Epidemiology and Genetics of Preserved Ratio Impaired SpiroMetry (PRISm): An Analysis of UK Biobank

Daniel H Higbee MBBS\textsuperscript{1,2}, Raquel Granell PhD\textsuperscript{1}, George Davey Smith FRS\textsuperscript{1}, James W Dodd MB ChB, PhD\textsuperscript{1,2}

1 MRC Integrative Epidemiology Unit (IEU), University of Bristol, Oakfield Grove, Bristol, BS8 2BN, UK.

2 Academic Respiratory Unit, University of Bristol, Southmead Hospital, Bristol, BS10 5NB UK

Corresponding Author: Dr James W. Dodd

dr.james.dodd@bristol.ac.uk

Tel +44117 4142012
Fax +44117 4149496

ORCID 0000-0003-4805-5759

JD & DH conceived the study and initial draft, all authors made substantial contributions to the analysis, drafting and final approval of the paper.

Funding: This work was supported by the Medical Research Council and the University of Bristol Integrative Epidemiology Unit (MC_UU_00011). MRC CARP Fellowship (Grant ref: MR/T005114/1)

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Rationale

Preserved ratio impaired spirometry (PRISm) is defined as Forced Expiratory Volume in one second <80% predicted with a FEV\textsubscript{1}/Forced Vital Capacity of >0.70. It may be a precursor of COPD and has been associated with increased respiratory symptoms and comorbidities. Current understanding of PRISm is based on small selective cohorts and genetic determinants remain unknown.

Objectives

Identify the prevalence, risk factors, co-morbidity, and genetics of PRISm in a large cohort to improve power and generalizability.

Methods

The UKBiobank, provides a large cohort of genotyped individuals. Regression analysis determined risk factors for and associated co-morbidities of PRISm. Genome wide association studies, conditional, and gene-based analyses were performed with a PheWAS and literature search conducted on the results.

Measurements and Main Results

PRISm has a similar prevalence to airflow obstruction (8% vs 7.8%) and is associated with increased cardiovascular co-morbidity, OR of 1.26 (95%CI 1.20-1.31) after adjusting for BMI, smoking status, age, diabetes, and sex. Genetic analysis discovered four associated single nucleotide polymorphisms and eight genes. Two SNPs have not been described with respiratory disease before. Other genetic loci identified are known to increase risk of asthma and COPD, and influence response to inhaled steroids.

Conclusions

PRISm affects 8% of adults of whom ~40% report cardiovascular comorbidity. Detailed analysis reveals that PRISm has shared genetic associations with asthma and COPD. These findings highlight the need for research into the potential therapeutic strategies and screening for individuals with PRISm, in particular whether this may reduce transition to COPD and reduce cardiovascular disease.
Introduction
Preserved Ratio Impaired Spirometry (PRISm), also referred to as ‘restrictive pattern’ or ‘unclassified’ spirometry, is defined as Forced Expiratory Volume in one second (FEV₁) percent predicted <80%, despite a normal FEV₁/Forced Vital Capacity (FVC) ratio >0.70. The true prevalence of PRISm is unknown with estimates ranging from just 4% up to 48% amongst adult varying by gender, ancestry and smoking history.¹, ² Interest in the potential clinical implications of PRISm come from longitudinal data which suggest that over 5 years, up to 50% of people with PRISm transition to COPD but also that 15% may return to ‘normal’ spirometry.³ PRISm is understudied, but available data suggests that it is associated with increased respiratory symptoms, important co-morbidities such as obesity, diabetes, cardiac disease and an increased overall mortality.³, ⁴ However, the epidemiology and genetics of PRISm remains unclear. Research is needed to clarify and better understand the nature of this abnormality in lung function.

Detailed epidemiological and genetic understanding of PRISm has to date been limited by analysis restricted to insufficient population numbers in cohorts rarely containing more than 1000 cases of PRISm. Generalisability of data has also been an issue with some cohorts restricted to smokers.², ⁴ This may introduce selection bias and confounding, limiting conclusions regarding PRISm as a potential pre-COPD state and as an independent risk for cardiovascular disease.⁵

There has been just one attempted Genome Wide Association Study (GWAS) in PRISm but numbers again were small (N=1257), restricted to ever smokers, and failed to find any significant single nucleotide polymorphisms (SNPs, p value <5x10⁻⁸).² Genetic markers are critical to our understanding of PRISm since they provide insight into disease susceptibility and pathogenesis. Lung function trajectories determined by genetic and early life factors are now
known to influence risk of COPD, \(^6, 7\) and the same may be true of PRISm. Genetic variants could therefore provide therapeutic targets for PRISm and associated co-morbidities.

