Peer reviewed version

Link to published version (if available):
10.1016/S2213-2600(21)00369-6

Link to publication record on the Bristol Research Portal
PDF-document

University of Bristol – Bristol Research Portal
General rights

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

Daniel H Higbee MBBS[1,2], Raquel Granell PhD[1], George Davey Smith FRS[1], James W Dodd PhD[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

james.dodd@bristol.ac.uk

Tel +44117 4142012

Fax +44117 4149496

ORCID 0000-00034805-5759

Research in context

Evidence before this study

All articles relating to PRISm were reviewed and identified using pubmed search terms "preserved ratio impaired spirometry", "PRISm", "PRISm spirometry", "restrictive spirometry" and "non-specific spirometry" ending January 2021.

There remains uncertainty around the epidemiology, clinical significance, and long-term impact of PRISm in the general population. Estimates of prevalence varies from 4% to 48%, partly due to heterogenous populations studied. Available cross-sectional studies suggest a possible association between PRISm and respiratory symptoms, increased healthcare utilisation, co-morbidities (including obesity, diabetes, cardiac disease) and increased overall mortality. Longitudinal data are limited but suggest that over 5 years, up to 50% of people with PRISm may transition to COPD but also that 15% may return to 'normal' spirometry.

The two largest cohorts reporting PRISm are COPDGene and the Rotterdam studies which had ~1200 and ~350 cases respectively, decreasing to just ~300 and <100 available longitudinally. These relatively small samples limit generalisability of findings, and COPDGene was restricted to ex/current smokers. As smoking has a strong association with PRISm and related co-morbidity, this confers a risk of selection bias.

Added value of this study

This is the largest, most comprehensive longitudinal analysis of PRISm to date. The cohort is drawn from >500,000 adults in UKBiobank reducing the impact of selection bias. The sample size also means that detailed sensitivity and subgroups analyses have been possible, advancing our understanding about risk factors and underlying mechanisms.

Our analysis shows that PRISm has a general population prevalence of 11% (38,639/351,874) and is strongly associated with breathlessness and cardiovascular disease. PRISm appears to be a distinct lung function trait and not simply a result of increased BMI, underlying asthma or smoking. While obesity is associated with PRISm (OR 2·4, (2.26 – 2.55)), concerns that PRISm is pre-dominantly due to extra-thoracic compression seem unlikely given mean difference in BMI vs. control is just 1.8kg/m^2 and 62% (24,091/38,639) of PRISm have a BMI <30. Sensitivity subgroups analysis (men, women, never smokers, ever smokers, non-overweight and those without asthma) show an association between PRISm and current smoking, female gender, obesity, and asthma diagnosis. This contrasts with previous work suggesting male gender and age are the key risks for PRISm. Additionally, all subgroups analyses found the associations with shortness of breath and co-morbidity remained.

Contrary to previous work, our study shows that only ~12% (241/1,973) of adults with PRISm go on to develop COPD over the next 9 years, with the majority (50% (987/1,973) reverting to normal spirometry.

This study confirms that PRISm is associated with an increased all-cause mortality, adjusted hazard ratio 1·61 (1·53 - 1·69).

Implications of all available evidence

PRISm is common lung function trait which is clearly associated with respiratory symptoms, cardiovascular co-morbidities and increased risk of death. 12% (241/1,973) of adults with PRISm in this cohort developed COPD, but for many PRISm appears to be a reversible state and therefore a potential target for therapeutic intervention.

Public health measures to reduce smoking and BMI may reduce the prevalence of PRISm.

Future epidemiological studies should be mindful of the risks of bias, including selection bias, small sample bias, and regression to the mean. Large populations cohorts such as UKBioank now mean genetic studies are possible and could be used to generate evidence about underlying pathogenesis and causality of co-morbid associations. Studies to determine how best to prevent and treat PRISm are warranted.

Abstract

Background

Preserved ratio impaired spirometry (PRISm) is defined as FEV_1 <80% predicted and FEV_1/FVC ≥0·70. It has been suggested that PRISm is associated with respiratory symptoms and is a precursor of COPD. However, this is based on relatively small selective cohorts with limited follow up. Our objective was to determine the prevalence, risk factors, clinical implications, and mortality of PRISm in a large adult general population.

Methods

UKBiobank provides a cohort of well phenotyped adults. Multivariable regression was used to determined risk factors for PRISm and associated co-morbidities. Longitudinal analysis of PRISm over time and risk factors for transition to COPD was performed in addition to 12 year survival analysis.

Findings

We found an 11% prevalence of PRISm (38,639/351,874 UKBB participants). PRISm is strongly associated with obesity, current smoking and asthma diagnosis. PRISm is associated with increased risk of breathlessness (Odds Ratio 2·0, 95%CI 1·91 - 2·14) and cardiovascular disease (1·71, 1·64 – 1·83 for heart attack) after adjustment. Longitudinal analysis showed 12% of people with PRISm go on to develop airflow obstruction consistent with COPD. PRISm is associated with an increased all-cause mortality, hazard ratio of 1·61 (1·53 - 1·69) vs controls after adjustment.

