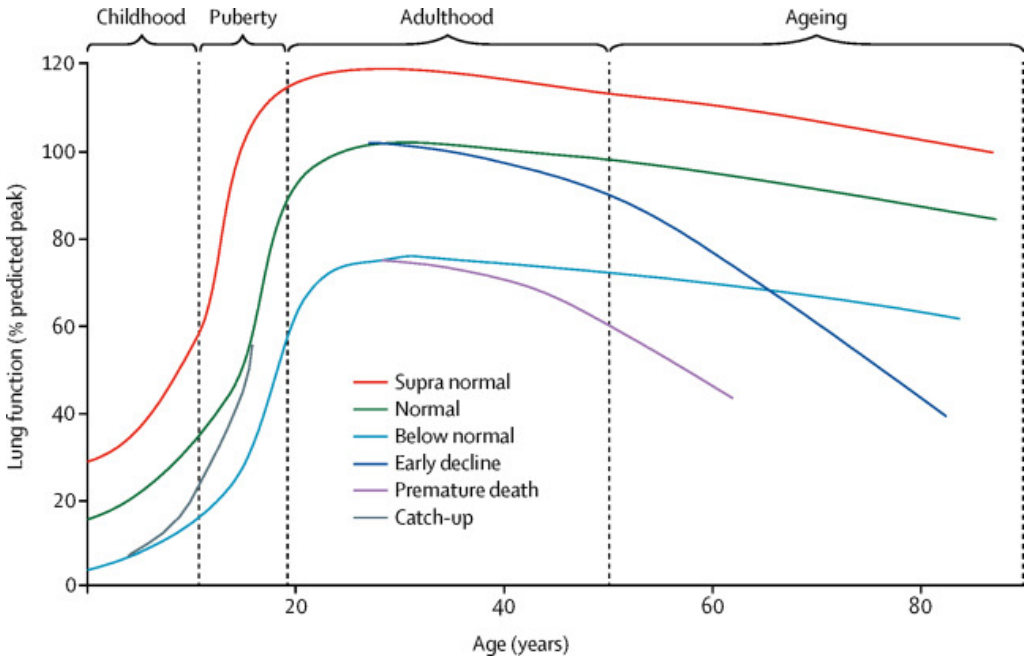


Figure 4 shows that there are a number of lung function trajectories, that differ from different starting points at birth, and then can display phenomena such as "catch-up" improvement in puberty, and early rapid decline

It is estimated that half of those with COPD will have had a normal rate of lung function decline but will have started declining from a lower lung function peak.[72] Factors influencing these trajectories are not fully known but include genetic risk factors.

To understand the pathogenesis of reduced lung function and lung disease, genetic studies are essential. Understanding which genetic variants cause reduced lung function could lead to better prediction of lung disease, and early instigation of treatment. Genetic variants that cause better lung function could potentially be used as treatment targets for reduced lung function. This could then reduce the risk of extra-pulmonary disease. It may be that genetic variants that cause reduced lung function also act to increase the risk of extra-pulmonary disease. If this is found it would explain an underlying causal mechanism and could be used for screening or treatments for extra-pulmonary disease.

1.6.1 Current understanding of genetic determinants of lung function, and the need for further research

Multiple studies have discovered genetic variants associated with lung function and lung disease.[73,74] The largest of these identified 279 Single Nucleotide Polymorphisms (SNPs) for the continuous lung function traits FEV_1, FVC, FEV_1/FVC and Peak Expiratory Flow (PEFR).[75] These SNPs were divided into deciles of a genetic risk score, in the highest decile there was a 82.4% risk of developing COPD whereas in the lowest it was 17.4%. This demonstrates the large impact genetic variation has on lung function and lung disease.

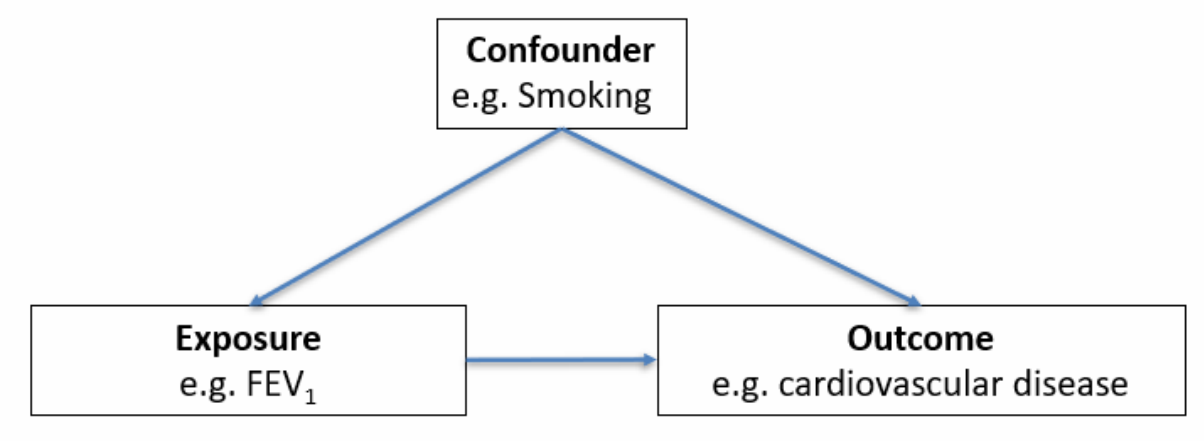
Thus far genetic studies have focused on discovering variants by examining continuous traits of spirometric measurements, as well as common respiratory diseases such as asthma and COPD. Studies of different lung function/disease phenotypes have found different genetic variants that influence lung function, so by expanding the phenotypes studied novel genetic variants may still be found.

There has only been one published genetic study of PRISm. It used a small number of cases and failed to find any variants. I decided to undertake a genetic study of PRISm using a large cohort to increase power and the chance of successfully finding associated genetic variants. As described in Chapter 7, this produced the first successful genetic study of PRISm, and found novel variants that influence lung function. It then allowed for studies to determine if there is a shared genetic predisposition between PRISm and its associated extra-pulmonary co-morbidities.

1.7 Limitations of traditional observational epidemiology

Traditional observational epidemiology has provided major advances in identifying causes of diseases and changed public health practices.[76] In a traditional observational study, a population is examined to see if an exposure e.g. FEV_1 is associated with an increased risk of an outcome e.g. cardiovascular disease, after adjustment for measured confounding factors e.g. smoking, see **Figure 5**.

Figure 5. A Directed Acyclic Graph (DAG) demonstrating a traditional observational epidemiological study



This figure shows a DAG for a traditional observational study. Although some known confounders can be measured and adjusted for, there is likely to always be residual confounding present.

Estimated associations are of value but are not an estimation of a causal effect. An association does not equate to causality, limiting conclusions and relevance to clinical practice. Additionally, traditional observational epidemiology is prone to certain sources of bias.[77] This includes; unmeasured confounding – when not all confounders can be identified and adjusted for,[78] and reverse causation – where the outcome can influence the exposure. It is essential to determine causality as this indicates the potential for intervention. Randomisation can be used to reduce these sources of bias and determine a causal estimate.[79,80] Therefore causal effects can be determined in randomised controlled trials (RCT's), which provide the highest grade of evidence.[81] There are many examples of promising disease-modifiable associations that are shown to not have any effect in RCTs due to bias in traditional epidemiological studies.[77,82,83] However, it is not feasible to randomise participants to some exposures of interest e.g. COPD or reduced lung function. Therefore, methods that determine causal estimates for exposures beyond the constraints of RCTs are required. This can be done using genetic epidemiology, particularly Genome-Wide Association Studies (GWAS) and Mendelian Randomization (MR).

1.8 The role of genetic epidemiology and Mendelian Randomisation

Genome-Wide Association Studies (GWAS) are performed to identify genetic variants associated with a disease or trait. This determines if diseases/traits are heritable, identifies biological pathways that cause them, and can provide therapeutic targets.[84-86] Genetic variants discovered can also be used as instrumental variables in Mendelian Randomisation studies.[7]

Mendelian Randomisation (MR) is a method that can overcome problems of unmeasured confounding and reverse causation typical of conventional observational epidemiology.[87] In a robust study, MR allows causal inference through the use of genetic variants as proxies for modifiable risk factors or health outcomes.[88] MR has multiple advantages, it uses genetic variants which are randomly allocated at conception so they can be exploited to simulate randomisation.[89] Genetic variants are not influenced by behavioural or environmental factors and are far less susceptible to bias from reverse causation. Additionally, the effects are equivalent to lifetime differences, reducing issues relating to transient fluctuations in exposures.[83] Chapter 2 will discuss genetic epidemiology, GWAS and Mendelian Randomisation in detail.

1.9 Thesis overview and research questions

This thesis aims to explore both physical and mental extra-pulmonary co-morbidity associated with reduced lung function, COPD, PRISm, and asthma.

Firstly, MR was used to determine if reduced lung function or COPD has a causal effect on extra-pulmonary diseases and traits: Alzheimer's Disease, cardiovascular disease, and cognitive function. Secondly, to surmount limitations in current PRISm research, I utilised UK Biobank to undertake a traditional epidemiological study of PRISm and its extra-pulmonary associations. Genetic epidemiology was used to gain a better understanding of whether PRISm is a heritable trait and any underlying pathogenesis for PRISm and its co-morbidities.

Lastly, the COVID-19 pandemic offered an important opportunity to investigate the effect of asthma on mental health during social isolation measures.

The research questions of this thesis are;

1. Can genetic epidemiology be used to determine the extra-pulmonary effects of lung function and COPD? Specifically;
    a. Does reduced lung function or COPD cause an increased risk of Alzheimer's disease?
    b. Does reduced lung function or COPD cause an increased risk of cardiovascular disease?
    c. Does reduced lung function of COPD reduce general cognitive function?
2. What is the prevalence, co-morbidity, and mortality of PRISm in a large generalizable cohort?
3. Is PRISm heritable, and can genetic studies give an insight into the pathogenesis of PRISm and its co-morbidities?
4. Is asthma associated with worse mental health and well-being during COVID-19 lockdown

CHAPTER 2. GENETIC EPIDEMIOLOGY METHODS

2.1 Genetic background

The human genome consists of 23 pairs of chromosomes that are randomly assigned at conception from the biological parents. Chromosomes are constructed of condensed deoxyribonucleic acid (DNA). DNA is a sequence of pairs of nucleic acids, which are made of a nucleobase; cytosine (C), guanine (G), adenine (A), and thymine (T), combined with a sugar and a phosphate group. Humans have approximately 3 billion pairs of DNA. Although ~99.9% of the DNA sequence between humans is identical, there remain >15 million places where the nucleobase differs at a specific location in more than 1% of the population.[90] These locations are called Single Nucleotide Polymorphisms (SNPs) and are known by their reference number or by their genomic position (Chromosome: position), for example, rs11113127 or 12:107228443. The different nucleobase possibilities at the SNP are the different alleles e.g. an A allele or a G allele.[91] Humans have ~21,000 genes that are made of sequences of DNA that encode the synthesis of rib0nucleic acid (RNA) or proteins. Different alleles can change the function of a gene, these are known as mutations and form the basis of genetic variation. More than 50,000 associations between alleles and common diseases or traits have been discovered by GWAS.[92]

2.2 GWAS

GWAS aim to identify SNPs associated with phenotypic variation. They can be performed in either a case-control population e.g. COPD vs control, or with continuous traits e.g. FEV_1. Participants donate a DNA sample, and alleles at SNPs are detected by a SNP array. Imputation can be performed to predict alleles at SNPs that haven't been directly tested for by the SNP array. GWAS are performed by conducting a regression analysis on the alleles at each SNP to determine if they are associated with a disease or trait. GWAS usually test millions of SNPs to increase the likelihood of finding a result in a candidate free approach There is a high potential for false positive results due to repeated testing, so stringent p-value thresholds are used, typically $p < 5^{-8}$.[93] The associated effect size of each

SNP discovered is usually small, so large numbers of individuals are tested to increase power. GWAS using sample sizes of over a million individuals are now being reported.[94,95]

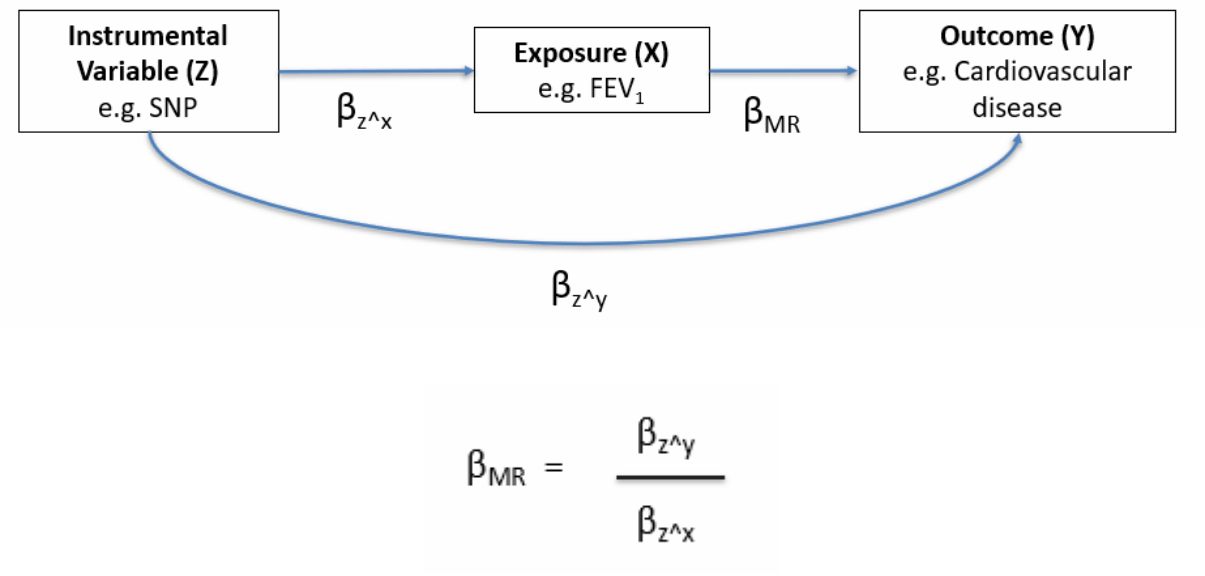
2.3 MR method

MR uses genetic variants for modifiable health traits as instrumental variables (IV) to allow for causal inference, if all the MR assumptions are met.[88] SNPs discovered in GWAS can be used as IV's. For example, in a population of cases and controls for coronary artery disease, if SNPs that reduce FEV_1 are found more commonly in those with coronary artery disease, then reduced FEV_1 has a causal effect on coronary artery disease in a robust MR study. An effect size can then be calculated.

2.3.1 Effect Estimate

Different statistical methods are used to calculate the causal effect depending on the number of exposure SNPs and whether the exposure and/or outcome is binary or continuous.[89] As shown in **Figure 6**, when using only one SNP a simple ratio can be calculated which is the regression coefficient of the SNP on the outcome ($\beta_{z\hat{}y}$) divided by the regression coefficient of the SNP on the exposure ($\beta_{z\hat{}x}$).[96]

Figure 6. A DAG showing how to calculate the effect of an exposure on an outcome



$$\beta_{MR} = \frac{\beta_{z\hat{}y}}{\beta_{z\hat{}x}}$$

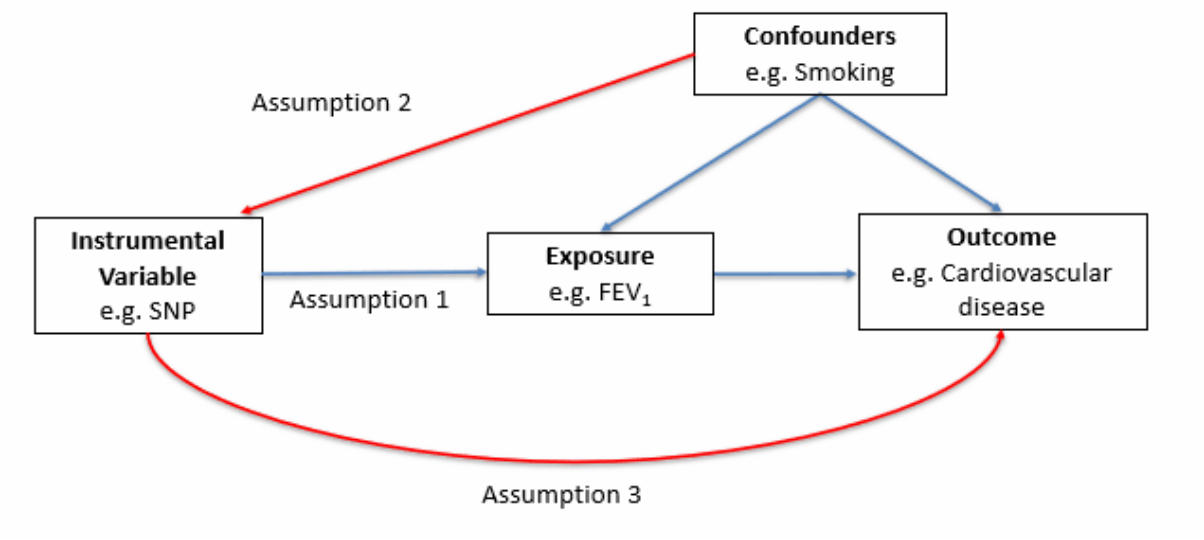
This figure shows a DAG for an MR study, and demonstrates how the effect of an exposure on an outcome is calculated

In this thesis I used multiple SNPs as instruments, so the primary method used was Inverse-Variance Weighted (IVW) estimate. This method combines the ratio estimates and standard errors for all the SNPs in a fixed-effects meta-analysis model.[97]

2.3.2 MR assumptions

For the model to work the IV must adhere to three assumptions, depicted in **Figure 7**.[98]

Figure 7. DAG to demonstrate the assumptions of IV analysis



The DAG in this figure demonstrates the three assumptions of an IV analysis, as is used in MR

Assumption 1. The IV must be strongly associated with the exposure. The relevance assumption.

Assumption 2. The IV-outcome association is not confounded. The independence assumption.

Assumption 3. The IV must only affect the outcome via the exposure. The exclusion restrictionassumption.

Assumption 1 can be directly tested by using SNPs that are known to be strongly associated with the exposure from GWAS (p-value $<5^{-8}$) and assessing the strength of the SNPs as instruments, for example with an F statistic.[99] If a SNP is affected by genetic confounding e.g. population stratification, then assumption 2 is violated. If a SNP is violating assumption 3 there is horizontal pleiotropy. SNPs can be excluded if they are strongly associated with known confounders, but there

can be unknown/unmeasured confounding, so it is not possible to directly test for assumption 2 or 3.

2.3.3 MR Sensitivity analyses

To avoid horizontal pleiotropy a biological method can be used, where only SNPs that have firm biological mechanisms for association with the exposure can be used, although this is not always practical, and pleiotropy could still be present. Pleiotropic SNPs may be seen as outliers, when SNPs exhibit a larger influence on the outcome than would be expected given the effect on the exposure. Statistical methods can identify and remove outlier SNPs causing horizontal pleiotropy. MR-Egger and other methods to detect horizontal pleiotropy are described below. Heterogeneity (the variability in causal estimates obtained for each SNP) is an indication of a potential violation of assumptions. This can be calculated and assessed with a Q statistic.[100] MR-Radial can be used which identifies and removes outlier SNPs that contribute more to heterogeneity than would be expected.[101]

A variety of visual plots can be used to identify outliers. A single-SNP plot, as shown in **Figure 8**, allows comparison of each SNPs effect size for the exposure on the outcome.

Figure 8. Single-SNP analysis plot with obvious outlier

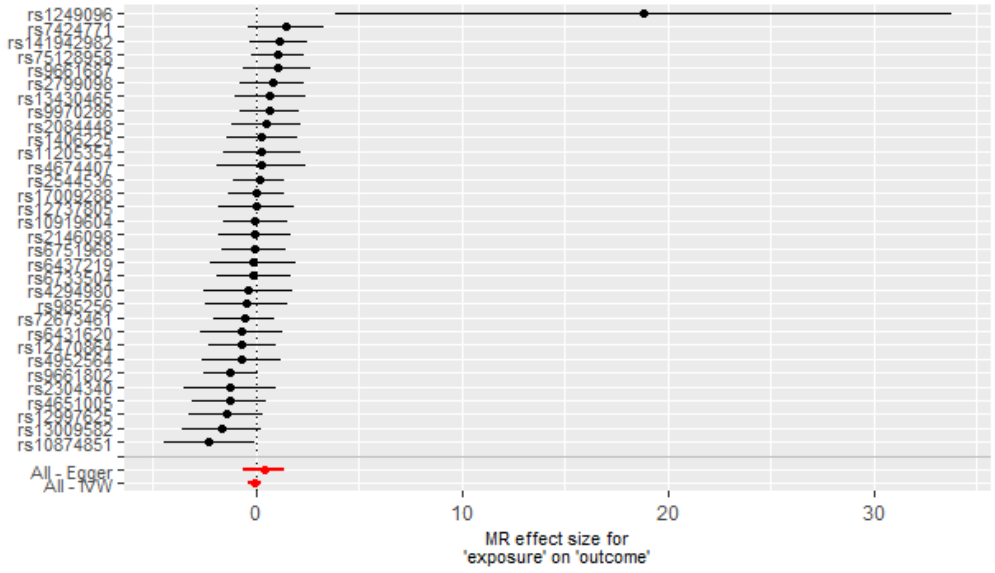
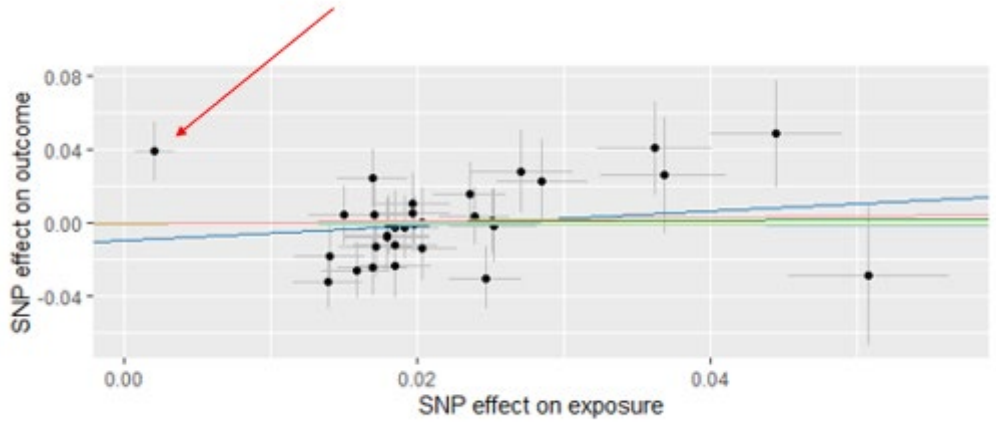


Figure 8 shows that SNP rs1249096 at the top of the y axis has a far larger effect size for the exposure on the outcome than the other SNPs and has a very large standard error. Therefore, it is an outlier, and a repeat analysis without it should be performed.

A scatter plot plots the SNPs effect on the exposure on the x-axis, and the same SNPs effect on the outcome of the y axis. This can also show outliers as in **Figure 9**.

Figure 9. Scatter plots showing each SNPs effect



In Figure 9 the red arrow indicates a SNP having a much smaller effect on the exposure than the other SNPs but is having a comparatively large effect on the outcome so could be considered an outlier (effect on the outcome due to an alternate pathway).

Other plots used include leave-one-out analysis, which is a forest plot of the IVW after selectively removing one SNP at a time. This will identify if individual SNPs are responsible for large changes in the causal estimate.[96] A funnel plot of instrument strength versus causal effect can be used to see if weak instruments are driving the causal estimate.[102]

IVW assumes that all variants are valid, so sensitivity analysis can be performed with MR-Egger, weighted median, and weighted mode that provide robust estimates even if this assumption is violated. As IVW assumes all instruments are valid, the regression intercept is constrained to zero, as indicated by the red line in **Figure 9**. MR Egger also estimates the causal effect from the gradient of the slope from a weighted regression of IV-outcome on IV-exposure, but the intercept is the average pleiotropic effect of the SNPs, and so is not constrained to zero.[103] The MR-Egger gradient can be seen as the blue line in **Figure 9**. This makes a more robust estimate of the causal effect accounting for potential horizontal pleiotropy. A weighted median estimate is robust when up to 50% of genetic

variants are invalid IV's and violation of assumptions 2 and 3 are allowed.[96] Weighted mode is robust when the largest number of similar estimates comes from valid instruments, even if the majority of instruments are invalid.[104] Steiger filtering can be used which removes SNPs that that may be affecting the exposure via a pathway that uses the outcome, ensuring the correct direction of effect is being examined i.e. SNP→Exposure→Outcome.[105] All these methods can be used in a MR paper to ensure the most reliable results.

2.3.4 One and Two sample MR

The SNP effect on the exposure and outcome can be measured in the same sample, which is known as one-sample MR. When the SNP effect on the exposure is measured in one sample, and the SNP effect on the outcome is measured in another sample, this is known as two-sample MR (2SMR).[102] It is not always possible to perform one-sample MR studies, if genotyped cohorts have not measured both the exposure and the outcome in the sample population. For this reason, only 2SMR is used in this thesis.

2.4 Benefits of MR

If the assumptions of MR are met, then it can be used to determine the causal effect of an exposure on an outcome, avoiding bias from residual confounding.[87] SNPs are not influenced by behavioural or environmental factors which minimizes reverse causation.[88] Additionally, the effects are equivalent to lifetime differences, reducing issues relating to transient fluctuations in exposures.[83] Genetic variants are randomly allocated at conception, so can be used to simulate randomisation.[89] If genetic and phenotypic data already exists, an MR study can be conducted quickly and cheaply.[83]

CHAPTER 3. EXAMINING THE POSSIBLE CAUSAL RELATIONSHIP BETWEEN LUNG FUNCTION, COPD AND ALZHEIMER'S DISEASE: A MENDELIAN RANDOMISATION STUDY

3.1 Publication and contributions

This chapter has been published.

Examining the possible causal relationship between lung function, COPD and Alzheimer's disease: a Mendelian randomisation study. Daniel Higbee, Raquel Granell, Esther Walton, Roxana Korologou-Linden, George Davey Smith, James Dodd. BMJ Open Respir Res. 2021 Jul;8(1):e000759.

DH and JD developed the concept. DH and RG designed the data analysis plan. EW provided advice regarding code for the analysis. RKL provided the outcome GWAS with UK Biobank cases removed, and advice regarding Steiger filtering. DH performed the data analysis and wrote the paper. GDS and JD provided supervision.

3.2 Introduction

Impairment of cognitive function is a well described extra-pulmonary complication of COPD, with a reported prevalence ranging from 10-61%. Cognitive impairment in COPD is associated with greater disability,[42] poorer medication compliance,[106] and risk of exacerbation and mortality.[42] Cognitive function is made up of different domains including motor skills and memory. When multiple cognitive domains are impaired and general function is permanently impaired, dementia is diagnosed. Around 25% of older adults with dementia also have COPD.[45] Poor pulmonary function in early life has been associated with increased odds of dementia later in life, even after adjustment for smoking.[107]

Alzheimer's disease (AD) is the most common type of dementia,[26] its association with COPD is less well defined than general cognitive ability, but reports of a potential link between COPD and AD were first described nearly 30 years ago.[27] Large retrospective observational case-control cohorts

have reported increased risk of AD in patients with both COPD and reduced lung function.[28,29] For example, Lutsey et al reviewed hospitalisation codes in the Atherosclerosis Risk In Communities Study for AD-related outcomes and reported an Odds Ratio (OR) of 1.24 (0.97–1.60) for AD-type dementia or mild cognitive impairment (MCI) in patients with COPD and OR 1.79 (1.24–2.58) for those with a restrictive impairment compared to controls.[29] If lung function and COPD have a causal effect on the risk of AD, then they could be modifiable risk factors.

My objective was to use MR to investigate if there is any evidence of a causal effect between the exposures of reduced lung function and liability to COPD, with the outcome of Alzheimer's disease.

3.3 Methods

3.3.1 Lung Function

I used data from the Shrine et al. GWAS as it is the largest currently available lung function GWAS in terms of both sample population size (N = 400,102) and number of reported SNPs (279 genome-wide significant SNPs, $p<5\times10^{-9}$).[75] The Shrine et al GWAS discovered SNPs for Forced Expiratory Volume in 1 second (FEV_1), Forced Vital Capacity (FVC), FEV_1/FVC ratio, and Peak Expiratory Flow (PEFR). 140 of the SNPs were previously reported and explained 5.0%, 3.4%, 9.2%, and 4.5% of the estimated heritability of FEV_1, FVC, FEV_1/FVC, and PEFR, respectively. The 139 new signals reported explained an additional 4.3%, 3.3%, 3.9%, and 3.3% of the estimated heritability, respectively. The SNPs discovered were strongly associated with risk of COPD (p= 6.64x10-63), with an Odds ratio of 1.55 (1.47-1.63) for COPD with each standard deviation of the risk score.[75,108] This satisfies the first assumptions of MR analysis, that the IV is strongly associated with the exposure. Further details of the study population can be found in the reference.[75]

3.3.2 Liability to COPD

I used 82 SNPs associated with COPD, as identified in Sarkonsakaplat et al. case-control GWAS.[109] This study was performed in 35,735 cases and 222,076 controls via the meta-analysis of 25 studies. COPD was defined by Global Initiative for Chronic Obstructive Lung Disease criteria; FEV_1/FVC <0.7 and FEV_1 <80% predicted. SNPs discovered explained up to 7% of phenotypic variance. Further details of the study population can be found in the reference.[109]

COPD is a binary trait; it is either present or not. As this was a two sample MR study, it is not known if the participants in the outcome GWAS had COPD or not. Therefore, I can only say that MR study demonstrates evidence of the effect of liability to COPD on Alzheimer's disease, rather than COPD itself.

80% and 77% of the Shrine et al and Sarkonsakaplat et al GWAS samples respectively were from the UKBiobank.[68] As described in 1.5.2, the UK Biobank is a large prospective cohort study where

>500,000 participants were recruited from 2006 – 2010 in the United Kingdom (54% female). Pre-bronchodilation lung function testing was performed by trained healthcare staff.

3.3.3 Alzheimer's disease

I used data from a meta-analysis of the International Genomics of Alzheimer's disease (IGAP) consortium,[110] Alzheimer's Disease Sequencing Project (ADSP),[111] and Psychiatric Genomics Consortium (PGC) totalling 24,807 AD cases and 55,058 controls.[112,113] All cases had a clinical diagnosis of AD. Some participants of the ADSP cohort were previously also included in IGAP, so ADSP individuals that were duplicates in IGAP and ADSP were excluded. This was based on the comparison of individual level genetic data.

There was no sample overlap between the exposure and outcome samples. All participants were of European ancestry.

3.3.4 Statistical Analysis

Statistical analysis was done using R Studio version 3.5.1. and the MRCIEU/TwoSampleMR R package.[114]

For all exposures SNPs LD-clumping was performed using European reference population and the ieugwasr:ld_clump tool. (kb = 10000, r^2 0.001). This filters SNPs that are close to each other (within 10000 base pairs) and SNPs that are highly correlated with one another (r^2 0.001), retaining the most significant SNP (lowest p-value), and removing those less significant. This ensures that the SNPs used are all independent signals. Palindromic SNPs (i.e. A/T and C/G SNPs) with intermediate allele frequencies were excluded to ensure the correct effect allele is being used. The remaining SNPs were harmonised.[98] This ensures that the effect allele from the exposure GWAS is the same as the allele being examined in the outcome GWAS. Steiger filtering was performed to remove SNPs that explained more variance in the outcome than the exposure.[105] This ensures the the correct direction of effect (exposure to outcome) is being examined. F-statistics of the SNPs used in the analysis were

calculated (F statistic = beta^2/standard error^2). The higher the F-statistic the lower the chance of weak instrument bias, and therefore the lower the chance of violating the first MR assumption.[99] An F-statistic of <10 would indicate a weak instrument.

3.3.5 Main Mendelian Randomisation Analysis

The analysis methods have been described in 2.3.3. Inverse Variance Weighting (IVW) was used for the main effect estimate. To account for the possibility of horizontal pleiotropy (SNPs influence exposure and/or outcome through independent pathways), MR Egger was performed. To minimise the effect of unbalanced instruments on an overall estimate of the mean, weighted median and mode MR methods were performed. To assess for horizontal pleiotropy a funnel plot was made by plotting the effect against its precision (beta against standard error). To ensure the results were not due to outliers with a large effect, a leave-one-out analysis was performed by re-estimating the total effect after sequentially excluding one SNP at a time and a single-SNP analysis, where the effect of each SNP was individually assessed via IVW analysis and represented in a forest plot.

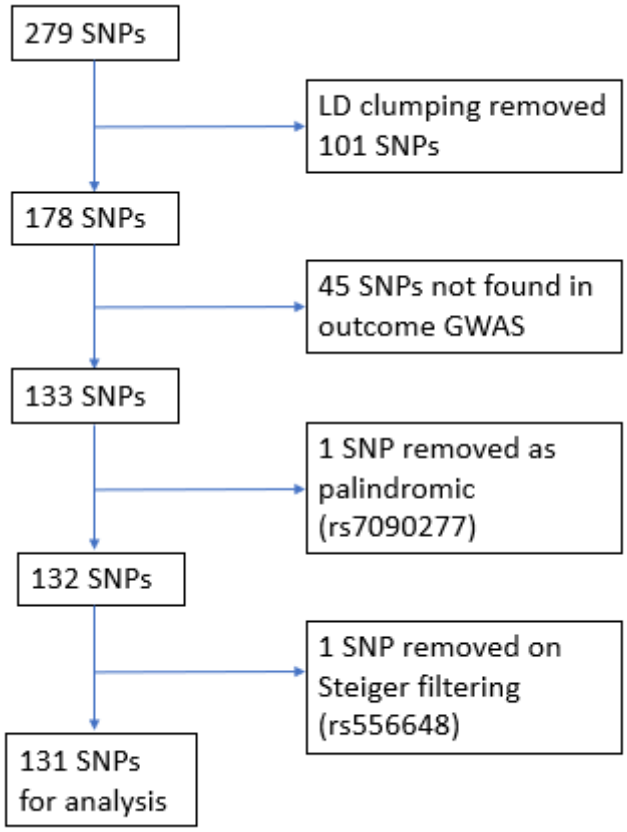
Heterogeneity (the variability in causal estimates obtained for each SNP) is an indication of the potential violation of assumptions. This was calculated and assessed with a Q statistic.

3.4 Results

3.4.1 Lung function traits as exposure (Shrine et al GWAS[75])

After clumping, extracting SNPs from outcome GWAS, Steiger filtering, and removal of palindromic SNPs, 131 SNPs were available for analysis. A flow chart is shown in **Figure 10**. F-statistic for lung function GWAS exposures were all >10 making weak instrument bias unlikely. The F-statistics for all lung function traits combined was = 114, FEV_1 = 72, FVC = 75 and FEV_1/FVC = 150.[75]

Figure 10. Flow Chart of SNPs used in analysis of all lung function traits effect on Alzheimer's Disease



This flow chart shows the exposure SNPs being clumped, harmonised and Steiger filtered prior to analysis

Minimal evidence for a causal effect of lung function (all traits combined) on Alzheimer's disease was found, (IVW Odds Ratio [OR]:1.02 per SD; 95% CI: 0.91-1.13; p-value = 0.68). **Figure 10** shows these results in a scatter plot. This result was further confirmed in a sensitivity analysis using both weighted median (OR:1.01 per SD; 95% CI:0.86-1.19, p-value 0.81), and weighted mode MR (OR 0.99 per SD;95% CI 0.78-1.19), p-value 0.81). The MR-Egger causal estimation produced similar results with an OR 1.05 per SD (95% CI 0.79-1.34, p-value 0.71). The confidence interval of the MR-Egger is wider than that of IVW, consistent with the lower statistical power of this test.