Our first objective was to utilise the UKBiobank to examine a cohort of individuals with PRISm, obstructive and normal spirometry.\(^8\) The large sample size of UKBiobank and broad recruitment based on age, not smoking history, increases power and reduces selection bias allowing improved generalizability to the wider population. Differences in demographics, symptom burden and associated co-morbidity in PRISm were compared to other spirometry states.

Our second objective was to use genetic epidemiology to determine associations of PRISm. A GWAS, conditional analysis, and gene-based analysis were performed to discover genetic associations of PRISm and compare to known genetic causes of other respiratory diseases.

**Method**

**Objective One - Observational Epidemiology**

UKBiobank includes 502,543 individuals aged between 40 and 69 at recruitment across the UK.\(^8\) Participants completed detailed health questionnaires and blood samples were taken for genotyping. 151,973 participants had pre-bronchodilator lung function testing that included \(\text{FEV}_1\) percent predicted. Obstructive spirometry was defined as \(\text{FEV}_1 < 80\%\) predicted and \(\text{FEV}_1/\text{FVC} < 0.70\) (Stage II-IV obstruction),\(^9\) and controls as \(\text{FEV}_1 \geq 80\%\) with \(\text{FEV}_1/\text{FVC} > 0.70\). Statistical analysis was performed using Stata 15.\(^{10}\) We used logistic regression analysis to examine for associations between risk factors for PRISm, and clinically relevant correlates of PRISm.
Objective Two – GWAS
A GWAS of PRISm and controls was performed using BOLT-LMM via the IEU GWAS pipeline adjusting for sex, body mass index (BMI), year of birth and smoking status.\(^{(11)}\) This uses a linear mixed model to account for both cryptic relatedness and population stratification. Only participants of European ancestry were included. 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.\(^{(11, 12)}\) Linkage Disequilibrium(LD)-clumping, which retains the most significant SNP at a locus to ensure all signals are independent, was performed with criteria \(kb = 1000\) and \(r^2 = 0.001\). LD-regression analysis was performed to assess correlation of PRISm with a GWAS examining continuous lung function measures.\(^{(13-15)}\)

Objective Two – Conditional analysis
GWAS methodology assumes the top SNP found explains the maximum variance in the region, and that close SNPs showing association only do so because they are in high LD. This assumption may be incorrect leading to SNPs not being detected. Conditional analysis mutually adjusts on the top SNP to test whether the other SNPs at the locus are significantly associated.\(^{(16)}\) We used Genome-wide Complex Trait Analysis software for a Conditional and Joint analysis (GCTA-COJO) on our GWAS results.\(^{(17)}\) A reference sample of individually genotyped participants of the Avon Longitudinal Study of Parent and Children (ALSPAC) was used to estimate LD patterns.\(^{(18)}\)

Objective Two – Gene-based analysis
Gene-based analysis was performed using Versatile Gene-based Association Study software (VEGAS2).\(^{(19)}\) Gene based analysis reduces the risk of type 1 error as ~20,000 genes are tested rather than millions of SNPs. It has greater power to detect genes that may contain multiple SNPs of marginal significance, which may be missed by SNP based GWAS. 21212 genes were tested so a Bonferroni corrected p-value threshold of <2.36x10\(^{-6}\) (=0.05/21212) was applied.
to correct for multiple testing. The gene-based analysis determines the SNP with the highest p value in each gene, and reports it as Top SNP.

Genes found were cross referenced with online resources to determine their chromosomal location and function.\(^{[20-22]}\) We performed a literature search for reported lung disease associations of significant genes found.\(^{[23, 24]}\) A Phenome Wide Association Study (PheWAS) was performed on SNPs from all three objectives to determine if they have associations with other lung function measures or disease.\(^{[21, 25]}\) Full PheWAS results are in the online additional material.

Objective Two methods were repeated after excluding participants who have asthma/COPD. This is reported in the online data supplement.