Interpretation

This large general adult population cohort shows that PRISm is common, clearly associated with respiratory symptoms, cardiovascular disease and increased risk of death. 12% (241/1,973) of people with PRISm progress to obstructive spirometry consistent with COPD, but for many PRISm appears to be a reversible and potentially treatable trait. Further studies to determine prevention and treatment of PRISm are warranted.

Funding

Introduction

Preserved Ratio Impaired Spirometry (PRISm), also referred to as 'restrictive pattern' or 'unclassified' spirometry, is defined as a Forced Expiratory Volume in one second (FEV_1) <80% predicted, despite a normal or preserved FEV_1/Forced Vital Capacity (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.[1-4] Clinical interest in PRISm come from data which suggest that over 5 years, up to 50% may transition to COPD but that 15% return to 'normal' spirometry.[3,5] 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.[6] Imaging studies suggest that PRISm may be associated with a degree of airway disease and emphysema which may affect progression to COPD.[7,8] Analyses from other 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.[3,5,9-11]

Definitive epidemiological understanding of PRISm has been limited by cohorts with relatively small patient numbers, rarely containing >1000 cases.[3,5] Some cohorts have used selected populations too e.g. only smokers, which limits generalisability.[2,5,12]. Duration of follow up is also often limited to ≤5 years.[3,5,13] This may introduce selection bias and confounding, limiting conclusions.[14]

Our first objective was to use the UKBiobank (UKBB) to examine a large adult general population to determine PRISm prevalence, risk factors and associated symptoms and co-morbidity.[15] 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 UKBB and broad recruitment based on age and the inclusion of non-smokers, reduces selection bias, this increases power and improves generalizability. The long follow up of UKBiobank participants compared to other cohorts also allows for accurate estimation of PRISm trajectories over time and survival analysis.

Methods
Baseline

UKBB includes 502,543 individuals aged between 40 and 69 at recruitment across the UK.[15] Participants were identified from the NHS register and were invited to assessment appointments by letter. No weighting mechanism for recruitment was used. Initial assessment took place from 2006-2010 – these data were used as the baseline timepoint. All participants were asked to perform pre-bronchodilator spirometry. See appendix 1 for full details. Only pre-bronchodilator spirometry was available, although medications were not withheld. We used previously derived variables of quality-controlled spirometry for "best measure" FEV_1 and FVC, that 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.[16]

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.[17] 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). We then examined clinically relevant correlates of PRISm (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 analysis. Statistical analysis was performed using Stata 15.[18]

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 was used, see appendix 1. 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. We 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. We determined what proportion of participants would be expected to revert to control due to regression to the mean using Stata package rtmci.[19] 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 phenomena whereby outlier results are more likely to be followed by results closer to the mean due to standard deviation in testing any complex trait, rather than due to a causal or pathophysiological effect.[20]

Sensitivity analyses were performed stratifying the sample by sex, BMI, asthma diagnosis and BMI, repeating the analysis with lower limit of normal definition of spirometry criteria, and using GOLD II-IV as the definition of obstruction. See supplement for details and results.

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

Survival analysis

UKBiobank obtained dates of death from NHS Digital and NHS Central registry. Death records up to February 2018 were available allowing us to perform survival analysis covering a period of 12 years. We 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. Survival analysis was performed in Stata.

Role of the Funder

Results

Prevalence of PRISm

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 for analysis at baseline (see supplementary information appendix 2). Table 1 shows a prevalence of 11·0% for PRISm (38,639/351,874) and 15.8% for stage I-IV airflow obstruction (55,592/351,874).

Risk factors for PRISm

55% of PRISm and controls were female (21,388/38,639 and 143,289/257,643) vs. 44% (24,570/55,592) with airflow obstruction. Current smokers were more common in PRISm

than controls (12% (4,787/38,639) vs 7% (20,458/257,643), p-value <0·0001), as was smoking pack/years of ever smokers (23 vs 16 pack/years, p-value <0·0001). Doctor diagnosed asthma was more common in PRISm than in controls (16% (1,436/8,472) vs 9%,(7,078/71,281) p-value <0·0001). BMI was shown to have a non-linear association with PRISm, violating an assumption of logistic regression. Therefore, for regression analysis BMI was categorised into three clinically relevant groups; Not overweight (BMI <25), Overweight (BMI ≥25 and <30) and Obese (BMI ≥30).