Figure 11. Scatter plot of the SNP-effect on lung function trait and SNP-effect on Alzheimer's disease

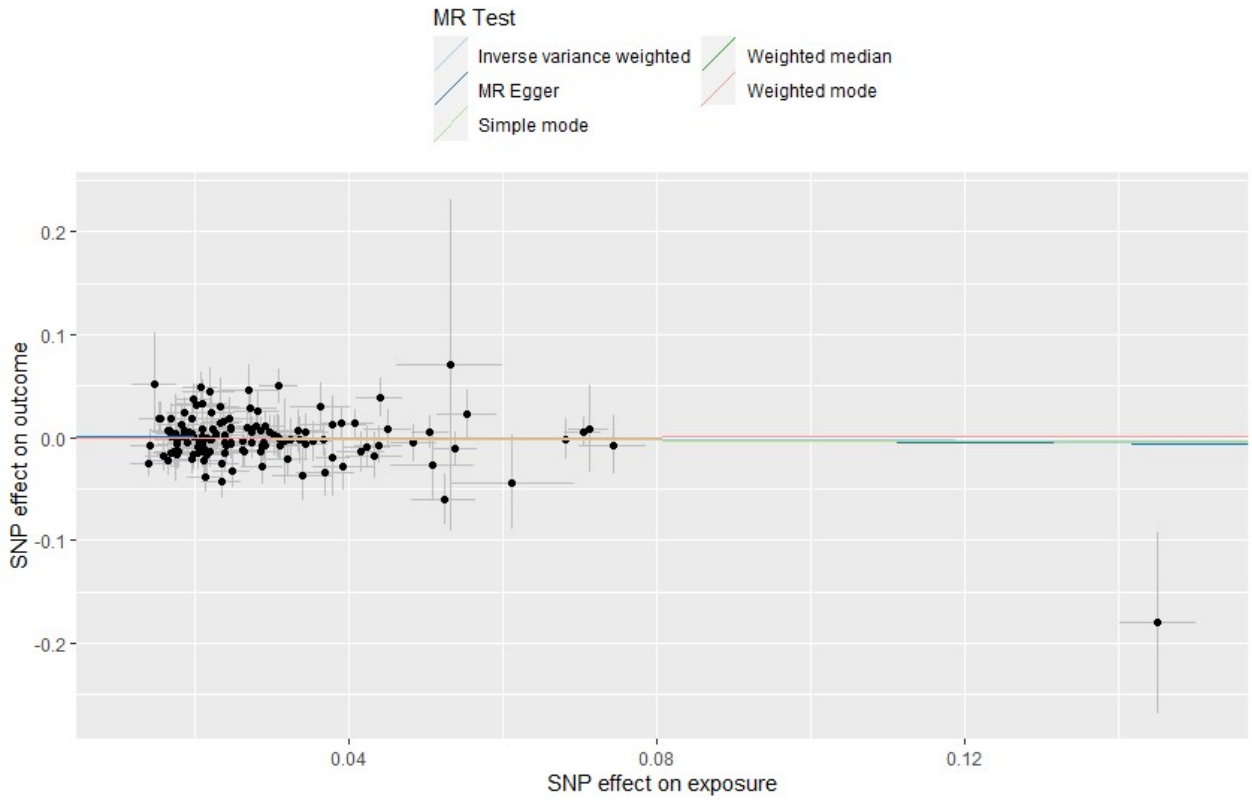


Figure 10 plots each individual SNP-exposure effect against SNP-outcome with the coloured lines representing each statistical test. Increasing lung function (exposure) does not have a consistent effect on Alzheimer's disease (outcome). This shows no effect of lung function on Alzheimer's disease. MR Egger intercept is close to zero indicating no unbalanced directional pleiotropy.

Table 1 shows that these results were consistent when analysing lung function traits FEV_1, FVC, and FEV_1/FVC individually with no good evidence of a causal association on Alzheimer's disease with confidence intervals crossing 1 for all statistical tests.

Table 1. Two-sample MR results of lung function traits[75] effect on risk of Alzheimer's disease[112]

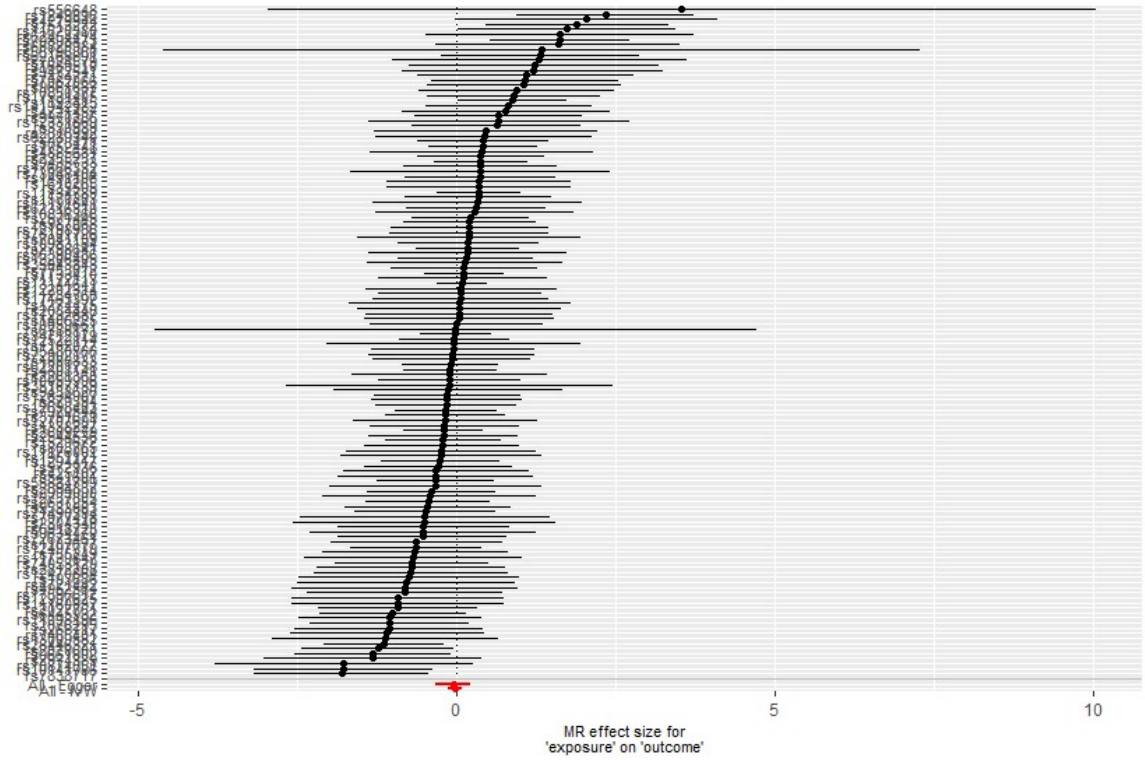
		Lung Function Trait (Exposure)			
		FEV_1, FVC, FEV_1/FVC, PEF	**FEV_1**	**FVC**	**FEV_1/FVC**
	No. SNPs used	131	42	46	73
IVW	OR of AD per SD	1.02	1.04	1.08	0.99
	(95% CI)	(0.91 – 1.13)	(0.82 – 1.32)	(0.85 – 1.37)	(0.88 – 1.13)
	p-value	0.68	0.73	0.51	0.97
	Q_p-value*	0.26	0.30	0.19	0.71
Weighted Median	OR per SD	1.01	1.15	1.14	0.95
	(95% CI)	(0.86 – 1.19)	(0.82 – 1.61)	(0.83 – 1.58)	(0.79 – 1.15)
	p-value	0.81	0.39	0.39	0.62
Weighted Mode	OR per SD	0.99	1.07	1.04	0.97
	(95% CI)	(0.78 – 1.26)	(0.60 - 1.90)	(0.61 – 1.78)	(0.74 – 1.26)
	p-value	0.97	0.80	0.86	0.84
MR-Egger	OR per SD	1.05	1.22	0.97	0.95
	(95% CI)	(0.79 – 1.34)	(0.57 - 2.59)	(0.36 – 2.62)	(0.69 – 1.31)
	p-value	0.71	0.59	0.96	0.77

*A test for heterogeneity. If this was <0.05 it would suggest heterogeneity

OR – Odd ratio; CI – Confidence Interval; IVW – Inverse Variance Weighting

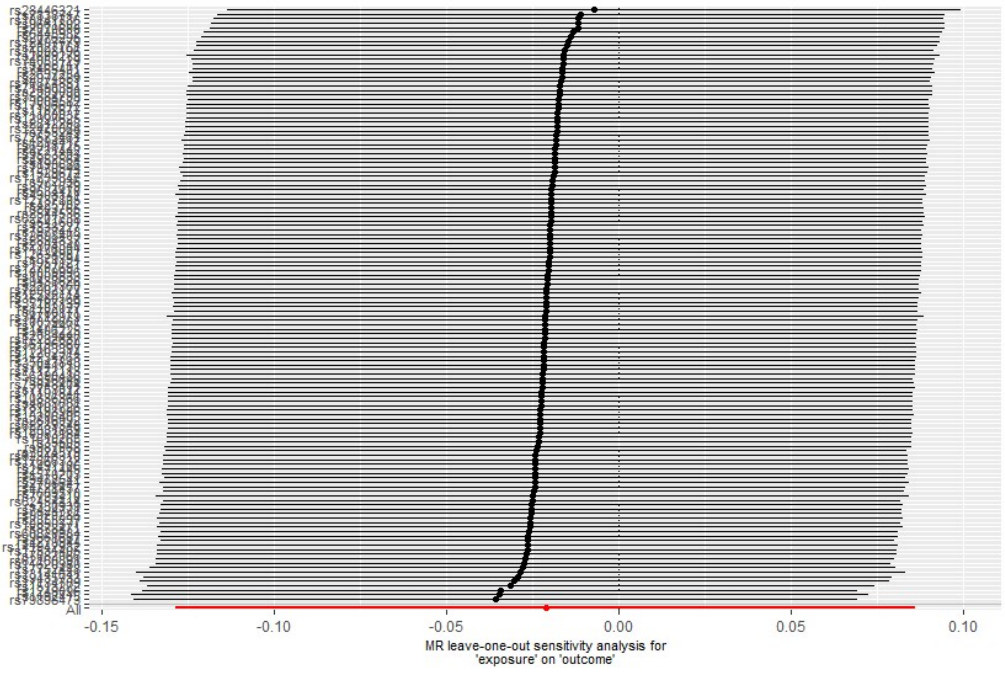
A single-SNP analyses was used to determine the effect of each lung function SNP on the odds of Alzheimer's disease (**Figure 11**). The SNP rs2070600 may be considered as an outlier due to its comparatively large effect on both lung function and AD. Polymorphisms in this SNP have been described as having a weak effect on Alzheimer's disease risk, which would violate the third MR assumption .[115] However, despite excluding this SNP from the analysis the results were similar (e.g. see leave-one-out plot in **Figure 12**).

Figure 12. Single SNP analysis of lung function traits on Alzheimer's disease



Single SNP plot showing no obvious outliers.

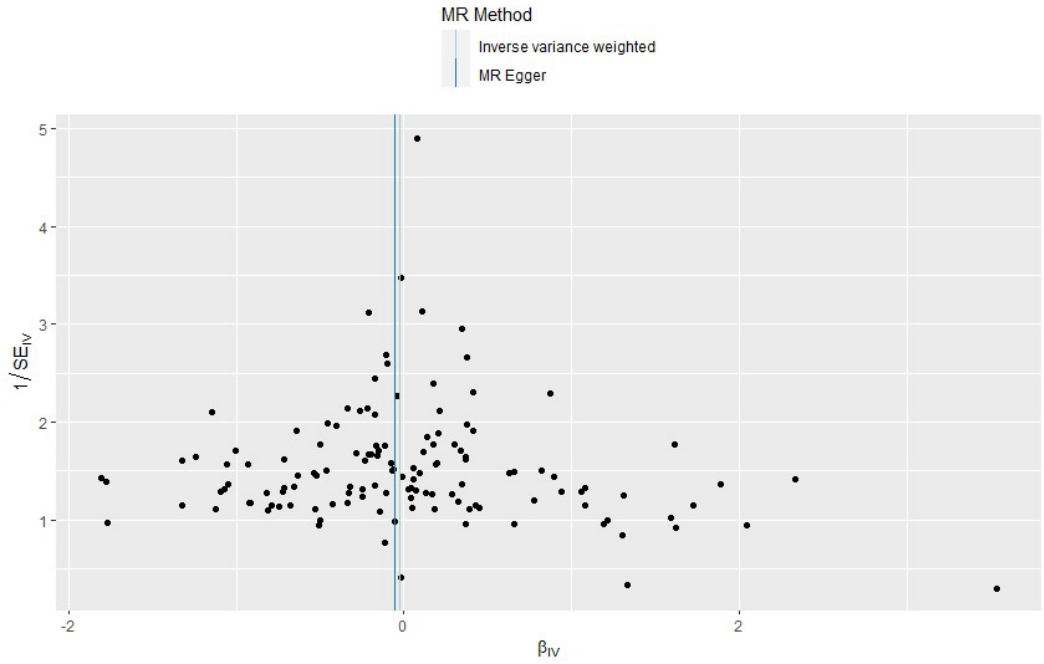
Figure 13. Leave-one out analysis of lung function on risk of Alzheimer's' Disease



There is no meaningful change in the effect estimate one leaving out each SNP sequentially. This means the estimate is not being driven by a single outlier causing horizontal pleiotropy

Each SNP beta was plotted against its inverse standard error (**Figure 13**) producing a funnel shape indicating no heterogeneity. In addition to these visual tests, there was little evidence of heterogeneity using a Q statistic when lung function traits were combined or assessed individually (**Table 1**. Q_p-value >0.51). MR-Egger intercept was <0.001, visually displayed in **Figure 10**, indicating there was no unbalanced horizontal pleiotropy.

Figure 14. Funnel plot. SNPs affecting lung function have their effect plotted against the inverse of their standard error



As the plot if funnel shaped it indicates to issues with horizontal pleiotropy

3.4.2 COPD as exposure (Sakornsakolpat GWAS[109])

After clumping, extracting SNPs from outcome GWAS, Steiger filtering, and removal of palindromic SNPs, 53 SNPs for liability to COPD were available for analysis in the Alzheimer's outcome GWAS. The remaining SNPs had an F-statistic of 54, making weak instrument bias unlikely. Results are displayed in **Table 2**.

Table 2. Two-sample MR results of COPD[194] effect on risk of Alzheimer's disease[206]

		COPD
	No. SNPs used	53
IVW	OR of AD per SD (95% CI)	0.97 (0.92 – 1.03)
	p-value	0.40
	Q_p-value	0.57
Weighted Median	OR per SD (95% CI)	0.97 (0.90 – 1.05)
	p-value	0.52
Weighted Mode	OR per SD (95% CI)	0.96 (0.86 – 1.08)
	p-value	0.56
MR-Egger	OR per SD (95% CI)	1.10 (0.93 – 1.31)
	p-value	0.23

OR – Odd ratio; CI – Confidence Interval; IVW – Inverse Variance Weighting

There was minimal evidence for an effect of liability to COPD on the risk of Alzheimer's disease (IVW OR: 0.97 per SD; 95% CI: 0.92 – 1.03; p-value 0.40). This result was further confirmed in the sensitivity analysis using both weighted median (OR: 0.97 per SD; 95% CI: 0.90-1.05; p-value = 0.52), and weighted mode MR (OR: 0.96 per SD; 95% CI: 0.86-1.08; p-value = 0.56). The MR-Egger causal estimation produced an OR 1.11 per SD (95% CI: 0.93-1.31; p-value 0.2), the only test to show a direction of effect of increasing COPD causing an increased risk of Alzheimer's disease. There was no evidence of heterogeneity, with a Q-p-value 0.57.

3.5 Discussion

3.5.1 Evidence before this study

These results indicate that there is minimal evidence of a causal effect of reduced lung function or liability to COPD on the risk of Alzheimer's disease. This contrasts with two large observational studies,[28,29] which report a strong association between COPD and Alzheimer's disease. The observed associations may be due to unmeasured confounding by risk factors common to both COPD and Alzheimer's disease such as smoking, physical inactivity, social deprivation, and lower educational attainment.[116]

Apolipoprotein e4 allele is the biggest risk factor for Alzheimer's disease whereas it is thought that COPD affects cognition via vascular effects. There is evidence that COPD and reduced lung function is associated with micro and macrovascular damage that could mediate the relationship.[46,117,118] Different causes of dementia can be hard to differentiate. The observational studies may have inadvertently included other forms of dementia other than Alzheimer's disease, such as vascular dementia. Vascular dementia may be causally linked to COPD and lung function, but this outcome was not included in the analysis which was restricted to Alzheimer's disease only.

Cognitive dysfunction and Mild Cognitive Impairment are well described in COPD.[45] It may be that this association is causal, but that patients do not progress from these states to Alzheimer's disease due to their lung disease.

3.5.2 Impact of this study

This analysis uses two-sample MR to explore a causal association between lung function, COPD, and Alzheimer's disease. The increasing incidence of Alzheimer's disease in Western society has been described as an epidemic.[119] COPD is responsible for 5% of global disability-adjusted life years and 5% of total deaths.[120] Consequently, prevention and treatment of both COPD and Alzheimer's disease is a global health priority. Although there have been efforts to search for causal mechanisms linking the two diseases, this analysis using multiple means of assessing causation would suggest

scientific attention and health prevention resources may be better focused on overlapping risk factors such as smoking, diet, and physical activity,[121,122] rather than attempts to reduce risk of AD by improving lung function or reducing liability to COPD alone.

3.5.3 Strengths and Limitations

I used a large number of SNPs that are strongly associated with lung function and COPD, that were discovered in large samples using stringent significance thresholds.[75,109] A large number of exposure SNPs were removed due to clumping and more were removed because they were not found in the outcome GWAS. At the time of the original study, I did not check for proxies for exposure SNPs not found in the outcome GWAS. For my thesis minor corrections, I have repeated the analysis using proxies $r^2 \geq 0.8$. This has not changed interpretation of the results for any analysis e.g. for all lung function traits with proxies IVW OR: 0.98 (95% CI: 0.87 – 1.09, p-value 0.72).

MR Steiger was used in the main analysis, although it should be used as a sensitivity test. As per Figure 10, only 1 SNP was removed with MR Steiger, so it is unlikely to have caused a meaningful change in the estimates in this study. MR Steiger is a useful test to ensure the correct direction of SNP effect, it has limitations.[105] MR Steiger may provide the wrong direction of causality for several reasons leading to SNPs being falsely removed. Unmeasured confounding can have large influences on inferring causal direction, which can adversely affect the MR Steiger. Measurement error of the phenotypes can cause MR Steiger to provide the wrong direction of causality. If there is a large imbalance of power between the exposure and outcome GWAS, it can falsely be interpreted by MR Steiger as the wrong direction of effect.

It is important to ensure that the assumptions of MR are met when dealing with SNPs for complex phenotypes like lung function and COPD. None of the sensitivity tests provided strong evidence for a violation of the MR assumptions. The 15q25 locus is known to have strong associations with smoking behaviour, which could bias the results.[123] When reviewing the SNPs in the 2SMR analysis, only 6 of

the 279 SNPs are in chromosome 15, and none of them are in the region of concern.[75] When reviewing the COPD GWAS, only 4 of the 82 SNPs are in chromosome 15, and only one SNP from the COPD GWAS is in this locus (rs55676755).[109] However, this SNP was not found in the outcome GWAS so was not included in the analysis. Therefore, none of the SNPs used were from the 15q25 locus. COPD is a clinical diagnosis with set spirometric thresholds, whereas in the discovery GWAS a diagnosis of COPD was made based on spirometric criteria alone (i.e.. by dichotomising continuous traits). As discussed in the methods, COPD is a binary trait, so the SNPs confer liability to COPD. As this is a 2SMR study, it is not possible to know how many participants in the outcome population had COPD. The dichotomisation of continuous traits in MR studies can make interpretation of the causal estimate less reliable, but MR has still been shown to be a valid test of the causal null hypothesis for a binary exposure.[124]

Survivor bias occurs when selection into a study is dependent patients being alive to enable recruitment. Additionally selective drop out due to death of certain populations under follow up can lead to bias. This bias should be considered in studies involving potentially fatal diseases of later life.[125] Potentially, patients with COPD would be less likely to be recruited to a GWAS, biasing the MR towards a null result. Observational studies performed by analysing health records may be less likely to be affected by this. It is not possible to directly test for this in an MR study.

As the SNPs were discovered in populations of those with European ancestry, the results may not be generalisable to other populations.

3.5.4 Conclusions

No strong evidence that lung function and liability to COPD are causally associated with an increased risk of Alzheimer's disease was found in this study. Previous observational studies showing an

association between impaired lung function or COPD and Alzheimer's disease are most likely biased by unmeasured confounding.

CHAPTER 4. LUNG FUNCTION AND CARDIOVASCULAR DISEASE: A TWO-SAMPLE MENDELIAN RANDOMISATION STUDY

4.1 Publications and contributions

This chapter has been published.

Lung Function and cardiovascular disease: a two-sample Mendelian randomisation study. Daniel Higbee, Raquel Granell, Eleanor Sanderson, George Davey Smith, James Dodd. European Respiratory Journal. 2021 Sep 9;58(3)2003196.

DH and JD developed the concept of the study. DH and ES developed the data analysis plan. Dh performed the data analysis. GDS and JD provided supervision.

4.2 Introduction

Impaired lung function measures such as FEV_1 and FVC are strongly associated with extra-pulmonary disease and are reported as independent predictors of cardiovascular disease.[30] Although research has often focused on the contribution of FEV_1 and obstructive airways disease to cardiovascular risk, FVC is a stronger predictor of survival, and appears to add value to the Framingham Risk Score for prediction of mortality.[31,32] However, it is unclear if there is a causal link between lung function and extra-pulmonary disease, or if the association is due to confounding factors such as cigarette smoking.

Observational studies have reported that COPD, decreased FEV_1, FVC and FEV_1/FVC ratio are all associated with an increased risk of coronary artery disease.[33,34] However results are inconsistent, with some studies reporting no association,[35] or that the association is limited to those with

abnormally high blood pressure.[36] There is also evidence suggesting that COPD and impaired lung function are associated with increased risk of stroke.[37]

Impaired lung function and associated lung diseases could have a direct detrimental effect on cardiovascular health via different biological pathways including systemic inflammation or oxidative stress.[38,126] However the mechanisms may vary between different lung function traits.[127]

My objective was to determine if impaired lung function causally increases the risk of two cardiovascular diseases, coronary artery disease and stroke.

4.3 Methods

4.3.1 Exposure – Lung function traits, Shrine et al. preliminary analysis [75]

For the initial analysis I used data from the largest currently available lung function GWAS, by Shrine et al. to undertake a preliminary 2SMR analysis. The Shrine et al. GWAS reported 279 genome wide significant SNPs ($p<5\times10^{-9}$) in a European ancestry population and was adjusted for age, age^2, height and smoking status. Full details are provided elsewhere.[75]

4.3.2 Outcomes

4.3.2.1 Coronary artery disease

I used the largest published GWAS of coronary artery disease. The CARDIOGRAMplusC4D GWAS used 60,901 cases of coronary artery disease (CAD) and 123,504 controls, 77% of whom were of European ancestry.[128] CAD was defined as a history of myocardial infarction, acute coronary syndrome, chronic stable angina, or coronary stenosis of >50%.

4.3.2.2 Ischaemic stroke

For stroke I used MEGASTROKE GWAS based on 34,217 cases of acute ischaemic stroke and 406,111 controls. This was the largest published GWAS of stroke. All participants were of European ancestry.[129] There was no overlap between the exposure and outcome population samples.

4.3.3 Statistical Analysis

Statistical analysis was done using R Studio version 3.6.1 with MRCIEU/TwoSampleMR and MRInstruments packages.[105,130]

F-statistics were calculated to assess exposure instruments strength.[99] Linkage disequilibrium clumping (LD-clumping) and Steiger filtering were performed.[105] Duplicate SNPs and palindromic

SNPs were removed, and all SNPs were harmonised. Proxies were identified when CAD was the outcome.

4.3.4 Main Mendelian Randomisation Analysis

Inverse Variance Weighting (IVW) was used for main effect estimate for 2SMR analyses. Weighted median, weighted mode and MR Egger were all used as sensitivity analysis.

4.4 Preliminary Results

4.4.1 Shrine et al. preliminary analysis

Analyses showed lung function had weak evidence of an effect on both the risk of coronary artery disease (**Table 3**) and ischaemic stroke (**Table 4**), with variable direction of effect and wide confidence intervals.

Table 3. Results of decreasing lung function traits on risk of coronary artery disease

		FEV_1, FVC, FEV_1/FVC, PEF	FEV_1	FVC	FEV_1/FVC
No. SNPs used		173	60	67	93
IVW	OR per SD	0·95	1·14	1·01	0·90
	(95% CI)	(0·88 – 1·04)	(0·94 – 1·37)	(0·86 – 1·18)	(0·82 – 0·99)
	Q_p-value*	3.7×10^{-17}	3.07×10^{-8}	$3\cdot12\times10^{-6}$	$1\cdot1\times10^{-6}$
Weighted Median	OR per SD	0·95	1·24	0·98	0·96
	(95% CI)	(0·86 – 1·05)	(1·01 – 1·53)	(0·82 – 1·16)	(0·85 – 1·07)
Weighted Mode	OR per SD	0·92	0.80	0·93	0·96
	(95% CI)	(0·78 – 1·08)	(0·43 – 1.05)	(0·66 – 1·31)	(0·81 – 1·15)
MR-Egger	OR per SD	0·90	1·06	0·94	1.10
	(95% CI)	(0·72 – 1·12)	(0·57 - 1·96)	(0·51 – 1·74)	(0·86 – 1·42)

Table 3 shows the Odds Ratio of coronary artery disease per SD decrease in lung function traits

Table 4. Results of decreasing lung function traits on risk of ischaemic stroke

		FEV_1, FVC, FEV_1/FVC, PEF	FEV_1	FVC	FEV_1/FVC
No. SNPs used		171	58	77	93
IVW	OR per SD	1·05	1·01	1·04	1·04
	(95% CI)	(0·98 – 1·13)	(0·87 – 1·18)	(0·91 – 1·19)	(0·95 – 1·13)
	Q_p-value	2.2×10^{-5}	0.005	0.04	0.01
Weighted Median	OR per decrease SD	1·05	1·05	1·11	1·02
	(95% CI)	(0.95 – 1·64)	(0·87 – 1·28)	(0·93 – 1·32)	(0·91 – 1·14)
Weighted Mode	OR per SD	1·08	1·07	1·06	1·02
	(95% CI)	(0·93 – 1·26)	(0·77 - 1·51)	(0·77 – 1·46)	(0·85 – 1·12)
MR-Egger	OR per SD	1·13	0·91	1·47	1·05
	(95% CI)	(0·93 – 1·36)	(0·55 - 1·49)	(0·87 – 2·48)	(0·84 – 1·31)

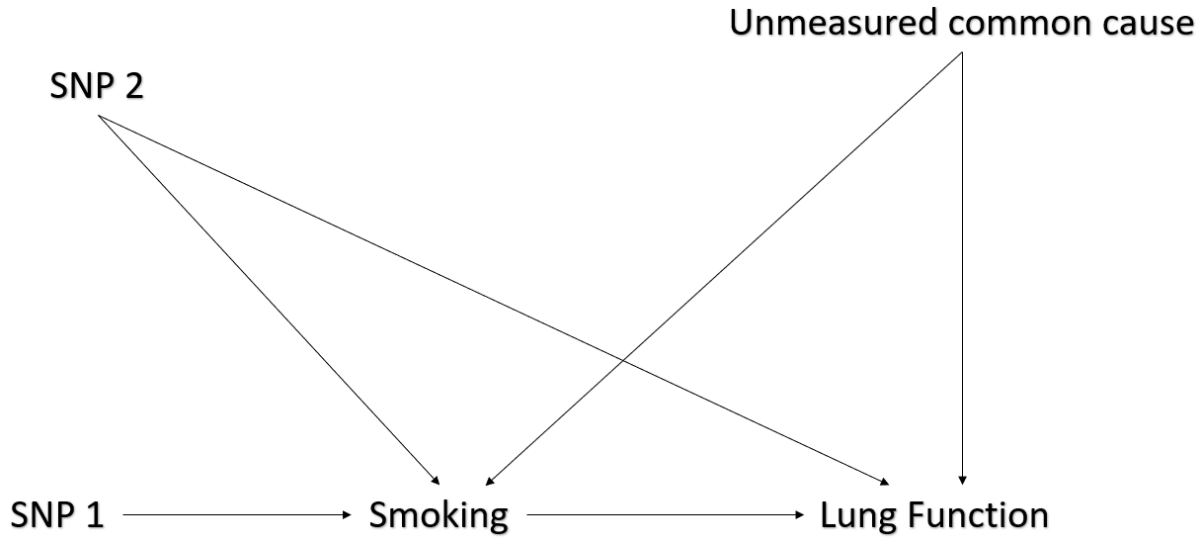
Table 4 shows the Odds ratio of ischaemic stroke per SD decrease in lung function traits

4.5 Collider bias

I was surprised to find no good evidence of an effect of lung function on cardiovascular disease given the wealth of observational data suggesting a strong association. It was suggested that collider bias could be influencing the results.

GWAS for lung function and COPD are often adjusted for heritable covariates such as height, smoking, and body mass index (BMI). These factors are also covariates for other disease, including cardiovascular disease. Adjustment of covariates in GWAS has been shown to affect the SNP-exposure estimate leading to bias in MR studies. **Figure 14** and legend explain some possible pathways.

Figure 15. Directed acyclic graph demonstrating possible pathways leading to bias



Smoking is used in this example, but it would be true of other heritable covariates, such as height or BMI. Figure 14 shows how adjusting for smoking can lead to SNP being incorrectly identified as having an association with lung function, as well as biasing the estimate of a SNPs association with lung function if the SNP was associated with both smoking and lung function

SNP 1 does not have a direct effect on lung function. However, it has an indirect effect via smoking. Smoking and lung function have unmeasured common causes. Therefore, smoking is a collider of the path of SNP 1 and lung function (SNP 1 → Smoking ← Unmeasured common causes → Lung function). Adjusting on a collider opens the path on the collider, which means that adjusting for smoking in the lung function GWAS could wrongly identify SNP 1 as having a direct effect on lung function.

SNP 2 has a direct effect on lung function, but also an indirect effect via smoking. Adjusting for smoking in the lung function GWAS will cause a biased estimate for the direct effect of SNP 2 on lung function (as it would be a combination of the true direct effect and the bias from the collider adjustment of smoking).

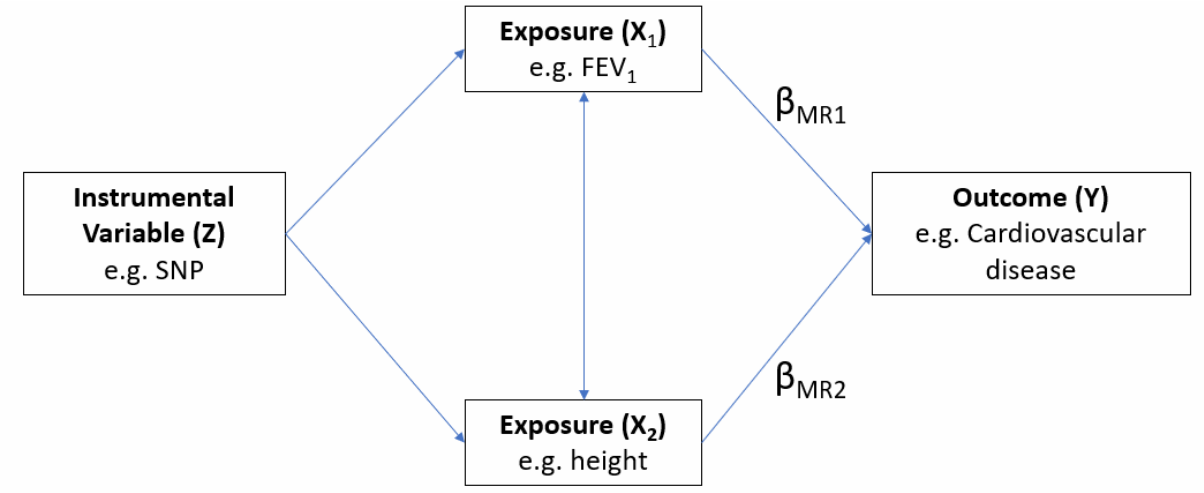
To eliminate the collider bias induced by covariate adjustment, the GWAS would need to be adjusted for all measured and unmeasured common causes. Not only is this impractical, but it is impossible as by definition some covariates are unmeasured. If there is any residual confounding of the exposure-

covariate/exposure-outcome/covariable-outcome relationship, then covariate adjustment can bias the MR estimate. This is particularly noticeable if there is bias of the covariable-outcome relationship. There is a high chance of this in this study where the covariates of height, smoking and BMI have a strong effect on cardiovascular disease. The degree of bias will depend on the underlying causal structure and can exist in many possibilities beyond what is shown in **Figure 14**. This reference explains in further detail.[131] To surmount this issue, a GWAS that has not been adjusted for heritable covariates can be used in a multivariable MR model.

4.5.1 Multivariable Mendelian Randomisation

Multivariable MR (MVMR) can be used to model the effect of two or more exposures on an outcome, as seen in **Figure 15**. This can be used in situations where a secondary exposure is a confounder, collider, mediator, or on the pleiotropic pathway between an exposure and an outcome.[132] MVMR estimates the direct effect of each exposure in the model, conditional on the other exposures. This has the benefit of determining the direct effect of each exposure trait, rather than the total effect a pleiotropic SNP may have via multiple exposures. In **Figure 15**, if non-multivariable MR was used for the pleiotropic SNP, the causal estimate calculated would be the total combined effect of both β_{MR1} and β_{MR2}. MVMR will estimate the direct effect of each exposure (β_{MR1} and β_{MR2} separately) on the outcome, after conditioning for the other exposure.[133]

Figure 16. DAG demonstrating how MVMR can be used to estimate the direct effect (βMR) of two exposures by conditioning their effects



This DAG demonstrates the how MVMR works for a pleiotropic SNP, allowing conditioning for each exposure and resulting in a direct effect estimate of each exposure on the outcome

I utilised this to avoid collider bias from using GWAS adjusted for heritable covariates of the exposure and outcome described above. I conducted a GWAS of lung function adjusted only for sex (sex cannot be a collider) and then used the results in a multivariable model conditioning for the covariates height, smoking and BMI. Each SNPs effect on lung function (X_1 in **Figure 15**), is conditioned with its effect on the covariate trait (X_2). This process occurs vice versa, so each SNP used from covariate trait GWAS is conditioned on its effect on lung function.

This was the first time an MR study had examined lung function in an MVMR model as an exposure in this way.

As MVMR reduces collider bias and gives an estimation of the direct effect of lung function, it is more reliable and accurate than the preliminary analysis using the Shrine et al. GWAS. Therefore the MVMR analysis should be considered the main analysis.

4.6 Exposures – MVMR

To avoid collider bias I required exposure SNPs discovered in GWAS that had not been adjusted for covariates. To find suitable exposure SNPs the UKBiobank was used.[68] 353,315 participants have

"best measures" of pre-bronchodilator FEV_1 and FVC, measured as absolute values in litres. I performed a GWAS on these individuals (adjusting for sex). I also performed a GWAS based on 55,907 cases of airflow obstruction (defined as FEV_1/FVC <0.70) and 297,408 controls (FEV_1/FVC ≥0.70). The SNPs discovered in this unadjusted GWAS were then used in a two-sample MVMR model conditioning with the SNPs effect on covariates of exposure and outcome: standing height, BMI, and current smoking. SNPs for these covariates were identified in pre-existing GWAS performed in the UKBiobank.[134]

4.7 MVMR Results

Using a threshold of $p<5\times10^{-8}$, after quality control and LD-clumping the unadjusted GWAS of lung function in UKBiobank identified 360 SNPs for FEV_1, 464 SNPs for FVC and 154 SNPs for FEV_1/FVC <0.70 explaining 3.6%, 4.8% and 0.9% of variance respectively. F-statistic for FEV_1 = 38, FVC = 40 and Ratio <0.70 = 36. For covariates, F-statistic for standing height, BMI and current smoking were 50, 39 and 32 respectively.

4.7.1.1 MVMR results – FVC and FEV_1 as exposure, CAD as outcome

Results are presented as per SD decrease in lung function trait. Analysis showed strong evidence of increased risk of CAD per SD decrease in FVC (OR:1.32 per SD; 95% CI; 1.19-1.46) as shown in **Table 5**. This effect did not attenuate after conditioning for BMI (1.41; 1.25-1.59) or current smoking (1.32; 1.19-1.47). The effect size weakened after conditioning for height (1.22; 1.03-1.44) and all covariables together (1.44; 1.18-1.76) but strong evidence for an effect remained.

Table 5. Multivariable MR results of FEV1 and FVC on Coronary Artery Disease and Ischaemic Stroke using UKBiobank lung function GWAS

Lung function trait	Condition	No. SNPs (LF/condition)	OR (95% CI)* for Coronary Artery Disease	No. SNPs (LF/condition)	OR (95% CI)* for Ischaemic Stroke
FEV_1	Nil	300/Nil	1.27 (1.12, 1.44)	291/Nil	1.11 (0.97-1.26)
FEV_1	Height	194/744	1.08 (0.89, 1.30)	193/741	1.01 (0.83, 1.22)
FEV_1	BMI	179/645	1.26 (1.08, 1.47)	185/660	1.03 (0.88, 1.20)
FEV_1	Smoking	274/15	1.26 (1.10, 1.44)	273/12	1.11 (0.95, 1.29)
FEV_1	Height/BMI/Smoking	80/432/391/4	1.28 (1.02, 1.61)	85/440/413/3	1.18 (0.94, 1.48)
FVC	Nil	391/Nil	1.32 (1.19-1.46)	384/Nil	1.12 (1.01-1.24)
FVC	Height	272/726	1.22 (1.03, 1.44)	273/728	1.04 (0.88, 1.24)
FVC	BMI	227/599	1.41 (1.25, 1.59)	227/607	1.05 (0.93, 1.19)
FVC	Smoking	359/15	1.32 (1.19, 1.47)	368/11	1.11 (1.00, 1.23)
FVC	Height/BMI/Smoking	105/406/388/4	1.44 (1.18, 1.76)	102/408/399/3	1.05 (0.86, 1.29)

*Per SD decrease in lung function trait

OR – Odds Ratio. 95% CI – 95% Confidence Interval. LF – Lung Function

Prior to any conditioning, there was evidence that reduced FEV_1 increases risk of CAD (OR: 1.27 per SD; 95% CI: 1.12-1.44). However, when conditioning for height the effect size decreases and the confidence interval widens (1.08; 0.89-1.30) **Table 5**. This is probably due to the pleiotropy in the MR analysis as the unadjusted GWAS would have discovered SNPs that affected lung function (LF) via height. Therefore, there is limited evidence of a direct effect of FEV_1 on cardiovascular risk. Conditioning for BMI (1.26; 1.08-1.47) and current smoking (1.26; 1.10-1.44) made minimal difference to the estimated effect. When conditioning for all covariables together, there is some evidence that lower FEV_1 increases the risk of coronary artery disease 1.28 (1.02, 1.61). However, the point estimate and upper limit of the confidence interval is higher for FEV_1 conditioned with all covariables than the estimate for unconditioned FEV_1, indicating a possible problem with the model. It is worth noting the reduction in SNPs for FEV_1, as well as all the covariables, available for analysis

due to SNPs being removed when clumping, or not being found in the other GWAS. This reduces the power making the analysis less reliable. Additionally in this model, each covariable conditions on all the other covariables. This will reduce the conditioning effect of each individual covariable.