Results

Objective One – Observational Epidemiology

12,238 of participants in UKBiobank meet the criteria for PRISm, 11904 participants had stage II-IV obstruction and there are 115,890 controls. Pearson’s chi-squared test shows that the prevalence of PRISm (8.0%) is marginally higher than obstruction stage II-IV pattern spirometry (7.8%, p-value 0.025). PRISm is strongly associated with current smoking (p value <0.001), the absolute percentage of current smokers in PRISm is double that of controls, and doubles again between PRISm and obstruction (Table 1). Current smoking is higher in PRISm with an Odds Ratio (OR) of 2.37 (95%CI 2.23-2.51) vs. control after adjustment for age, sex and BMI. Mean BMI values are 2.2kg/m\(^2\) higher amongst the PRISm group that controls. Regression analysis shows a small OR of 1.08 (95%CI 1.07-1.08) per SD increase in BMI for PRISm compared to controls after adjustment for age, sex and smoking.
Self-reported rates of asthma were higher in PRISm than controls (18% vs 10%, p value <0.001). Rates of self-reported COPD equivalent between PRISm and obstruction (4.97% vs 4.51%, p value 0.1).

PRISm participants report a higher burden of respiratory symptoms than controls, with 7.1% reporting shortness of breath walking on ground level compared to 2.7% controls, OR 2.03 (95%CI 1.88-2.26) after adjustment by age, sex and BMI. This was measured by participants being asked, “Do you get short of breath walking with people of your own age on level ground?” Co-morbidities are more frequent in PRISm with >10% diagnosed with diabetes, OR 1.67 (95%CI 1.56-1.79) in PRISm vs. controls after adjustment for age, sex and BMI. Cardiovascular co-morbidity in UKBiobank is defined as patients self-reporting any diagnosis of hypertension, angina, stroke, or myocardial infarction. 39.4% of PRISm reported cardiovascular disease vs. 26.7% of controls, OR of 1.26 (95%CI 1.20-1.31) after adjusting for BMI, smoking status, age, diabetes, and sex. There does not appear to be a difference between the prevalence of cardiovascular co-morbidity between PRISm vs. participants with obstructive pattern spirometry despite the difference in smoking rates, sex, and age (cardiovascular co-morbidity 39.4% PRISm, 38.9% Obstruction, p value 0.59).
Table 1. Characteristics of PRISm, Obstructive and Control participants in UK Biobank

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>PRISm N=12,238 (8.0%)</th>
<th>Control N=115,807 (76.2%)</th>
<th>Pval* PRISm vs. Control</th>
<th>Stage II-IV Obstruction N=11,904 (7.8%)</th>
<th>Pval* PRISm vs. Obstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Years) Mean (SD)</td>
<td>63.7 (7)</td>
<td>62.5 (7)</td>
<td>&lt;0.001</td>
<td>66.5 (7)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI (kg/m²) Mean (SD)</td>
<td>29.6 (5)</td>
<td>27.4 (5)</td>
<td>&lt;0.001</td>
<td>27.5 (5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female (%)</td>
<td>58.6</td>
<td>60.4</td>
<td>&lt;0.001</td>
<td>47.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>FEV₁ % predicted Mean (SD)</td>
<td>72.1 (7)</td>
<td>101.4 (12)</td>
<td>&lt;0.001</td>
<td>64.9 (12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Never smoker (%)</td>
<td>59.7</td>
<td>74.4</td>
<td>&lt;0.001</td>
<td>37.5</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ex-smoker (%)</td>
<td>25.8</td>
<td>17.4</td>
<td>&lt;0.001</td>
<td>33.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Current smoker (%)</td>
<td>14.3</td>
<td>7.7</td>
<td>&lt;0.001</td>
<td>28.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pack/years † Mean (SD)</td>
<td>37.4 (19)</td>
<td>31.3 (15)</td>
<td>&lt;0.001</td>
<td>42.0 (20)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SOB walking on ground (%)</td>
<td>7.1</td>
<td>2.7</td>
<td>&lt;0.001</td>
<td>9.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Self-reported asthma</td>
<td>18.4%</td>
<td>10.4%</td>
<td>&lt;0.001</td>
<td>35.3%</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Self-reported COPD</td>
<td>4.97%</td>
<td>0.08%</td>
<td>&lt;0.001</td>
<td>4.51%</td>
<td>0.105</td>
</tr>
<tr>
<td>Diabetes (%)</td>
<td>10.7</td>
<td>4.1</td>
<td>&lt;0.001</td>
<td>6.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Any cardiac history‡ (%)</td>
<td>39.4</td>
<td>26.7</td>
<td>&lt;0.001</td>
<td>38.9</td>
<td>0.599</td>
</tr>
<tr>
<td>Weight loss in last year (%)</td>
<td>14.2</td>
<td>15.1</td>
<td>0.014</td>
<td>15.2</td>
<td>0.050</td>
</tr>
</tbody>
</table>