Multivariable logistic regression examining age, sex, BMI categories, smoking status (never/ex/current) and doctor diagnosed asthma association 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 and effect 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). 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 1. Forest plot showing factors associated with PRISm vs control spirometry

Table 1. 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%	7·0%	<0·0001	15·1%	<0·0001
Doctor diagnosed asthma	16.8%	9.9%	<0·0001	27·4%	<0·0001
Doctor diagnosed COPD	1·7%	0.3%	<0·0001	6·7%	<0·0001
Diabetes (%)	8·6%	3·7%	<0·0001	4·7%	<0·0001
Heart attack (%)	3·7%	1·5%	<0·0001	3·1%	0·103
Angina (%)	4·6%	2·2%	<0·0001	3·9%	<0·0001
High blood pressure (%)	33·3%	24·3%	<0·0001	28·8%	<0·0001
Stroke (%)	2·0%	1·0%	<0·0001	1·9%	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

PRISm symptoms and co-morbidities

There was a higher prevalence of breathlessness in PRISm 17% (2,219/12,506) vs 6% (6,120/86,389) 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 (1·91 – 2·14, p-value <0·0001). Diabetes was more common in PRISm than controls or airflow obstruction (8·6% (3,350/38,520) vs 3·7% (9,660/247,531) vs 4·7% (2,650/52,806) 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 at least double the prevalence of angina (4·6% (1,790/38,550) vs 2·2% (5,711/257,324), p-value <0·0001), heart attack (3·3% (1,300/38,550) vs 1·5% (3,901/257,324), p-value <0·0001) and stroke (2·0% (772/38,550) vs 1·0% (2,712/254,612), 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.

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. 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. The mean FEV_1 for the cohort at follow up was higher vs. baseline (3·0 litres vs 2·8 litres). The mean annual FEV_1 decline for individual participants between baseline and follow

up of -28mls/yr. Prevalence of PRISm at follow up was lower than at baseline PRISm (7·1% (1,746/24,404) vs 11·0% (38,639/351,874)) but similar for airflow obstruction (15·5 (3,792/24,404) vs 15·8% (55,592/351,874), p value 0.28). See supplement, Table E2.

PRISm trajectories

The median time between baseline and follow up was 9.0 years. 1,973 participants with PRISm at baseline were included in follow up (Figure 1). 37·7% (745/1,973) had persistent PRISm, 50% (987/1,973) reverted to normal control spirometry and 12·3% (241/1,973) transitioned to airflow obstruction. More participants with PRISm at baseline transitioned to a different lung function state (62·4% (1,228/1,973)) vs controls (12·2% (2,283/19,195)) and airflow obstruction (33·7% (1,098/3,268)). 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.

Persistent PRISm vs. PRISm to control trajectories had higher mean BMI (29 vs 27), median pack/years (21 vs 18), more diabetes (14% (95/649) vs 9% (87/984)) and shortness of breath (15% (105/617) vs 7% (68/902)). Persistent PRISm vs. PRISm to COPD trajectories were younger, mean age (62 vs 65 yrs) and had a higher mean BMI (28 vs 26). For full results see table E3.

Figure 2. Lung function trajectories from Baseline to Follow up

Persistent PRISm vs PRISm to control trajectories

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 (Table 2). Persistent PRISm had a high baseline and mean change in BMI (0·8 kg/m^2 and 0·24 respectively), whilst those that transitioned to control had a mean change of -0·48. There was a clinically significant difference in smoking, with higher pack/years (20 vs 16, p-value <0·0001) at baseline for persistent PRISm.

Multivariable multinomial logistic regression analysis revealed strong evidence of negative association between doctor diagnosed asthma with PRISm changing to control vs persistent PRISm (Relative Risk Ratio 0·67 (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 (RRR 0·86, 0·81 – 0·91, p-value <0·0001).

Table 2. Baseline demographics of participants by PRISm trajectory at follow up

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

Persistent PRISm vs PRISm progressing to airflow obstruction trajectories

At baseline participants with persistent PRISm vs. PRISm to airflow obstruction were younger (53·6 vs 56·7 years, p-value <0·0001). Although there were more current smokers and less 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 (28·8 vs 27·3, p-value <0·0001) and higher FEV$_1$/FVC ratios (76% vs 73%, p-value <0·0001) vs PRISm to airflow obstruction at baseline. Change in BMI between differed between those with persistent PRISm (mean change in BMI of 0·24) vs, PRISm to airflow obstruction (mean change of -0·58).

Multivariable multinomial regression analysis showed change from PRISm to airflow obstruction was strongly positively associated with increased age, RRR 1·07 (95%CI 1·04 – 1·10) and a doctor diagnosis of asthma RRR 1.91 (1.17 – 3.13). Change in BMI (per mg/kg^2) increase showed a strong negative (RRR 0·86, 0·79 – 0·95, p-value 0.0022) of PRISm changing to airflow obstruction vs persistent PRISm after adjustment.

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, see supplementary information.

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 change from control to airflow obstruction. For full results see supplementary information.

Survival Analysis

12,810 deaths were recorded during follow up, 2.8% (7,202/250,441) of controls, 5.0% (1,911/36,728) of PRISm and 6.7% (3,697/51,895) 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 (p values <0·0001 between groups).