4.7.1.2 MVMR analysis – FEV_1 and FVC as exposure, ischaemic stroke as outcome

There is little evidence to suggest that lower FEV_1 increases the risk of ischaemic stroke (OR: 1.11 per SD; 95% CI: 0.97-1.26) **Table 5**. The magnitude of the effect decreased further when conditioning for both height and BMI, although the direction remained consistent. There is evidence that decreased FVC increases risk of ischaemic stroke (1.23; 1.01-1.24) but the effect size and strength of evidence attenuates after conditioning for height or BMI or all covariables together (1.16; 0.98-1.38 and 1.05; 0.93-1.19 and 1.05; 0.86-1.29 respectively).

4.7.1.3 MVMR analysis – FEV_1/FVC ratio <0.7 as exposure, CAD and ischaemic stroke as outcomes

Steiger filtering removed 87 SNPs for FEV_1/FVC ratio <0.70 with CAD as the outcome and 96 SNPs with ischaemic stroke as the outcome. I found very little evidence of an effect of liability to airflow obstruction on cardiovascular disease (CVD) as can be seen in **Table 6**.

Table 6. Multivariable MR results of and FEV1/FVC <0.7 on Coronary Artery Disease and Ischaemic Stroke using UKBiobank lung function GWAS

Trait	Condition upon	No SNPs (LF/condition)	OR (95% CI)* for Coronary Artery Disease	No. SNPs (LF/condition)	OR (95% CI)* for Ischaemic Stroke
FEV1/FVC <0.7	Nil	50/Nil	1.00 (0.60, 1.67)	39/Nil	0.96 (0.52, 1.79)
FEV1/FVC <0.7	Smoking	49/17	1.00 (0.83, 1.21)	38/13	0.98 (0.82, 1.16)

*Per SD increase in liability to ratio <0.7

4.8 Discussion

This MVMR study provides evidence that a one standard deviation lower FVC *causes* approximately a 20% increased risk of CAD. This finding provides further evidence for causality of previous observational associations.[33,34] These results are unlikely to be affected by reverse causation or confounding factors due to the use of SNPs as instrumental variables. This effect was not seen in the preliminary non-MVMR analysis because of collider bias introduced in the model by covariate adjustment in the Shrine et al. discovery GWAS. The main analysis used MVMR which is a robust tool when a secondary exposure acts as a confounder, a mediator, a pleiotropic pathway, and/or a collider.[132]

Although historically, most observational studies of cardiovascular morbidity have focused on FEV_1 and COPD, there was little evidence of a causal association between FEV_1 and liability to obstructive ratio on CVD risk. These results mirror findings that FVC is a stronger predictor of overall survival than FEV_1.[31] The findings suggest that the observed association between low FEV_1, obstruction, and increased risk of CVD is unlikely to be causal. In healthy individuals, FEV_1 and FVC are highly correlated. Therefore, I hypothesise that the unknown underlying biological mechanism linking lung function and cardiovascular disease may be specific to FVC reduction.

Finding modifiable risk factors for CAD is important, however, most therapies designed to improve lung function (such as inhaled bronchodilators) have a temporary and limited impact on FVC and so are unlikely to be sufficient to modify cardiovascular risk. Available treatments which do target decline in FVC are for specific and rare lung diseases such as pulmonary fibrosis.[135]

There are numerous strengths to this study. First, it utilises large numbers of instrumental variables, far more than were available in previous MR studies.[136] Secondly, a large exposure sample population and multiple robust methods. In a robust study, MR avoids unmeasured confounding and reverse causation, problems typical of conventional observational epidemiology and estimates causality by the use of randomly assigned genetic instrumental variables.[87,137,138] In addition, this

study benefited from using MVMR to condition for these covariates avoiding collider bias that could have contributed to the weak evidence found in the preliminary analysis using the Shrine et al. GWAS.[75] MVMR estimates the direct, rather than total effect of an exposure allowing us to show that much of the effect of FEV_1 on CAD risk was due to pleiotropic SNPs affecting FEV_1 via height (an established determinant of cardiovascular risk). Finally, this is the first study to use SNPs for FEV_1/FVC <0.7 ratio. MR has assumptions and is vulnerable to certain biases if not used properly. The sensitivity analysis in the pre-liminary analysis using plots, MR Egger, weighted median, and mode did not indicate any violation of assumptions.

4.8.1 Limitations

The exposure GWAS and the outcome GWAS (MEGASTROKE) used only those of European heritage. The CARDIOGRAMplusC4D GWAS was 23% non-European heritage. Lung function SNPs discovered in European ancestral populations in the Shrine et al. GWAS have been shown to have a smaller effect in non-European populations.[75] I did not test to see if the SNPs I used in the MVMR model had a similar smaller effect in non-European populations. However as both the Shrine et al. GWAS and my unadjusted GWAS used a high proportion of the same sample population and examined similar traits, it is likely that in a non-European population the effects of the unadjusted SNPs would be smaller. Therefore, these results may not be generalisable to non-European populations. As the SNPs in the MVMR analysis were discovered and effects estimated in the same population, the effects could have been over-estimated due to "Winner's Curse" phenomena.[139] There was a reduction in the number of instruments available for analysis following LD-clumping, removal of duplicates, and extraction from exposure and outcome GWAS. This reduces the strength of the instruments which may have reduced the power to show an effect of FEV_1 or FEV_1/FVC <0.7 ratio. In the MVMR analysis I used FEV_1/FVC <0.7 ratio as an exposure because this is a commonly used threshold of obstructive lung function. Using FEV_1/FVC ratio as a continuous trait has inherent issues in MR analysis. High FEV_1/FVC ratio is a sign of restriction and a low FEV_1/FVC ratio defines airflow

obstruction, both of which are pathological states that could affect cardiovascular disease, making interpretation of the continuous variable challenging. Most MR analysis assumes a linear effect, which would be violated when using FEV_1/FVC as a continuous trait. Dichotomization of continuous traits in MR studies can make interpretation of the causal estimate less reliable, but MR can still be a valid test of the causal null hypothesis for a binary exposure.[124] An assumption of MR is that SNPs only affect the outcome via the exposure. To ensure that the SNPs were not affecting the outcomes via smoking I checked to see if any of the lung function SNPs were found in the 15q25 locus (a locus known to have strong associations with smoking behaviour).[140] In the MVMR analysis for FEV_1 only one SNP (rs72736802) is from the locus, none from the FVC analysis. Therefore, I do not think this will affect the results. Lung function is a complex trait and SNPs affect LF via differing pathological processes.[75] The differing processes may vary in their impact on the risk of co-morbidities, perhaps reflected in the assessments of heterogeneity. It is possible the study was limited by the number of ischaemic stroke cases in the outcome population. If there is a causal effect of lung function on ischaemic stroke, it is likely to only occur with large decline in lung function as seen with CAD.

4.8.2 Implications

There are several important implications of these findings. First is that it is FVC not obstructive lung function that is causally associated with coronary artery disease. This suggests that attention should be focused on understanding the mechanisms by which FVC causes CAD. Secondly, given there are limited FVC specific therapies, it is most likely that future interventions to improve CAD outcomes through modifying FVC are most likely to be achieved through environmental/ behavioural public health interventions designed to achieve optimal lung development and preventing lung function decline. Third, FVC is a widely and routinely collected clinical measure (spirometry), this study supports the call for FVC measurements to be evaluated as part of cardiovascular prognostication / secondary prevention risk assessments.

It remains uncertain if lung function has a causal effect on the risk of ischaemic stroke. The MVMR models show very little evidence that reduced lung function increases the risk of ischaemic stroke. Larger outcome sample sizes may become available as genetic consortia grow which could provide more conclusive results. Future studies are needed to determine the mechanism by which FVC causes increased coronary artery disease.

This study highlights issues of collider bias in MR studies when using GWAS adjusted for heritable covariates of the exposure-outcome relationship. MVMR should be considered for future MR studies that may be affected by these issues.

4.8.3 Conclusions

There is strong evidence that reduced FVC is independently and causally associated with coronary artery disease. Although the mechanism remains unclear, FVC may play an important contribution in the assessment of cardiovascular risk. Further studies are needed to test whether interventions to improve or maintain FVC may also modify cardiovascular risk. FEV_1 and obstructive lung function do not appear to cause increased cardiovascular events, confounding and collider bias may explain previous observational and MR findings of a causal association.

CHAPTER 5. LUNG FUNCTION, COPD AND COGNITIVE FUNCTION: A MULTIVARIABLE AND TWO SAMPLE MENDELIAN RANDOMISATION STUDY

5.1 Publication and contributions

This chapter has been published.

Lung function, COPD, and cognitive function: a multivariable and two sample Mendelian randomisation study. Daniel Higbee, Raquel Granell, Gibran Hemani, George Davey Smith, James Dodd. BMC Pulmonary Medicine. 2021 Jul 22;21(1):246.

DH and JD developed the concept. DH planned and performed the analysis plan. GH helped format an external GWAS data for use. GDS and JD provided supervision.

5.2 Introduction

Cognitive function impacts important physical and mental health outcomes including mortality and educational attainment.[39] It exists on a continuum from normal cognitive function to the potentially reversible state of mild cognitive impairment (MCI), which can lead to irreversible dementia.[40,41] The theory of cognitive reserve suggests that those that reach higher levels of cognitive function are more resilient to the development of cognitive ageing or impairment and dementia.[141,142] There are very limited therapeutic options that effectively increase cognitive function or treat MCI, so finding modifiable risk factors to enable people to attain and maintain maximum cognitive function is important.

Co-morbid lung disease and cognitive impairment is associated with worse health outcomes, quality of life, and increased healthcare utilisation.[42] The association between lung function and cognition may be due to shared risk factors seen more commonly in those with lung disease e.g. smoking.[43,44]

However associations independent of these factors mediated through plausible causal pathological pathways such as hypoxia, hypercapnia, or chronic lung disease associated inflammation may also cause extra-pulmonary end organ damage.[45,48]

Neuroimaging provides further evidence of a relationship between lung function and cognition. After adjustment for smoking, reduced lung function remains associated with white matter brain lesions, [46] and a 'dose response' like relationship is seen between severity of lung function deficit and risk of cognitive impairment.[47]

My objective was to use MR and MVMR to determine if lung function or liability to COPD causes lower general cognitive function. If the relationship is shown to be causal, then interventions to treat lung function could also be used to reduce subsequent cognitive decline or dementia.

5.3 Methods

5.3.1 Exposure populations

I used the Shrine et al. lung function GWAS and the COPD GWAS performed by Sarkonsakaplat et al. as exposure populations. These GWAS are detailed in section 3.3.

The lung function GWAS was adjusted for covariates of lung function and cognition e.g. height and smoking.[43,143] As discussed in chapter 4 this adjustment can lead to collider bias as SNPs can be related to the covariates e.g. height, or other adverse risk factors.[83] This can result in misleading SNP effect estimates and subsequent bias in MR studies.[144] To avoid these types of bias I performed MVMR using exposure SNPs discovered in GWAS that had not been adjusted for covariates. To find suitable exposure SNPs I performed a GWAS (adjusting for sex) based on 353,315 UK BioBank participants for "best measure" FEV_1 and FVC. The SNPs discovered in the unadjusted GWAS were used in a two-sample MVMR model conditioning on SNPs for covariates: standing height, body mass index (BMI), current smoking. Educational attainment is a significant determinant of cognitive

function, but it was not adjusted for in the outcome GWAS.[145] Therefore, I used SNPs identified in a GWAS of educational attainment in the UK Biobank. Educational attainment was determined by the by asking UK Biobank participants, "At what age did you complete your continuous full time education?". Participants were asked this via a touchscreen during the initial assessment. By conditioning for educational attainment, I determined the direct effect of lung function on cognitive function. SNPs for lung function and cognitive function were found in pre-existing GWAS performed in the UK Biobank.[134]

5.3.2 Outcome Population

Data from a meta-analysis of the Cohort of Heart and Aging Research in Genomic Epidemiology (CHARGE) and Cognitive Genomics Consortium (COGENT) was used.[39] UK Biobank participants were excluded from the analysis to ensure no overlap of exposure and outcome populations. This resulted in an outcome population of 132,452. The CHARGE and COGENT cohorts had used a wide range of cognitive tasks to test for different cognitive domains. The authors of the outcome GWAS constructed a general cognitive function phenotype using these different cognitive tasks. The general cognitive function phenotype required the consortiums to have tested at least three different cognitive domains. As different tasks were used across the different sample populations, a consistent method of extracting general cognitive function was used. Principal component analysis was applied to the cognitive test scores to derive a measure of general cognitive function. The authors avoided taking more than one cognitive score from any individual cognitive test. A GWAS meta-analysis adjusting for age, sex and population stratification was performed using the derived measure of general cognitive function. Therefore, the outcome used in the analysis represents general cognitive function and is not domain specific. For full details of the tasks and samples see the supplementary information of the referenced GWAS paper.[39]

Both exposure and outcome populations were in those of European ancestry.

5.3.3 Statistical Analysis

Statistical analysis was done using R Studio version 3.6.1 with MRCIEU/TwoSampleMR and MRInstruments packages.[105,130]

F-statistics were calculated to assess exposure instruments' strength (F statistic = beta^2/standard error^2) to determine the likelihood of weak instrument bias. [99] For all exposures SNPs, LD-clumping was performed using European reference population and the ieugwasr:ld_clump tool and Steiger filtering was performed.[105] Duplicate SNPs were removed. Palindromic SNPs (i.e. A/T and C/G SNPs) with intermediate allele frequencies were excluded from the analysis to ensure that the correct allele was being used. All SNPs were harmonised to ensure the exposure and outcome alleles were the same.

Inverse Variance Weighting (IVW) was used for the main effect estimate for all analyses. MR Egger, weighted median, and mode were all used as sensitivity tests to account for horizontal pleiotropy. Heterogeneity was calculated and assessed with a Q statistic, presented as a Q_P-value.

5.3.3.1 MR Radial

I used MR Radial as a sensitivity analysis. MR Radial is a way of using a statistical method to identify and remove outliers, as opposed to visually identifying outliers using graphs as discussed in the previous chapters. MR Radial excluded SNPs that contributed more than 5% heterogeneity to the model and re-estimates the IVW [146] If SNPs are causing considerable heterogeneity this could indicate a violation of the second and third MR assumptions. I only became aware of MR radial after I had completed the studies in the previous chapters.

5.4 Results

5.4.1 Two Sample MR, effect of lung function on cognitive function

The analysis using the adjusted lung function GWAS in a 2SMR lung function analysis is reported in **Table 7**.[75] The F statistic for all SNPs combined for each lung function trait were: All traits = 111, FEV_1 = 69, FVC = 70, FEV_1/FVC = 148, making weak instrument bias unlikely.

The effect size is given as the beta (β) value, which is the coefficient from the regression analysis, and is the change in the general cognitive function score per standard deviation (SD) decrease in lung function measure. If the beta is negative, it means there is a decrease in cognitive function with a decrease in lung function.

The predictive causal effects show reduced FEV_1 and FVC reduced cognition across all tests, but the evidence was weak (**Table 7**) as demonstrated by the SE being larger than the corresponding beta and the high p-values. All lung function measures combined did not show a consistent direction of effect across the tests used. There was strong evidence of heterogeneity of effect based on the Q_P-value, especially when assessing all measures combined. However, there were no visual outliers as assessed by leave-one-out, single-SNP and funnel plots.

Table 7. 2SMR, decreasing lung function effect on cognitive function

			Cognitive function		
Lung Function Measure	**No. of SNPs**	**Test used**	**Beta (SE)**	**P-value**	**Q_P-value**
FEV_1, FVC, FEV_1/FVC, PEF	173	IVW	-0.002 (0.02)	0.86	1.44×10^{11}
		Weighted Median	0.02 (0.02)	0.44	
		Weighted Mode	-0.004 (0.03)	0.91	
		MR Egger	-0.005 (0.03)	0.90	
FEV_1	59	IVW	-0.05 (0.04)	0.44	1.8×10^{-8}
		Weighted Median	-0.02 (0.04)	0.46	
		Weighted Mode	-0.05 (0.08)	0.51	
		MR Egger	-0.09 (0.1)	0.43	
FVC	68	IVW	-0.004 (0.03)	0.87	0.01
		Weighted Median	-0.02 (0.04)	0.51	
		Weighted Mode	-0.03 (0.07)	0.65	
		MR Egger	-0.16 (0.10)	0.13	
FEV_1/FVC	93	IVW	0.01 (0.02)	0.52	0.01
		Weighted Median	0.02 (0.02)	0.52	
		Weighted Mode	0.007 (0.05)	0.87	
		MR Egger	0.02 (0.04)	0.67	

SE – Standard Error. Q_*P*-value – A measure of Heterogeneity (*P*-value <0.05 provides strong evidence of heterogeneity). Negative beta indicates decreasing cognitive function.

5.4.2 Two Sample MR, effect of COPD on cognitive function

The F statistic for the COPD SNPs combined is 52, making weak instrument bias unlikely. There is a consistent direction of effect that increased liability to COPD causes lower cognition, however, the strength of evidence is weak (**Table 8**).

Table 8. 2SMR, COPD effect on cognitive function

			Cognitive Function		
Lung Function Measure	**No. of SNPs**	**Test used**	**Beta (SE)**	**P-value**	**Q_*P*-value**
COPD	67	IVW	-0.008 (0.008)	0.35	0.005
		Weighted Median	-0.01 (0.01)	0.16	
		Weighted Mode	-0.03 (0.02)	0.28	
		MR Egger	-0.01 (0.02)	0.66	

Negative beta indicates decreasing cognitive function.

5.4.3 MVMR Analysis

Using a threshold of $p<5\times10^{-8}$, quality control and clumping the unadjusted GWAS of lung function in UK Biobank produced 360 SNPs for FEV_1 and 464 SNPs for FVC explaining 3.6% and 4.8% of variance, respectively. I tested the effect of the covariate's height, BMI, educational attainment, and current smoking on the outcome of cognitive function. Height (Beta 0.05 (SE 0.01), p-value <0.001), BMI (-0.1 (0.01), p-value <0.001) and educational attainment (0.3 (0.07), p-value <0.001) showed strong evidence for an effect on cognitive function. The SNPs discovered in the GWAS of current smoking only had very weak evidence of an effect on cognitive function (-0.04 (0.2), p-value 0.84), so they were not included in the analysis. The F-statistic were FEV_1 = 38, FVC = 40, height = 50, BMI=39 and age completed full time education = 1350, making weak instrument bias unlikely.

Results were calculated per SD decrease in lung function measure and are displayed in **Table 9**. There was strong evidence that reduced FEV_1 (-0.06 (0.03), p-value <0.001) and FVC (-0.06 (0.01), p-value <0.001) causes lower general cognition, however when using MVMR and conditioning with educational attainment the evidence became weak. Evidence also became weak after conditioning for height. This is probably due to the pleiotropy in the MR analysis as the unadjusted GWAS would have identified SNPs that affected LF via height. Educational attainment is likely to be a mediator on the pathway of lung function and cognitive function (lung function → educational attainment → cognitive function) rather than a confounder (lung function ← educational attainment → cognitive function). Therefore, by using MVMR it removes the mediating effect of lung function on cognitive function via educational attainment. This demonstrates a potential pathway how lung function could affect cognitive function. However, as conditioning for height gives weak evidence of an effect of lung function on cognitive function it remains unclear if there is a true causal relationship between lung function and cognitive function, via educational attainment or other pathways.

Table 9. MVMR Analysis results. Effect of decreasing Lung function trait on cognitive function with and without conditioning for covariates

			Cognitive Function	
Lung function measure	**Condition**	**No SNPs (Trait/Condition)**	**Beta (SE)**	**P-value**
FEV_1	None	298/0	-0.06 (0.03)	<0.001
FEV_1	Height	212/299	-0.03 (0.03)	0.29
FEV_1	BMI	180/629	-0.07 (0.03)	0.03
FEV_1	Educational attainment	274/32	-0.03 (0.03)	0.33
FVC	None	381/0	-0.06 (0.01)	<0.001
FVC	Height	278/314	-0.01 (0.03)	0.62
FVC	BMI	224/595	-0.05 (0.02)	0.004
FVC	Educational attainment	335/31	-0.01 (0.02)	0.35

SE – Standard Error. Negative beta indicates decreasing cognitive function.

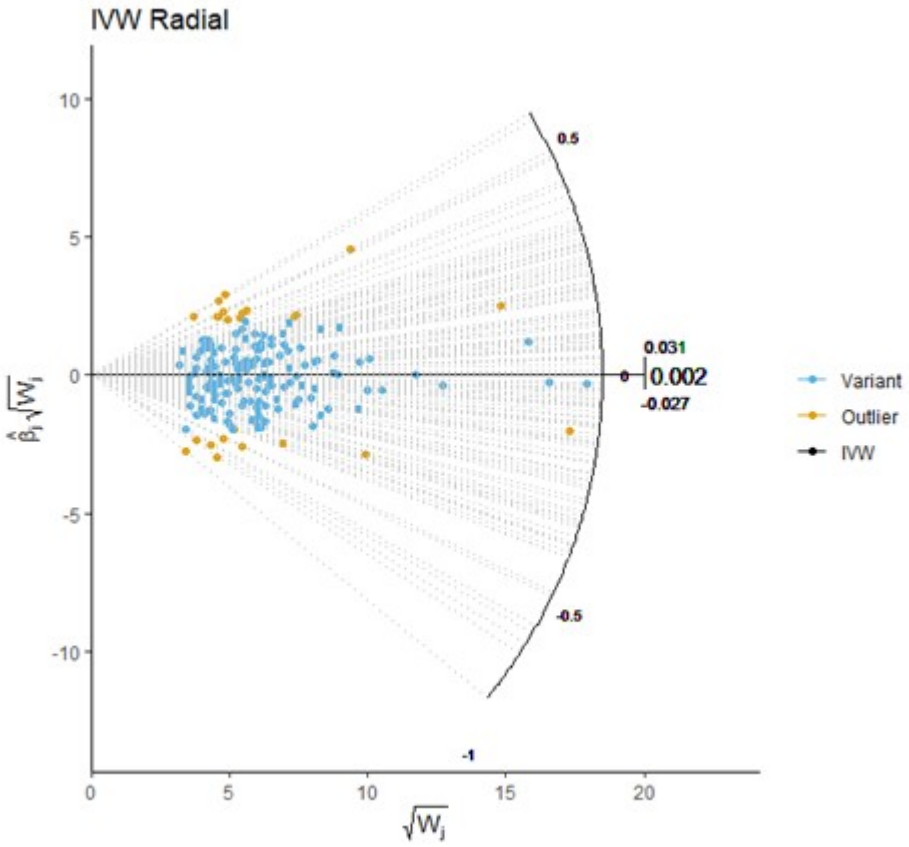
5.4.4 MR Radial

Exclusion of outliers using MR-Radial only minimally changed effect estimates (**Table 10 and Figure 16**).

Table 10. Results of 2 Sample MR all lung function SNPs (Shrine et al GWAS) effect on cognition using MR Radial after outlier exclusion

	All traits (Beta (SE))	P-value
IVW (2^{nd} order weights)	-0.002 (0.01)	0.90
MR Egger	-0.02 (0.03)	0.58

Figure 17. MR Radial plots for 2 sample MR all lung function SNPs (Shrine et al GWAS) effect on cognition



Yellow dots represent outlier SNPs removed as they explained over 5% heterogeneity. Blue dots represent SNPs kept in IVW Radial analysis. IVW causal estimate is close to zero with SE higher than corresponding beta and high p-values indicating no strong evidence of an effect.

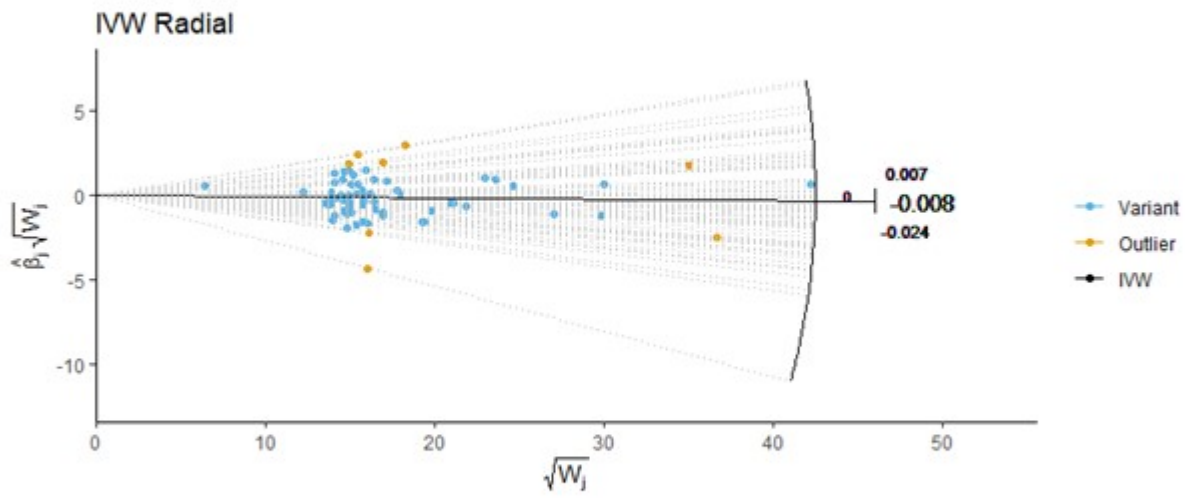
5.4.1 Results for 2SMR COPD effect on cognition using MR Radial

There was a minimal change in effect estimates after using MR-Radial to exclude outliers and recalculation of IVW (**Table 11 and Figure 17**).

Table 11. Results of 2SMR COPD effect on cognition using MR Radial after outlier exclusion

	All traits (Beta (SE))	P-value
IVW (2nd order weights)	-0.008 (0.007)	0.30
MR Egger	-0.001 (002)	0.95

Figure 18. MR Radial plot for 2SMR COPD effect on cognition



The radial plot shows no evidence of effect after removal of outliers

5.5 Discussion

The analyses show weak evidence that lung function or liability to COPD causes lower general cognitive function. Most of my MR causal estimates show the same direction of effect as observational studies, with lower lung function and liability to COPD causing lower general cognitive

function. However, the evidence is weak. The previously observed association in traditional epidemiology papers may be secondary to residual confounding and collider bias. There may be a genuine association between lung function and COPD with general cognitive function, but the relationship is unlikely to be causal.[45,47,52,147-149]

This analysis suggests that shared risk factors are likely explanations for the observed association between lung function and cognition, for example cigarette smoking. The short term effects of smoking on cognitive function are complex with acute nicotine consumption improving smokers' cognition, and nicotine abstinence decreasing cognition.[43] Longitudinal research has shown that lower childhood IQ is associated with an increased risk of smoking, and smokers have significantly worse cognition scores in old age than ex- or never-smokers.[44] Therefore, public health measures to reduce rates of smoking could improve both lung function and cognitive function.

This study examines the effect of lung function and COPD on general cognitive function in a general adult population. COPD has been shown to be associated with general cognitive function, but often more specific patterns of cognitive impairment including attention, memory, learning, and motor function domains.[45] Therefore it may be that COPD does have a causal effect on these domains, not detected in the general cognitive function analysis.

Both cognition and lung function follow life course trajectories, influenced by a combination of genetic and life course factors.[30] Genetic determinants of lung development and disease are increasingly recognised.[150] In addition to shared environmental risk for lung function and cognitive function, shared genetic risks may be found. Genes involved in growth factors and Vitamin A regulation have been found to affect lung function. Both growth patterns and Vitamin A levels may have a role in cognitive function.[143,151] It may be that genetic pleiotropy can determine both lung function and cognitive trajectories.

5.5.1 Strengths and Limitations

MR eliminates many confounders in observational epidemiology.[88,137] I used a large number of robust SNPs for lung function and SNPs that influence liability to COPD, which have been well validated in large samples.[75,109] MVMR was utilised to decrease the risk of bias and allowed us to condition for the effect of educational attainment. The effects of instrumental variables were assessed in a huge outcome population of similar age and the same ancestral population as the discovery GWAS.

It may be that the conclusion of only weak evidence of lung function affecting cognitive function is due to the MR study being underpowered. The unconditioned results of lower FEV_1 and FVC showed strong evidence of an effect on cognitive function. The strength of evidence weakened on conditioning for education attainment, however as discussed in results this is because educational attainment is a mediating factor, suggesting lung function could affect cognitive function via an effect on educational attainment. Conditioning with height led to the evidence becoming weak too, and height is unlikely to be a mediating factor. However, repeating the study in future as more lung function SNPs are discovered could show different results. Due issues with demography and assortative mating, it has been shown that effect estimates for demographic traits such as height have been over estimated.[152] Therefore, the conditioning effect of height may have been falsely high. Repeating the analysis with height SNPs found in within-family GWAS may give different results.

It is worth noting the SNPs for the exposure of COPD refer to 'liability' rather than a confirmed diagnosis of COPD as no SNP guarantees COPD, but MR is still a valid test of the causal null hypothesis for a binary exposure, if the MR assumption are met.[124] The mean age in the outcome sample was 56, lower than the average age of COPD diagnosis. The effects of SNPs for binary traits may be underestimated in the outcome sample, as participants have not yet developed COPD or its co-morbidities. However, the exposure SNPs were discovered and validated in large populations that

had a similar mean age to the outcome population.[75,109] Therefore the SNPs' estimated effects on COPD liability should be accurate in the outcome population. If the mean age of the outcome population was older, it could introduce a survivor bias if a proportion of those with the SNPs had died from COPD and could not be recruited for the outcome GWAS.[125]

One of the proposed mechanisms whereby lung function and COPD may cause reduced cognition is via cerebrovascular pathology.[153] The outcome population excluded those with a history of clinical stroke, this is unlikely to have excluded the proposed COPD-specific brain changes, but would have excluded those with large vessel vascular damage causing changes in cognitive function.

Estimates of cognitive function can be distorted by factors including demography, (when a populations' genetic variance is related to geographical location), assortative mating (partners are chosen due to phenotypes e.g. higher cognitive function, rather than randomly), and dynastic effects (phenotypic expression of parents genotype affects offspring phenotype e.g. parents with higher education giving educational books to their children).[154] This can be corrected by using within-family GWAS, a possible area for future studies.[155] However, MR studies have tended to over-estimate the effect of anthropomorphic traits which then attenuate when using within-family studies, and in this study I found weak evidence of an effect. Therefore, repeating the study using SNPs from within-family GWAS is likely to produce even weaker evidence.

The populations used by the GWAS for exposure SNPs only used European ancestry participants. These results may therefore not be generalizable to non-European ancestral populations. I did not have a replication cohort for the SNPs used in the MVMR analysis. The effects of the SNPs could have been over-estimated due to the "Winner's Curse" phenomenon.[139] This occurs in GWAS when effect estimates are biased towards the SNPs with the strongest association. However, if present this phenomenon would bias results towards strong evidence of an effect, whereas I found weak evidence of an effect.

5.5.2 Future research

The outcome GWAS was performed with global cognitive function in the general population as a continuous outcome. Much research has focused on whether lung function and lung disease causes mild cognitive impairment (MCI).[156] Normal cognitive function and MCI exist on a spectrum, but this study is unable to fully assess whether reduced lung function or lung disease cause MCI. If a GWAS of MCI becomes available MR could be used in future studies.

The results indicate that lung function alone does not cause lower cognition in the general population. However, lower lung function and lung disease have been shown to be associated with reduced cognitive function, but this is most likely due to shared risk factors. Research should focus on reducing exposure to these shared risk factors and optimising the management of co-morbidities in those with chronic lung disease.

5.5.3 Conclusion

This study provides evidence that lung function and COPD do not cause reduced cognition in the general population. Previous observational studies suggesting a causal link may have been biased by residual confounding. The observed associations between reduced lung function, COPD, and cognitive function remain important. Research should now focus on the management of cognitive impairment in these groups, rather than targeting lung function in order to improve cognition in the general population.

CHAPTER 6. PRESERVED RATIO IMPAIRED SPIROMETRY (PRISM): A UKBIOBANK COHORT STUDY

6.1 Publication and contributions

This chapter has been published.

Prevalence, risk factors, and clinical implications of preserved ratio impaired spirometry: a UK Biobank cohort analysis. Daniel Higbee, Raquel Granell, George Davey Smith, James Dodd. Lancet Respir Med. 2021 Nov 2; S2213-2600(21)00369-6.

DH and JD conceived the project. DH determined and performed the analysis plan. RG gave advice regarding the analysis plan and some code required. DH wrote the paper. GDS and JD supervised.

6.2 Introduction

Preserved Ratio Impaired Spirometry (PRISm), also referred to as 'restrictive pattern' or 'unclassified' spirometry, is defined as a FEV_1 <80% predicted, despite a normal or preserved FEV_1/FVC ratio ≥0.70. The true population prevalence of PRISm is unknown with estimates from 4% to 48% depending on gender, ancestry, geographical location, and smoking history.[58-61] Clinical interest in PRISm comes from data that suggest over 5 years, up to 50% may transition to COPD but that 15% return to 'normal' spirometry.[57,60] If PRISm is a pre-cursor of COPD, it would be an appealing target for interventions to prevent COPD, a leading cause of global mortality.[157] Imaging studies suggest that PRISm may be associated with a degree of airway disease and emphysema which may affect progression to COPD.[158,159] Analyses from other cohort studies show an association between PRISm and respiratory symptoms, increased healthcare utilisation, extra-pulmonary co-morbidities such as obesity, diabetes, cardiac disease, and increased overall mortality.[57,60,62-64]

Definitive epidemiological understanding of PRISm has been limited by cohorts with relatively small patient numbers, rarely containing >1000 cases.[57,60] Some cohorts have used selected populations e.g. only smokers, which limits generalisability.[57,59,65]. Duration of follow-up is often limited to ≤5 years[57,60,66], limiting conclusions about the long term trajectory of PRISm.[67]

My first objective was to use the UK Biobank to examine a large adult general population to determine PRISm prevalence, risk factors and associated symptoms and co-morbidity.[68] The second objective was to use follow-up data to examine the longitudinal outcomes of PRISm including transition to other spirometric states and mortality. The large sample size of UK Biobank and broad recruitment based on age and the inclusion of non-smokers increases power and improves generalizability. The long follow-up of UK Biobank participants compared to other cohorts also allows for accurate estimation of PRISm trajectories over time and survival analysis.

6.3 Methods

6.3.1 Baseline

UK Biobank includes 502,543 individuals aged between 40 and 69 at recruitment across the UK.[68] Participants were identified from the NHS register and were invited to assessment appointments by letter. No weighting mechanism for recruitment was used. The initial assessment took place from 2006-2010 and this data were used as the baseline timepoint. All participants were asked to perform pre-bronchodilator spirometry. Only pre-bronchodilator spirometry was available, although participants were not told to withhold their normal medications. Previously derived variables of quality-controlled spirometry were used for "best measure" FEV_1 and FVC, which excludes participants that do not have acceptable spirometry. Patients with no known smoking status or weight were excluded. FEV_1 percent predicted was calculated as per GLI-2012 values using RSpiro R package in R studio 3.6.1.[160]

PRISm was defined as FEV_1 <80% predicted and FEV_1/FVC ≥0.70. Airflow obstruction was defined using the GOLD criteria for Stage I-IV obstruction, FEV_1/FVC <0.70.[18] Controls were defined by FEV_1 ≥80% with FEV_1/FVC ≥0.70.