*Z-score from regression for continuous outcomes, Pearson’s Chi squared for categorical outcomes; † (ex & current smokers); ‡ angina, MI, CVE, HTN

Objective Two – GWAS

12,321,875 imputed SNPs were tested in the GWAS. LD score regression showed very strong negative correlations with continuous lung function measures from a large GWAS,[15] FEV₁ (rg -0.93, p value 2.39⁻³⁷), FVC (rg -0.89, p value 1.13⁻³⁵), with strong negative correlation for FEV₁/FVC ratio (-0.19, p value 0.004).

953 SNPs met the GWAS significance threshold of p value <5x10⁻⁸, all of which were on chromosome 17 (Figure 1). After LD-clumping two SNPs remained (p value <5x10⁻⁸): rs378392 (chr17:43687542) and rs553682853 (chr17:44759766) with OR of PRISm 0.992 (0.990 -0.994) and 0.990 (0.988-0.992) with C and G allele respectively.

PheWAS did not show that either SNP has previously been associated with lung function or lung disease. They have been described in association with other traits, see supplementary tables, including circulating cells. Rs378392 is an intron in the ribosomal protein S26.
pseudogene 8 gene (*RPS26P8*), a processed pseudogene in the 17q21 locus.\(^{[21]}\) This gene has been associated with FEV\(_1\) and FVC in previous studies of non-UK Biobank populations.\(^{[26]}\)

**Figure 1.** Each point represents one of the 755,702 SNPs that reached p value <0.05 in our GWAS

**Objective Two – Conditional analysis**
Conditional analysis found one p value<5x10\(^{-8}\) SNP, rs1991556 in chromosome 17 (17:44083402). Per G allele it confers OR (95%CI) 0.91 (0.88–0.94) of PRISm. PheWAS showed that this SNP has a very strong association with FEV\(_1\) (p value 4.3x10\(^{-25}\)) and FVC (p value 7.5x10\(^{-18}\)) percent predicted, some evidence of association with self-reported emphysema/chronic bronchitis (p value 7.8x10\(^{-4}\)) and age of asthma diagnosis (p value 9.7x10\(^{-4}\)). It is found in the *MAPT* gene discussed below.

**Objective Two – Gene-based analysis**
Gene-based analysis found one p value<5x10\(^{-8}\) SNP, rs7213474 (17:79962718), an intron variant in the Alveolar Soft Part Sarcoma Gene (*ASPCR1*) gene on chromosome 17.\(^{[20]}\) Extracting from the GWAS it gives OR (95% CI) 1.006 (1.004 – 1.008) per T allele. PheWAS showed it has previously been strongly associated with FEV\(_1\), FVC and PEFR (all p value <0.5x10\(^{-8}\)).

8 genes reached significance threshold, (Table 2) the 6 most significant are all on chromosome 17. All genes discovered are protein coding.
Table 2. Results of Gene based analysis. Genes p-value <2.36x10^{-6}