Figure 3. Kaplan-Meier survival estimates based on spirometric group at baseline.

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.

Discussion

This study in UKBB shows that PRISm is common at 11% (38,639/351,874), this is lower than some estimates in smoker, but similar to a cohort which included never smokers.[3] Longitudinal analysis showed 62·1% (1,228/1,973) of PRISm changed to different lung function states over 8 years. After regression to the mean was considered, rates of persistent PRISm and reverting from PRISm back to control were similar at ~40%. We found considerably lower rates of progression from PRISm to airflow obstruction than has previously been reported.[3,5] This is likely due to recruitment based on smoking and selection bias. 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. We believe that this difference is unlikely to be explained solely by extra-thoracic restriction as 62% (24,901/38,639) of those with PRISm had BMI <30, and restricting analysis to only those with BMI <25 showed that 7·6% (8,823/38,639) 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.[22]

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.[23] 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.[24,25]

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.[26]

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.[27,28] We 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.

Limitations
UKBiobank collects only pre-bronchodilator spirometry, although bronchodilator medication is not withheld if prescribed, post bronchodilator spirometry was not available. Post-bronchodilator spirometry is not required for diagnosis of PRISm, but differences between pre and post-bronchodilator spirometry have been reported for PRISm and airflow obstruction.[4,29] Post-bronchodilator spirometry may reduce numbers classified as having PRISm and airflow obstruction spirometry in UKBB, but by performing sensitivity analysis using LLN criteria, and classifying airflow obstruction as GOLD II-IV we are likely to 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, we note the mean annual decline in FEV_1 is similar to a normal population, and the prevalence of airflow obstruction was the same at both time points. 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% (857/24,404). 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. UKBiobank 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.[30] Additionally, due to its large size and heterogeneity of exposure methods, associations between exposures and health outcomes are generalizable to other populations.[31] We had 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. We used spirometry to define airflow obstruction. Airflow obstruction is not the same as COPD which remains a clinical diagnosis assuming spirometric criteria are fulfilled. We do not have access to 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 UKBB so are unlikely to influence results. Our sample was 100% European ancestry. Therefore, these results may not be generalisable to non-European ancestral populations. We do not know if these results are generalizable to people under the age of 40. Finally, traditional observational epidemiological analysis such as this could be affected by collider bias.[32] 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 transition to other lung function states. If this did occur, then the observed associations may be induced by the statistical model.

<u>Future Research</u>
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,[2] but discovery of genetic markers for PRISm in larger cohorts could help explain underlying pathological mechanisms for PRISm and be used for Mendelian Randomization studies to determine if the observed association with co-morbidity is causal.

<u>Conclusion</u>

This analysis of UKBB shows a general adult population prevalence of PRISm of 11% (38,639/351,874). 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% (987/1,973) returning to normal lung function and 12% (241/1,973) 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.

Contribution

JD & DH conceived the study and initial draft, all authors made substantial contributions to

the analysis, drafting and final approval of the paper. DH and RG verify the underlying data.

References

1.   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.

2.   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-.

3.   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.

4.   Schwartz A, Arnold N, Skinner B, et al. Preserved Ratio Impaired Spirometry in a Spirometry Database. *Respiratory care* 2021; **66**(1): 58-65.

5.   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.

6.   WHO T. https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/. 2020.

7.   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.

8.   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.

9.   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).

10.   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.

11.   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.

12.   Barnes PJ, Vestbo J, Calverley PM. The Pressing Need to Redefine "COPD". *Chronic Obstr Pulm Dis* 2019; **6**(5): 380-3.

13.   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.

14.   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.

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

16.   Lytras T. 2020. https://github.com/thlytras/rspiro2020).

17.   Global Iniative for Chronic Obstructive Pulmonary Lung Disease, 2019.

18.   STATACorp. Stata Statistical Software: Release 15. 2017.

19.   Linden A. RTMCI: Stata module to estimate regression to the mean effects with confidence intervals. In: S457757 SSC, editor.; 2013.

20.   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.

21.   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.

22.   Maclay JD, MacNee W. Cardiovascular disease in COPD: mechanisms. *Chest* 2013; **143**(3): 798-807.

23.   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.

24.   Burney PGJ, Hooper R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax* 2011; **66**(1): 49.

25.   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.

26.   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.

27.   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.

28.   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.

29.   Mahmoud O, Dudbridge F, Smith GD, Munafo M, Tilling K. Slope-Hunter: A robust method for index-event bias correction in genome-wide association studies of subsequent traits. *bioRxiv* 2020: 2020.01.31.928077.

30.	Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLoS genetics* 2017; **13**(10): e1006944-e.