Demographic differences between PRISm vs. controls, PRISm vs airflow obstruction, were examined. P-values were calculated using Z-score for continuous outcomes and Pearson's Chi-squared for categorical outcomes. Multivariable logistic regression analysis was performed for risk factors associated with PRISm (age, sex, BMI, diagnosis of asthma, smoking status, trunk fat mass/percentage). Clinically relevant correlates of PRISm were examined (cardiovascular disease, diabetes, shortness of breath) adjusting for confounders (age, sex, BMI, smoking status, hypertension). If data were missing, it was not imputed, and individuals were excluded from the analysis. Statistical analysis was performed using Stata 15.[161]

6.3.2 Follow-up

From 2014-2019 those that lived close to an assessment centre were invited for a repeat visit, with repeat spirometry. Only participants that had been included in baseline were examined in follow-up. The highest measures of FEV_1 and FVC from acceptable spirometry were used. Those without height, body mass index (BMI) and smoking status recorded at follow-up were excluded. Participants with PRISm at baseline and follow-up were classified as having persistent PRISm. I examined baseline demographic differences between change from PRISm to control or airflow obstruction vs. persistent PRISm. Multivariable multinomial logistic regression analysis was performed to determine the risk of age, BMI, smoking status, sex, and doctor diagnosis of asthma with change from PRISm to control or airflow obstruction vs. persistent PRISm. Results are presented as relative risk (RR). I determined what proportion of participants would be expected to revert to control due to regression to the mean using Stata package rtmci.[162] This is a statistical effect of all longitudinal studies, especially those that follow a pathological subset population identified at baseline. It is a well-recognised phenomenon in testing any complex trait whereby outlier results are more likely to be followed by results closer to the mean. This is due to standard deviation, rather than due to a causal or pathophysiological effect.[163]

Sensitivity analyses were performed stratifying the sample by sex, asthma diagnosis, and BMI. I repeated the analysis with lower limit of normal definition of spirometry criteria as using set thresholds can be less reliable in older populations. I also repeated the analysis using GOLD II-IV as the definition of obstruction as some previously published work has done this to capture those more likely to have clinical impacts of COPD.

I also repeated the analysis examining those that transitioned from control spirometry and airflow obstruction at baseline to other spirometric states at follow-up.

6.3.3 Survival analysis

UK Biobank obtained dates of death from NHS Digital and NHS Central registry. Death records up to February 2018 were available allowing me to perform survival analysis covering 12 years. I conducted an unadjusted Kaplan-Meir survival analysis, and both univariate and multivariate Cox's proportional hazard model, adjusting for smoking status, BMI, age, and sex.

6.4 Results

6.4.1 Prevalence of PRISm

353,315 participants had "best measure" FEV_1 and FVC. 1,441 were excluded for missing smoking status and/or BMI. This left 351,874 participants for analysis at baseline (see **Figure 19**). **Table 12** shows a prevalence of 11.0% for PRISm and 15.8% for stage I-IV airflow obstruction. Due to the large sample sizes, clinically similar results often showed strong evidence of a statistical difference when examining p-values, for example the mean difference in age between PRISm and control is 0.4 years, but the when examining for a statistical difference, the p-value < 0.0001.

Table 12. Baseline demographics of participants with PRISm, Control and Airflow Obstruction

Demographic at baseline	**PRISm N = 38,639**	**Control N= 257,643**	**P-value* PRISm vs Control**	**Stage I-IV Obstruction N = 55,592**	**P-value* PRISM vs I-IV Obstruction**
Age (Years) Mean (SD)	56.4 (7)	56.0 (7)	<0.0001	59.1 (7)	<0.0001
BMI (kg/m^2) Mean (SD)	29.1 (5)	27.2 (4)	<0.0001	26.8 (4)	<0.0001
Female (%)	55.4%	55.6%	0.33	44%	<0.0001
FEV_1 % predicted Median (IQR)	74% (68 - 77)	98% (90 - 106)	<0.0001	79% (67 - 90)	<0.0001
FVC % predicted Median (IQR)	76% (71 -81)	99% (91 - 108)	<0.0001	94% (83 - 106)	<0.0001
FEV_1/FVC Median (IQR)	0.75 (0.72 – 0.78)	0.77 (0.74 – 0.80)	<0.0001	0.64 (0.62-0.68)	<0.0001
Never smoker (%)	51.2%	56.8%	<0.0001	40.8%	<0.0001
Ex-smoker (%)	36.4%	35.3%	0.0002	39.9%	<0.0001
Current smoker (%)	12.4%	7.9%	<0.0001	19.1%	<0.0001
Pack/years Median (IQR) †	23 (13 -36)	16 (8 - 27)	<0.0001	27 (15 - 41)	<0.0001
SOB walking on ground (%)	17.7%	7.1%	<0.0001	15.2%	<0.0001
Doctor diagnosed asthma	16.9%	9.9%	<0.0001	24.5%	<0.0001
Doctor diagnosed COPD	1.7%	0.4%	<0.0001	6.7%	<0.0001
Diabetes (%)	8.7%	3.8%	<0.0001	4.8%	<0.0001
Heart attack (%)	3.4%	1.5%	<0.0001	3.2%	0.103
Angina (%)	4.6%	2.2%	<0.0001	3.9%	<0.0001
High blood pressure (%)	33.4%	24.3%	<0.0001	28.9%	<0.0001
Stroke (%)	2.0%	1.1%	<0.0001	2.0%	0.68

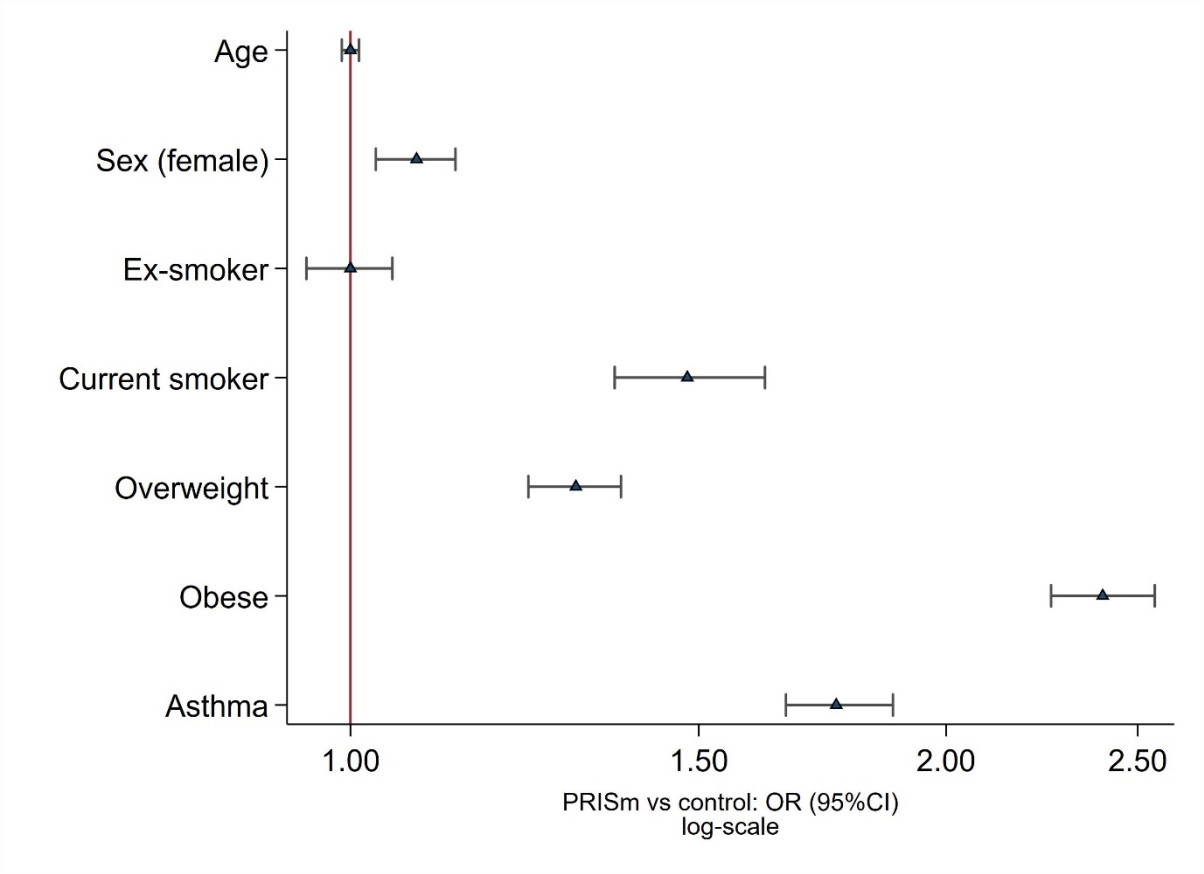
*P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. † For ex and current smokers only. SD – Standard Deviation. IQR – Interquartile range

6.4.2 Risk factors for PRISm

55% of PRISm and controls were female vs. 44% with airflow obstruction. Current smokers were more common in PRISm than controls (12.4% vs 7.9%, p-value <0.0001), and the smoking pack/years of ever smokers was higher (23 vs 16 pack/years, p-value <0.0001). Doctor diagnosed asthma was more common in PRISm than in controls (16.9% vs 9.9%, p-value <0.0001). BMI was shown to have a non-linear association with PRISm, violating an assumption of logistic regression. Therefore, I changed BMI from a continuous variable in the regression analysis to categorical as determined by three clinically relevant groups; Not overweight (BMI <25), Overweight (BMI ≥25 and <30), and Obese (BMI ≥30).

Multivariable logistic regression examining the association of age, sex, BMI categories, smoking status (never/ex/current) and doctor diagnosed asthma with PRISm vs controls was performed. Female gender was statistically associated with PRISm OR 1.08 (95% CI 1.03 -1.13, p-value 0.0010), with strong evidence of association found for overweight OR 1.30 (1.23 – 1.37. p-value <0.0001), obesity OR 2.40 (2.26 – 2.55. p-value <0.0001), current smoking OR 1.48 (1.36 – 1.62, p-value <0.0001) and doctor diagnosis of asthma OR 1.76 (1.66 – 1.88, p-value <0.0001), see **Figure 18**. When examining the association of trunk fat mass (per Kg) and trunk fat percentage (per %) an association with PRISm was also seen (OR 1.08 (1.08 – 1.09) and 1.06 (1.06 – 1.07) respectively, p-values <0.0001).

Figure 19. Forest plot showing factors associated with PRISm vs control spirometry



OR derived from multivariable logistic regression

This figure is a forest plot comparing possible risk factors for PRISm vs control spirometry

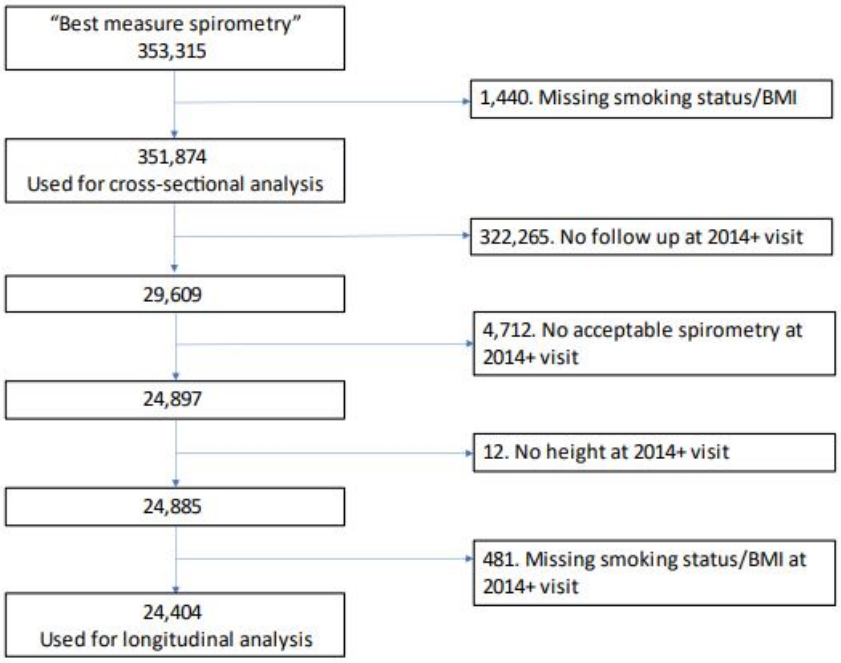
6.4.3 PRISm symptoms and co-morbidities

There was a higher prevalence of breathlessness in PRISm 17.7% vs 7.1% in controls, p-value <0.0001. After adjustment for BMI, age, smoking status, and asthma diagnosis PRISm remained associated with increased breathlessness OR 2.0 (95%CI 1.91 – 2.14, p-value <0.0001). Diabetes was more common in PRISm than controls or airflow obstruction (8.7% vs 3.8% vs 4.8% respectively, p-value's <0.0001) this remained after adjustment for BMI, age, and sex vs. controls OR 1.79 (1.72 – 1.87, p-value <0.0001). Cardiovascular co-morbidity was higher in PRISm vs. control, with approximately double the prevalence of angina (4.6% vs 2.2%, p-value <0.0001), heart attack (3.4% vs 1.5%, p-value <0.0001) and stroke (2.0% vs 1.1%, p-value <0.0001). Prevalence of hypertension and angina were also higher in PRISm vs. airflow obstruction (p-value <0.0001). After adjustment for hypertension, diabetes, BMI, age, smoking status, and sex, PRISm remained associated with an increased risk of stroke OR 1.4 (95%CI 1.36 – 1.61), angina OR 1.47 (95%CI 1.35 – 1.60) and heart attack OR 1.71 (95%CI 1.64 – 1.83) vs. controls.

6.4.4 Longitudinal analysis of PRISm

Follow-up data was available for 29,609 participants. 4,712 did not have acceptable spirometry. 493 were excluded for not having a recorded height, smoking status, or BMI at follow-up leaving 24,404 participants for analysis.

Figure 20. CONSORT diagram



CONSORT diagram showing the number of participants used for baseline and follow up analysis in the PRISm analysis

Participants with follow-up data were younger, less overweight, with better lung function and lower rates of smoking at baseline compared to the population that did not have follow-up, rates of asthma diagnosis were similar (see **Table 13**).

Table 13. Baseline demographics of participants that only had baseline data vs participants included in follow up

Demographic at baseline	No Follow up (N = 327470)	Follow up (N = 24404)	P-value
Age (Years) Mean (SD)	56.7 (8)	54.5 (7)	<0.0001
BMI (kg/m^2) Mean (SD)	27.4 (5)	26.7 (4)	<0.0001
Female (%)	54.0%	51.0%	<0.0001
FEV_1 % predicted Median (IQR)	93% (83 – 103)	96% (86 – 104)	<0.0001
FVC % predicted Median (IQR)	97% (87 – 106)	99% (90 – 108)	<0.0001
FEV_1/FVC Median (IQR)	76% (72 – 80)	77% (73 – 80)	<0.0001
Never smoker (%)	53.2%	59.8%	<0.0001
Ex-smoker (%)	36.3%	33.8%	<0.0001
Current smoker (%)	10.5%	6.3%	<0.0001
Pack/years Median (IQR) †	19 (10 – 32)	15 (8 – 26)	<0.0001
Doctor diagnosed asthma	12.9%	13.0%	0.71
Doctor diagnosed COPD	1.4%	0.88%	<0.0001

The mean FEV_1 for the cohort at follow-up was higher vs. baseline (3.0 litres [SD 0.8] vs 2.8 litres [SD 0.8]). The mean annual FEV_1 decline for individual participants between baseline and follow-up of -30mls/yr [SD 46]. Prevalence of PRISm at follow-up was lower than at baseline PRISm (7.1% vs 11.0%). Full comparison of lung function states cross-sectionally at the follow-up time point are in **Table 14**.

Table 14. Demographics of participants by PRISm trajectory at follow up

Demographic at follow up	PRISm to PRISm N = 745	PRISm to Control N = 987	P-value*	PRISm to Airflow Obstruction N = 241	P-value†
Age (Years) Mean (SD)	62 (7)	62 (7)	0.86	65 (8)	<0.0001
BMI (kg/m^2) Mean (SD)	29.0 (6)	27.5 (5)	0.0012	26.8 (5)	<0.0001
Mean change in BMI (SD)	0.24 (2)	-0.48 (2)	<0.0001	-0.58 (2)	<0.0001
Female (%)	56%	52%	0.12	58%	0.43
FEV_1 % predicted Median (IQR)	74% (69 – 77)	87% (84 – 90)	<0.0001	73% (63 – 80)	0.0014
FVC % predicted Median (IQR)	75% (70 – 79)	89% (84 – 96)	<0.0001	58% (76 – 81)	<0.0001
FEV_1/FVC Median (IQR)	76% (73 – 78)	77% (74 – 80)	<0.0001	68% (65 – 69)	<0.0001
Mean change in FEV_1 (mls)	-210mls	265mls	<0.0001	-269mls	0.004
Mean change in FVC (mls)	-260mls	304mls	<0.0001	-43mls	<0.0001
Never smoker (%)	59%	61%	0.31	56%	0.15
Ex-smoker (%)	36%	35%	0.66	38%	0.62
Current smoker (%)	5%	3%	0.14	8%	0.43
Pack/years Median (IQR)**	21 (12 – 35)	18 (10 – 26)	0.0004	19 (8 – 35)	0.23
SOB walking on ground (%)	17.0%	7.5%	<0.0001	13%	0.44
Diabetes (%)	14.6%	8.8%	0.0090	8%	0.039
Heart attack (%)	3%	3%	0.68	3%	0.65
Angina (%)	4%	2%	0.12	2%	0.45
High blood pressure (%)	27%	27%	0.89	30%	0.57
Stroke (%)	1%	1%	0.83	4%	0.21
Years between phases. Median (IQR)	9 (8 – 10)	9 (7 -10)	0.010	9 (8 – 10)	0.26

*P-value comparing those with PRISm at baseline and follow up, with those that transitioned from PRISm at baseline to control at follow up. †P-value comparing those with PRISm at baseline and follow up, with those that transitioned from PRISm at baseline to airflow obstruction at follow up. P-values Calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. ** For ex and current smokers only

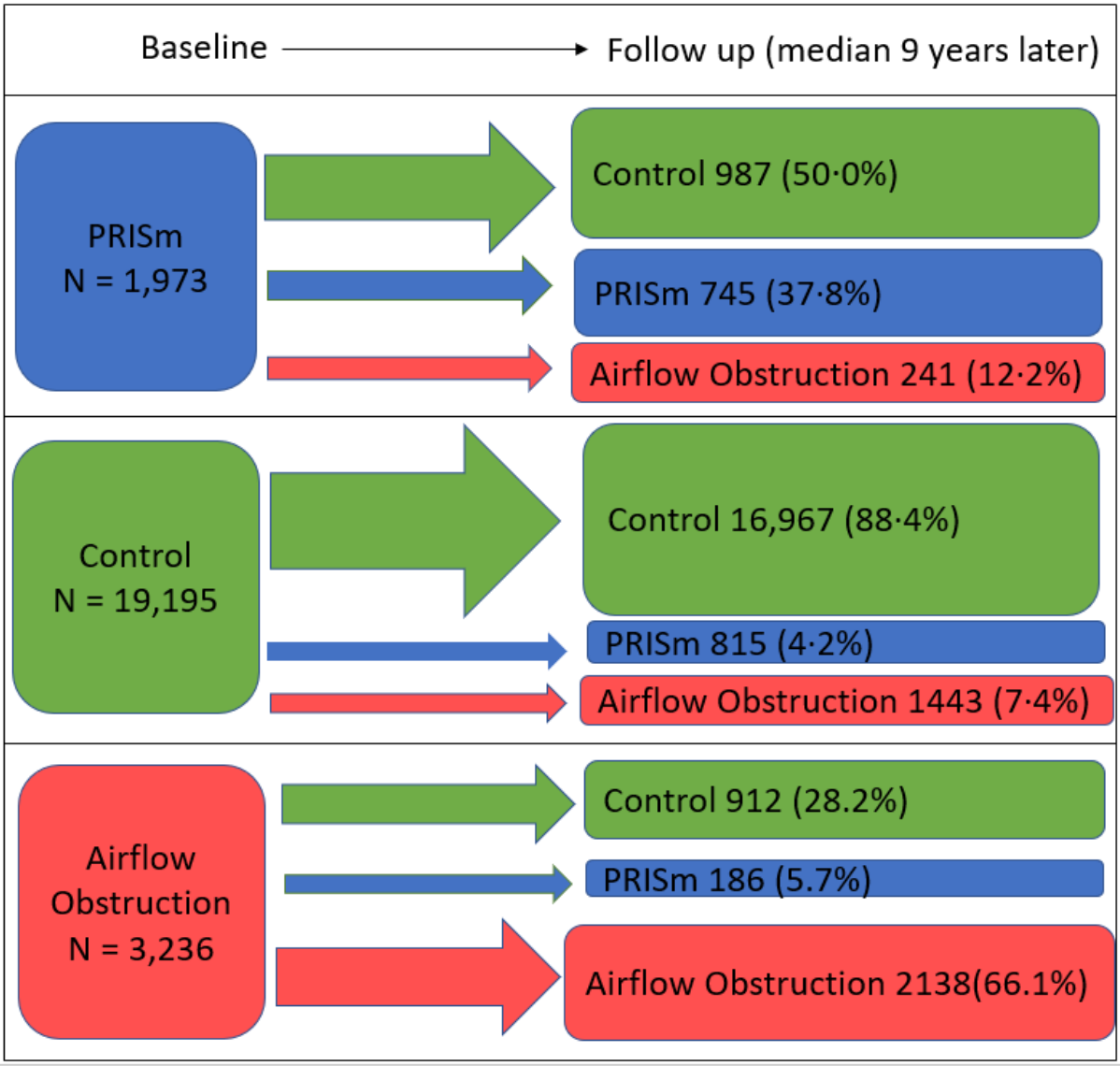
6.4.5 PRISm trajectories

The median time between baseline and follow-up was 9.0 years [IQR 8.0 – 10.0]. 1,973 participants with PRISm at baseline were included in follow-up (**Figure 19**). As shown in **Figure 20**, 37.8% had persistent PRISm, 50.0% reverted to normal control spirometry and 12.2% transitioned to airflow obstruction. More participants with PRISm at baseline transitioned to a different lung function state (62.2%) vs controls (11.6%) and airflow obstruction (34.0%). Those that transitioned from PRISm to control (i.e. normal spirometry) had nominal evidence of slightly shorter times between baseline and follow-up than those with persistent PRISm (median 9 years [IQR 7 -10] vs 9 years [IQR 8 – 10], p-value 0.010).

Regression to the mean analysis of PRISm and controls at baseline estimated that 11.8% (95%CI 11.4 -12.2) of PRISm would be expected to revert to control if one follow-up analysis is performed. If this is taken into account, then rates of persistent PRISm and reversion to control would be similar.

As per **Table 14**, cross-sectionally at follow up time point, participants with persistent PRISm had higher mean BMIs than participants who transitioned from PRISm to control spirometry (29 [SD 6] vs 27 [SD 5]), median pack/years (21 [IQR 12-35] vs 18 [10 – 26]), more diabetes (14.6% vs 8.8%), and shortness of breath (17.0% vs 7.5%). At follow up participants with persistent PRISm vs. PRISm to COPD trajectories were younger, mean age (62 [SD 7] vs 65 [SD 8] years) and had a higher mean BMI (28 kg/m^2 [5] vs 26 [5]).

Figure 21. Lung function trajectories from Baseline to Follow-up



This figure demonstrates the trajectories of the different lung function states between baseline and follow up. PRISm is a much more transient state than control and airflow obstruction. Only a minority of those with PRISm developed COPD on follow up

6.4.6 Persistent PRISm vs PRISm to control trajectories

As can be seen in **Table 15**, those with persistent PRISm had reduced FEV_1 and FVC % predicted at baseline compared to those that reverted to control, however with a median difference of ~2% predicted it is not clinically meaningful. Persistent PRISm had a high baseline and mean change in BMI (0.8 and 0.24 kg/m^2 [SD 2.33] respectively), whilst those that transitioned to control had a mean

change of -0.48 kg/m^2. There was a clinically significant difference in smoking at baseline was observed between participants with persistent PRISm and participants with PRISm that transitioned to control, with higher pack/years in the group with persistent PRISm (20 pack/yrs [IQR 12-34] vs 16 [8-26], p-value <0.0001).

Multivariable multinomial logistic regression analysis showed strong evidence of a negative association between doctor diagnosed asthma with PRISm changing to control vs persistent PRISm (RR 0.67 (95% CI 0.47 – 0.96, p-value 0.030). Change in BMI, per mg/kg^2 increase, was also strongly negatively associated with PRISm changing to control vs persistent PRISm after adjustment (RR 0.86 (95% CI 0.81 – 0.91, p-value <0.0001).

Table 15. Baseline demographics of participants by PRISm trajectory

Demographic at baseline	**PRISm to PRISm** **N = 745**	**PRISm to Control Spirometry** **N = 987**	**P-value***	**PRISm to Stage I-IV Obstruction** **N = 241**	**P-value†**
Age (Years) Mean (SD)	53.6 (7)	53.8 (7)	0.56	56.7 (7)	<0.0001
BMI (kg/m^2) Mean (SD)	28.8 (5)	28.0 (4)	0.0011	27.3 (4)	0.0001
Female (%)	56%	52%	0.12	58%	0.51
FEV_1 % predicted Median (IQR)	74% (69 – 77)	76% (71 – 78)	0.0044	74% (69-77)	0.85
FVC % predicted Median (IQR)	76% (71 – 80)	77% (77 – 81)	0.0023	80% (73 – 83)	<0.0001
FEV_1/FVC Median (IQR)	76% (73 – 79)	76% (73 – 79)	0.76	73% (71 – 75)	<0.0001
Never smoker (%)	57%	59%	0.41	52%	0.17
Ex-smoker (%)	34%	34%	0.99	35%	0.79
Current smoker (%)	8%	6%	0.12	12%	0.056
Pack/years Median (IQR) **	20 (12 – 34)	16 (8 – 26)	<0.0001	19 (9 – 34)	0.136
SOB walking on ground (%)	9%	7%	0.28	16%	0.083
Doctor diagnosed asthma	18%	13%	0.017	25%	0.078
Doctor diagnosed COPD	1%	1%	0.43	1%	0.57

*P-value comparing those with PRISm at baseline and follow-up (persistent PRISm), with those that transitioned from PRISm to control at follow-up. †P-value comparing persistent PRISm with those that transitioned from PRISm to airflow obstruction at follow-up. ** For ex and current smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes

6.4.7 Persistent PRISm vs PRISm progressing to airflow obstruction trajectories

At baseline participants with persistent PRISm vs. PRISm to airflow obstruction were younger (53.6 years [SD 7.5] vs 56.7 years [SD 7.7], p-value < 0.0001). Although there were more current smokers and fewer never smokers at baseline in those that transitioned from PRISm to airflow obstruction, statistical evidence was weak (p-value > 0.05). Those with persistent PRISm had higher BMI's (28.8kg/m^2 [SD 5] vs 27.3 [SD 4], p-value < 0.0001) and higher FEV_1/FVC ratios (76% [IQR 73-79] vs 73% [71-75], p-value < 0.0001) vs PRISm to airflow obstruction at baseline. Change in BMI differed between those with persistent PRISm (mean change in BMI of 0.24kg/m^2 [2.33]) vs, PRISm to airflow obstruction (mean change of -0.58kg/m^2 [SD 2.34]).

Multivariable multinomial regression analysis showed a change from PRISm to airflow obstruction was strongly positively associated with increased age, RR 1.07 (95%CI 1.04 – 1.10, p-value < 0.0001) and a doctor diagnosis of asthma RR 1.91 (95% CI, 1.17 – 3.13, p-value 0.010). Change in BMI (per mg/kg^2) increase showed a strong negative (RR 0.86, (0.79 – 0.95, p-value 0.0022) of PRISm changing to airflow obstruction vs persistent PRISm after adjustment.

6.4.7.1 Sensitivity Analysis

The sensitivity analysis (stratifying by sex, BMI, asthma, smoking status) showed similar rates of persistent PRISm (ranging from 32 – 39%), PRISm changing to control (48 – 63%) and PRISm changing to airflow obstruction (5 – 15%) across the sensitivity analyses (**Table 16**).

Table 16. PRISm trajectory in different subgroups

PRISm trajectory	**Main analysis (N= 1,973)**	**PRISm LLN (N = 1532)**	**COPD as stage II-IV (1,973)**	**Men only (N = 905)**	**Women only (N = 1068)**	**Never smokers (N = 1138)**	**Ever smokers (N = 835)**	**Non-asthmatics (N = 923)**	**BMI <25 (N= 547)**
PRISm to PRISm	38%	32%	38%	37%	39%	38%	38%	37%	35%
PRISm to control	50%	63%	50%	52%	48%	51%	49%	53%	50%
PRISM to COPD	12%	5%	9%	11%	13%	11%	14%	10%	15%

6.4.7.2 Transition from control spirometry at baseline

Regression analysis showed that female sex, being overweight, obesity, and current smoking were all associated with transition from control spirometry at baseline to PRISm, whereas doctor diagnosis of asthma was not. Doctor diagnosed asthma was strongly associated with a change from control to airflow obstruction (**Table 17**).

Table 17. Multivariable multinomial logistic regression of baseline demographics association with control to prism and control to airflow obstruction vs. persistent control trajectories

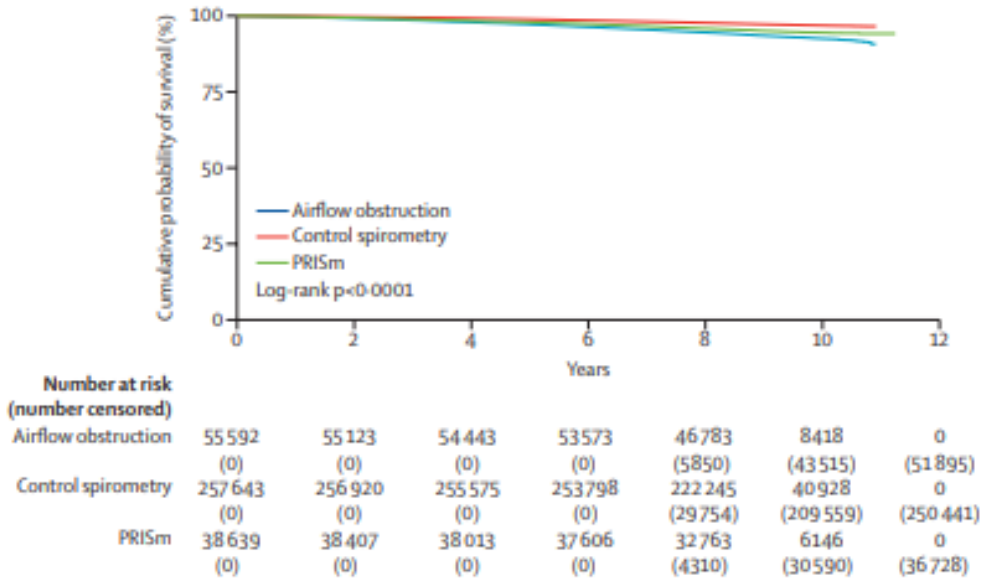
Demographic	**RRR (95% CI) Control to PRISm**	**RRR (95% CI) Control to airflow obstruction**
Age	1.00 (0.99 – 1.02) pval 0.61	1.06 (1.04 - 1.07) pval <0.0001
Sex (Female)	1.23 (1.11 – 1.74) pval 0.039	0.97 (0.83 - 1.12) pval 0.67
Overweight (BMI ≥25 and <30)	1.39 (1.11 – 1.74) pval 0.0040	0.65 (0.55 - 0.76) pval <0.0001
Obese (BMI ≥30)	1.82 (1.40 - 2.37) pval <0.0001	0.57 (0.45 - 0.71) pval <0.0001
Ex-smoker	0.88 (0.71 – 1.10) pval 0.26	1.18 (1.01 - 1.38) pval 0.041
Current smoker	2.46 (1.78 - 3.39) pval <0.0001	2.13 (1.60 - 2.85) pval <0.0001
Doctor diagnosis of asthma	1.26 (0.94 – 1.69) pval 0.12	1.45 (1.15 - 1.81) Pval 0.0015

RRR – Relative risk ratio.

6.5 Survival Analysis

12,810 deaths were recorded during follow-up, 2.8% of controls, 5.0% of PRISm, and 6.7% of airflow obstruction. All 351,874 participants were included in survival analysis which showed 3 deaths per 1000 individuals per year in the control, 6 in PRISm, and 7 for airflow obstruction. An unadjusted Kaplan-Meier plot (**Figure 21**) was produced which showed strong evidence of difference in survival between the groups, log rank p-value <0.0001.

Figure 22. Kaplan-Meier survival estimates based on spirometric group at baseline



The Kaplan-Meier graph shows a clear difference in survival between the three lung function states

PRISm vs controls Hazard Ratio for all-cause mortality was 1.79 (CI 1.70 – 1.88, p-value <0.0001). After adjustment for smoking status, age (grouped as quintiles), sex, BMI (categorised) this reduced to 1.61 (CI 1.53 – 1.69, p-value <0.0001). Assumptions were checked with log-log plots.

6.6 Discussion

This study in UK Biobank shows that PRISm is common at 11% in this population. This is lower than some estimates in cohorts of smokers, but similar to a cohort that included never smokers.[60] Longitudinal analysis showed 62.2% of those with PRISm changed to a different lung function state over 9 years. After regression to the mean was considered, rates of persistent PRISm and reverting from PRISm back to control were similar at ~40%. Considerably lower rates of progression from PRISm to airflow obstruction were found compared to previous reports.[57,60] This is likely due to other cohorts' recruiting based on smoking. Restricting analysis to ever smokers is likely to bias other factors associated with smoking that can influence PRISm transition to airflow obstruction e.g., age, sex, BMI, and asthma which could further confound results. As age also had a strong association with PRISm progressing to airflow obstruction, cohorts with older populations are also likely to see higher rates of impaired spirometry, especially with longitudinal follow-up.

There was strong statistical evidence of an association between BMI, particularly obesity, and both incident and persistent PRISm. This difference is unlikely to be explained solely by extra-thoracic restriction as 62% of those with PRISm had BMI <30 kg/m^2 and restricting the analysis to only those with BMI <25 kg/m^2 showed that 7.6% had PRISm. It may be that BMI is contributing to PRISm risk via a different pathway such as metabolic and inflammatory effects of adipose tissue itself.[164]

The high prevalence of cardiovascular disease and diabetes in PRISm even after adjustment for confounders is important. COPD may have a direct causal effect on extra-pulmonary disease, for example through systemic inflammation or oxidative stress.[126] Therefore it is conceivable this could also occur in PRISm. In addition, reduced FVC (as seen in PRISm), has been shown to be associated with death and causally linked with risk of coronary artery disease.[31,165]

Despite the variable state of PRISm over time, survival analysis showed strong evidence of an increased risk of death. This may be due to increased co-morbidities in PRISm, but further work is needed to determine if this is causal and/or whether it would be a modifiable risk factor. However,

even if no causal pathway were to be found between PRISm and co-morbidity, this could be due to shared environmental or genetic factors and studies such as those screening for diabetes and cardiovascular disease in PRISm would still be of interest.

The strong association of current smoking to incident PRISm, persistent PRISm, and progression to airflow obstruction shows that smoking cessation is important. PRISm is a variable state, and so it is possible that quitting smoking will improve the chance of reversion to control spirometry and prevent progression to COPD. Imaging studies quantifying smoking-associated features such as emphysema, airway wall thickness and air trapping may be used to predict more rapid lung function decline in those with PRISm.[166]

The observed relationship between asthma and PRISm may be complicated by self-report rather than objectively confirmed diagnoses. For example, people with PRISm and respiratory symptoms may manifest with 'asthma like symptoms' and be incorrectly diagnosed. However, there are plausible mechanisms by which asthma may contribute to both PRISm and airflow obstruction via small airways obstruction and gas trapping.[167,168] I performed a sensitivity analysis by excluding participants with asthma which suggested that neither lung function trait of PRISm or its association with co-morbidities are solely due to an asthma misdiagnosis or effect.

The large sample size of UKBiobank, which has recruited participants based on age, instead of smoking history, has allowed the largest and most generalisable study of PRISm to date and a more powerful analysis of its relationship with comorbidities. This is the first study to estimate the effect of regression to the mean, an important source of bias in longitudinal studies, especially when examining an outlier group. The follow-up period of this study is particularly long with a median of 9 years between data sets reducing the risk of short-term changes. Having mortality records for up to 12 years after recruitment also allows for accurate estimation of mortality associations with PRISm.