<table>
<thead>
<tr>
<th>Chr</th>
<th>Gene</th>
<th>nSNPs</th>
<th>Pvalue for Gene</th>
<th>Top SNP</th>
<th>Top SNP Pvalue</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>CRHR1</td>
<td>281</td>
<td>9.9 x 10^{-07}</td>
<td>rs113173628</td>
<td>2.40 x 10^{-07}</td>
</tr>
<tr>
<td>17</td>
<td>KANSL1</td>
<td>88</td>
<td>9.9 x 10^{-07}</td>
<td>rs11233322</td>
<td>1.60 x 10^{-07}</td>
</tr>
<tr>
<td>17</td>
<td>MAPT</td>
<td>73</td>
<td>9.9 x 10^{-07}</td>
<td>rs118087478</td>
<td>1.60 x 10^{-07}</td>
</tr>
<tr>
<td>17</td>
<td>MAPT-AS1</td>
<td>26</td>
<td>9.9 x 10^{-07}</td>
<td>rs111962225</td>
<td>3.30 x 10^{-07}</td>
</tr>
<tr>
<td>17</td>
<td>MAPT-IT1</td>
<td>5</td>
<td>9.9 x 10^{-07}</td>
<td>rs62056785</td>
<td>3.90 x 10^{-07}</td>
</tr>
<tr>
<td>17</td>
<td>NSF</td>
<td>69</td>
<td>9.9 x 10^{-07}</td>
<td>rs1378358</td>
<td>8.70 x 10^{-07}</td>
</tr>
<tr>
<td>20</td>
<td>ZGPAT</td>
<td>81</td>
<td>1.9 x 10^{-06}</td>
<td>rs4809327</td>
<td>6.10 x 10^{-08}</td>
</tr>
<tr>
<td>6</td>
<td>ZKSCAN4</td>
<td>19</td>
<td>1.9 x 10^{-06}</td>
<td>rs9986596</td>
<td>6.40 x 10^{-08}</td>
</tr>
</tbody>
</table>

Three of the genes with the strongest signal; CRHR1 (Corticotrophin-Releasing Hormone Receptor 1), KANSL1 (KAT8 Regulatory NSL Complex Subunit 1) and MAPT (Microtubule Associated Protein Tau) are all in the same 17q21 locus. Variants in this locus are the strongest known genetic determinants of early-onset asthma and are associated with frequent exacerbation\(^{(27)}\) and persistence of symptoms.\(^{(28)}\) Variants in CRHR1 and MAPT have both been shown to affect lung function response to corticosteroid.\(^{(29)}\) An integrative analysis combining results of a COPD GWAS with an expression quantitative trait locus analysis from lung tissue samples reported MAPT as the top candidate gene for causing COPD.\(^{(30)}\) KANSL1 has been associated with extremes of FEV\(_1\), with the strongest signal seen in never smokers.\(^{(31)}\)

NSF gene (N-Ethylmaleimide Sensitive Factor) also reached significance in chromosome 17. It contains rs553682853 which was discovered in our GWAS. NSF has been associated with FEV\(_1\) and FVC in UKBB previously.\(^{(21)}\)

A PheWAS was conducted on the top SNP in the ZGPAT gene (Zinc Finger CCCH-Type And G-Patch Domain Containing), p value 6.10 x 10^{-08}. PheWAS showed strong associations for rs4809327 with FEV\(_1\) percent predicted, FVC and wheeze in the last year (p value <0.5x10^{-8}) and has weaker associations with self-reported asthma (p value 5.6x10^{-5}). Literature review
found evidence from an epigenetic paper that CpG sites in ZGPAT have been shown to have an inverse causal effect on FEV$_1$ and COPD.$^{(32)}$

The top SNP in ZKSCAN4 on chromosome 6, rs9986596, has been associated with FEV$_1$ percent predicted, FVC and PEFR (all p value <0.5x10$^{-8}$) and asthma diagnosis, p value 2.7x10$^{-6}$.

With the exception of CRHR1, all genes discovered are known to be expressed in human lung tissue.$^{(20)}$ CRHR1 is found in mouse lung tissue and mutations in it can cause pulmonary alveolar haemorrhage, abnormal lung morphology, emphysema and respiratory distress.$^{(33)}$

**Discussion**

**Evidence before this study**

COPD is a common and heterogenous diseases, defined by airway obstructive spirometry.$^{(9)}$

The current diagnostic criteria of COPD excludes many individuals with smoking histories and a high burden of respiratory symptoms.$^{(34, 35)}$ Many of these excluded individuals have PRISm, and although they do not meet the diagnostic spirometric criteria for COPD they exhibit similar changes such as small airways disease, emphysema and gas trapping.$^{(4, 36, 37)}$

Little focus has been given to whether asthma contributes to PRISm, despite high rates of self-reported asthma in other PRISm cohorts (21% in COPDgene$^{(2)}$) and in this UKBB cohort reported asthma among PRISm was 18.4%.$^{(4)}$ There are many plausible ways asthma could contribute to both PRISm and airflow obstruction. Asthma causes small airways obstruction, CT changes of airways disease and gas trapping, and is a significant risk factor for COPD.$^{(38)}$

Research thus far has been produced by cohorts with small PRISm samples, with some excluding never smokers. This has limited generalizability of findings and prevented genetic variants from being discovered.
Added value of this study
The sample size and detailed data of the UKBiobank cohort enables reliable assessment of individuals with PRISm. Estimates of prevalence and associations of PRISm compared to normal and obstructive spirometry are well powered and generalizable. The increased sample size also allows for genetic discoveries.