Cohort study of Preserved Ratio Impaired Spirometry (PRISm) in the UKBiobank

Contents

Appendix 1. UK Biobank variables

Table E1. UKBiobank variables used

Variable	ID number	% With variable at Phase 1	Weblink
FEV_1 best measure	20150	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20150
FEV_1	3063	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=3063
FVC best measure	20151	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20151
FVC	3062	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=3062
Acceptability spirometry	3061	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=3061
Age	21003	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=21003
Sex	31	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=31
Standing Height	50	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=50
BMI	21001	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=21001

Ethnic background	2100	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=21000
Smoking status	20166	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20116
Pack years smoking	20161	31	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20161
Trunk fat	23128	98	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=23128
Trunk fat percentage	23127	98	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=23127
Short of breath walking on ground level	4717	33	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=4717
Doctor diagnosed asthma	22127	26	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=22127
Doctor diagnosed COPD	22130	26	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=22130
Doctor diagnosed diabetes	2443	99	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=2443
Cardiovascular outcomes	6150	99	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=6150
Phase dates	53	100	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=53
Date of death	40000	NA	https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=40000

Shortness of breath, diabetes, smoking history and cardiovascular diseases were self-reported by patients using a touchscreen questionnaire at time of attending recruitment centre. Doctor diagnosis of asthma/COPD was reported using an online questionnaire after recruitment. Trunk fat mass and percentage were calculated using impedance measurements at recruitment centres.

Spirometry

All participants were requested to perform pre-bronchodilator spirometry, using Vitalograph Pneumotrac 6800. The participants were asked to record two to three blows (lasting for at least 6 seconds) within a period of about 6 minutes. The computer compared the reproducibility of the first two blows and, if acceptable (defined as a <5% difference in FVC and FEV_1), indicated that the third blow was not required. The highest measures of FEV_1 and FVC from acceptable blows was used.

Appendix 2. Flow chart of participants used in analysis

Figure E1. Flow chart of participants used in analysis

Appendix 3. Comparison of baseline and follow up participants

Table E2. Baseline demographics of baseline and follow up participants

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

P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – interquartile range

Appendix 4. Cross-sectional Phase 2 characteristics of participants by PRISm trajectory

Table E3. Phase 2 demographics of participants by PRISm trajectory

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 (7)	<0·0001
BMI (kg/m^2) Mean (SD)	29·0 (5)	27·5 (4)	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 (%)	15%	7%	<0·0001	13%	0·44
Diabetes (%)	13%	9%	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

Appendix 5. Demographic tables using GOLD Stage II-IV definition of airflow obstruction

Only pre-bronchodilator spirometry is available in UKBiobank. Although post-bronchodilator spirometry is not required for diagnosis of PRISm, differences between pre and post-bronchodilator spirometry have been reported for PRISm and airflow obstruction.[4,29] Therefore, we repeated the cross-sectional and longitudinal analysis with more stringent criteria, defining airflow obstruction as GOLD II-IV, and also by using lower limit of normal criteria for PRISm, normal and obstructive spirometry. Additionally, we repeated the analysis only in those that do not have a doctor diagnosis of asthma, to remove all those likely to have a large degree of bronchodilator reversibility.

Table E4. Baseline demographics of participants with PRISm, control and GOLD Stage II-IV airflow obstruction

Demographic at baseline	PRISm N = 38639	Control Spirometry N= 257643	P value* PRISm vs Control	Stage II-IV Obstruction N = 29656	P Value* PRISM vs II-IV Obstruction
Age (Years) Mean (SD)	56 (7)	55 (7)	<0·0001	59 (7)	<0·0001
BMI (kg/m^2) Mean (SD)	29·1 (5)	27·2 (4)	<0·0001	27·3 (4)	<0·0001
Female (%)	55%	55%	0·68	44%	<0·0001
FEV$_1$ % predicted Median (IQR)	74% (68 – 77)	98% (90 – 106)	<0·0001	68% (58 – 74)	<0·0001
FVC % predicted Mean (SD)	76% (71 – 81)	99% (91 – 108)	<0·0001	83% (75 – 90)	<0·0001
FEV$_1$/FVC Median (IQR)	0·75 (0·72 – 0·78)	0·78 (0·75 – 0·81)	<0·0001	0·61 (0·58 – 0·67)	<0·0001
Never smoker (%)	51%	56%	<0·0001	36%	<0·0001
Ex-smoker (%)	36%	35%	0·0002	40%	<0·0001
Current smoker (%)	12%	7%	<0·0001	23%	<0·0001
Pack/years Median (IQR) †	23 (13 – 26)	16 (8 – 27)	<0·0001	30 (17 – 44)	<0·0001
SOB walking on ground (%)	17·7%	7·0%	<0·0001	21·9%	<0·0001
Doctor diagnosed asthma	16%	9%	<0·0001	35%	<0·0001
Doctor diagnosed COPD	1·7%	0·3%	<0·0001	12·0%	<0·0001
Diabetes (%)	8·6%	3·7%	<0·0001	6·3%	<0·0001
Heart attack (%)	3·7%	1·5%	<0·0001	4·0%	<0·0001
Angina (%)	4·6%	2·2%	<0·0001	5·0%	0·005
High blood pressure (%)	33·3%	24·3%	<0·0001	32·4%	0·02
Stroke (%)	2·0%	1·0%	<0·0001	2·4%	<0·0001

*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

Comparing Table 1 and Table E4 the prevalence of shortness of breath, heart attack, angina and stroke is higher in the airflow obstruction group than PRISm when using GOLD II-IV as definition or airflow obstruction.