6.6.1 Limitations

UK Biobank collects only pre-bronchodilator spirometry, post-bronchodilator spirometry was not available. Post-bronchodilator spirometry is not required for a diagnosis of PRISm, but differences between pre and post-bronchodilator spirometry have been reported for PRISm and airflow obstruction.[61,169] If post-bronchodilator spirometry was used, it is likely to have reduced the numbers classified as having PRISm and airflow obstruction spirometry in UK Biobank potentially changing some conclusions. One study of 18,509 participants showed a prevalence of 22.3% PRISm pre-bronchodilator, that reduced to 17.7% post-bronchodilator. It would be valuable for a study of PRISm to be performed in subjects who had post-bronchodilator spirometry. I hope that by performing sensitivity analysis using LLN criteria and classifying airflow obstruction as GOLD II-IV it may have eliminated a proportion of individuals whose FEV_1/FVC ratio would have normalised post-bronchodilation. There was a lower prevalence of PRISm and a higher mean FEV_1 at follow-up compared to baseline. However, the mean annual decline in FEV_1 is similar to a normal population. Participants that have follow-up data were younger with lower rates of smoking. The rate of current smoking at follow-up was low at 6.3%. This is a potential source of bias. Recruitment to follow-up was based solely on participants proximity to assessment centres. Participants with health problems may be less inclined to repeatedly attend. Due to increased mortality associated with PRISm and airflow obstruction survivor bias may play a role, although the proportion of the cohort that died was low. It may be those participants living closer to recruitment centres have healthier lung function. UK Biobank has been shown to have a "healthy volunteer" bias as no weighted sampling was undertaken. Despite this, research has shown that established associations between risk factors and outcomes are comparable to studies with more representative sample populations.[170] Additionally, due to its large size and heterogeneity of exposure methods, associations between exposures and health outcomes are generalizable to other populations.[171] There were two time points available for analysis. More time points would allow for a more nuanced understanding of change over time, increased power and precision of estimates, and regression to the mean analysis. Spirometry was used to defined airflow

obstruction. Airflow obstruction is not the same as COPD which remains a clinical diagnosis assuming spirometric criteria are fulfilled. There is no more detailed lung physiology such as lung volumes or gas transfer, but they are not necessary for the diagnosis of PRISm. Interstitial lung diseases are very rare and comprise <0.1% of UK Biobank so are unlikely to influence results. The sample was 100% European ancestry. Therefore, these results may not be generalisable to non-European ancestral populations. There results may not be generalizable to people under the age of 40. Finally, traditional observational epidemiological analysis such as this could be affected by collider bias.[172] By stratifying lung function, a continuous trait influenced by multiple exposures, into conditional phenotypes, lung function can become a collider. This can induce associations between exposures for both lung function and other outcomes e.g. cardiovascular disease. Time-varying covariates can become colliders. For example, participants could decide to quit smoking due to a lung disease, which could affect the transition to other lung function states. If this did occur, then the observed associations may be induced by the statistical model.

6.6.2 Future Research

Studies assessing structural, functional lung changes and genetics of PRISm are now needed. For example, the frequency and severity of small airways dysfunction in PRISm will provide further insight into underlying pathophysiology and future risk of COPD. Importantly, small airways obstruction may be amenable to treatment. Genetic studies of PRISm have so far failed to find associated variants,[59] but discovery of genetic markers for PRISm in larger cohorts could help explain underlying pathological mechanisms for PRISm and be used for Mendelian Randomisation studies to determine if the is evidence that the observed association with co-morbidity is causal.

6.6.3 Conclusion

This analysis of UK BIOBANK shows a general adult population prevalence of PRISm of 11%. PRISm is associated with breathlessness, diabetes, and cardiovascular co-morbidity, and death even after adjustment for shared risk factors including smoking. PRISm is often a transient state with 50%

returning to normal lung function and 12% progressing to airflow obstruction over ~9 years. While PRISm is strongly associated with asthma, BMI and smoking, these factors do not appear to entirely account for this lung function trait and the mechanisms remain unclear.

CHAPTER 7. GENOME WIDE ASSOCIATION STUDY OF PRESERVED RATIO IMPAIRED SPIROMETRY (PRISM) IN UK BIOBANK

7.1 Publications and Contributions

This chapter is currently being submitted for publication.

DH and JD conceived the study. DH planned and performed the data analysis and wrote the paper. RG, GDS and JD provided supervision. Alvin Lirio (University of Lecester) performed the LD correlation studies. Fergus Hamilton assisted with the LDSC analysis.

7.2 Introduction

Preserved Ratio Impaired Spirometry (PRISm), also referred to as 'restrictive pattern' or 'unclassified' spirometry, is defined as FEV_1 percent predicted <80%, and FEV_1/FVC ratio ≥0.70. It was thought that PRISm was a pre-cursor of COPD, with some data suggesting that over 5 years, up to 50% may transition to COPD while 15% return to 'normal' spirometry.[57,60] However this work may have been affected by selection and small cohort bias. My research in a larger more generalizable cohort has shown that PRISm can be a transient state, with only 12% of those with PRISm developing airflow obstruction. However, clinical interest remains as PRISm is associated with respiratory symptoms, co-morbidities such as obesity, diabetes, cardiac disease, and an increased overall mortality.[57,60]

There are several factors associated with the risk of PRISm including Body Mass Index (BMI) and smoking. Whether PRISm has genetic risk factors is not yet known. There has been just one attempted Genome Wide Association Study (GWAS) of PRISm, but numbers were small (N=1257), restricted to ever smokers, and failed to find any significant (p-value $<5x10^{-8}$) SNPs. [59] Lung function traits, trajectories and diseases are known to be associated with genetic variants, [30,75] and the same may be true of PRISm.

If found, genetic variants could provide insight into pathogenesis of PRISm, and potentially provide therapeutic targets. Genetic variants for PRISm could be used for Mendelian Randomisation studies

to determine if there is evidence of causality of associations of PRISm. It may be that PRISm and its associated co-morbidities have shared genetic risk factors.

My objective was to perform a case-control GWAS of PRISm to determine if PRISm is heritable and report novel associated SNPs. For the discovery cohort I used the UK Biobank, and for replication I used cohorts within the SpiroMeta and CHARGE consortiums. I focused on variants that had not been described in the Shrine GWAS of continuous lung function, the largest lung function GWAS to date.[75] I undertook a Phenome Wide Association Study (PheWAS) and literature review of any successfully replicated SNPs.

7.3 Method

7.3.1 Study Populations

UK Biobank includes 502,543 individuals aged between 40 and 69 years at recruitment across the UK.[68] Only previously derived variables of quality-controlled pre-bronchodilator spirometry were used for "best measure" FEV_1 and FVC, excluding participants without acceptable spirometry. Only those of European ancestry were included to reduce population stratification and maintain a large sample size. Patients with no known smoking status or weight were excluded. FEV_1 percent predicted was calculated as per GLI-2012 values using RSpiro R package in R studio 3.6.1.[160] PRISm was defined as FEV_1 % predicted <80% and FEV_1/FVC ratio ≥0.70, controls as FEV_1 % predicted ≥80% and FEV_1/FVC ratio ≥0.70. Genotyping was performed by the UK Biobank using the Axiom UK BiLEVE array and the Axiom Biobank array (Affymetrix) and imputed to the Haplotype Reference Consortium (HRC) panel.[173]

Replication GWAS was performed individually in 13 cohorts from the SpiroMeta consortium. Although not formally tested, there was no overlap between the discovery and replication samples. Replication was only performed in those of European ancestry.

7.3.2 GWAS

A GWAS of PRISm and controls was performed using BOLT-LMM logistic regression via the IEU GWAS pipeline adjusting for sex, body mass index (BMI), age, and smoking status.[174] BOLT-LLM uses a linear mixed model to account for both cryptic relatedness and population stratification. SNPs used all met the criteria: Mean Allele Frequency >0.01, Genotyping rate >0.015, Hardy-Weinberg equilibrium p-value <0.0001, r^2 threshold of 0.10.[174,175] Mean allele frequency refers to the frequency of alleles in the population (i.e. A and a), genotype rate is the frequency of genotypes in the population (i.e. AA, Aa, aa). Therefore, the criteria ensures rare alleles and genotypes are excluded. The Hardy-Weinberg equilibrium excludes genotypes that do not occur at a rate expected in a normal population,

reducing the risk of population stratification. The r^2 threshold means that closely correlated SNPs are removed. Stringent linkage disequilibrium clumping (r^2 0.001, kb 10,000) was applied to SNPs reaching significance threshold of p-value 5^{-08}, this retains the most significant SNP at a locus to ensure all signals are independent. Only SNPs not reported in the largest published GWAS of lung function to date were then investigated in the replicating cohorts.

Replicating cohorts performed a logistic regression of the remaining SNPs in those with PRISm and control spirometry. Adjustment for age, BMI, sex, smoking status (either pack/years or as above), and population substructure by either principal components or using linear mixed models was done. Meta-analysis of the replicating cohorts was performed using an inverse-weighted random-effects model with the STATA command metan. A priori determined significance thresholds were tier 1 (p-value $<5^{-08}$ in UK Biobank, and in replication p-value = 0.05/number stage 1 SNPs) and tier 2 (p-value $<5^{-08}$ in UK Biobank, and in replication p-value < 0.05). Both tiers required SNPs to have consistent direction of effect in discovery and replication analysis.

7.3.3 Analysis of GWAS and replicated SNPs

I conducted a PheWAS of replicated SNPs reporting associations of p-value $<5^{-8}$. LD-correlation analysis was performed to assess the correlation of all SNPs tested in the discovery PRISm GWAS with other GWAS including continuous lung function traits, moderate to severe asthma, asthma-COPD overlap, respiratory tract infections and eosinophils. I also examined the LD-correlation between PRISm and extra-pulmonary traits including type II diabetes, BMI, hypertension, and myocardial infarction. LD-correlation studies are reported with rg, a measure of the genetic correlation between PRISm and the other trait. A rg of 1 implies identical genetic influences, an rg of 0 implies the genetic effect on one trait is independent of the other. Replicated SNPs were cross referenced with online resources to determine their function and expression.[176,177] Phenoscanner was used to determine the nearest gene to replicated SNPs.[176] The previously documented genes function was found using Genecards, with results in **Table 19**.[178] Online resources were used to see if

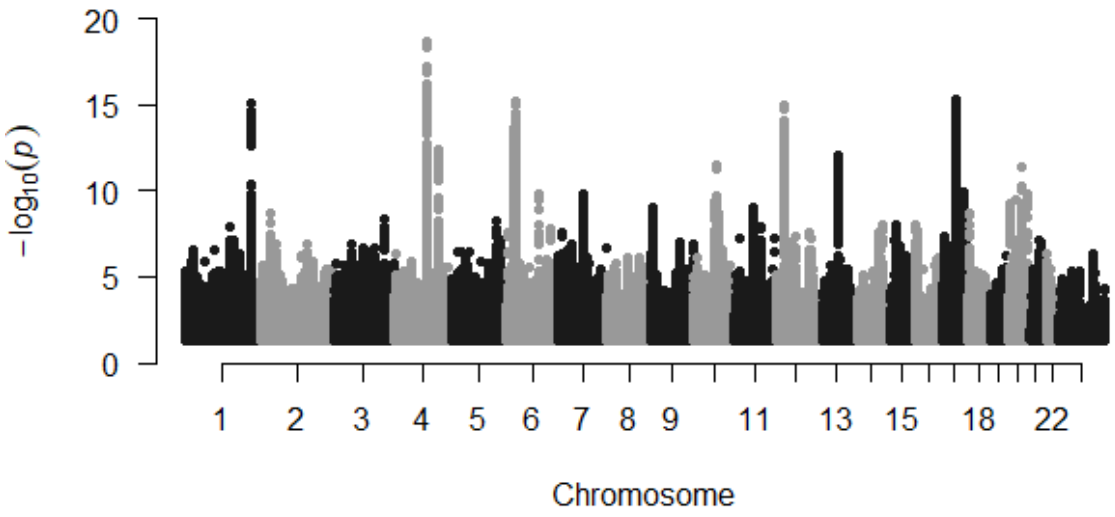
SNPs are known to be expressed in mouse lung tissue and if nearest genes are associated with Mendelian inherited diseases.[179,180]

7.4 Results

7.4.1 Discovery GWAS

353,315 participants had "best measure" FEV_1 and FVC. 1,440 were excluded for missing smoking status and/or BMI. This left 351,874 participants, including 38,639 PRISm cases and 257,643 controls. A GWAS of 296,282 individuals was performed (after removing 6,490 without genotypes). 12,321,875 SNPs (imputed or genotyped) were tested. 8,233 reached p-value <5×10-8 as shown in **Figure 22**. After LD-clumping (r2 0.001, kb 10,000) 34 SNPs from 18 chromosomes remained. I then attempted to replicate the 27 SNPs from 16 chromosomes that had not been previously reported as being associated with lung function.[75]

Figure 23. Manhattan plot of discovery PRISm GWAS

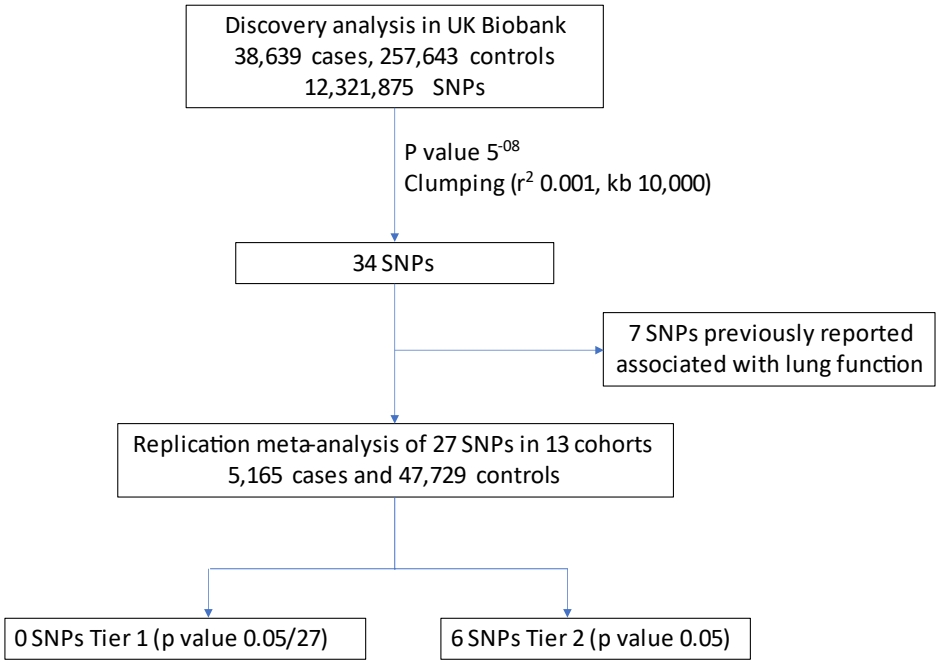


Each circular point in the Manhattan plot represents a SNP tested in the GWAS. The figure shows which chromosome the SNPs are on and their significance threshold for PRISm.

7.4.2 Replication analysis

13 cohorts performed a replication analysis. Replicating cohorts used proxies if SNPs were not found in their panel (r^2 ≥0.8). The meta-analysis totalled 5,165 PRISm cases and 47,729 controls. No SNPs were replicated to tier 1 threshold, but 6 SNPs were replicated to tier 2 threshold. A flow chart showing the process can be seen in **Figure 23**.

Figure 24. Flow chart of analysis



Flow chart showing the process of SNP discovery and replication in the PRISm GWAS

7.4.3 Analysis of GWAS

LD-correlation analysis showed an extremely strong negative correlation between the genetic causes of increasing risk of PRISm and increasing continuous FEV_1 (rg -0.95, p-value <0.001) and FVC (rg -0.93, p-value <0.001). There was strong correlation with peak flow rate (rg -0.65, p-value <0.001) and asthma-COPD overlap (rg 0.52, p-value <0.001), moderate-severe asthma (rg 0.31, p-value

<0.001). Despite the definition of PRISm requiring a preserved ratio, there was a moderate correlation between PRISm and FEV_1/FVC (rg-0.23, p-value <0.001). Type II diabetes showed a moderate correlation with PRISm (rg 0.12, p-value 0.007). The other extra-pulmonary traits all had some evidence of correlation, but the strength of correlation is less meaningful with rg's ≤0.08 as shown in **Table 18**.

Table 18. Results of LD-correlation study between PRISm and both pulmonary and extra-pulmonary GWAS

Trait	**rg with PRISm**	**SE**	**P-value**
FEV_1	-0.96	0.01	<0.001
FVC	-0.93	0.01	<0.001
PEFR	-0.65	0.02	<0.001
FEV_1/FVC	-0.23	0.02	<0.001
Asthma-COPD overlap	0.52	0.04	<0.001
Mod-severe asthma	0.31	0.05	<0.001
Respiratory infection	0.18	0.6	0.003
Eosinophils	0.06	0.02	0.012
Type II Diabetes	0.12	0.03	0.007
BMI	-0.04	0.02	0.031
Systolic hypertension	0.08	0.02	<0.001
Diastolic hypertension	0.05	0.02	0.035
Myocardial infarction	0.07	0.03	0.007

7.4.4 Analysis of replicated SNPs

Table 19 shows the replicated SNPs. LD hub[181] was used to see if the replicated SNPs are in high LD with any previously reported SNPs for lung function in the largest published GWAS to date by Shrine et al.[75] Of the six replicated SNPs, three (rs6923462/rs2240885/rs9421040) are in high LD (r^2 ≥0.7) with a SNP reported as a tier 1 SNP by Shrine et al, and two replicated SNPs (rs780151/rs10278266) are in high LD with tier 3 SNPs from the same GWAS.[75] Rs11113127 is not in the 1000G reference panel of LD hub so it is not possible to report r^2, but it is in close proximity (0.4Mb) to SNPs that showed p-value 1^{-06} for FVC in UK Biobank and the meta-analysis of the Shrine GWAS.

PheWAS showed that all SNPs discovered have been associated with lung function traits in UK Biobank. Rs780151 is associated with anthropomorphic traits in UK Biobank including weight, trunk

fat mass, and basal metabolic rate. Rs2240885 is associated with BMI. Rs2240885 has been associated with risk of type II diabetes in East Asian populations.[182] Rs6923462 has been associated with basal metabolic rate, trunk mass and trunk fat free mass. Rs10278266 has been associated with fasting blood sugar level in both UK Biobank and a non-UK Biobank GWAS.[183]

GTEX[177] examining for SNP association with eQTL showed that rs2240885 is associated with change in gene expression in the *CLUAP1* gene. *CLUAP1* gene has been shown to affect cilia biogenesis by affecting function within the multiple intra-flagellar transport complex B.

None of the SNPs found have been expressed in mouse lungs.[179] None of the nearest genes to the SNPs have been shown to affect Mendelian inherited diseases.[180]

Table 19. Replicated SNPs results from discovery and replication GWAS

SNP	Gene	Gene function	CHR:BP	A1F	A1/A0	UK BioBank			Replication meta		
						OR	95% CI	P-value	OR	95% CI	P-value
rs9431040	*HLX*	Enables sequence-specific DNA binding. Effects organ development and T Helper cell differentiation	1:221152299	0.28	T/C	0.992	0.990 – 0.994	9.2^{-16}	0.945	0.901 – 0.991	0.02
rs6923462	*BMP6*	Encodes a ligand of TNF-beta. Effects iron homeostasis, bone and fat development	6:7801112	0.75	T/C	0.993	0.991 – 0.995	2.9^{-8}	0.912	0.851 – 0.977	0.009
rs10278266	*DGKB*	Regulators secondary messenger diacylglycerol , involved in cellular processes.	7:14943333	0.79	A/G	0.994	0.992 – 0.996	2.5^{-8}	0.924	0.875 – 0.976	0.005
rs780151	*ZMIZ1*	Encodes a protein inhibitor.This regulates transcription factors including an androgen receptor.	10:80931481	0.58	G/A	1.005	1.003 - 1.007	3.5^{-9}	1.062	1.005 – 1.123	0.033
rs11113127	*RIC8B*	Enables G-protein alpha-subunit binding activity, affecting receptor signaling pathway.	12:107597518	0.38	T/C	0.995	0.993 – 0.997	2.4^{-8}	0.945	0.901 – 0.991	0.02
rs2240885	*SLX4*	Encodes a protein that functions as an assembly component for multiple structure-specific endonucleases required for repair DNA lesions.	16:3647098	0.78	G/A	0.994	0.991 – 0.996	9.4^{-9}	0.938	0.889 – 0.989	0.018

A1F – Allele 1 Frequency. A1 – Allele 1 or effect allele. A0 – Allele 0 or reference allele. OR – Odds Ratio. 95% CI – 95% confidence interval. CHR:BP – chromosome:base pair position

7.5 Discussion

This is the first GWAS of PRISm to successfully discover and replicate SNPs, showing that there is a heritable component for the development of PRISm. Modifiable risk factors for PRISm such as smoking and obesity remain important, but not all PRISm risk will be amenable to medical intervention. COPD is no longer viewed as a single organ self-inflicted smoking disease, but a result of genetic and early life factors that determine lung function trajectories.[30,75] PRISm is also a heterogenous condition with variable trajectories that can change to COPD in a minority of cases. Given PRISm and COPD both have genetic risk factors affecting lung function traits, this could partially explain the transition between them.

PRISm has been shown to be associated with systemic co-morbidities such as diabetes, heart disease, and increased risk of mortality in observational studies. One PRISm SNP reported has been associated with type II diabetes, and there was moderate LD-correlation with a type II diabetes GWAS, suggesting that genetic pleiotropy could contribute to co-morbidity as opposed to a causal effect. Similarly, a few SNPs discovered have been associated with BMI and weight which are strong risk factors for PRISm, indicating a shared genetic risk.

All my successfully replicated SNPs are likely to represent loci that have been shown or suspected to be associated with lung function measures as continuous traits. The LD-correlation analysis shows extremely strong correlation between PRISm and continuous lung function traits. This is not surprising given PRISm is an artificial construct of spirometry, as opposed to a disease defined by a unique pathogenic mechanism. Similarly, previous GWAS of COPD based on spirometric criteria have discovered loci that have been described as associated with diverse range of lung diseases including asthma and idiopathic pulmonary fibrosis.[109] This is likely to reflect the genetic heterogeneity of lung diseases, as well as the similar traits used for diagnosis and analysis.

Lung function is a continuous trait, and dichotomisation of continuous traits reduces power of GWAS. Despite a large sample, a small number of SNPS have been discovered that are likely to all represent loci previously discovered for lung function. GWAS of larger samples of PRISm vs control are unlikely to discover SNPs that are unique to PRISm, or novel to lung function. This is compounded by the fact that PRISm is likely to represent heterogenous subtypes. It may be that better phenotyping of PRISm from a more homogenous population could lead to more specific SNP discovery e.g. those with PRISm and a BMI <25 as they may represent a distinct pathological mechanism. Given the high proportion that transition from PRISm to other lung function states over time, it would seem sensible that if future genetic studies were undertaken, those with persistent PRISm are used as cases. However, this will limit total sample size reducing power.

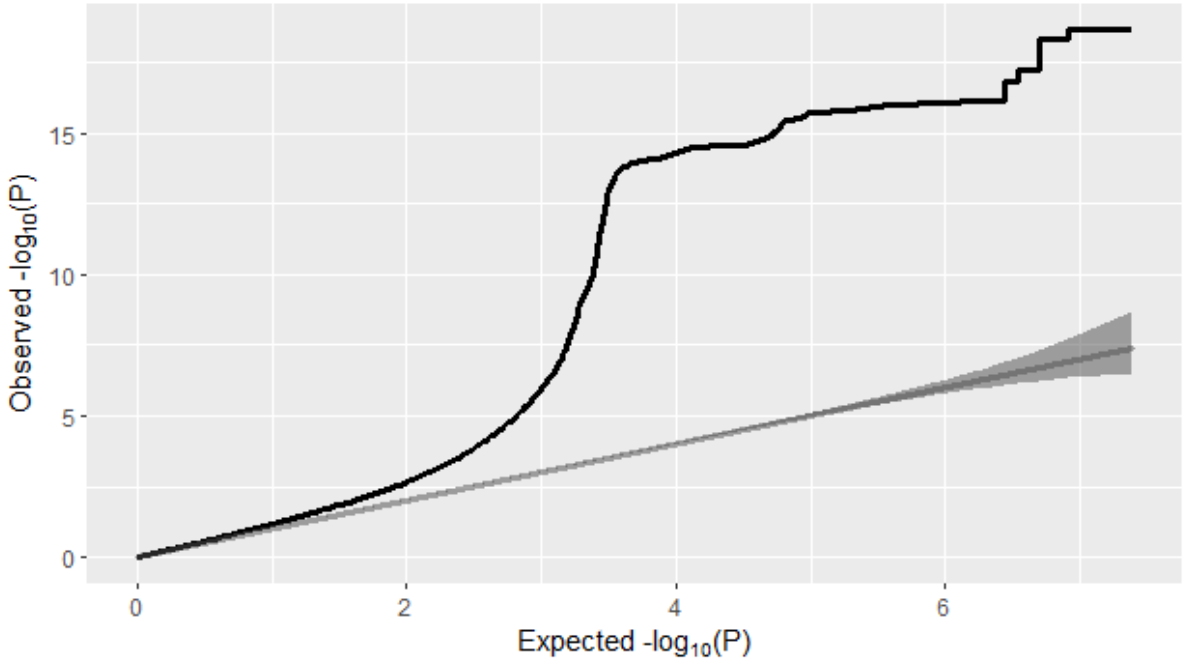
The GWAS was only performed in those of European heritage, these results may not be generalizable to other ancestral populations. The discovery GWAS was performed using pre-bronchodilator values, although medication was not withheld prior to spirometric testing. Although post-bronchodilator values are not required for PRISm diagnosis, there is evidence that spirometric values can change in those with PRISm post-bronchodilation.[169] None of the SNPS found met Bonferroni corrected p-values on replication meta-analysis, potentially leading to the reporting of false positive results.

7.5.1 Genomic Inflation

After completing the first draft of this paper, an external co-author suggested I adjust for genomic inflation. Genomic inflation can lead to false positive SNPs discovery. It can be caused by population stratification, high LD between SNPs, or strong association between the SNPs and the phenotypes tested. A Quantile-Quantile (QQ) plot and lambda value can be used to assess for genomic inflation visually and numerically respectively. A QQ plot is made by plotting SNPs observed p-value against their expected p-value and the lambda is calculated from this. If a straight line is shown on the QQ plot, and the lambda is below 1.1 it indicates no genomic inflation. As can be shown in **Figure 24**, the

QQ plot for the PRISm discovery GWAS has a large deviation from the straight line (in light grey) and the lambda value was 2.0, which is a very high value.

Figure 25. QQ plot of PRISm GWAS



The plot deviates considerably from the predicted line. The associated lambda was 2.0. This indicates many more SNPs discovered than expected and probable genomic inflation.

At the time of writing the first draft of this paper I thought that the genomic inflation was purely driven by the large number of SNPs discovered below the p-value $<5^{-08}$ threshold. Most of these SNPs removed by clumping and not reported, so not a concern (as discussed in 7.4.1 clumping reduced the number of significant SNPs from 8,223 to 34). **Figure 22** shows that there was a hit in the major histocompatibility complex region of chromosome 6 which can introduce many pre-clumping hits due to its high gene density, polymorphism, and linkage disequilibrium.[184]

To further assess and adjust for genomic inflation I used the LD score regression package. This further filtered the SNPs in the discovery GWAS which reduced the lambda to a less concerning 1.25. The LDSC intercept value (a measure of confounding due to population stratification) was low at 1.02. To control for genomic inflation due to population stratification I adjusted the GWAS for the

LDSC intercept value. After clumping the LDSC intercept adjusted GWAS and removing SNPs already reported in the Shrine et al. GWAS, 26 SNPs reached significance threshold. 20 of these SNPs are the same that I reported prior to LDSC intercept adjustment. The 6 SNPs that are differ between the pre and post adjustment results are all in high LD ($r^2 \geq 0.8$) with each other and therefore likely represent the same signals. Therefore the conclusions drawn from my earlier results are still valid despite the high lambda.

7.5.2 Replication method

The same external co-author informed me the replication method I used is not standard. I performed the replication meta-analysis just in the replicating cohorts. This is still of value, and can be reported, but is quite stringent as the replicating cohorts have a much smaller population so are less likely to replicate the SNPs. Instead I intend to combine the replicating cohorts with the discovery cohort in a meta-analysis, and report SNPs that meet p-value $<5^{-08}$. I still hope to undertake this analysis. It is likely to increase the number of SNPs replicated furthering the strength and novelty of the study.

7.6 Conclusion

This is the first GWAS of PRISm to successfully identify SNPs, showing that PRISm has genetic risk factors. It may be that genetic pleiotropy pre-disposes to PRISm and associated anthropomorphic traits and diseases. All SNPs reported have been associated with lung function before and there was an extremely strong correlation with continuous lung function trait GWAS. Further GWAS of PRISm is unlikely to reveal neither novel SNPs for lung function nor SNPs unique to PRISm, unless homogenous subpopulations are used.

CHAPTER 8. THE IMPACT OF ASTHMA ON MENTAL HEALTH AND WELL-BEING DURING COVID-19 LOCKDOWN

8.1 Disclaimer

This paper on which this chapter is based was written during and immediately after the first lockdown for COVID-19. Since then, there have been dramatic changes in the understanding and treatments of COVID-19 as well as evolving social isolation measures, including further lockdowns. However, this paper remains a robust exploration of the impact of asthma on mental health during a period of social isolation. I have included it as my final research chapter as there is no genetic epidemiology element, it focusses on asthma rather than reduced lung function (or states defined by reduced lung function like COPD or PRISm), and it relates to a specific unusual period. It was unplanned, given COVID-19 did not exist in the UK until I was 1 year into my PhD. Despite these differences it is an exploration of extra-pulmonary disease associated with a chronic lung disease so is well suited to this thesis. I begin with a background discussion of asthma, its associated co-morbidities and COVID-19 as these issues have not been discussed previously.

8.2 Publication and contributions

This chapter has been published as a letter.

The impact of asthma on mental health and wellbeing during COVID-19 lockdown. Daniel Higbee, George Nava, Alex Kwong, James Dodd, Raquel Granell. European Respiratory Journal. 2021 Jul 29;58(1):2004497.

DH, JD, and RG conceived the project and determined the data analysis plan. RG performed the data analysis. DH, GN and RG wrote the paper. AK provided advice regarding utilising the data. JD provided overall supervision.

8.3 Background

8.3.1 Asthma

Asthma was first described as a collection of symptoms in the 4th century BC by the ancient Greeks characterised by spasms of breathing difficulty.[185] Asthma affects 300 million people worldwide and is a heterogenous condition characterised by symptoms, such as wheezing, shortness of breath, chest tightness, and cough that vary with onset, frequency and intensity.[186] Airway inflammation and mucus hypersecretion leads to bronchial hyperresponsiveness and airflow obstruction.[187]

8.3.2 Asthma and co-morbidity

Asthma is in the top 10 most prevalent conditions associated with co-morbidity, being associated most frequently with hypertension, pain, anxiety & depression. [188] Anxiety and depression are more common in people with multi-morbidity which can impact their ability to manage their conditions, with negative consequences. [189] The National Registry of Asthma Deaths audited the notes of patients that died from asthma.[190] They found that psychological co-morbidity contributed to the risk of asthma death, with 16% of those that died having depression and mental health issues.

8.3.2.1 Asthma & Anxiety

Studies examining the association of asthma and anxiety have been conducted in different sample populations and controlled for different confounding factors. Results vary so the true nature of the association is uncertain.

A meta-analysis found that across 15 studies comprising 15,443 patients there was 22% prevalence of anxiety in asthmatic people.[191] 11 of the studies comprising 10,223 people under the age of 18 found that young people with asthma had a ~30% higher number anxiety disorders (Cohen's d = 0.37, 95%CI: 0.24-0.50) and anxiety symptoms (d = 0.29, 95%CI: 0.19-0.39) compared to those without asthma. This effect is not just seen in youth, a case-controlled study using patients with a

mean age of 40-41 years compared 96 asthmatics with controls and found that asthma was associated with Odds Ratio (OR) of 3.03 (1.41–6.84) of lifetime anxiety.[192]

The effect is maintained, to varying degrees across racial groups. A large study of US citizens found that stratified by ethnicity there was an increased risk of generalised anxiety was seen in Latinos and black patients but not in Asians.[193] These findings differ from a large Taiwanese study that found that asthma was an independent risk factor for anxiety with an adjusted hazard ratio of 1.8 (1.66-1.95). [194]

When asthma and anxiety do co-exist research has shown that they produce a worse quality of life, [195-198] and increased healthcare utilisation compared to asthma without anxiety. [199]

However, the association between asthma and anxiety is not seen in all studies and may be confounded by a number of factors. One study found that after controlling for confounding factors, asthma was no longer associated with increased odds of anxiety disorders. [200] A case-controlled study found that in those with well controlled asthma, as evidenced by a low Asthma Control Questionnaire score, there was no increased risk of anxiety. [201]

8.3.2.2 Asthma & Depression

The relationship between asthma and depression is also uncertain. Some studies have found an increase in depression amongst asthmatics [202-204] with one study (Cases = 1453, Controls = 5812) reporting a Hazards Ratio (HR) of 1.81 (1.14–2.89) for major depression and 1.74 (1.27–2.37) for any depressive disorder. A large prospective multi-national cohort (N= 85,088) estimated an OR of 1.6 (1.4-1.8).[205] A meta-analysis of 51 paper found a higher risk of depression was observed in patients with asthma Relative Risk (RR) 1.58 (1.44 – 1.74).[206] The severity of asthma control may affect the relationship. A small (N = 140) case-controlled study of well-controlled asthmatics found no correlation[201] between asthma and worse scores for depression compared to healthy controls. A meta-analysis of 8 papers found that severity of asthma was not a moderator for the asthma-depression risk.[204] The direction of affect is not certain and bi-directionality has been explored. A

large cohort study found an increased risk of developing asthma after a major depressive disorder [207] and a meta-analysis of prospective studies found that depression predicted the development of asthma, but notably not the other way around.[208]

Having co-existent depression worsens quality of life,[196,209] and asthma control [197] with a meta-analysis of 10 papers (N = 31,432 adults with asthma) showing that depression increased the risk of hospitalisations (RR 1.07 (1.04-1.11) and A&E visits (RR 1.06 (1.02-1.11) due to asthma.[210]

8.3.2.3 Underlying Mechanisms between asthma and mental health coi-morbidity

Asthma may cause these psychological disorders, the psychological disorders may cause asthma, or there may be a shared common pathway. What the underlying pathological mechanisms could be remains uncertain, but various pathways have been hypothesised for both shared and directional causality. Interactions between behavioural, neural, endocrine, and immune processes suggest that psychological factors play an active role in causing asthma.[211] However, there is also research demonstrating biological pathways whereby asthma can cause anxiety, particularly via the hypothalamus-pituitary axis.[212] This effect can be influenced by the presence of depression. From a behavioural perspective, physical manifestations of anxiety and panic can worsen asthma symptoms and reduce treatment response via hyperventilation, changes in self-management behaviour, and physician response.[213]

8.3.2.4 Coronavirus

The novel severe acute respiratory syndrome coronavirus 2 (COVID-19) global pandemic has resulted in hundreds of millions of infections and millions of deaths.[214] I began my PhD in January 2018, before the pandemic. During the first and second wave of COVID-19 in the UK I returned to my previous clinical role as a respiratory and general medicine registrar. My primary supervisor also had a huge increase in his clinical duties. During both the first and second wave I was involved in the recruitment of patients for the multi-armed Randomised Evaluation of COVID-19 Therapy (RECOVERY Trial) leading to multiple treatment findings and publications of which I am a co-

author.[215-219] Although this reduced research time, it also prompted research questions and opportunities.

To reduce transmission of COVID-19 in the population and decrease pressure on hospitals, social isolation measures were introduced. It was assumed that people living with pre-existing respiratory diseases would be at higher risk of developing severe COVID-19 illness, despite a lack of available evidence at the time.[220] This meant that there were prolonged periods when recommended social isolation measures were stricter for those with respiratory disease. Cross-sectional studies showed a high prevalence of anxiety and depression during COVID-19 across ages and countries.[221,222] Research to assess the impact of having asthma on mental health during the first lockdown was required, especially given the debate about the risk-benefit of further lockdowns underway at the time of research. Available sample populations had not included longitudinal data (including pre-pandemic measures) or detailed clinical history, prior physical or psychological assessment, meaning that conclusions about the effect of COVID-19 on people with asthma were impossible.

8.4 Introduction

To mitigate the risk of harm of COVID-19 to those with chronic respiratory disease, and to reduce transmission in the general population, measures to restrict social interaction including social distancing, self-isolation, and quarantining were introduced. Similar measures were used during the Severe Acute Respiratory Syndrome (SARS) epidemics leading to negative mental health impacts even in those not exposed to the virus.[223] Unlike SARS, COVID-19 has caused a high proportion of the global population to endure social isolation measures and financial instability, increasing the potential for psychological harm. The UK, and other countries around the world, have had repeated social isolation measures and may re-instate them. This is an ongoing threat to mental health.

Asthma is a common respiratory disease of co-morbidity affecting ~10% of the UK's population and is associated with anxiety and depression.[188,224] The presence of psychological co-morbidity with asthma decreases quality of life, increases healthcare utilisation, and contributes to the risk of asthma death.[190,198,199]

Cross-sectional studies have shown a high prevalence of anxiety and depression during COVID-19 across ages and countries.[221,222] Available sample populations thus far have not included longitudinal data (including pre-pandemic measures) or detailed clinical history, prior physical or psychological assessment, meaning that conclusions about the effect of COVID-19 on people with asthma have been impossible.

My objective was to determine if people with asthma experienced worse anxiety, depression and/or wellbeing during the first lockdown for COVID-19 compared to people without asthma. The secondary objectives were to explore the additional impact of the COVID-19 pandemic on factors associated with the disease and lockdown. I hypothesized that physical symptoms such as having a suspected COVID-19 infection or difficulty breathing, as well as social factors such as worry of losing a job, could impact people with asthma more severely than people without asthma.