PRISm in UKBiobank is common at 8%, and slightly more prevalent than stage II-IV obstruction. This is lower than many estimates, but is similar to another cohort that included never smokers.\(^{(3)}\) The importance of smoking as a risk factor was replicated in our analysis. It is unclear why some smokers develop PRISm and others COPD. The pack/year of those with COPD is higher but it does not fully explain the risk of PRISm given the transition between normal lung function, PRISm and COPD seen in longitudinal work and the non-smoking related genetic risk factors we found. BMI does not appear to be as clinically relevant as previously described, providing further evidence that PRISm is not simply due to extra-pulmonary restriction related to obesity.\(^{(3)}\) The high prevalence of cardiovascular disease and diabetes in PRISm even after adjustment for confounders is important. COPD may have a causal effect on extra-pulmonary disease, perhaps via systemic inflammation or oxidative stress,\(^{39}\) it is conceivable this could occur in PRISm too. Even if no casual pathway between PRISm and its co-morbidity is found, future studies to determine if screening for diabetes and cardiovascular disease in PRISm are warranted.

Our study is the first to utilise conditional analysis and gene-based analysis in PRISm and the first to successfully identify genetic variants. GWAS identified two SNPs associated with PRISm that are novel for lung function and disease. Conditional analysis and gene-based analysis found genetic variants with reported associations for wheeze, asthma, COPD, and inhaled steroid response in asthma and COPD. 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.\textsuperscript{(6, 7)} Our findings that PRISm and COPD have shared genetic risk factors could partially explain the transition between them. Variants pre-disposing to asthma, especially as a child, suggest that early life factors and disease could lead to a trajectory that includes PRISm.

Limitations
UKBiobank has only conducted cross sectional lung function tests, so we are unable to add to longitudinal follow up, but our genetic findings contribute to our understanding of why PRISm can progress to COPD. Post-bronchodilator spirometry has not been performed in UKBiobank which is the gold standard for COPD diagnosis. However, post-bronchodilator spirometry is not used for diagnosis of PRISm, and the effect of bronchodilation in PRISm is not known. We used FEV\textsubscript{1} predicted <80\% and FEV\textsubscript{1}/FVC <0.7 to define obstructive spirometry. Obstructive spirometry is not the same as COPD which remains a clinical diagnosis assuming spirometric criteria are fulfilled. Using lower limits of normal (LLN), FEV\textsubscript{1}/FVC ratio and post-bronchodilator spirometry may have reduced numbers classified as having obstructive spirometry. By eliminating stage 1 obstruction from the analysis, we are likely to have eliminated many individuals whose FEV\textsubscript{1}/FVC ratio was below 0.7, but within the LLN. Many of our conclusions are drawn from comparing PRISm to normal spirometry, which would not be affected by bronchodilation or the use of LLN FEV\textsubscript{1}/FVC. We do not have access to 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 UKBiobank so are unlikely to influence results. We do not have access to a second cohort for a replication analysis of our genetic discoveries, which potentially means our results are limited by “Winners Curse”.\textsuperscript{(40)}
Future Research
Research assessing underlying structural and functional lung changes of PRISm is warranted. Particularly the frequency and severity of small airways dysfunction, perhaps via FEF_{25-75} or measures of resistance from forced oscillation testing, may provide further insight into underlying pathology in PRISm and contribute to COPD prediction. Small airways obstruction may be amenable to treatment with bronchodilators, reversing spirometry to normal or preventing obstruction. Given our findings, research into the initiation and optimisation of inhaled steroids in PRISm may be warranted, especially in those with pre-existing asthma. Lung volume measurements may not produce targets for currently available treatments but could aid risk stratification and understanding of PRISm pathogenesis. Trials examining the utility of screening for diabetes and cardiovascular risk could help reduce morbidity.