Table E5. Baseline demographics of participants by PRISm trajectory at follow up using GOLD Stage II-IV airflow obstruction

Demographic at baseline	PRISm to PRISm N = 745	PRISm to Control Spirometry N = 987	P value*	PRISm to Stage II-IV Obstruction N = 185	P value†
Age (Years) Mean (SD)	53·6 (7)	53·8 (7)	0·56	56·6 (7)	<0·0001
BMI (kg/m^2) Mean (SD)	28·8 (5)	28·0 (4)	0·0011	27·4 (4)	0·0015
Female (%)	56%	52%	0·12	58·9%	0·43
FEV$_1$ % predicted Median (IQR)	74% (69 – 77)	76% (71 – 78)	0·0044	74% (70 – 77)	0·19
FVC % predicted Median (IQR)	76% (71 – 80)	77% (77 – 81)	0·0023	79% (73 – 83)	0·0055
FEV$_1$/FVC Median (IQR)	0·76 (0·73 – 0·79)	0·76 (0·73 – 0·79)	0·76	0·73 (0·71 – 0·75)	<0·0001
Never smoker (%)	57%	59%	0·41	49%	0·034
Ex-smoker (%)	34%	34%	0·99	37%	0·45
Current smoker (%)	8%	6%	0·12	14%	0·017
Pack/years Median (IQR) **	20 (12 – 34)	196 (8 – 26)	<0·0001	23 (8 – 34)	0·26
SOB walking on ground (%)	9%	7%	0·28	18%	0·062
Doctor diagnosed asthma	18%	13%	0·017	23%	0·21
Doctor diagnosed COPD	1%	1%	0·43	1%	0·78

*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. ** For ex and current smokers only. P-values Calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – Standard Deviation. IQR – Inter Quartile Range

Comparing Table 2 and Table E5 there is no change in interpretation of the demographic differences between the PRISm trajectories when using GOLD II-IV as definition of airflow obstruction.

Appendix 6. Demographic tables, only those that do not have doctor diagnosis of asthma

Table E6. **Phase 1 characteristics of participants with PRISm, control spirometry and airflow obstruction – only those that do not have asthma**

Demographic at baseline	PRISm N = 7036	Control Spirometry N = 64203	P value PRISm vs control	Stage I-IV Airflow Obstruction N = 9011	P value. PRISm vs I-IV airflow obstruction
Age (Years) Mean (SD)	55·9 (7)	55·8 (7)	0·33	59·0 (7)	<0·0001
BMI (kg/m²) Mean (SD)	28·3 (5)	26·2 (4)	<0·0001	26·0 (4)	<0·0001
Female (%)	56·6	56·0	0·30	43·0	<0·0001
FEV$_1$ % predicted Median (IQR)	75 (70 – 78)	99 (91 – 107)	<0·0001	83 (73 – 93)	<0·0001
FVC % predicted Median (IQR)	77 (72 – 81)	100 (92 – 108)	<0·0001	99 (87 – 109)	<0·0001
FEV$_1$/FVC Median (IQR)	76 (73 – 79)	78 (75 – 81)	<0·0001	67 (63 – 69)	<0·0001
Never smoker (%)	54·6%	59·3%	<0·0001	46·0%	<0·0001
Ex-smoker (%)	35·5%	34·8%	0·27	42·0%	<0·0001
Current smoker (%)	8·3%	5·8%	<0·0001	13·1%	<0·0001
Pack/years Median (IQR)*	20 (11 – 33)	14 (7 – 24)	<0·0001	22·5 (12 – 37)	0·0003
SOB walking on ground (%)	10·6%	4·3%	<0·0001	8·0%	<0·0001
Doctor diagnosed COPD	1·4%	0·2%	<0·0001	6·1%	<0·0001
Diabetes (%)	5·7%	2·6%	<0·0001	3·3%	<0·0001
Heart attack (%)	2·5%	1·0%	<0·0001	2·0%	0·031
Angina (%)	3·2%	1·5%	<0·0001	2·0%	0·0021
High blood pressure (%)	29·2%	21·2%	<0·0001	25·1%	<0·001
Stroke (%)	1·1%	0·6%	<0·0001	1·1%	0·64

*Ex and never smokers only.

P-values Calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – interquartile range

Comparing Table 1 and Table E6, the only change in direction after limiting analysis to those who do not have a doctor diagnosis of asthma is the prevalence of angina is higher in PRISm than airflow obstruction.