If there is a more severe worsening of mental health or wellbeing among people with asthma, it could be used to help identify those at risk for support. Associated factors may be modifiable reducing the impact. Results may also help balance public health decisions regarding re-instigation of lockdown measures.

8.5 Methods

8.5.1 Study Samples

Data from the Avon Longitudinal Study of Parent and Children (ALSPAC), a British longitudinal cohort of children born in 1991 and their parents, was utilised to compare associations of asthma in young and middle-aged adults.[225]

ALSPAC is an ongoing longitudinal population-based study that recruited pregnant women residing in Avon in the south-west of England with expected delivery dates between 1st April 1991 and 31st December 1992.[225,226] The cohort consists of 13,761 mothers and their partners (referred to as ALSPAC-G0), and their 14,901 children (ALSPAC-G1).[227] The cohort has been extensively investigated with at regular time points allowing us to compare pre-COVID-19 levels of mental and physical disease to anxiety and depression assessed during lockdown. This study uses data from 3737 ALSPAC-G0 and 2942 ALSPAC-G1 who completed an online questionnaire in April 2020 covering COVID-19 symptoms, mental health, lifestyle, finances, and family life.[228] This questionnaire was designed to quantify the impact of COVID-19, how it is transmitted among the general population, and learn more about the psychosocial and economic impact of the government's lockdown strategy. The first lockdown was announced in the UK on the 23rd of March.

8.5.2 Self-reported Current Asthma

Current asthma was assessed in the COVID-19 surveys with the question - Participant has asthma yes/no- at mean age 28 years in ALSPAC-G1 and at mean 59 years in ALSPAC-G0. In ALSPAC-G1 the proportion of participants with asthma correlated well with previous reports of current asthma at 23 years (92% overlap); in ALSPAC-G0 the overlap with previous reports of asthma was 72% with 'asthma ever' at 39 years.

8.5.3 Pre-COVID-19 vs COVID-19 Mental Health

The measures used in the COVID-19 survey examine symptoms in the preceding 2 weeks, thus represent mental health in lockdown. Depressive symptoms in both ALSPAC-G0 and ALSPAC-G1 samples were measured using the Short Mood and Feelings Questionnaire (SMFQ).[18] Scores range

between 0-26 with higher scores indicating higher depressive symptoms. Anxiety symptoms were measured using the Generalised Anxiety Disorder Assessment (GAD-7).[19] Scores range between 0-21 with higher scores indicating higher anxiety symptoms. Mental wellbeing was measured using the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS).[20] Scores range between 14-70, with higher scores indicating better mental wellbeing.

In ALSPAC-G1 the same scores measured at the COVID-19 survey (mean age=28) were available pre-pandemic: WEMWBS-14 for wellbeing at 24 years, SMFQ-13 score for depression at 26 years, and GAD-7 score for anxiety at 22 years.

In the older ALSPAC-G0 population (mean age=59), mental wellbeing was not previously assessed. The Edinburgh Postnatal Depression Scale score was used for depression [229] at 52 years (scoring ranges from 0-30) and State-Trait Anxiety Inventory score for anxiety [230] at 39 years (scoring ranges from 20-80), as pre-pandemic measures for adjustment.

8.5.4 Symptoms, Change in Activities, and Worries during lockdown

In participants with and without asthma I explored factors relating to COVID-19 and lockdown available in the ALSPAC COVID questionnaire including symptoms (e.g. breathlessness), changes in activities (e.g. self-isolation), worries during lockdown (e.g. job losses), and the differences in those with confirmed or suspected COVID-19.

8.5.5 Statistical analyses

Analysis was conducted in Stata (version 15).[161]

P-values were reported from Pearson Chi-square tests when comparing categorical characteristics and Z-test from logistic regression when comparing continuous characteristics in people with asthma vs. people without asthma. Two-sample Kolmogorov-Smirnov tests for equality of distribution functions were used to compare mental health scores between the younger population with asthma (ALSPAC-G1) vs. the older population with asthma (ALSPAC-G0).

Poisson and logistic regression models were used to estimate the effect of asthma on wellbeing, anxiety and depression, and other factors relating to COVID-19 and lockdown. Missing data was not imputed as a different analysis using the same data resources did not find it changed results.[231] Analysis was conducted separately for ALSPAC-G0 and ALSPAC-G1.

For ALSPAC-G0 adjustment was performed for sex, age, smoking/vaping status and overweight. Adjustment for sex used ALSPAC variable kz021, taken from Cohort profile data file version 2b[232]. Age and smoking/vaping status were adjusted using two participant reported variables taken from the COVID questionnaire (variables covid1yp_1004 and covid1yp_3053 from datset COVID1_YP_1b dat for age and smoking/vaping status respectively). Overweight was a variable created from participant reported BMI (overweight defined as BMI >25) from the COVID questionnaire, variable covid1yp_1004.

For ALSPAC G0 adjustment was performed using variables from self-reported data from the COVID questionnaire. Adjustment for age used covid1m_9650 for mothers and covid1p_9650 for partners. For smoking/vaping status covid1m_3053 for mothers and covid1p_3053 for partners. For overweight adjustment a variable calculated from BMI was used. For mothers covid1m_1004 and covid1p_1004 for partners. Mothers' variables come from data set COVID1_Mum_1b and partners from data set COVID1_G0partner_1b.

Analyses were also adjusted for pre-existing mental health as described in detail above.

Demographics for these variables are presented in Table 20 below.

Exponentiated Poisson regression estimates, exp(b), were reported as symptom count ratios (SCRs), known as incident rate ratios. These can be interpreted as a percentage increase/decrease of the outcome score in asthmatics, compared to non-asthmatics given as (SCR-1)×100 [233,234].

2-sample Z-test were used to compare the asthma effects in the younger population (ALSPAC-G1) vs. the older population (ALSPAC-G0).

8.6 Results

8.6.1 Demographics

410 (13.9%) participants in ALSPAC-G1 (mean age 28) and 400 (10.7%) in ALSPAC-G0 (mean age 59 years) responded yes to having asthma in the COVID-19 survey. Full details in **Table 20**. Participants with asthma reported being overweight more (28% vs. 19% $p = 1.4\times10^{-5}$ in ALSPAC-G0 and 39% vs 33% p = 0.02 in ALSPAC-G1) and taking regular medications (72% vs. 40% $p=1.8\times10^{-32}$ in ALSPAC-G0 and 89% vs. 60% $p = 1.4\times10^{-29}$ in ALSPAC-G1), compared to participants without asthma. Furthermore, in ALSPAC-G0 there were more females (79% vs. 72% p = 0.01) and keyworkers (38% vs. 31% p = 0.01) among people with asthma, compared to those without.

Table 20. Demographics in G1-Index children and G0 parents' cohorts

	G1-Index Children					**G0 Parents**				
	No Asthma at 28 years		Asthma at 28 years			No Asthma at 59 years		Asthma at 59 years		
	N=2532 (86.1%)		N=410 (13.9%)			N=3337 (89.3%)		N=400 (10.7%)		
	Total	Mean (SD)	Total	Mean (SD)	P-value†	Total	Mean (SD)	Total	Mean (SD)	P-value†
Age in years	2532	27.6 (0.5)	410	27.6 (0.5)	0.18	3337	58.7 (4.8)	400	57.8 (5.0)	2.13E-04
	Demographics at 28 years					Demographics at 59 years				
	Total	N (%)	Total	N (%)	P-value†	Total	N (%)	Total	N (%)	P-value†
Gender (male)	2530	727 (28.7)	410	112 (27.3)	0.56	3337	917 (27.5)	400	85 (21.3)	0.008
Smoking/vaping	2479	494 (19.9)	403	85 (21.1)	0.59	3228	297 (9.2)	378	27 (7.1)	0.19
Overweight	2517	471 (18.7)	407	114 (28)	1.36E-05	3309	1079 (32.6)	399	154 (38.6)	0.016
Diabetes	2528	11 (0.4)	408	5 (1.2)	0.04	3324	130 (3.9)	397	21 (5.3)	0.19
Keyworker	2393	1303 (54.5)	392	215 (54.8)	0.88	3122	967 (31)	375	141 (37.6)	0.01
Taking regular medications	2530	1014 (40.1)	410	293 (71.5)	1.83E-32	3334	1997 (59.9)	400	355 (88.8)	1.42E-29

† Pearson Chi-square Test for categorical variables, Z-test from logistic regression for continuous variables

8.6.2 Pre-COVID-19 vs COVID-19 Mental Health

The younger ALSPAC-G1 participants with asthma reported worse pre-existing wellbeing scores at 24 years (mean WEMWBS-14 score 47.5 (SD 9.4) vs. 49.0 (8.7) p=0.01) and worse pre-existing depression scores at 26 years (mean SMFQ-13 score 8.1 (6.6) vs. 6.6 (6.2) p=1.6×10-4), compared to participants with no asthma. Little or no evidence for worse pre-existing anxiety scores using GAD-7 score at 22 years (p=0.22) was found. However, in the COVID-19 survey G1 participants with asthma reported worse wellbeing (42.2 (8.8) vs. 44.4 (8.4) p=1.9×10-6), more depression (8.3 (6.2) vs. 6.7 (5.3) p=1.6×10-7) and more anxiety (8.3 (5.4) vs. 6.9 (5.0) p=1.9×10-6), compared to younger people with no asthma (see **Table 21**). In the older ALSPAC-G0 population, there was little evidence of people with asthma reporting worse pre-COVID-19 lockdown depression or anxiety (p≥0.06), however they did report worse depression (3.6 (4.4) vs. 2.9 (3.6) $p=4.5\times10^{-4}$) and anxiety (4.2 (4.9) vs. 3.4 (4.0) $p=2.3\times10^{-4}$) at the lockdown COVID-19 survey, compared to older people with no asthma (see **Table 21**).

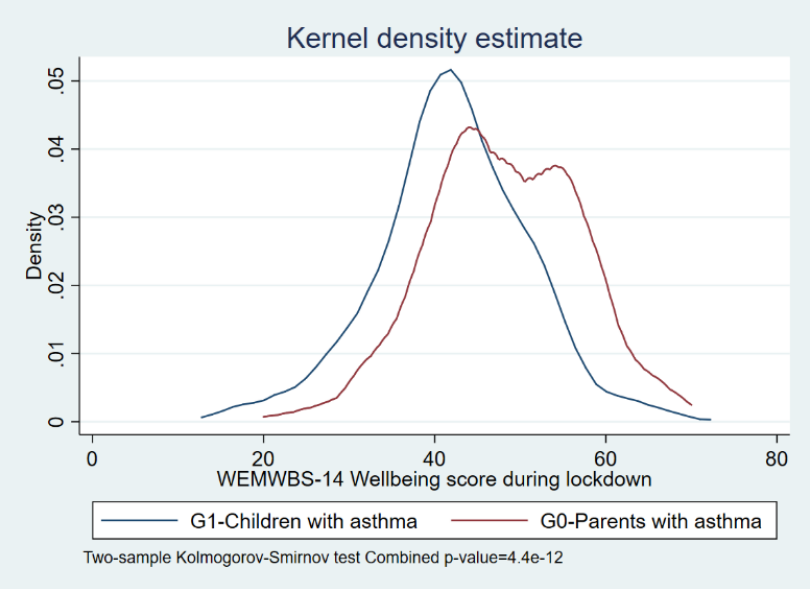
Table 21. Lockdown and pre-existing mental health: asthma vs. no asthma

G1-Index Children					
	No Asthma at 28 years		**Asthma at 28 years**		
	N=2532 (86.1%)		N=410 (13.9%)		
	Total	Mean (SD)	Total	Mean (SD)	P-value†
Pre-existing Mental Health					
WEMWBS-14 Wellbeing score at 24 years	1959	49.0 (8.7)	313	47.5 (9.4)	0.007
SMFQ-13 Depression score at 26 years	1979	6.6 (6.2)	313	8.1 (6.6)	1.56E-04
GAD-7 Anxiety score at 22 years	1615	4.5 (4.4)	245	4.9 (4.4)	0.22
Mental Health during lockdown					
WEMWBS-14 Wellbeing score at 28 years	2399	44.4 (8.4)	392	42.2 (8.8)	1.93E-06
SMFQ-13 Depression score at 28 years	2148	6.7 (5.3)	362	8.3 (6.2)	1.62E-07
GAD-7 Anxiety score at 28 years	2113	6.9 (5.0)	362	8.3 (5.4)	1.90E-06
G0 Parents					
	No Asthma at 59 years		Asthma at 59 years		
	N=3337 (89.3%)		N=400 (10.7%)		
	Total	Mean (SD)	Total	Mean (SD)	P-value†
Pre-existing Mental Health					
EPDS-10 Depression score at 52 years	2591	6.3 (5.3)	307	6.9 (5.3)	0.06
STAI-20 Anxiety score at 39 years	2831	35.0 (10.2)	332	35.7 (10.6)	0.23
Mental Health during lockdown					
WEMWBS-14 Wellbeing score at 59 years	3087	48.1 (8.4)	364	48.1 (8.8)	0.96
SMFQ-13 Depression score at 59 years	3076	2.9 (3.6)	364	3.6 (4.4)	4.49E-04
GAD-7 Anxiety score at 59 years	3138	3.4 (4.0)	371	4.2 (4.9)	2.32E-04

† Z-test from logistic regression

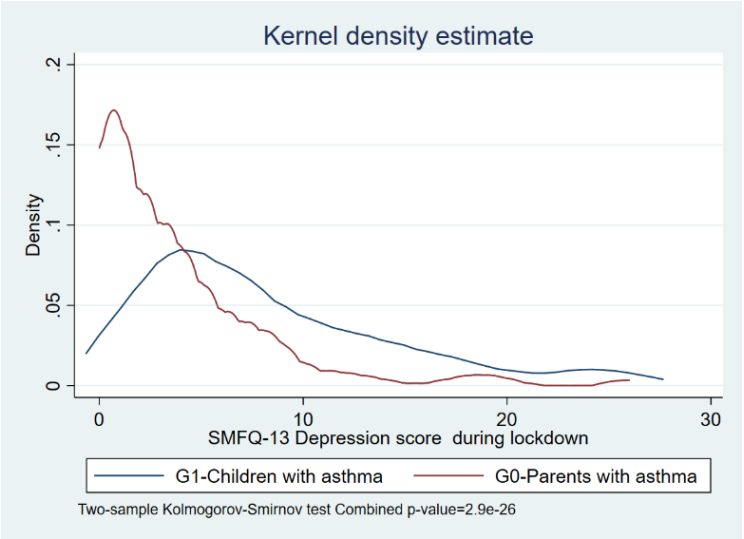
Well-being, depression, and anxiety levels were all worse in G1 asthmatics compared to G0 asthmatics during COVID-19 lockdown (p-value <0.001, see **Figures 25, 26, 27**). This was tested using the Two-sample Kolmogorov-Smirnov test which used for equality of distribution functions to compare mental health scores between the younger population with asthma (ALSPAC-G1) vs. the older population with asthma (ALSPAC-G0).

Figure 26. WEMWBS-14 Wellbeing score during lockdown in G0-ALSPAC children with asthma vs G1-ALSPAC parents with asthma



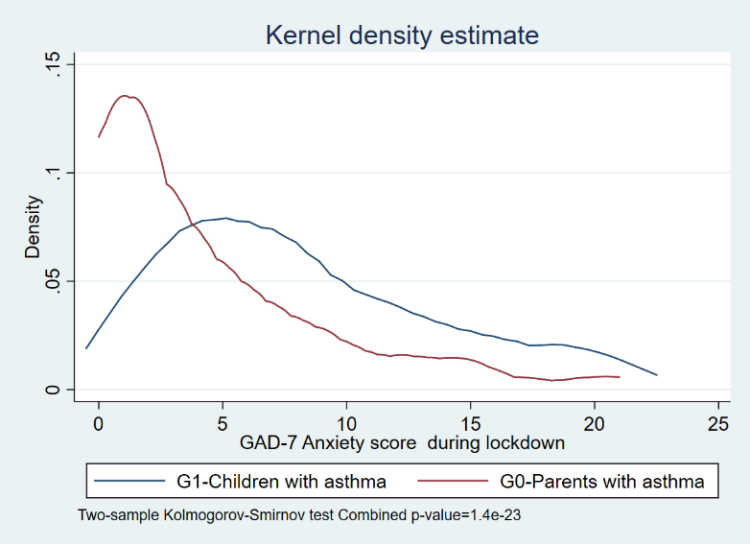
This plot and corresponding p-value showing strong evidence for a difference in well-being scores between the G1 and G0 groups during lockdown

Figure 27. SMFQ-13 Depression score during lockdown in G0-ALSPAC children with asthma vs G1-ALSPAC parents with asthma



This plot and corresponding p-value showing strong evidence for a difference in depression scores between the G1 and G0 groups during lockdown

Figure 28. GAD-7 Anxiety score during lockdown in G0-ALSPAC children with asthma vs G1-ALSPAC parents with asthma



This plot and corresponding p-value showing strong evidence for a difference in anxiety scores between the G1 and G0 groups during lockdown

In addition to the absolute mental health scores in asthmatics vs. non-asthmatics, the expected difference in mental health scores during COVID-19 between asthmatics and non-asthmatics was examined after adjusting for pre-existing mental health, gender, age, smoking, and being overweight as shown in **Tables 22** and **23**. Asthma was associated with a 13% increase in depression score in ALSPAC-G1 (adjusted SCR 1.13 95%CI (1.04,1.22), p=0.005 in ALSPAC-G1) and 15% increase in ALSPAC-G0 (1.15 (1.00,1.31), p=0.05), compared to those without asthma. Anxiety scores in people with asthma increased by 14% in ALSPAC-G1 (1.14 (1.04,1.26), p=0.005) and by 16% in ALSPAC-G0 (1.16 (1.02,1.32), p=0.02 in ALSPAC-G0), compared to no asthma. Well-being score in people with asthma decreased by 3% (0.97 (0.95,1.00), p=0.02) in ALSPAC-G1, compared to those without asthma (NA in ALSPAC-G0 as no previous well-being score).

Table 22. Effect of current asthma on outcomes specific to lockdown adjusted for age, gender, smoking, overweight and pre-existing mental health in G1-Index Children

	Effect of Current Asthma at 27 years on Lockdown in G1-Index Children							
	R^2	Crude Model-restricted	Wald Test P-value	N	R^2	Adjusted Model*	Wald Test P-value	N
Count Outcomes		SCR (95% CI)				SCR (95% CI)		
WEMWBS-14 Wellbeing score (14-70)	0.001	0.96 (0.93,0.98)	0.001	2153	0.05	0.97 (0.95,1.00)	0.016	2153
SMFQ-13 Depression score (0-26)	0.007	1.26 (1.14,1.38)	3.03E-06	1963	0.16	1.13 (1.04,1.22)	0.005	1963
GAD-7 Anxiety score (0-21)	0.003	1.17 (1.06,1.29)	0.002	1557	0.08	1.14 (1.04,1.26)	0.005	1557
Length of self-isolation in weeks (0-13)	0.002	1.32 (0.95,1.84)	0.09	2144	0.02	1.27 (0.91,1.77)	0.15	2144
Worried about getting COVID-19 score (1-5)	0.002	1.14 (1.09,1.19)	6.27E-09	2186	0.01	1.12 (1.08,1.17)	7.55E-08	2186
Worried about losing their job score (1-5)	0.001	1.12 (1.03,1.21)	0.005	1980	0.005	1.10 (1.02,1.19)	0.016	1980
Binary Outcomes		OR (95%CI)				OR (95%CI)		
Confirmed or suspected COVID19	0.00	0.76 (0.54,1.08)	0.13	2218	0.00	0.76 (0.53,1.08)	0.13	2218
Shortness of breath or difficulty in breathing	0.02	2.81 (1.51,5.21)	0.001	2136	0.05	2.42 (1.29,4.55)	0.01	2136
Difficulty sleeping	0.00	1.34 (1.01,1.79)	0.046	2170	0.03	1.23 (0.92,1.66)	0.17	2170
Participant has or is self-isolating	0.00	1.08 (0.81,1.42)	0.61	2202	0.01	1.03 (0.78,1.36)	0.84	2202
Eating change since lockdown	0.00	1.20 (0.94,1.54)	0.13	2210	0.02	1.12 (0.88,1.44)	0.36	2210

† Adjusted by pre-existing mental health, overweight, smoking, gender, and age; * scale is 1 (not-at-all) to 5 (very); ‡ WEMWBS score not available for G0-ALSPACSCR= symptom count ratios from Poisson regression models; OR=odds ratios from logistic regression models

Table 23. Effect of current asthma on outcomes specific to lockdown adjusted for age, gender, smoking, overweight and pre-existing mental health in G0-parents

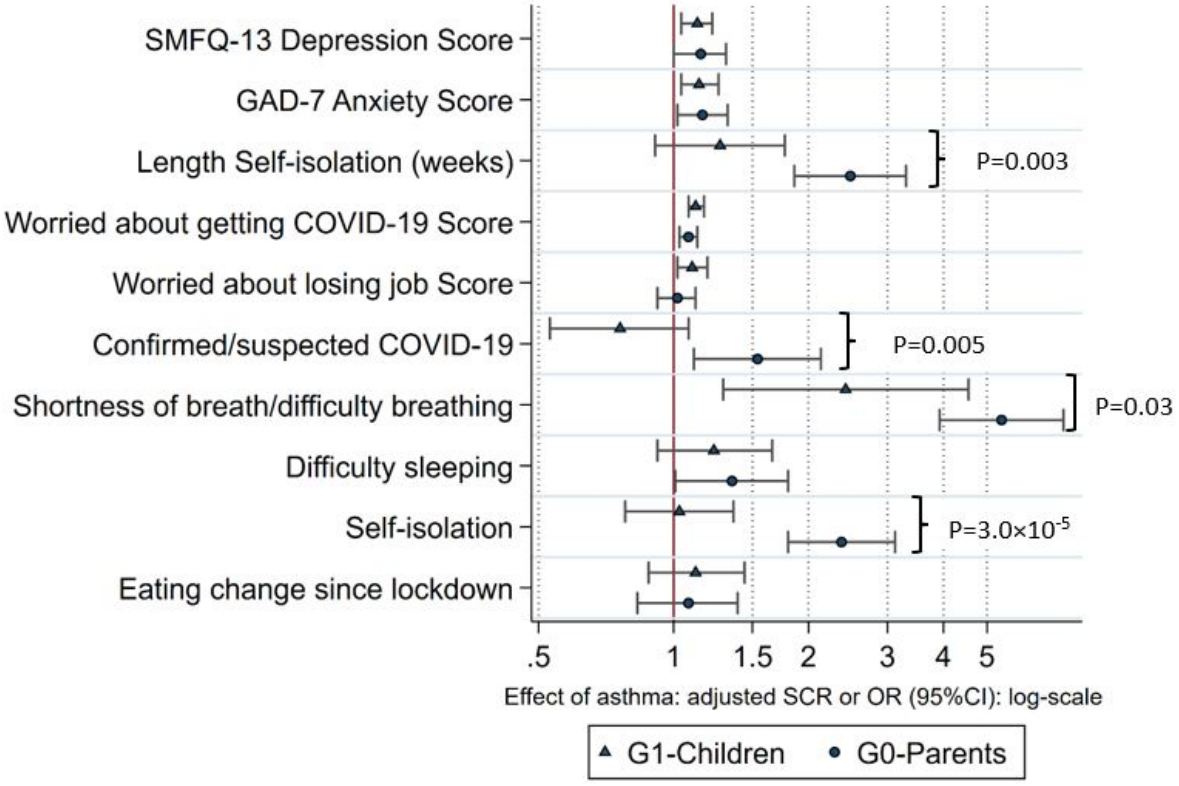
	Effect of Current Asthma at 59 years on Lockdown in G0-Parents							
	R^2	Crude Model-restricted	Wald Test P-value	N	R^2	Adjusted Model*	Wald Test P-value	N
Count Outcomes		SCR (95%CI)				SCR (95%CI)		
SMFQ-13 Depression score (0-26)	0.002	1.23 (1.06,1.43)	0.007	2652	0.14	1.15 (1.00,1.31)	0.05	2652
GAD-7 Anxiety score (0-21)	0.003	1.26 (1.10,1.45)	9.97E-04	2940	0.11	1.16 (1.02,1.32)	0.02	2940
Length of self-isolation in weeks (0-13)	0.024	2.52 (1.89,3.36)	2.87E-10	2689	0.05	2.48 (1.86,3.30)	5.8E-10	2689
Worried about getting COVID-19 score (1-5)	0.001	1.09 (1.04,1.14)	9.44E-05	2758	0.004	1.08 (1.03,1.13)	4.1E-04	2758
Worried about losing their job score (1-5)	0.000	1.04 (0.94,1.15)	0.41	1784	0.01	1.02 (0.92,1.12)	0.73	1784
Binary Outcomes		OR (95%CI)				OR (95%CI)		
Confirmed or suspected COVID19	0.00	1.63 (1.18,2.25)	0.003	2770	0.01	1.54 (1.11,2.13)	0.01	2770
Shortness of breath or difficulty in breathing	0.07	5.75 (4.21,7.84)	3.01E-28	2748	0.10	5.39 (3.92,7.41)	3.4E-25	2748
Difficulty sleeping	0.00	1.45 (1.09,1.93)	0.01	2685	0.04	1.35 (1.01,1.80)	0.04	2685
Participant has or is self-isolating	0.01	2.39 (1.82,3.13)	3.92E-10	2769	0.02	2.37 (1.80,3.12)	8.5E-10	2769
Eating change since lockdown	0.00	1.20 (0.94,1.54)	0.150	2771	0.04	1.08 (0.83,1.39)	0.57	2771

† Adjusted by pre-existing mental health, overweight, smoking, gender, and age. In the parents, WEMWBS-14 Wellbeing score is not available at a previous timepoint. We used previous STAI score at 39 years to adjust current GAD-7 anxiety at 59 years. For all other variables we used previous EPDS score for depression at 52 years to adjust for pre-existing mental health. * Scale is 1 (not-at-all) to 5 (very)

SCR= symptom count ratios from Poisson regression models; OR=odds ratios from logistic regression models

Asthma was associated with a similar change of anxiety and depression scores during COVID-19 in both generations (Z test p-values >0.80. See **Figure 28**, **Table 26**).

Figure 29. Effect of current asthma on factors specific to lockdown adjusted for age, gender, smoking status, overweight and pre-existing mental health.



P-values are from Z-test for 2-sample (2-sided).

Figure 28 is a Forest plot comparing the effect of asthma on different outcomes in G1 and G0 cohorts. The brackets indicate outcomes where there was strong evidence of a difference between the two generations.

8.6.3 Symptoms, Change in Activities, and Worries during lockdown

Since lockdown, ALSPAC-G1 participants with asthma reported more shortness of breath or difficulty in breathing, more difficulty sleeping, longer self-isolation, more eating and sleep changes, more worry about getting COVID-19 and losing their job compared to those without asthma (**Table 24**).

Table 24. Lockdown characteristics: asthma vs. no asthma in G1-Index Children

	G1 Index Children				
	No Asthma at 28 years		**Asthma at 28 years**		
	N=2532 (86.1%)		N=410 (13.9%)		
	Total	**N (%)**	**Total**	**N (%)**	**P-value†**
Symptoms during lockdown					
Confirmed COVID-19 by test	2529	47 (1.9)	410	5 (1.2)	0.36
Confirmed or suspected COVID-19	2529	406 (16.1)	410	60 (14.6)	0.47
Shortness of breath or difficulty in breathing in last week	2425	44 (1.8)	391	25 (6.4)	5.47E-08
Contact with confirmed COVID-19 in last 2 weeks	2527	109 (4.3)	410	14 (3.4)	0.40
Contact with confirmed or suspected COVID-19 in last 2 weeks	2527	278 (11.0)	410	47 (11.5)	0.78
Difficulty sleeping in last week	2479	486 (19.6)	396	98 (24.7)	0.018
Severe fatigue in last week	2459	82 (3.3)	396	16 (4)	0.47
Went to A&E about symptoms last week	2532	3 (0.1)	410	0 (0.0)	0.49
Sought medical attention last week	2532	4 (0.2)	410	2 (0.5)	0.17
Change of activities during lockdown					
Participant has or is self-isolating	2459	621 (25.3)	400	114 (28.5)	0.17
Sleep change since lockdown	2461	1379 (56)	401	248 (61.8)	0.029
Exercise change since lockdown	2468	1843 (74.7)	401	289 (72.1)	0.27
Alcohol drinking change since lockdown	1942	1216 (62.6)	297	191 (64.3)	0.57
Eating change since lockdown	2474	1226 (49.6)	401	226 (56.4)	0.011
	Total	**Mean (SD)**	**Total**	**Mean (SD)**	**P-value‡**
Length of self-isolation in weeks	2394	0.92 (2.54)	383	1.36 (3.32)	0.003
Worries during lockdown					
Worried about getting COVID-19 *	2431	2.82 (1.17)	402	3.24 (1.17)	8.71E-11
Worried about losing their job*	2211	2.13 (1.30)	359	2.32 (1.43)	0.010

† Pearson Chi-square Test; ‡ Z-test from logistic regression; * scale is 1 (not-at-all) to 5 (very)

The older ALSPAC-G0 participants with asthma also reported more shortness of breath, difficulty sleeping, longer self-isolation, more sleep and eating changes, more worry about getting COVID-19, and more loneliness, compared to those without asthma. Additionally, the older G0-ALSPAC

participants with asthma reported more confirmed or suspected COVID-19 and more self-isolation (**Table 25**).

Table 25. Lockdown characteristics: asthma vs. no asthma in G0 Parents

	G0 Parents				
	No Asthma at 59 years		**Asthma at 59 years**		
	N=3337 (89.3%)		N=400 (10.7%)		
	Total	**N (%)**	**Total**	**N (%)**	**P-value†**
Symptoms during lockdown					
Confirmed COVID-19 by test	3318	22 (.7)	396	11 (2.8)	2.25E-5
Confirmed or suspected COVID-19	3318	390 (11.8)	396	70 (17.7)	0.001
Shortness of breath or difficulty in breathing in last week	3295	210 (6.4)	391	103 (26.3)	6.64E-41
Contact with confirmed COVID-19 in last 2 weeks	3327	50 (1.5)	396	4 (1)	0.44
Contact with confirmed or suspected COVID-19 in last 2 weeks	3327	166 (5)	396	20 (5.1)	0.96
Difficulty sleeping in last week	3202	656 (20.5)	387	102 (26.4)	0.008
Severe fatigue in last week	3186	46 (1.4)	384	9 (2.3)	0.18
Went to A&E about symptoms last week	3336	2 (0.1)	399	3 (0.8)	3.54E-04
Sought medical attention last week	3336	11 (.3)	399	1 (.3)	0.79
Change of activities during lockdown					
Participant has or is self-isolating	3222	538 (16.7)	374	121 (32.4)	1.29E-13
Sleep change since lockdown	3216	1227 (38.2)	377	172 (45.6)	0.005
Exercise change since lockdown	3219	2250 (69.9)	378	267 (70.6)	0.77
Alcohol drinking change since lockdown	2609	1054 (40.4)	291	133 (45.7)	0.08
Eating change since lockdown	3222	1182 (36.7)	379	166 (43.8)	0.007
	Total	**Mean (SD)**	**Total**	**Mean (SD)**	**P-value‡**
Length of self-isolation in weeks	3135	0.75 (2.52)	353	1.97 (4.02)	6.14E-14
Worries during lockdown					
Worried about getting COVID-19 score*	3206	3.32 (1.16)	376	3.68 (1.21)	3.29E-08
Worried about losing their job score*	2100	1.95 (1.24)	266	2.09 (1.35)	0.09

† Pearson Chi-square Test, ‡ Z-test from logistic regression

* Scale is 1 (not-at-all) to 5 (very)

When comparing G1 with G0 participants with asthma during COVID-19, the older G0 reported more confirmed or suspected COVID-19 (G1 vs G0 Z-test $p = 0.005$), more shortness of breath ($p = 0.03$), more self-isolation ($p = 3.0\times10^{-5}$), and longer self-isolation ($p = 0.003$), compared to participants with asthma in the younger ALSPAC-G1 population. (**Figure 28, Table 26**).

Table 26. Comparison of asthma-effects in ALSPAC-G0 vs. ALSPAC-G1

	G0-Parents		G1-Children		Coef_G1-Coef_G0	SE_G1-SE_G0		
Count Outcomes	**Coef1**	**SE1**	**Coef1**	**SE1**	**Coef_diff**	**SE_diff**	**Z=Coeff_diff/SE_diff**	**Z test P-value**
SMFQ-13 Depression score (0-26)	0.14	0.07	0.12	0.04	-0.02	0.08	-0.25	0.80
GAD-7 Anxiety score (0-21)	0.15	0.07	0.13	0.05	-0.02	0.09	-0.23	0.82
Length of self-isolation in weeks (0-13)	0.91	0.15	0.24	0.17	-0.67	0.23	-2.96	0.003
Worried about getting COVID-19 score (1-5)	0.08	0.02	0.12	0.02	0.04	0.03	1.41	0.16
Worried about losing their job score (1-5)	0.02	0.05	0.10	0.04	0.08	0.06	1.25	0.21
Binary Outcomes	**Coef2**	**SE2**	**Coef2**	**SE2**	**Coef_diff**	**SE_diff**	**Z=Coeff_diff/SE_diff**	**Z test P-value**
Confirmed or suspected COVID19	0.43	0.17	-0.27	0.18	-0.70	0.25	-2.83	0.005
Shortness of breath or difficulty in breathing	1.68	0.16	0.88	0.32	-0.80	0.36	-2.24	0.025
Difficulty sleeping	0.3	0.15	0.21	0.15	-0.09	0.21	-0.42	0.67
Participant has or is self-isolating	0.86	0.14	0.03	0.14	-0.83	0.20	-4.19	3.00E-05
Eating change since lockdown	0.07	0.13	0.11	0.13	0.04	0.18	0.22	0.83

Coef1, SE1 are the logarithmic form of the SRC and corresponding standard error; Coef2, SE2 are the logarithmic form of the OR and corresponding standard error

8.7 Discussion

Surveys suggest that the mental health and wellbeing of populations across the world have been affected by the COVID-19 pandemic.[221,222] It was hypothesised that asthma was associated with a greater mental health impact during COVID-19 lockdown, and that this difference could be related to the psychological effects of the outbreak rather than the physical symptoms of asthma. This longitudinal study provides evidence that levels of anxiety, mood, and wellbeing in the whole ALSPAC population worsened during the pandemic. However, having asthma was associated with a decline to lower levels of mental health, with the younger asthmatics reaching the lowest levels. The data highlights that a range of psychological factors might be mediating this difference and that these differ by age.

During the first lockdown, those with asthma in the younger ALSPAC-G1 population had higher levels of depression & anxiety and lower well-being scores than those without asthma. These differences were also evident pre-lockdown for well-being and depression but not anxiety. Worse scores for anxiety and depression in the older ALSPAC-G0 participants with asthma, compared to participants without asthma were identified; these differences were not present pre-outbreak with the available scores (different scores to the ones used at the COVID-19 survey). The change in anxiety and depression scores of asthmatics vs non-asthmatics between the two generations was similar.

The mean WEMWBS-14 score for the younger ALSPAC-G1 participants with asthma dropped from 47.5 to 42.2 across the COVID-19 outbreak, whereas the participants without asthma dropped from 49.0 to 44.4. This decrease in well-being is clinically meaningful for both groups, with mean scores declining to below the cut-off for identifying patients at risk of depression in those with asthma.[235] The minimal clinically important difference for the GAD-7 anxiety score has been estimated as 4 [236] and a score of 10+ has a high sensitivity and specificity for the diagnosis of generalised anxiety disorder.[237] The mean magnitude of increase (post minus pre-COVID-19 score) in the GAD-7 score was 3.4 (mean absolute score during lockdown 8.27) for those with asthma and 2.4 (6.9) for those

without asthma in ALSPAC-G1. Therefore, although this does not reach a meaningful clinical change, with repeated lockdowns it may.

The older ALSPAC-G0 participants with asthma reported worse breathlessness and had more confirmed or suspected COVID-19 than the younger ALSPAC-G1 participants with asthma when compared to their peers without asthma. There were no pre-lockdown measurements of breathlessness available, so it is not possible to say whether the high levels of breathlessness in participants with asthma are pre-existing. If symptoms of respiratory disease were the cause of the anxiety, then it would be expected the older ALSPAC-G0 participants with asthma to be the most anxious group. Whilst there was no direct record of asthma control in the data, a British Lung Foundation survey has estimated that 24.6% of people with asthma triggered by pollution reported an improvement in their symptoms since lockdown.[238] Furthermore, control of asthma in paediatric patients has generally improved since the outbreak.[239] This reduction in asthma symptoms may relate to an interruption of transmission of other viruses that cause exacerbations, decline in pollution levels, or improved medication adherence. This supports the findings and suggests that patients with asthma are not more anxious because of their physical symptoms.