Conclusion
PRISm affects 8% of adults, reporting a high burden of breathlessness and a ~40% cardiovascular comorbidity rate. Genetic variants that increase the risk of obstructive lung diseases in childhood and adults are strongly associated with PRISm. This provides novel biological mechanisms for the observed transition between PRISm and obstructive spirometry and has implications for prevention, diagnosis, and treatment of PRISm.

Acknowledgements
This research was conducted using UKBiobank resource (project number 55521). We would like to thank all UKBiobank participants and all staff involved in UKBiobank.

References


8. https://www.ukbiobank.ac.uk/.


26. NHGRI-EBI. [https://www.ebi.ac.uk/gwas/](https://www.ebi.ac.uk/gwas/).


Online Data Supplement

Epidemiology and Genetics of Preserved Ratio Impaired SpiroMetry (PRISm): An Analysis of UK Biobank

Daniel H Higbee, Raquel Granell PhD, George Davey Smith FRS, James W Dodd
Contents
Appendix 1. UKBiobank data fields used
Appendix 2. Flow chart of analysis
Appendix 3. Details of asthma and COPD diagnosis in UKBiobank and results of genetic analysis done excluded those with asthma and COPD
Appendix 4. Supplementary tables

Appendix 1. UKBiobank data fields used

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>UKBiobank data field (and link)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forced expiratory volume in 1-second, predicted percentage</td>
<td>20154 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20154">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20154</a>]</td>
</tr>
<tr>
<td>Forced expiratory volume in 1-second (FEV1), Best measure</td>
<td>2010 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20150">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20150</a>]</td>
</tr>
<tr>
<td>Forced vital capacity (FVC), Best measure</td>
<td>20150 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20151">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20151</a>]</td>
</tr>
<tr>
<td>Smoking status</td>
<td>20116 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20116">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20116</a>]</td>
</tr>
<tr>
<td>Pack years of smoking</td>
<td>20161 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20161">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20161</a>]</td>
</tr>
<tr>
<td>Year of birth</td>
<td>34 [<a href="http://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=34">http://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=34</a>]</td>
</tr>
<tr>
<td>Age at recruitment</td>
<td>21102 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21022">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21022</a>]</td>
</tr>
<tr>
<td>Sex</td>
<td>31 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=31">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=31</a>]</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td>21001 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21001">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=21001</a>]</td>
</tr>
<tr>
<td>Diabetes diagnosed by a doctor</td>
<td>2443 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2443">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2443</a>]</td>
</tr>
<tr>
<td>Self-reported asthma</td>
<td>2002 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002</a>]</td>
</tr>
<tr>
<td>Self-reported COPD</td>
<td>2002 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002</a>]</td>
</tr>
<tr>
<td>Doctor diagnosed asthma</td>
<td>22127 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22127">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22127</a>]</td>
</tr>
<tr>
<td>Doctor diagnosed COPD</td>
<td>22130 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22130">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22130</a>]</td>
</tr>
<tr>
<td>Diagnosis – ICD 10</td>
<td>41270 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41270">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41270</a>]</td>
</tr>
<tr>
<td>Diagnosis secondary ICD10</td>
<td>41204 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41204">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=41204</a>]</td>
</tr>
<tr>
<td>Vascular/heart problems diagnosed by doctor</td>
<td>6150 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6150">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6150</a>]</td>
</tr>
<tr>
<td>Shortness of breath walking on level ground</td>
<td>4717 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4717">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4717</a>]</td>
</tr>
<tr>
<td>Weight change compared with 1 year ago</td>
<td>2306 [<a href="http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2306">http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2306</a>]</td>
</tr>
</tbody>
</table>
Appendix 2. Flow chart of analysis

Appendix 3. Details of asthma and COPD diagnosis in UKBiobank and results of genetic analysis done excluded those with asthma and COPD

Asthma and COPD Diagnosis

Table E2. Asthma and COPD diagnosis in UKBiobank

<table>
<thead>
<tr>
<th>Condition</th>
<th>PRISm</th>
<th>Control</th>
<th>Obstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self reported asthma*</td>
<td>2,262 (18.49%)</td>
<td>12,101 (10.45%)</td>
<td>4,209 (35.36%)</td>
</tr>
<tr>
<td>Self reported COPD</td>
<td>608 (4.97%)</td>
<td>92 (0.08%)</td>
<td>536 (4.51%)</td>
</tr>
<tr>
<td>ICD codes asthma†</td>
<td>1562 (12.77%)</td>
<td>6,299 (5.44%)</td>
<td>2,585 (21.72%)</td>
</tr>
<tr>
<td>ICD codes COPD</td>
<td>626 (5.12%)</td>
<td>1,019 (0.88%)</td>
<td>2,527 (21.23%)</td>
</tr>
</tbody>
</table>

*We reported asthma diagnosis using UKBiobank variable 20002, “Non-cancer illness code, self-reported”. Participants stated their illnesses, if they were uncertain of the type of illness, it was described to the interviewer (a trained nurse) who attempted to place it in the coding tree.