Table E7. Baseline demographics of participants by PRISm trajectory, only those that do not have asthma

Demographic at phase 1	PRISm to PRISm N = 340	PRISm to Control Spirometry N = 488	P value*	PRISm to Stage I-IV Airflow Obstruction N = 95	P value†
Age (Years) Mean (SD)	53·9 (7)	54·4 (6)	0·31	57·5 (8)	0·0001
BMI (kg/m²) Mean (SD)	28·7 (5)	27·8 (4)	0·022	26·3 (4)	0·0002
Female (%)	56%	54%	0·67	60%	0·44
FEV$_1$ % predicted Median (IQR)	74 (70 – 76)	76 (71 – 78)	0·039	75 (69 – 78)	0·89
FVC % predicted Median (IQR)	76 (71 – 80)	77 (73 – 81)	0·065	79 (73 – 83)	0·0080
FEV$_1$/FVC Median (IQR)	76 (73 – 79)	76 (74 – 79)	0·55	73 (71 – 75)	<0·0001
Never smoker (%)	57%	61%	0·99	52%	0·34
Ex-smoker (%)	35%	36%	0·55	36·8	0·97
Current smoker (%)	7%	6%	0·35	12%	0·103
Pack/years Median (IQR) **	21 (14 – 36)	15 (9 – 25)	0·0010	15 (8 – 27)	0·015
SOB walking on ground (%)	7%	5%	0·402	9%	0·76
Doctor diagnosed COPD (%)	1%	1%	0·92	0%	0·35

*P value comparing those that had PRISm at baseline and follow up with those that changed from PRISm at baseline to control at follow up. † P value comparing those with PRISm at baseline and follow up, with those that had PRISm at baseline and airflow obstruction at follow up. **Ex and never smokers only.

P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 2 with table E7, there is no change in interpretation of differences in baseline demographics between PRISm trajectories when restricting analysis to those that do not have a doctor diagnosis of asthma.

Appendix 7. Demographic tables using Lower Limit of Normal (LLN)

For this sensitivity analysis PRISm was defined as FEV_1 % predicted < LLN and $FEV_1/FVC \geq$ LLN, control as FEV_1 % predicted $\geq$ LLN and $FEV_1/FVC \geq$ LLN, and airflow obstruction as $FEV_1/FVC <$ LLN.

Table E8. Baseline demographics with spirometric criteria defined using LLN criteria

Demographic at baseline	PRISm N = 31074	Normal Spirometry N= 289014	P value PRISm vs Control	Airflow obstruction N = 31786	P Value PRISM vs Airflow Obstruction
Age (Years) Mean (SD)	56·1 (8)	56·5 (8)	<0·0001	56·7 (8)	<0·0001
BMI (kg/m^2) Mean (SD)	29·2 (6)	27·2 (5)	<0·0001	26·5 (5)	<0·0001
Female (%)	52·2%	54·4%	<0·0001	50·0%	<0·0001
FEV$_1$ % predicted Median (IQR)	70% (64 – 68)	97% (84 – 105)	<0·0001	73% (61 – 85)	<0·0001
FVC % predicted Median (IQR)	74% (68 – 79)	99% (90 – 107)	<0·0001	93% (80 – 105)	<0·0001
FEV$_1$/FVC Median (IQR)	74% (71 – 78)	77% (74 – 80)	<0·0001	63% (58 – 66)	<0·0001
Never smoker (%)	48·8%	55·6%	<0·0001	40·2%	<0·0001
Ex-smoker (%)	37·0%	36·1%	0·0008	36·5%	0·18
Current smoker (%)	14·2%	8·3%	<0·0001	23·3%	<0·0001
Pack/years Median (IQR) *	25 (14 – 39)	17 (9 – 29)	<0·0001	28 (15 – 41)	<0·0001
SOB walking on ground (%)	19·6%	7·5%	<0·0001	18·4	0·038
Doctor diagnosed asthma (%)	19·6%	10·5%	<0·0001	35·0%	<0·0001
Doctor diagnosed COPD (%)	3·7%	0·5%	<0·0001	9·3%	<0·0001
Diabetes (%)	9·0%	4·0%	<0·0001	4·1%	<0·0001
Heart attack (%)	3·7%	1·72%	<0·0001	2·7%	<0·0001
Angina (%)	4·8%	2·5%	<0·0001	3·9%	<0·0001
High blood pressure (%)	33%	25%	<0·0001	26·2%	<0·0001
Stroke (%)	2·2%	1·2%	<0·0001	1·9%	<0·0001

*For ex and current smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 1 with Table E8 the mean age changes becoming larger in those with normal spirometry than PRISm, however the mean difference is small at 0.4 years.