Other factors that differed in the population with asthma were identified which might provide more detailed insight into the origins of the increased anxiety. Participants with asthma in the younger ALSPAC-G1 group reported greater concerns about catching COVID-19 and worry of losing their jobs when compared to those without asthma. Participants with asthma in the older ALSPAC-G0 group reported increased frequency and length of self-isolation, increased loneliness, and worry of catching COVID-19. On the 16th of March 2020 British Prime Minister Boris Johnson announced that vulnerable groups of people will need to shield from social contact for 12 weeks.[240] Such groups at increased risk of severe illness from COVID-19 included those with mild-moderate asthma.[241] This guidance might result in participants with asthma feeling more isolated and having greater concerns for their physical and financial health than those without asthma. The implementation of social

isolation measures and the way that they are communicated to the public may have significantly affected the mental health of patients with asthma more so than those without asthma.

8.7.1 Strengths & Limitations

The strength of this study lies in the detailed data available from the large ALSPAC population. The longitudinal data has provided not just a snapshot of the state of mind of the population during the crisis but has allowed an assessment of its development from a pre-pandemic baseline state. The detailed questionnaires provided details about specific concerns during the first lockdown.

There are limitations to the data. Asthma diagnoses were self-reported, so may over or under-represent the true asthma population within the group. Whilst this might confound the effect of the physical symptoms of asthma on mental health during the pandemic, it will not affect the psychological outcome of those carrying the label of an asthma diagnosis that is central to the hypothesis. Furthermore, there is evidence of the validity of self-reported asthma in ALSPAC.[242]

The ALSPAC-G0 pre-pandemic assessments of anxiety and depression were made using different assessment tools to those used in the COVID-19 assessment, but it was still possible to assess the mental state of this population over time.

It is difficult to say whether the increase in anxiety and decrease in mental wellbeing in the participants with asthma translates to a significant pathological effect. There is no data regarding healthcare utilisation of the cohorts pre- and post-lockdown. It is uncertain whether people have sought help for their asthma or mental health issues, or whether there has been reduced contact with mental health services for those patients already in the system.

8.7.2 Future Research

This study highlights higher levels of anxiety and persistently worse markers of depression and mental wellbeing in participants with asthma during the COVID-19 pandemic. This is married with an

increased frequency and length of self-isolation and more prominent concerns for physical and financial health. People with asthma were thought to be at higher risk of severe COVID-19 infections and it is vital to protect the physical health of this population. Yet there was mixed evidence regarding this risk,[243] and the psychological health of this population must not be forgotten. It is also important to further clarify the origins of mental health issues in people with asthma. This information should remind the healthcare profession to screen people with asthma for symptoms of anxiety and depression. It will also help to inform government policies which whilst intended to protect the population, are not without negative consequences.

8.7.3 Conclusion

In conclusion, this study highlights the increased prevalence of anxiety and reduced mental wellbeing in a population with asthma following the start of the COVID-19 pandemic. Clinicians should be made aware of this, and further investigation is required to help inform national policies to try to prevent it.

CHAPTER 9. SUMMARY OF KEY FINDINGS, MY LEARNING, FUTURE RESEARCH, AND CONCLUSIONS

9.1 Introduction

Reduced lung function and lung diseases are common and are frequently found in combination with a wide range of extra-pulmonary diseases.[3,6,20,191,244,245] This co-morbidity is increasing, particularly with an ageing population.[188,246-249] Traditional observational epidemiology has been used to describe co-morbidity. However, some conditions such as PRISm, have been neglected. Additionally, the impact of widely studied diseases may change depending on circumstances, such as asthma and mental health during social isolation. Given the limitations of this traditional observational epidemiology, genetic epidemiology can be used to provide an understanding of the causality of these associations and provide insight into the underlying causes of lung disease.[7,79,96] If the associations are causal, then research to discover underlying mechanisms should be performed with the potential to treat lung function as a modifiable risk factor for extra-pulmonary disease.

In this final chapter, I will summarise the findings of my papers, suggest future areas for research, and describe what I learned in the process. I will then write about the limitations of genetic epidemiology and discuss how different epidemiological methods can be used together to provide robust evidence.

9.1 Examining the possible relationship between Lung function, COPD and Alzheimer's' Disease: A Mendelian Randomisation Study

9.1.1 Summary

Observational studies have described an association between reduced lung function, COPD, and an increased risk of Alzheimer's disease. If causal, then measures to improve lung function and decrease COPD could reduce the risk of Alzheimer's Disease. This was the first 2SMR study to examine for a causal effect of lung function and COPD on the risk of Alzheimer's disease. Little

evidence for a causal effect was found. Observational studies describing an association are likely to be affected by residual confounding.

9.1.2 Future Research

Measures trying to establish an underlying mechanism between lung function/disease and Alzheimer's may be best avoided as is if there is no causal pathway then it could waste research efforts. However, lung function and lung disease are more strongly associated with a different type of dementia, known as vascular dementia. Residual confounding is a particular issue when examining lung function/COPD and vascular dementia as smoking is a strong risk factor for both, and both are more common in those with lower educational attainment.[250,251] Unfortunately at present it is not possible to undertake an MR study into this as there has only been one GWAS of vascular dementia which had only 67 cases.[252] If the UK Biobank continues to link diagnosis from hospital/community records to their participants, then in the future it may be possible to perform a one sample MR study examining this association could be performed.

9.1.3 My learning

This was the first study of any kind I performed during my PhD. It required me to become familiar with coding programmes, enabling me to import, store and analyse huge data sets. I came to understand the methodology of MR and how to perform MR studies. The challenges of trying to communicate complex relatively novel genetic epidemiology techniques to the readers of a clinical respiratory journal when trying to publish the paper became clear. Additionally, despite virtually all authorities agreeing that using terms such as "significant" based on p-value threshold of 0.05 is incorrect, it is still a common practice in clinical journals.[253,254] The use of more accurate terms in my paper such as "weak evidence" or "minimal evidence" was criticised and led to mis-interpretation of the results by peer reviewers. Despite increasing the challenge of publishing papers, I shall not regress to conclusions based on binary interpretation of an arbitrary p-value threshold for the purpose of easier publication.

9.2 Lung function and cardiovascular disease: A two-sample Mendelian Randomisation study

9.2.1 Summary

Lung function and COPD have been strongly associated with cardiovascular disease including coronary artery disease and strokes.[10,12,21,33-36,38,46,48,49,126,127,153,255-260] I used 2SMR and an MVMR model to test if lung function or COPD had a causal effect on the risk of coronary artery disease and/or ischaemic stroke. This found that reduced FVC, but not FEV_1, increased the risk of coronary artery disease.

9.2.2 My Learning

I initially conducted this study using a GWAS adjusted for heritable covariates of lung function and cardiovascular disease in a two-sample model. The lack of evidence of a causal effect seemed surprising given the wealth of observational data. I was then introduced to the concept of collider bias and its role in biasing GWAS and MR studies. Although challenging concepts to understand at first, I have gained essential knowledge into these potential pitfalls. I then learned about MVMR and how this bias could be avoided by using unadjusted GWAS and conditioning for heritable covariates. This led to me applying for access to UK Biobank data and performing a GWAS of FEV_1 and FVC adjusted only for sex. Using the results, I learned how to perform MVMR studies. Publishing this paper took many attempts and re-writes but it was valuable to learn how to communicate with a clinical audience about collider bias and MVMR.

9.2.3 Future research

The finding that there is evidence that FVC has a causal effect on the risk of coronary artery disease has multiple clinical implications. It has been suggested previously that FVC can add weight to the Framingham risk score prediction of cardiovascular disease and all-cause mortality.[32] This could be investigated further with other risk prediction models to guide preventative treatment.

FVC could be a modifiable risk factor for cardiovascular disease. Treatments to improve FVC may decrease the risk of heart attacks, a leading cause of morbidity and mortality. Further work to identify the underlying mechanism could reveal potential drug targets. MR could be used to achieve this. Using an MVMR model and conditioning on SNPs for potential mediators, such as inflammatory cytokines, could confirm suitable treatment targets.

Cohorts are trying to understand lung function trajectories to predict COPD and extra-pulmonary disease but tend to focus on FEV_1.[71,72,261] FVC should be prioritised given there is both traditional observational and MR evidence of a causal effect on coronary artery disease, a major cause of morbidity and mortality. Public health interventions to achieve the highest FVC possible and to reduce decline may prevent coronary artery disease.

9.3 Lung function, COPD and cognitive function: A multivariable and two sample Mendelian Randomisation Study

9.3.1 Summary

Lung function and COPD have both been associated with cognitive function.[27,53,147,148,262-264] I performed a two-sample MR studies using lung function and COPD SNPs as the exposures. This showed no strong evidence of a causal effect of lung function/COPD on cognitive function. I then performed an MVMR study which showed that although decreasing FEV_1 and FVC were both strongly associated with a decrease in cognitive function, the evidence became weak after adjusting for either height or educational attainment. Therefore, observational research showing an association between lung function and COPD with cognitive function may be affected by residual confounding and other sources of bias.

9.3.2 Future research

The outcome used for this study was general cognitive function in a normal population. Although lung function and COPD have been associated with general cognitive function, more research has focused on their association with mild cognitive impairment (MCI). Although not featured in this thesis, during this PhD I wrote an editorial discussing COPD and co-morbid MCI.[156] Unfortunately there are no published GWAS of MCI. This may be partly due to logistical challenges. People with MCI may not have capacity to consent to take part in a GWAS, and as it is a transient state it makes recruitment more challenging. Were a GWAS of MCI to become available, performing another MR study using the same exposure populations would provide valuable insight into this damaging extra-pulmonary morbidity.

9.3.3 My learning

Having attended an MR conference, I learned about techniques that iteratively remove outliers using models such as MR PRESSO and MR radial.[146,265] I had been relying on identifying outliers visually

from scatter plots, single SNP analysis, and leave-one-out-analysis. In this paper I used MR radial to exclude outliers. Although it did not change the interpretation of the results, I am pleased to utilise these techniques.

Cognitive function is a complex trait that can be affected by demography, assortative mating, and dynastic effects. These issues can lead to bias in GWAS effect estimates which can then cause bias in MR studies. I only became familiar with these concepts when writing up this study towards the end of my PhD, however, they can be limitations of genetic epidemiology and MR more generally. I discuss them in more detail later in this chapter.

9.4 Preserved Ratio Impaired Spirometry (PRISm): A UKBiobank cohort study

9.4.1 Summary

Prior to my study, it was thought that ~50% of those with PRISm would transition to COPD.[57,60] However scientific knowledge was derived from small cohorts affected by selection bias. Using the UK Biobank allowed me access to a large well-phenotyped cohort, with a long follow-up time, that was less affected by selection bias.

This showed that PRISm was common at 11% with obesity, current smoking, and asthma diagnosis being strongly associated risk factors. Those with PRISm have an increased risk of breathlessness, with high rates of diabetes and cardiovascular co-morbidity. On follow-up, only 12% developed COPD, with the remainder approximately evenly split into those with persistent PRISm and those that changed to normal spirometry (once regression to the mean is considered). Despite PRISm being a relatively transient state, follow-up rates of all-cause mortality are higher for PRISm compared to control, over an 18-year period and after adjustment for sex, smoking, BMI, and age.

9.4.2 Future research

The underlying pathophysiology of PRISm is not clear. Given it is defined by an artificial construct of lung function it will contain numerous subtypes rather than any single unifying pathology. Radiological studies of PRISm have looked for small airway changes and other early COPD signs in those with PRISm. However, much of this work has been performed in COPDGene, and so the results may not be generalizable to non-smokers.[158,166] A non-COPDGene imaging study comparing those with chronic bronchitis, chronic bronchitis with PRISm, and GOLD I and II has been performed in China.[159] However, 82% of those with PRISm were current smokers with a mean pack/year of 32. This will limit generalisability, particularly when examining radiological changes that are related to progression to COPD. Studies examining radiographic changes, such as airway wall thickness, lung density, lung volume, and emphysema in more generalizable cohorts are required.

It may be that small airway obstruction is contributing to PRISm. Tests for small airway obstruction such as impulse oscillometry testing or FEF_{25-75} (Forced Expiratory Flow over the middle half on an FVC) could provide answers.[266] Most PRISm research has been carried out in older populations. Using birth cohorts with longitudinal data to see study prevalence, co-morbidities of PRISm may provide a much better understanding of the pathogenesis PRISm.

Obesity is a significant risk factor for PRISm. It is probable that one sub-type of PRISm is purely caused by extra-thoracic restriction due to obesity. Researching those with PRISm without obesity using imaging and in-depth lung function tests will uncover a subtype with an unknown pathological mechanism that may have treatable traits.

Many cohorts and consortiums are trying to understand FEV_1 trajectory and COPD prediction. This thesis has demonstrated the association of PRISm with morbidity and mortality, which aligns with observational evidence that FVC rather than FEV_1 of FEV_1/FVC is a predictor of all-cause mortality,[10,31]. Additionally, this thesis has used Mendelian Randomisation to show evidence for a causal effect of FVC, but not FEV_1, on the risk of coronary artery disease. This points towards the need for better understanding of FVC and PRISm trajectories.

9.4.3 My learning

Prior to my PhD, I was unaware of the term PRISm but was familiar with the concept of restrictive spirometry. I performed an informal literature review to increase my insight into the current scientific knowledge of PRISm and identify the fundamental issues. I completed a first draft of a paper about PRISm in UK Biobank cross-sectionally with a GWAS. When submitting the first draft the editor of a journal informed that I had used a UK Biobank FEV_1 variable that was only measured in healthy never smokers, and smokers with symptoms. Therefore, the paper was flawed. Although this was frustrating and caused a huge time delay, I became aware of the pitfalls in analysing large databases collated by external groups. It was only after the re-drafting that I realised UK Biobank had enough data to perform longitudinal and survival analysis, which lead to a much stronger paper.

Although I had heard of regression to the mean, exploring ways to account for it has increased my knowledge of this common phenomenon and how to manage it. The Lancet Respiratory had five peer reviewers give feedback on the paper including a thorough statistical review. Through this I became aware of numerous issues with statistical models and how to test and correct them. This included linear assumption of logistic regression models, censoring of survival analysis, using medians and IQR for non-normally distributed data. This knowledge will help me to both produce and peer review epidemiological papers in the future.

9.5 Genome-Wide Association Study of Preserved Ratio Impaired Spirometry

9.5.1 Summary

Prior to my study, there was only one published attempted GWAS of PRISm, which had not discovered any SNPs.[59] I conducted a GWAS of PRISm hoping to uncover information about the heritability of PRISm, underlying mechanisms, and discover SNPs that could be used for MR studies. However, few SNPs were successfully replicated, and none are likely to be specific to PRISm or represent novel loci for lung function. Additionally, LD-correlation study showed an extremely strong correlation between the PRISm GWAS, and GWAS for continuous lung function traits.

PheWAS showed that there may be shared genetic risks for PRISm and its co-morbidities. The LD-correlation study provides further evidence for this with a moderate correlation between the GWAs for PRISm and a GWAS of type II diabetes.

9.5.2 Future research

Future genetic studies of PRISm are unlikely to discover loci that are specific to PRISm or novel for lung function. GWAS of well-phenotyped sub-types of PRISm could still be of value, but genotyped populations of these subtypes large enough to power discovery and replication studies may not exist. Better use of resources would be to perform GWAS of continuous lung function traits in larger populations. This can be achieved by incorporating more consortiums into meta-analysis. Increased understanding of the genetic contribution to lung trajectories could allow for novel treatment targets and prevention of PRISm and COPD.

9.5.3 My learning

This project has taught me the importance of replication analysis for GWAS, and how to conduct them. I was fortunate to have the support of Professor Tobin and the SpiroMeta and CHARGE consortiums which enabled the replication analysis. Liaising with multiple global cohorts that used different methodologies and reporting terms made it initially challenging to perform the meta-

analysis. The limited findings reduced the impact and further research plans but was valuable for me to learn about the limitations of GWAS. The knowledge I gained from this project has meant I can contribute to discussions between 23andme and the UK Severe Asthma Registry regarding a GWAS of the severe asthma population.

9.6 The impact of asthma on mental health and wellbeing during COVID-19 lockdown

9.6.1 Summary

This study was designed to determine if asthma was associated with worse mental health outcomes during the COVID-19 lockdown.

During the first lockdown, those with asthma in the younger ALSPAC-G1 population had higher levels of depression & anxiety and lower well-being scores than those without asthma. Worse scores for anxiety and depression were found in the older ALSPAC-G0 participants with asthma, compared to participants without asthma. Participants with asthma in the younger ALSPAC-G1 group reported greater concerns about catching COVID-19 and worry of losing their jobs when compared to those without asthma. Participants with asthma in the older ALSPAC-G0 group reported increased frequency and length of self-isolation, increased loneliness, and worry of catching COVID-19.

At the time of the first lockdown, there was intense speculation about the negative impacts of lockdown. This was the first study to model asthma as the exposure and have mental health measures from before and during the lockdown. Having this data across two generations with an array of physical symptoms and psychosocial concerns were additional factors that made this study unique. I was extremely pleased that the study was cited by www.gov.uk in their communication about weighing the risk-benefit of lockdowns demonstrating the studies clinical utility and impact.

9.6.2 Future research

Since this study was written evidence has shown that those with asthma are probably not at increased risk of severe illness or death due to COVID-19.[267,268] Many treatments for asthma such as steroid inhalers and tablets are now used routinely to treat COVID-19 and effective vaccines are widely available.[218,269] Although another lockdown was instigated since this study, it seems less likely another will be required for COVID-19, and those with asthma may now be less concerned about contracting the disease. However, pandemics with as-yet-unknown pathogens affecting the respiratory system are a possibility, which could lead to social restrictions.

Studies have shown shared genetic risk factors for asthma with anxiety and major depressive disorders.[270] A MR study has shown some evidence of causal effect of asthma on the risk of depression and anxiety.[271] Therefore, the association between asthma and mental health may be causal. Larger GWAS of more phenotyped asthma populations could give further insight into shared heritability and underlying pathways between asthma and mental health. 2SMR studies using MVMR to determine the direct effect of asthma on mental health outcomes are warranted.

However, even if future evidence shows association is not found to be causal, supporting the mental health of those with asthma, particularly during social isolation, should be considered by health policymakers and clinicians.

9.6.3 My learning

This study showed me the value of “from birth” cohorts with extensive phenotyping. Few cohorts in the world would have been able to produce the breadth of data ALSPAC made, nor have the data available at such speed. Due to the multiple comparisons performed e.g. before/during lockdown, between those with/without asthma, and comparing this between the generations, a wide range of statistical tests was required. This included tests that were previously unknown to me, such as

exponentiated Poisson regression and Two-sample Kolmogorov-Smirnov tests. Communicating this to an audience was a challenging and valuable experience.

9.7 Traditional observational epidemiology and the benefits of MR

Traditional observational epidemiology is limited by sources of bias such as residual confounding and reverse causation. It cannot determine causality, just describe associations. It is unfeasible to randomise participants of a trial to exposures such as COPD or reduced lung function. MR does not suffer from the same limitations as observational epidemiology. Genetic variants are randomly assigned at conception and can be exploited as proxies for traits to provide evidence for causality. However, as I learned whilst producing my papers, there are limitations of genetic epidemiology which can lead to potential sources of bias.

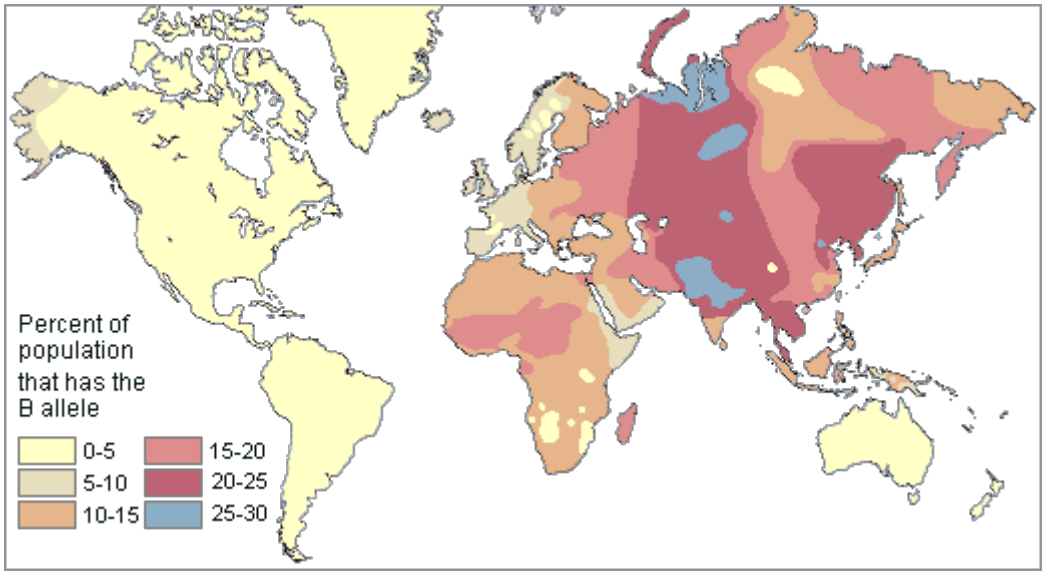
9.8 Limitations of genetic epidemiology and MR

Research attention has been focused on dealing with problems of horizontal pleiotropy as discussed in chapter 4, and implemented in chapters 6, 7, and 8. However, it is important to consider factors that could bias the instrument and the outcome.

9.8.1 Demography, assortative mating, and dynastic effects

Demography, the composition of the human population, is influenced by genetic variation which occurs because of geographical location. This may be due to geological, political, or cultural divides preventing mixing of populations.

Figure 30. Showing the global frequency of allele B for blood type [272]



World map showing the natural variation in frequency of a common allele causing B blood type across different geographical areas

Complex traits can be strongly affected by demography as demonstrated in **Figure 29**. If this variation is not accounted for it can lead to population stratification in a GWAS and bias the results. If the effect estimate is biased, this bias is carried through into MR studies and will bias estimations of causal effects. Assortative mating describes the process whereby reproductive partners are selected due to certain traits e.g. cognitive function or educational attainment. Therefore, although SNPs may be assigned randomly at conception from biological parents, the process of parents mating is not random, leading to changes in allele frequencies in populations. If this is not accounted for when using different populations it could lead to bias and reduces generalizability of studies outside the populations examined.

Dynastic effects describe the indirect effects of the parents' genotype on their offspring, mediated by the parent's phenotype.[273] This is a particular issue in traits such as educational attainment and age having first child, but also affects height, BMI, and smoking. For example, parents that provide a

strong learning environment for their children are likely to have children with higher educational attainment. A GWAS performed in the children could falsely identify SNPs associated with educational attainment, for exaggerate the effect size of these SNPs. SNPs that increase parents smoking may be falsely identified in their offspring as SNPs that causes wheezing, even if the offspring did not smoke but had wheeze induced by a smoky environment.

9.8.2 Within family MR

Issues with population stratification, demography, assortative mating, and dynastic effects can all lead to biased estimates for SNP-phenotype association of commonly studied traits. How this affects MR studies will depend on the research question but could bias causal estimates. To surmount this problem within-family GWAS and MR could be used. One model is using studies performed within siblings. When comparing siblings, the genetic difference between them reflects true random allocation of alleles and should remove most potential environmental confounding.[155] Another model is to use MVMR and condition the effect of the offspring's SNPs with the parents' SNPs. The effects of assortative mating can tested for by using spousal pairs, where the trait of one spouse is the exposure, and the other spouses' trait is the outcome.[273]

These study models are becoming increasingly possible as large data banks containing offspring, spouses, and siblings become increasingly available and should be utilised.

9.8.3 Limited genetic data for different ancestral populations

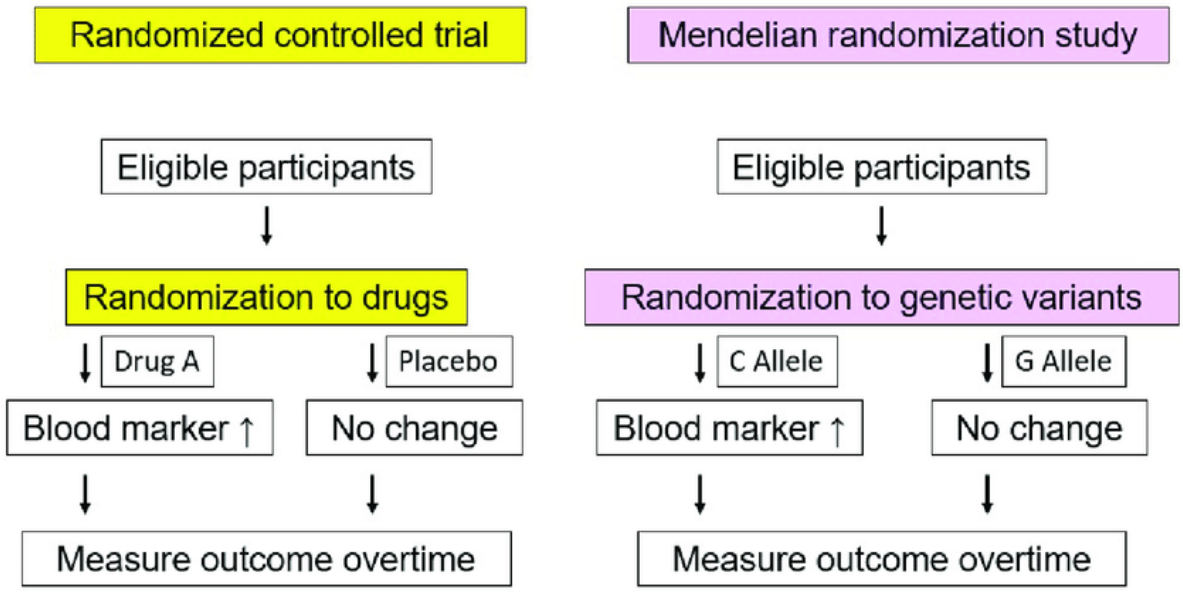
In 2009 over 96% of participants in published GWAS were of European ancestral populations.[274] By 2016 this decreased to 80% but was almost entirely driven by increasing genotyping of Asians. The proportion of GWAS including those of African and Hispanic descent has barely changed. Genetic variants have different effects in different ancestral populations. For example, 25% of variants discovered in European populations that affect BMI are found to have different strengths of association in >20% of non-European ancestral groups.[275] Therefore, genetic variants discovered in

European populations cannot necessarily be exploited as IVs in other ancestral populations and results of MR studies may not be generalizable to non-European populations. Clearly, there is a need to increase participation in GWAS from non-European ancestral groups.

9.8.4 Mendelian Randomisation and RCTs

Mendel's second law, the law of random assortment, states that the inheritance of one trait is independent of the inheritance of another.[276] Therefore, traits caused by genetic variation are randomly assigned in the population and are not related to confounders.[87] This is exploited in MR, and it was thought that this allowed MR studies to be analogous to randomised trials as shown in **Figure 30**.

Figure 31. Schematic demonstrating how it was thought that MR and RCT's were analogous[277]



Schematic shows how a MR can be thought of as potentially analogous to an RCT, using the example of a drug/allele that cause a change in a blood maker

However, there are fundamental differences between these study designs.

- In MR the IV is used both as a randomisation tool and as the exposure, whereas in a RCT randomisation determines the groups that receive the different exposures.
- Due to demography and assortative mating the alleles may not be randomly distributed in the population compromising the randomization process.
- The effect of the allele may differ between individuals due to dynastic effects.
- MR studies are designed to assess whether an exposure-outcome relationship is causal. A RCT doesn't just test causality, but it is also designed to evaluate if the effect of the exposure on the outcome causes a clinically significant change. Generally, most SNPs have a small effect on the exposure and this effect is present from conception. Whereas a RCT will tend to investigate a large exposure effect, over a relatively short period of time, starting considerably later in participants' lives which may well relate better to clinical interventions.

These differences can explain discrepancies between my finding that FVC causally affects the risk of coronary artery disease, and the lack of evidence from trials of inhalers and other therapeutics reduce mortality. It may be that interventions instigated later in life (rather than from birth) may not have an impact. The MR model estimated the increased odds of coronary artery disease based on a standard deviation of FVC, which in the discovery sample equated to nearly a litre, an amount far beyond any benefit delivered by current therapies.

Despite these differences, MR studies can be used in conjunction with RCTs to great effect. RCTs are often expensive and time-consuming, so determining the causality of an association if a robust MR study is possible before undertaking an RCT could save resources. MR has been used to identify side effects e.g. MR studies showed that lipid-lowering drugs could increase the risk of diabetes, and this was then found in some RCTs of the drugs.[278] By estimating the amount an exposure needs to change to affect an outcome, MR can be used to guide the minimum clinically significant difference and the inclusion/exclusion criteria of a RCT.[279]

9.9 Final summary

In this thesis, I have used traditional epidemiology to describe extra-pulmonary associations of asthma and PRISm. I have used genetic epidemiology to show that not all reported extra-pulmonary associations of lung function/disease are likely to be causal, but certain lung function traits do causally affect the risk of certain extra-pulmonary disease. The causal relationships warrant further investigations as to their underlying mechanisms. Lung function could be used as a screening tool or modifiable risk factor for the extra-pulmonary diseases it causally affects.

Determining the causality of extra-pulmonary associations of lung function and lung disease is essential. Traditional observational epidemiology, genetic epidemiology, and randomised controlled trials all suffer from different sources of bias and weaknesses. Genetic epidemiology has undergone huge advances in recent years, both due to the wealth of genetic data now available, and evolving understanding of bias and novel methods to counteract them. Traditional and genetic epidemiology should be used in combination to enhance the strength of evidence in a process of triangulation to gain more reliable insights into the extra-pulmonary effects of lung function and lung disease.

REFERENCES

1. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020; **396**(10258): 1204-22.
2. Li X, Cao X, Guo M, Xie M, Liu X. Trends and risk factors of mortality and disability adjusted life years for chronic respiratory diseases from 1990 to 2017: systematic analysis for the Global Burden of Disease Study 2017. 2020; **368**: m234.
3. Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. *The European respiratory journal* 2009; **33**(5): 1165-85.
4. Brilleman SL, Purdy S, Salisbury C, Windmeijer F, Gravelle H, Hollinghurst S. Implications of comorbidity for primary care costs in the UK: a retrospective observational study. *Br J Gen Pract* 2013; **63**(609): e274-e82.
5. Morrison D, Agur K, Mercer S, Eiras A, González-Montalvo JI, Gruffydd-Jones K. Managing multimorbidity in primary care in patients with chronic respiratory conditions. *NPJ Prim Care Respir Med* 2016; **26**: 16043.
6. Sin DD, Anthonisen NR, Soriano JB, Agusti AG. Mortality in COPD: Role of comorbidities. *The European respiratory journal* 2006; **28**(6): 1245-57.
7. Burgess S, Foley CN, Zuber V. Inferring Causal Relationships Between Risk Factors and Outcomes from Genome-Wide Association Study Data. *Annual review of genomics and human genetics* 2018; **19**: 303-27.
8. NICE. Nintedanib fro treating progressive fibrosing interstitial lung diseases. 2021.
9. Han MK, Agusti A, Celli BR, et al. From GOLD 0 to Pre-COPD. *American journal of respiratory and critical care medicine* 2021; **203**(4): 414-23.
10. Magnussen C, Ojeda FM, Rzayeva N, et al. FEV1 and FVC predict all-cause mortality independent of cardiac function - Results from the population-based Gutenberg Health Study. *International journal of cardiology* 2017; **234**: 64-8.
11. Honda Y, Watanabe T, Shibata Y, et al. Impact of restrictive lung disorder on cardiovascular mortality in a general population: The Yamagata (Takahata) study. *International journal of cardiology* 2017; **241**: 395-400.
12. Sin DD, Wu L, Man SF. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest* 2005; **127**(6): 1952-9.
13. Liou TG, Kanner RE. Spirometry. *Clinical Reviews in Allergy & Immunology* 2009; **37**(3): 137-52.
14. Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *The European respiratory journal* 2012; **40**(6): 1324-43.
15. Culver BH. How should the lower limit of the normal range be defined? *Respiratory care* 2012; **57**(1): 136-45; discussion 43-5.
16. Lambrecht BN, Hammad H. The airway epithelium in asthma. *Nature medicine* 2012; **18**(5): 684-92.
17. Decramer M, Janssens W, Miravitlles M. Chronic obstructive pulmonary disease. *Lancet (London, England)* 2012; **379**(9823): 1341-51.

18. Global Iniative for Chronic Obstructive Pulmonary Lung Disease, 2019.
19. Agustí A, Hogg JC. Update on the Pathogenesis of Chronic Obstructive Pulmonary Disease. *The New England journal of medicine* 2019; **381**(13): 1248-56.
20. Cavaillès A, Brinchault-Rabin G, Dixmier A, et al. Comorbidities of COPD. *Eur Respir Rev* 2013; **22**(130): 454-75.
21. Feary JR, Rodrigues LC, Smith CJ, Hubbard RB, Gibson JE. Prevalence of major comorbidities in subjects with COPD and incidence of myocardial infarction and stroke: a comprehensive analysis using data from primary care. *Thorax* 2010; **65**(11): 956-62.
22. Laforest L, Roche N, Devouassoux G, et al. Frequency of comorbidities in chronic obstructive pulmonary disease, and impact on all-cause mortality: A population-based cohort study. *Respiratory medicine* 2016; **117**: 33-9.
23. Dementia: A public health priority. The World Health Organization.
24. Behrman S, Valkanova V, Allan CL. Diagnosing and managing mild cognitive impairment. *The Practitioner* 2017; **261**(1804): 17-20.
25. Wimo A, Guerchet M, Ali GC, et al. The worldwide costs of dementia 2015 and comparisons with 2010. *Alzheimer's & dementia : the journal of the Alzheimer's Association* 2017; **13**(1): 1-7.
26. Silva MVF, Loures CdMG, Alves LCV, de Souza LC, Borges KBG, Carvalho MdG. Alzheimer's disease: risk factors and potentially protective measures. *J Biomed Sci* 2019; **26**(1): 33-.
27. Incalzi RA, Gemma A, Marra C, Muzzolon R, Capparella O, Carbonin P. Chronic obstructive pulmonary disease. An original model of cognitive decline. *The American review of respiratory disease* 1993; **148**(2): 418-24.
28. Liao KM, Ho CH, Ko SC, Li CY. Increased Risk of Dementia in Patients With Chronic Obstructive Pulmonary Disease. *Medicine* 2015; **94**(23): e930.
29. Lutsey PL, Chen N, Mirabelli MC, et al. Impaired Lung Function, Lung Disease and Risk of Incident Dementia. *American journal of respiratory and critical care medicine* 2018.
30. Agusti A, Noell G, Brugada J, Faner R. Lung function in early adulthood and health in later life: a transgenerational cohort analysis. *The Lancet Respiratory medicine* 2017; **5**(12): 935-45.
31. Burney PGJ, Hooper R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax* 2011; **66**(1): 49.
32. Lee HM, Le H, Lee BT, Lopez VA, Wong ND. Forced vital capacity paired with Framingham Risk Score for prediction of all-cause mortality. *European Respiratory Journal* 2010; **36**(5): 1002.
33. Schroeder EB, Welch VL, Couper D, et al. Lung Function and Incident Coronary Heart Disease: The Atherosclerosis Risk in Communities Study. *American Journal of Epidemiology* 2003; **158**(12): 1171-81.
34. Kim JJ, Kim DB, Jang SW, et al. Relationship between airflow obstruction and coronary atherosclerosis in asymptomatic individuals: evaluation by coronary CT angiography. *The international journal of cardiovascular imaging* 2018; **34**(4): 641-8.
35. Cuttica MJ, Colangelo LA, Dransfield MT, et al. Lung Function in Young Adults and Risk of Cardiovascular Events Over 29 Years: The CARDIA Study. *Journal of the American Heart Association* 2018; **7**(24): e010672.
36. Engstrom G, Hedblad B, Valind S, Janzon L. Increased incidence of myocardial infarction and stroke in hypertensive men with reduced lung function. *Journal of hypertension* 2001; **19**(2): 295-301.