†ICD codes were using linked hospital ICD 10 codes.

Doctor diagnosed asthma or COPD uses UKBiobank variable 22127 and 22130 respectively. Participants were asked "Has a doctor ever told you that you have had any of the conditions below?" and asthma and COPD were listed.
Exclusion based on criteria strongly influenced by factors liable to cause an outcome can generate collider bias. Therefore, we present the results of genetic analysis after exclusion of participants with asthma and COPD here in the supplementary information.

Objective Two - GWAS

Exclusion of participants with asthma and COPD left 9408 cases and 109724 controls for analysis. GWAS produced no SNPs at \( p \) value <5x10\(^{-8}\), but slightly reducing the threshold too \( p \) value <5x10\(^{-7}\) produced 356 SNPs, pre-dominantly on chromosome 17. After clumping SNP rs189777036 remained which is in complete linkage disequilibrium (\( D' = 1.0, r^2 =1.0 \)) with rs378392, therefore the same signal was found in the GWAS despite exclusion of those with pre-existing airways disease.\(^{(40)}\)

Objective Two - Conditional analysis

Following exclusion of participants with asthma and COPD, repeat conditional analysis with a threshold of \( p \) value <5x10\(^{-7}\) found rs886249. This SNP is in near complete linkage disequilibrium with rs1991556 (\( D' = 1.0, r^2 = 0.94 \)), which was found prior to exclusion of those with asthma and COPD. Therefore, the same signal was found in the GWAS despite exclusion of those with pre-existing airways disease.

Objective Two - Gene-based analysis

After excluding all participants with asthma and COPD and re-running the gene-based analysis, all 6 genes on chromosome 17 still reached the significance threshold with the same top SNPs reported. No Top SNP reached \( p \)-value <5\(^{-08}\).

Appendix 4.

Table E3. SNPs found via all Objective Two methods

<table>
<thead>
<tr>
<th>SNP</th>
<th>Discovery</th>
<th>Chr:Pos</th>
<th>Gene</th>
<th>Function</th>
<th>EA</th>
<th>OA</th>
<th>EAF</th>
<th>beta</th>
<th>SE</th>
<th>pval</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs378392</td>
<td>GWAS</td>
<td>17:43687542</td>
<td>RPS26P8</td>
<td>Intron</td>
<td>C</td>
<td>G</td>
<td>0.72</td>
<td>-0.0078</td>
<td>0.001</td>
<td>1.4(^{-8})</td>
</tr>
<tr>
<td>rs553682853</td>
<td>GWAS</td>
<td>17:44759766</td>
<td>NSF</td>
<td>Intron</td>
<td>G</td>
<td>A</td>
<td>0.84</td>
<td>-0.0098</td>
<td>0.001</td>
<td>4.1(^{-8})</td>
</tr>
<tr>
<td>rs1991556</td>
<td>Conditional</td>
<td>17:44083402</td>
<td>MAPT</td>
<td>Intron</td>
<td>G</td>
<td>A*</td>
<td>0.77</td>
<td>-0.0917</td>
<td>0.016</td>
<td>1.0(^{-8})</td>
</tr>
<tr>
<td>rs7213474</td>
<td>Gene-based</td>
<td>17:79962718</td>
<td>ASPSCR1</td>
<td>Intron</td>
<td>T*</td>
<td>C*</td>
<td>0.33*</td>
<td>0.0065*</td>
<td>0.001*</td>
<td>1.3(^{-8})</td>
</tr>
</tbody>
</table>

*Extracted from GWAS as not supplied by discovery method

eQTL = expression quantitative trait loci

CHR – Chromosome. Pos – Position. Build GRCh37
Figure 1. Each point represents one of the 755,702 SNPs that reached p value <0.05 in our GWAS.