37. Hozawa A, Billings JL, Shahar E, Ohira T, Rosamond WD, Folsom AR. Lung function and ischemic stroke incidence: the Atherosclerosis Risk in Communities study. *Chest* 2006; **130**(6): 1642-9.
38. Corlateanu A, Covantev S, Mathioudakis AG, Botnaru V, Cazzola M, Siafakas N. Chronic Obstructive Pulmonary Disease and Stroke. *Copd* 2018; **15**(4): 405-13.
39. Davies G, Lam M, Harris SE, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature communications* 2018; **9**(1): 2098-.
40. Sanford AM. Mild Cognitive Impairment. *Clinics in geriatric medicine* 2017; **33**(3): 325-37.
41. Jongsiriyanyong S, Limpawattana P. Mild Cognitive Impairment in Clinical Practice: A Review Article. *American journal of Alzheimer's disease and other dementias* 2018; **33**(8): 500-7.
42. Chang SS, Chen S, McAvay GJ, Tinetti ME. Effect of coexisting chronic obstructive pulmonary disease and cognitive impairment on health outcomes in older adults. *Journal of the American Geriatrics Society* 2012; **60**(10): 1839-46.
43. Campos MW, Serebrisky D, Castaldelli-Maia JM. Smoking and Cognition. *Current drug abuse reviews* 2016; **9**(2): 76-9.
44. Corley J, Gow AJ, Starr JM, Deary IJ. Smoking, childhood IQ, and cognitive function in old age. *Journal of psychosomatic research* 2012; **73**(2): 132-8.
45. Dodd JW. Lung disease as a determinant of cognitive decline and dementia. *Alzheimer's research & therapy* 2015; **7**(1): 32.
46. Liao D, Higgins M, Bryan NR, et al. Lower pulmonary function and cerebral subclinical abnormalities detected by MRI: the Atherosclerosis Risk in Communities study. *Chest* 1999; **116**(1): 150-6.
47. Brunner EJ, Welch CA, Shipley MJ, Ahmadi-Abhari S, Singh-Manoux A, Kivimäki M. Midlife Risk Factors for Impaired Physical and Cognitive Functioning at Older Ages: A Cohort Study. *The journals of gerontology Series A, Biological sciences and medical sciences* 2017; **72**(2): 237-42.
48. Austin V, Crack PJ, Bozinovski S, Miller AA, Vlahos R. COPD and stroke: are systemic inflammation and oxidative stress the missing links? *Clinical science (London, England : 1979)* 2016; **130**(13): 1039-50.
49. Miller J, Edwards LD, Agusti A, et al. Comorbidity, systemic inflammation and outcomes in the ECLIPSE cohort. *Respiratory medicine* 2013; **107**(9): 1376-84.
50. Ajala O, Zhang Y, Gupta A, Bon J, Sciurba F, Chandra D. Decreased serum TRAIL is associated with increased mortality in smokers with comorbid emphysema and coronary artery disease. *Respiratory medicine* 2018; **145**: 21-7.
51. Fuschillo S, Martucci M, Donner CF, Balzano G. Airway bacterial colonization: the missing link between COPD and cardiovascular events? *Respiratory medicine* 2012; **106**(7): 915-23.
52. Wen XH, Li Y, Han D, Sun L, Ren PX, Ren D. The relationship between cognitive function and arterial partial pressure O2 in patients with COPD: A meta-analysis. *Medicine* 2018; **97**(4): e9599.
53. Dodd JW, Getov SV, Jones PW. Cognitive function in COPD. *The European respiratory journal* 2010; **35**(4): 913-22.
54. Martinez-Pitre PJ, Sabbula BR, Cascella M. Restrictive Lung Disease. StatPearls. Treasure Island (FL): StatPearls Publishing

55. Liou TG, Kanner RE. Spirometry. *Clin Rev Allergy Immunol* 2009; **37**(3): 137-52.
56. Aaron SD, Dales RE, Cardinal P. How Accurate Is Spirometry at Predicting Restrictive Pulmonary Impairment? *Chest* 1999; **115**(3): 869-73.
57. Wan ES, Fortis S, Regan EA, et al. Longitudinal Phenotypes and Mortality in Preserved Ratio Impaired Spirometry in the COPDGene Study. *American journal of respiratory and critical care medicine* 2018; **198**(11): 1397-405.
58. Mannino DM, McBurnie MA, Tan W, et al. Restricted spirometry in the Burden of Lung Disease Study. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 2012; **16**(10): 1405-11.
59. Wan ES, Castaldi PJ, Cho MH, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. *Respiratory research* 2014; **15**(1): 89-.
60. Wijnant SRA, De Roos E, Kavousi M, et al. Trajectory and mortality of preserved ratio impaired spirometry: the Rotterdam Study. *The European respiratory journal* 2020; **55**(1): 1901217.
61. Schwartz A, Arnold N, Skinner B, et al. Preserved Ratio Impaired Spirometry in a Spirometry Database. *Respiratory care* 2021; **66**(1): 58-65.
62. Heo IR, Kim HC, Kim TH. Health-Related Quality of Life and Related Factors in Persons with Preserved Ratio Impaired Spirometry: Data from the Korea National Health and Nutrition Examination Surve. *Medicina (Kaunas, Lithuania)* 2020; **57**(1).
63. Park HJ, Byun MK, Rhee CK, Kim K, Kim HJ, Yoo KH. Significant predictors of medically diagnosed chronic obstructive pulmonary disease in patients with preserved ratio impaired spirometry: a 3-year cohort study. *Respiratory research* 2018; **19**(1): 185.
64. Jankowich M, Elston B, Liu Q, et al. Restrictive Spirometry Pattern, Cardiac Structure and Function, and Incident Heart Failure in African Americans. The Jackson Heart Study. *Annals of the American Thoracic Society* 2018; **15**(10): 1186-96.
65. Barnes PJ, Vestbo J, Calverley PM. The Pressing Need to Redefine "COPD". *Chronic Obstr Pulm Dis* 2019; **6**(5): 380-3.
66. Sood A, Petersen H, Qualls C, et al. Spirometric variability in smokers: transitions in COPD diagnosis in a five-year longitudinal study. *Respiratory research* 2016; **17**(1): 147.
67. Lundbäck B, Backman H, Calverley PMA. Lung Function through the PRISm. Spreading Light or Creating Confusion? *American journal of respiratory and critical care medicine* 2018; **198**(11): 1358-60.
68. https://www.ukbiobank.ac.uk/.
69. Agusti A, Faner R. Chronic Obstructive Pulmonary Disease Pathogenesis. *Clinics in Chest Medicine* 2020; **41**(3): 307-14.
70. Kotz D, Wesseling G, Huibers M, Schayck O. Efficacy of confrontational counselling for smoking cessation in smokers with previously undiagnosed mild to moderate airflow limitation: Study protocol of a randomized controlled trial. *BMC public health* 2007; **7**: 332.
71. Agusti A, Faner R. Lung function trajectories in health and disease. *The Lancet Respiratory medicine* 2019; **7**(4): 358-64.
72. Lange P, Celli B, Agustí A, et al. Lung-Function Trajectories Leading to Chronic Obstructive Pulmonary Disease. *The New England journal of medicine* 2015; **373**(2): 111-22.
73. Hall R, Hall IP, Sayers I. Genetic risk factors for the development of pulmonary disease identified by genome-wide association. 2019; **24**(3): 204-14.

74. Kim KW, Ober C. Lessons Learned From GWAS of Asthma. *Allergy Asthma Immunol Res* 2019; **11**(2): 170-87.
75. Shrine N, Guyatt AL, Erzurumluoglu AM, et al. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nature Genetics* 2019; **51**(3): 481-93.
76. Smith GD, Ebrahim S. Epidemiology—is it time to call it a day? *International journal of epidemiology* 2001; **30**(1): 1-11.
77. Belbasis L, Bellou V. Introduction to Epidemiological Studies. *Methods Mol Biol* 2018; **1793**: 1-6.
78. Data MITC, Danziger J, Zimolzak AJ. Residual Confounding Lurking in Big Data: A Source of Error. Secondary Analysis of Electronic Health Records. Cham (CH): Springer

Copyright 2016, The Author(s). 2016: 71-8.
79. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008; **123**(1): 15-33.
80. Sheehan NA, Didelez V. Epidemiology, genetic epidemiology and Mendelian randomisation: more need than ever to attend to detail. *Hum Genet* 2020; **139**(1): 121-36.
81. Collins R, Bowman L, Landray M, Peto R. The Magic of Randomization versus the Myth of Real-World Evidence. *The New England journal of medicine* 2020; **382**(7): 674-8.
82. Taubes G. Epidemiology faces its limits. *Science (New York, NY)* 1995; **269**(5221): 164-9.
83. Davey Smith G, Paternoster L, Relton C. When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health?Mendelian Randomization and Clinical Practice and Public HealthEditorial. *JAMA* 2017; **317**(6): 589-91.
84. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nature reviews Genetics* 2019; **20**(8): 467-84.
85. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012; **90**(1): 7-24.
86. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American journal of human genetics* 2017; **101**(1): 5-22.
87. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology* 2003; **32**(1): 1-22.
88. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Human Molecular Genetics* 2014; **23**(R1): R89-R98.
89. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Statistical methods in medical research* 2017; **26**(5): 2333-55.
90. NHGRI. Introduction to Genetics. https://www.genome.gov/About-Genomics/Introduction-to-Genomics.
91. Brookes AJ. The essence of SNPs. *Gene* 1999; **234**(2): 177-86.
92. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research* 2017; **45**(D1): D896-d901.
93. Fadista J, Manning AK, Florez JC, Groop L. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European journal of human genetics : EJHG* 2016; **24**(8): 1202-5.

94. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 2018; **50**(8): 1112-21.
95. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019; **51**(2): 237-44.
96. Pagoni P, Dimou NL, Murphy N, Stergiakouli E. Using Mendelian randomisation to assess causality in observational studies. *Evid Based Ment Health* 2019; **22**(2): 67-71.
97. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic epidemiology* 2013; **37**(7): 658-65.
98. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *Am J Clin Nutr* 2016; **103**(4): 965-78.
99. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *International journal of epidemiology* 2011; **40**(3): 755-64.
100. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Statistics in medicine* 2015; **34**(21): 2926-40.
101. Bowden J, Del Greco M F, Minelli C, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *International journal of epidemiology* 2018; **48**(3): 728-42.
102. Zheng J, Baird D, Borges M-C, et al. Recent Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep* 2017; **4**(4): 330-45.
103. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology* 2015; **44**(2): 512-25.
104. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International journal of epidemiology* 2017; **46**(6): 1985-98.
105. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS genetics* 2017; **13**(11): e1007081.
106. Incalzi RA, Gemma A, Marra C, Capparella O, Fuso L, Carbonin P. Verbal Memory Impairment in COPD: Its Mechanisms and Clinical Relevance. *Chest* 1997; **112**(6): 1506-13.
107. Gilsanz P, Mayeda ER, Flatt J, Glymour MM, Quesenberry CP, Jr., Whitmer RA. Early Midlife Pulmonary Function and Dementia Risk. *Alzheimer disease and associated disorders* 2018; **32**(4): 270-5.
108. Wain L, Shrine N, Guyatt A, et al. A weighted genetic risk score based on 279 signals of association with lung function predicts Chronic Obstructive Pulmonary Disease. *European Respiratory Journal* 2018; **52**(suppl 62): OA2188.
109. Sakornsakolpat P, Prokopenko D, Lamontagne M, et al. Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. *Nature Genetics* 2019; **51**(3): 494-505.
110. Lambert J-C, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* 2013; **45**: 1452.

111. dbGaP. Genotypes and Phenotypes. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000572.v7.p42019).
112. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* 2019; **51**(3): 404-13.
113. Medicine USo. 2019. https://www.med.unc.edu/pgc/pgc-workgroups/alzheimers-disease-workgroup/.
114. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* 2018; **7**.
115. Daborg J, von Otter M, Sjölander A, et al. Association of the RAGE G82S polymorphism with Alzheimer's disease. *Journal of neural transmission (Vienna, Austria : 1996)* 2010; **117**(7): 861-7.
116. Crous-Bou M, Minguillón C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimer's research & therapy* 2017; **9**(1): 71-.
117. McKay GJ, McCarter RV, Hogg RE, et al. Simple non-mydriatic retinal photography is feasible and demonstrates retinal microvascular dilation in Chronic Obstructive Pulmonary Disease (COPD). *PloS one* 2020; **15**(1): e0227175.
118. Dodd JW, Chung AW, van den Broek MD, Barrick TR, Charlton RA, Jones PW. Brain structure and function in chronic obstructive pulmonary disease: a multimodal cranial magnetic resonance imaging study. *American journal of respiratory and critical care medicine* 2012; **186**(3): 240-5.
119. Johnson N, Davis T, Bosanquet N. The Epidemic of Alzheimer's Disease. *PharmacoEconomics* 2000; **18**(3): 215-23.
120. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)* 2012; **380**(9859): 2095-128.
121. Xu W, Tan L, Wang HF, et al. Meta-analysis of modifiable risk factors for Alzheimer's disease. *Journal of neurology, neurosurgery, and psychiatry* 2015; **86**(12): 1299-306.
122. Postma DS, Bush A, van den Berge M. Risk factors and early origins of chronic obstructive pulmonary disease. *The Lancet* 2015; **385**(9971): 899-909.
123. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010; **42**(5): 441-7.
124. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* 2018; **33**(10): 947-52.
125. Vansteelandt S, Dukes O, Martinussen T. Survivor bias in Mendelian randomization analysis. *Biostatistics* 2017; **19**(4): 426-43.
126. Maclay JD, MacNee W. Cardiovascular disease in COPD: mechanisms. *Chest* 2013; **143**(3): 798-807.
127. Ramalho SHR, Shah AM. Lung function and cardiovascular disease: A link. *Trends in Cardiovascular Medicine* 2020.
128. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature genetics* 2015; **47**(10): 1121-30.
129. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* 2018; **50**(4): 524-37.

130. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *International journal of epidemiology* 2018; **48**(3): 713-27.
131. Hartwig FP, Tilling K, Davey Smith G, Lawlor DA, Borges MC. Bias in two-sample Mendelian randomization when using heritable covariable-adjusted summary associations. *International journal of epidemiology* 2021.
132. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single sample and two-sample summary data settings. 2018: 306209.
133. Sanderson E. Multivariable Mendelian Randomization and Mediation. *Cold Spring Harbor perspectives in medicine* 2021; **11**(2).
134. IEU. https://gwas.mrcieu.ac.uk/. 2020.
135. Bonella F, Stowasser S, Wollin L. Idiopathic pulmonary fibrosis: current treatment options and critical appraisal of nintedanib. *Drug design, development and therapy* 2015; **9**: 6407-19.
136. Au Yeung SL, Borges MC, Lawlor DA. Association of Genetic Instrumental Variables for Lung Function on Coronary Artery Disease Risk: A 2-Sample Mendelian Randomization Study. *Circulation Genomic and precision medicine* 2018; **11**(4): e001952.
137. Smith GD, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? 2005; **330**(7499): 1076-9.
138. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ (Clinical research ed)* 2018; **362**: k601.
139. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. *Epidemiology (Cambridge, Mass)* 2008; **19**(5): 649-51; discussion 57-8.
140. Furberg H, Kim Y, Dackor J, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics* 2010; **42**(5): 441-7.
141. Soldan A, Pettigrew C, Cai Q, et al. Cognitive reserve and long-term change in cognition in aging and preclinical Alzheimer's disease. *Neurobiol Aging* 2017; **60**: 164-72.
142. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology* 2012; **11**(11): 1006-12.
143. Cohen-Manheim I, Doniger GM, Sinnreich R, et al. Body Mass Index, Height and Socioeconomic Position in Adolescence, Their Trajectories into Adulthood, and Cognitive Function in Midlife. *J Alzheimers Dis* 2017; **55**(3): 1207-21.
144. Aschard H, Vilhjálmsson Bjarni J, Joshi Amit D, Price Alkes L, Kraft P. Adjusting for Heritable Covariates Can Bias Effect Estimates in Genome-Wide Association Studies. *The American Journal of Human Genetics* 2015; **96**(2): 329-39.
145. Leggett A, Clarke P, Zivin K, McCammon RJ, Elliott MR, Langa KM. Recent Improvements in Cognitive Functioning Among Older U.S. Adults: How Much Does Increasing Educational Attainment Explain? *J Gerontol B Psychol Sci Soc Sci* 2019; **74**(3): 536-45.
146. Bowden J, Spiller W, Del Greco MF, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *International journal of epidemiology* 2018; **47**(4): 1264-78.
147. Suglia SF, Wright RO, Schwartz J, Wright RJ. Association between lung function and cognition among children in a prospective birth cohort study. *Psychosomatic medicine* 2008; **70**(3): 356-62.

148. Liesker JJ, Postma DS, Beukema RJ, et al. Cognitive performance in patients with COPD. *Respiratory medicine* 2004; **98**(4): 351-6.
149. Bajaj MK, Burrage DR, Tappouni A, Dodd JW, Jones PW, Baker EH. COPD patients hospitalized with exacerbations have greater cognitive impairment than patients hospitalized with decompensated heart failure. *Clinical interventions in aging* 2019; **14**: 1-8.
150. Portas L, Pereira M, Shaheen SO, et al. Lung Development Genes and Adult Lung Function. *American journal of respiratory and critical care medicine* 2020; **202**(6): 853-65.
151. Wołoszynowska-Fraser MU, Kouchmeshky A, McCaffery P. Vitamin A and Retinoic Acid in Cognition and Cognitive Disease. *Annual review of nutrition* 2020; **40**: 247-72.
152. Howe LJ, Nivard MG, Morris TT, et al. Within-sibship GWAS improve estimates of direct genetic effects. *bioRxiv* 2021: 2021.03.05.433935.
153. Spilling CA, Bajaj M-PK, Burrage DR, et al. Contributions of cardiovascular risk and smoking to chronic obstructive pulmonary disease (COPD)-related changes in brain structure and function. *International journal of chronic obstructive pulmonary disease* 2019; **14**: 1855-66.
154. Trejo S, Domingue BW. Genetic nature or genetic nurture? Introducing social genetic parameters to quantify bias in polygenic score analyses. *Biodemography Soc Biol* 2018; **64**(3-4): 187-215.
155. Brumpton B, Sanderson E, Heilbron K, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nature communications* 2020; **11**(1): 3519.
156. Higbee DH, Dodd JW. Cognitive impairment in COPD: an often overlooked co-morbidity. *Expert review of respiratory medicine* 2020.
157. WHO T. https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/. 2020.
158. Young KA, Strand M, Ragland MF, et al. Pulmonary Subtypes Exhibit Differential Global Initiative for Chronic Obstructive Lung Disease Spirometry Stage Progression: The COPDGene® Study. *Chronic Obstr Pulm Dis* 2019; **6**(5): 414-29.
159. Wei X, Ding Q, Yu N, et al. Imaging Features of Chronic Bronchitis with Preserved Ratio and Impaired Spirometry (PRISm). *Lung* 2018; **196**(6): 649-58.
160. Lytras T. 2020. https://github.com/thlytras/rspiro2020).
161. STATACorp. Stata Statistical Software: Release 15. 2017.
162. Linden A. RTMCI: Stata module to estimate regression to the mean effects with confidence intervals. In: S457757 SSC, editor.; 2013.
163. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *International journal of epidemiology* 2005; **34**(1): 215-20.
164. Wijnant SRA, Lahousse L, Brusselle GG. The global significance of PRISm: how data from low- and middle-income countries link physiology to inflammation. *European Respiratory Journal* 2020; **55**(4): 2000354.
165. Higbee DH, Granell R, Sanderson E, Davey Smith G, Dodd JW. Lung function & cardiovascular disease. A Two Sample Mendelian Randomisation Study. *The European respiratory journal* 2021.
166. Pompe E, Strand M, van Rikxoort EM, et al. Five-year Progression of Emphysema and Air Trapping at CT in Smokers with and Those without Chronic Obstructive Pulmonary Disease: Results from the COPDGene Study. *Radiology* 2020; **295**(1): 218-26.
167. Bonini M, Usmani OS. The role of the small airways in the pathophysiology of asthma and chronic obstructive pulmonary disease. *Ther Adv Respir Dis* 2015; **9**(6): 281-93.

168. Hartley RA, Barker BL, Newby C, et al. Relationship between lung function and quantitative computed tomographic parameters of airway remodeling, air trapping, and emphysema in patients with asthma and chronic obstructive pulmonary disease: A single-center study. *The Journal of allergy and clinical immunology* 2016; **137**(5): 1413-22.e12.
169. Backman H, Eriksson B, Hedman L, et al. Restrictive spirometric pattern in the general adult population: Methods of defining the condition and consequences on prevalence. *Respiratory medicine* 2016; **120**: 116-23.
170. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ (Clinical research ed)* 2020; **368**: m131.
171. Fry A, Littlejohns T, Sudlow C, Doherty N, Allen N. OP41 The representativeness of the UK Biobank cohort on a range of sociodemographic, physical, lifestyle and health-related characteristics. *Journal of epidemiology and community health* 2016; **70**(Suppl 1): A26-A.
172. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when selection bias can substantially influence observed associations. *International journal of epidemiology* 2018; **47**(1): 226-35.
173. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* 2016; **48**(10): 1279-83.
174. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet* 2018; **50**(7): 906-8.
175. Ruth Mitchell GH, Tom Dudding, Laura Corbin, Sean Harrison, Lavinia Paternoster. UK BioBank Genetic Data: MRC-IEU Quality Control, version 2. 2019.
176. Staley JR, Blackshaw J, Kamat MA, et al. PhenoScanner: a database of human genotype-phenotype associations. *Bioinformatics (Oxford, England)* 2016; **32**(20): 3207-9.
177. https://www.gtexportal.org/home/. (accessed 20/07/2020.
178. Science WIo. www.genecards.org. (accessed 20/07/2020.
179. http://www.informatics.jax.org/. (accessed 20/07/2020.
180. 2021. https://omim.org/.
181. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics (Oxford, England)* 2017; **33**(2): 272-9.
182. Spracklen CN, Horikoshi M, Kim YJ, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 2020; **582**(7811): 240-5.
183. Manning AK, Hivert MF, Scott RA, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**(6): 659-69.
184. Mungall AJ, Palmer SA, Sims SK, et al. The DNA sequence and analysis of human chromosome 6. *Nature* 2003; **425**(6960): 805-11.
185. Pavord ID, Beasley R, Agusti A, et al. After asthma: redefining airways diseases. *Lancet (London, England)* 2018; **391**(10118): 350-400.
186. GINA. The Global Strategy for Asthma Management and Prevention; 2017.
187. Mims JW. Asthma: definitions and pathophysiology. *International forum of allergy & rhinology* 2015; **5 Suppl 1**: S2-6.

188. Morrison D, Agur K, Mercer S, Eiras A, González-Montalvo JI, Gruffydd-Jones K. Managing multimorbidity in primary care in patients with chronic respiratory conditions. *NPJ primary care respiratory medicine* 2016; **26**: 16043-.
189. Bosley CM, Fosbury JA, Cochrane GM. The psychological factors associated with poor compliance with treatment in asthma. *The European respiratory journal* 1995; **8**(6): 899-904.
190. Physicians TRCo. Why asthma still kills. The National Review of Asthma Deaths. In: Partnership HQI, editor.; 2014.
191. Dudeney J, Sharpe L, Jaffe A, Jones EB, Hunt C. Anxiety in youth with asthma: A meta-analysis. *Pediatric pulmonology* 2017; **52**(9): 1121-9.
192. Del Giacco SR, Cappai A, Gambula L, et al. The asthma-anxiety connection. *Respiratory medicine* 2016; **120**: 44-53.
193. Oh H, Stickley A, Singh F, Koyanagi A. Self-reported asthma diagnosis and mental health: Findings from the Collaborative Psychiatric Epidemiology Surveys. *Psychiatry research* 2019; **271**: 721-5.
194. Lee YC, Lee CT, Lai YR, Chen VC, Stewart R. Association of asthma and anxiety: A nationwide population-based study in Taiwan. *Journal of affective disorders* 2016; **189**: 98-105.
195. Geraldo José Cunha Â, Zbonik Mendes A, Dias Wanderley de Carvalho F, Aparecida Ribeiro de Paula M, Gonçalves Brasil T. The impact of asthma on quality of life and anxiety: a pilot study. *Journal of Asthma* 2018: 1-6.
196. Lomper K, Chudiak A, Uchmanowicz I, Rosinczuk J, Jankowska-Polanska B. Effects of depression and anxiety on asthma-related quality of life. *Pneumonologia i alergologia polska* 2016; **84**(4): 212-21.
197. Ciprandi G, Schiavetti I, Rindone E, Ricciardolo FL. The impact of anxiety and depression on outpatients with asthma. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology* 2015; **115**(5): 408-14.
198. Urrutia I, Aguirre U, Pascual S, et al. Impact of anxiety and depression on disease control and quality of life in asthma patients. *The Journal of asthma : official journal of the Association for the Care of Asthma* 2012; **49**(2): 201-8.
199. Di Marco F, Verga M, Santus P, et al. Close correlation between anxiety, depression, and asthma control. *Respiratory medicine* 2010; **104**(1): 22-8.
200. Goodwin RD, Fergusson DM, Horwood LJ. Asthma and depressive and anxiety disorders among young persons in the community. *Psychological Medicine* 2004; **34**(8): 1465-74.
201. Letitre SL, de Groot EP, Draaisma E, Brand PL. Anxiety, depression and self-esteem in children with well-controlled asthma: case-control study. *Archives of disease in childhood* 2014; **99**(8): 744-8.
202. Chen MH, Su TP, Chen YS, et al. Higher risk of developing major depression and bipolar disorder in later life among adolescents with asthma: a nationwide prospective study. *Journal of psychiatric research* 2014; **49**: 25-30.
203. Goldney RD, Ruffin R, Fisher LJ, Wilson DH. Asthma symptoms associated with depression and lower quality of life: a population survey. *The Medical journal of Australia* 2003; **178**(9): 437-41.
204. Yonas MA, Marsland AL, Emeremni CA, Moore CG, Holguin F, Wenzel S. Depressive symptomatology, quality of life and disease control among individuals with well-characterized severe asthma. *The Journal of asthma : official journal of the Association for the Care of Asthma* 2013; **50**(8): 884-90.

205. Scott KM, Von Korff M, Ormel J, et al. Mental disorders among adults with asthma: results from the World Mental Health Survey. *General hospital psychiatry* 2007; **29**(2): 123-33.
206. Lu Z, Chen L, Xu S, et al. Allergic disorders and risk of depression: A systematic review and meta-analysis of 51 large-scale studies. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology* 2018; **120**(3): 310-7.e2.
207. Shen TC, Lin CL, Liao CH, Wei CC, Sung FC, Kao CH. Major depressive disorder is associated with subsequent adult-onset asthma: a population-based cohort study. *Epidemiology and psychiatric sciences* 2017; **26**(6): 664-71.
208. Gao YH, Zhao HS, Zhang FR, et al. The Relationship between Depression and Asthma: A Meta-Analysis of Prospective Studies. *PloS one* 2015; **10**(7): e0132424.
209. Mangold R, Salzman GA, Williams KB, Hanania NA. Factors associated with depressive symptoms in uncontrolled asthmatics. *The Journal of asthma : official journal of the Association for the Care of Asthma* 2018; **55**(5): 555-60.
210. Zhang L, Zhang X, Zheng J, et al. Co-morbid psychological dysfunction is associated with a higher risk of asthma exacerbations: a systematic review and meta-analysis. *Journal of thoracic disease* 2016; **8**(6): 1257-68.
211. Di Marco F, Santus P, Centanni S. Anxiety and depression in asthma. *Current opinion in pulmonary medicine* 2011; **17**(1): 39-44.
212. Chen E, Miller GE. Stress and inflammation in exacerbations of asthma. *Brain, behavior, and immunity* 2007; **21**(8): 993-9.
213. Deshmukh VM, Toelle BG, Usherwood T, O'Grady B, Jenkins CR. Anxiety, panic and adult asthma: a cognitive-behavioral perspective. *Respiratory medicine* 2007; **101**(2): 194-202.
214. University JH. COVID-19 Dashboard. https://coronavirus.jhu.edu/map.html (accessed 27/08/2020.
215. Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet (London, England)* 2020; **396**(10259): 1345-52.
216. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet (London, England)* 2021; **397**(10285): 1637-45.
217. Azithromycin in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet (London, England)* 2021; **397**(10274): 605-12.
218. Horby P, Lim WS, Emberson JR, et al. Dexamethasone in Hospitalized Patients with Covid-19. *The New England journal of medicine* 2021; **384**(8): 693-704.
219. Horby P, Mafham M, Linsell L, et al. Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19. *The New England journal of medicine* 2020; **383**(21): 2030-40.
220. Wang R, Bikov A, Fowler SJ. Treating asthma in the COVID-19 pandemic. *Thorax* 2020.
221. Odriozola-González P, Planchuelo-Gómez Á, Irurtia MJ, de Luis-García R. Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry research* 2020; **290**: 113108.

222. Qiu J, Shen B, Zhao M, Wang Z, Xie B, Xu Y. A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. *General Psychiatry* 2020; **33**(2): e100213.
223. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. *Emerg Infect Dis* 2004; **10**(7): 1206-12.
224. Sears MR. Descriptive epidemiology of asthma. *Lancet (London, England)* 1997; **350 Suppl 2**: Sii1-4.
225. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International journal of epidemiology* 2013; **42**(1): 97-110.
226. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013; **42**(1): 111-27.
227. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Res* 2019; **4**: 51.
228. Northstone K, Haworth S, Smith D, Bowring C, Wells N, Timpson N. The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Questionnaire data capture April-May 2020 [version 1; peer review: awaiting peer review]. *Wellcome Open Research* 2020; **5**(127).
229. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry : the journal of mental science* 1987; **150**: 782-6.
230. Spielberger C, Gorsuch R, Lushene R, Vagg PR, Jacobs G. Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2); 1983.
231. Kwong ASF, Pearson RM, Adams MJ, et al. Mental health during the COVID-19 pandemic in two longitudinal UK population cohorts. *medRxiv* 2020: 2020.06.16.20133116.
232. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Res* 2019; **4**: 51-.
233. Glaser B, Gunnell D, Timpson NJ, et al. Age- and puberty-dependent association between IQ score in early childhood and depressive symptoms in adolescence. *Psychol Med* 2011; **41**(2): 333-43.
234. Hill G. Data Analysis using Regression and Multilevel/Hierarchical Models. Cambridge, UK: Cambridge University Press; 2007.
235. Bianco. Performance of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) as a screening tool for depression in UK and Italy
236. Toussaint A, Hüsing P, Gumz A, et al. Sensitivity to change and minimal clinically important difference of the 7-item Generalized Anxiety Disorder Questionnaire (GAD-7). *Journal of affective disorders* 2020; **265**: 395-401.
237. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine* 2006; **166**(10): 1092-7.
238. BLF. Nearly 2 million people with lung conditions notice improved symptoms as a result of drop in air pollution. https://www.blf.org.uk/media-centre/press-releases/nearly-2-million-people-with-lung-conditions-notice-improved-symptoms-as (accessed 27/08/2020.

239. Papadopoulos NG, Custovic A, Deschildre A, et al. Impact of COVID-19 on Pediatric Asthma: Practice Adjustments and Disease Burden. *The journal of allergy and clinical immunology In practice* 2020; **8**(8): 2592-9.e3.
240. Johnson B. https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-16-march-2020 (accessed 27/08/2020.
241. PHE. Guidance on social distancing for everyone in the UK https://www.gov.uk/government/publications/covid-19-guidance-on-social-distancing-and-for-vulnerable-people/guidance-on-social-distancing-for-everyone-in-the-uk-and-protecting-older-people-and-vulnerable-adults (accessed 27/08/2020.
242. Cornish RP, Henderson J, Boyd AW, Granell R, Van Staa T, Macleod J. Validating childhood asthma in an epidemiological study using linked electronic patient records. *BMJ Open* 2014; **4**(4): e005345.
243. Bousquet J, Jutel M, Akdis CA, et al. ARIA-EAACI statement on Asthma and COVID-19 (June 2, 2020). *Allergy* 2020.
244. Lu Y, Mak KK, van Bever HP, Ng TP, Mak A, Ho RC. Prevalence of anxiety and depressive symptoms in adolescents with asthma: a meta-analysis and meta-regression. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* 2012; **23**(8): 707-15.
245. Negewo NA, Gibson PG, McDonald VM. COPD and its comorbidities: Impact, measurement and mechanisms. *Respirology (Carlton, Vic)* 2015; **20**(8): 1160-71.
246. Uijen AA, van de Lisdonk EH. Multimorbidity in primary care: prevalence and trend over the last 20 years. *The European journal of general practice* 2008; **14 Suppl 1**: 28-32.
247. Multimorbidity: clinical assessment and management. In: Excellence NIfHaC, editor.; 2016.
248. Organization TWH. Multimorbidity: Technical Series on Safer Primary Care Geneva, 2016.
249. Farmer C, Fenu E, O'Flynn N, Guthrie B. Clinical assessment and management of multimorbidity: summary of NICE guidance. 2016; **354**: i4843.
250. Kalaria RN, Akinyemi R, Ihara M. Stroke injury, cognitive impairment and vascular dementia. *Biochim Biophys Acta* 2016; **1862**(5): 915-25.
251. Assari S, Chalian H, Bazargan M. Race, Ethnicity, Socioeconomic Status, and Chronic Lung Disease in the U.S. *Res Health Sci* 2020; **5**(1): 48-63.
252. Schrijvers EM, Schürmann B, Koudstaal PJ, et al. Genome-wide association study of vascular dementia. *Stroke* 2012; **43**(2): 315-9.
253. Sterne JA, Davey Smith G. Sifting the evidence-what's wrong with significance tests? *BMJ (Clinical research ed)* 2001; **322**(7280): 226-31.
254. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019; **567**(7748): 305-7.
255. Baum C, Ojeda FM, Wild PS, et al. Subclinical impairment of lung function is related to mild cardiac dysfunction and manifest heart failure in the general population. *International journal of cardiology* 2016; **218**: 298-304.
256. Franssen FM, Soriano JB, Roche N, et al. Lung Function Abnormalities in Smokers with Ischemic Heart Disease. *American journal of respiratory and critical care medicine* 2016; **194**(5): 568-76.
257. Liang Y, Wang M, Xu X, Li N, Zhou Q, He B. Reduced Forced Expiratory Volume in 1 Second Percentage Predicted Is Associated With Diffuse Coronary Atherosclerosis in

Hospitalized Patients Undergoing Coronary Angiography. *The American journal of the medical sciences* 2018; **355**(4): 307-13.
258. Mannino DM, Thorn D, Swensen A, Holguin F. Prevalence and outcomes of diabetes, hypertension and cardiovascular disease in COPD. *The European respiratory journal* 2008; **32**(4): 962-9.
259. Soler EP, Ruiz VC. Epidemiology and risk factors of cerebral ischemia and ischemic heart diseases: similarities and differences. *Current cardiology reviews* 2010; **6**(3): 138-49.
260. Vivodtzev I, Tamisier R, Baguet J-P, Borel JC, Levy P, Pépin J-L. Arterial stiffness in COPD. *Chest* 2014; **145**(4): 861-75.
261. Belgrave DCM, Granell R, Turner SW, et al. Lung function trajectories from pre-school age to adulthood and their associations with early life factors: a retrospective analysis of three population-based birth cohort studies. *The Lancet Respiratory medicine* 2018; **6**(7): 526-34.
262. Finkel D, Reynolds CA, Emery CF, Pedersen NL. Genetic and environmental variation in lung function drives subsequent variation in aging of fluid intelligence. *Behav Genet* 2013; **43**(4): 274-85.
263. Schou L, Ostergaard B, Rasmussen LS, Rydahl-Hansen S, Phanareth K. Cognitive dysfunction in patients with chronic obstructive pulmonary disease--a systematic review. *Respiratory medicine* 2012; **106**(8): 1071-81.
264. Vasilopoulos T, Kremen WS, Grant MD, et al. Individual differences in cognitive ability at age 20 predict pulmonary function 35 years later. *Journal of epidemiology and community health* 2015; **69**(3): 261-5.
265. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nature Genetics* 2018; **50**(5): 693-8.
266. Chiu HY, Hsiao YH, Su KC, Lee YC, Ko HK, Perng DW. Small Airway Dysfunction by Impulse Oscillometry in Symptomatic Patients with Preserved Pulmonary Function. *The journal of allergy and clinical immunology In practice* 2020; **8**(1): 229-35.e3.
267. Carli G, Cecchi L, Stebbing J, Parronchi P, Farsi A. Is asthma protective against COVID-19? *Allergy* 2021; **76**(3): 866-8.
268. Eger K, Bel EH. Asthma and COVID-19: do we finally have answers? *The European respiratory journal* 2021; **57**(3).
269. Ramakrishnan S, Nicolau DV, Jr., Langford B, et al. Inhaled budesonide in the treatment of early COVID-19 (STOIC): a phase 2, open-label, randomised controlled trial. *The Lancet Respiratory medicine* 2021; **9**(7): 763-72.
270. Zhu Z, Hasegawa K, Camargo CA, Jr., Liang L. Investigating asthma heterogeneity through shared and distinct genetics: Insights from genome-wide cross-trait analysis. *The Journal of allergy and clinical immunology* 2020: S0091-6749(20)30966-0.
271. Budu-Aggrey A, Joyce S, Davies NM, et al. Investigating the causal relationship between allergic disease and mental health. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* 2021; **51**(11): 1449-58.
272. A.E.Mourant. The distribution of the Human blood groups and other polymorphisms. 2nd ed; 1976.
273. Davies NM, Howe LJ, Brumpton B, Havdahl A, Evans DM, Davey Smith G. Within family Mendelian randomization studies. *Human Molecular Genetics* 2019; **28**(R2): R170-R9.
274. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature* 2016; **538**(7624): 161-4.

275. Carlson CS, Matise TC, North KE, et al. Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study. *PLOS Biology* 2013; **11**(9): e1001661.
276. Salanti G, Sanderson S, Higgins JP. Obstacles and opportunities in meta-analysis of genetic association studies. *Genet Med* 2005; **7**(1): 13-20.
277. Tada H, Nohara A, Kawashiri M-a. Serum Triglycerides and Atherosclerotic Cardiovascular Disease: Insights from Clinical and Genetic Studies. *Nutrients* 2018; **10**(11): 1789.
278. Swerdlow DI, Preiss D, Kuchenbaecker KB, et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *The Lancet* 2015; **385**(9965): 351-61.
279. Ference BA, Holmes MV, Smith GD. Using Mendelian Randomization to Improve the Design of Randomized Trials. *Cold Spring Harbor perspectives in medicine* 2021; **11**(7).