Table E9. Baseline demographics of participants by PRISm trajectory at follow up, using LLN

Demographic at baseline	PRISm to PRISm N = 493	PRISm to Normal Spirometry N = 960	P value*	PRISm to airflow obstruction N = 79	P value†
Age (Years) Mean (SD)	53.0 (7)	52.4 (7)	0.36	56.1 (8)	0.0013
BMI (kg/m^2) Mean (SD)	28.7 (6)	28 .3(5)	0.11	27.4 (4)	0.038
Female (%)	51%	49%	0.42	54%	0.61
Never smoker (%)	59%	60%	0.501	38%	0.005
Ex-smoker (%)	32%	35%	0.21	41%	0.138
Current smoker (%)	9%	8%	0.35	22%	0.0010
Pack years Median (IQR)**	25 (15 – 38)	17 (9 – 30)	0.001	26 (18 – 34)	0.84
FEV$_1$ % predicted Median (IQR)	71% (66 – 75)	72% (68 – 75)	0.0005	69% (64 – 74)	0.047
FVC % predicted Median (IQR)	74% (69 – 79)	75% (71 – 80)	0.031	77% (71 – 83)	0.057
FEV$_1$/FVC Median (IQR)	74% (72 – 78)	75% (72 – 78)	0.087	71% (68 – 72)	<0.0001
SOB walking on ground (%)	10%	8%	0.67	31%	0.0014
Doctor diagnosed asthma	22%	16%	0.064	24%	0.83
Doctor diagnosed COPD	2%	2%	0.98	8%	0.014

*P value comparing those that had PRISm at baseline and follow up with those that changed from PRISm at baseline to control at follow up. † P value comparing those with PRISm at baseline and follow up, with those that had PRISm at baseline and airflow obstruction at follow up. **Ex and never smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 2 with Table E9 there is no change in interpretation of differences in baseline demographics between PRISm trajectories when performing analysis with LLN criteria.

Appendix 8. Demographic tables stratified by sex and smoking status

Table E10. Baseline demographics, only men

Demographic at baseline	PRISm N = 17251	Control Spirometry N= 114354	P value PRISm vs Control	Stage I-IV Obstruction N = 31022	P Value PRISM vs I-IV Obstruction
Age (Years) Mean (SD)	56·2 (8)	56·1 (8)	0·0064	59·2 (8)	<0·0001
BMI (kg/m²) Mean (SD)	29·7 (5)	28·0 (4)	<0·0001	27·2 (4)	<0·0001
FEV$_1$ % predicted Median (IQR)	74% (69 – 76)	98% (90 – 107)	<0·0001	79% (67 – 89)	<0·0001
FVC % predicted Median (IQR)	76% (70 – 80)	99% (91 – 107)	<0·0001	94% (82 – 105)	<0·0001
FEV$_1$/FVC Median (IQR)	75% (73 – 78)	78% (74 – 81)	<0·0001	66% (62 – 68)	<0·0001
Never smoker (%)	45·2%	52%	<0·0001	37·5%	<0·0001
Ex-smoker (%)	41·3%	38·4%	<0·0001	42·4%	0·017
Current smoker (%)	13·5%	9·4%	<0·0001	20·1%	<0·0001
Pack/years Median (IQR) *	26 (15 – 40)	18 (10 – 30)	<0·0001	29 (15 – 44)	<0·0001
SOB walking on ground (%)	15·4%	5·2%	<0·0001	13·7%	0·0048
Doctor diagnosed asthma	15·2%	8·5%	<0·0001	26·5%	<0·0001
Doctor diagnosed COPD	2·2%	0·4%	<0·0001	7·1%	<0·0001
Diabetes (%)	11·6%	5·1%	<0·0001	6·2%	<0·0001
Heart attack (%)	5·9%	2·8%	<0·0001	4·7%	<0·0001
Angina (%)	6·8%	3·3%	<0·0001	5·2%	<0·0001
High blood pressure (%)	37·4%	27·6%	<0·0001	32%	<0·0001
Stroke (%)	2·5%	1·4%	<0·0001	2·4%	0·31

*Ex and current smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 1 and table E10, there is no change in interpretation of demographic differences between PRISm, control and stage I-IV obstruction when restricting analysis to men.

Table E11. Baseline demographics, only women

Demographic at baseline	PRISm N = 21388	Control Spirometry N = 143289	P value PRISm vs Control	Stage I-IV Obstruction N = 24570	P Value PRISM vs I-IV Obstruction
Age (Years) Mean (SD)	57·0 (8)	56·1 (8)	<0·0001	59·1 (7)	<0·0001
BMI (kg/m²) Mean (SD)	29·0 (6)	27·1 (5)	<0·0001	26·1 (5)	<0·0001
FEV$_1$ % predicted Median (IQR)	74% (69 – 78)	98% (90 – 106)	<0·0001	78% (66 – 89)	<0·0001
FVC % predicted Median (IQR)	77% (72 – 81)	100% (92 – 108)	<0·0001	94% (82 – 106)	<0·0001
FEV$_1$/FVC Median (IQR)	76% (73 – 79)	78% (75 – 80)	<0·0001	66% (63 – 67)	<0·0001
Never smoker (%)	56·0%	60·4%	<0·0001	45·1%	<0·0001
Ex-smoker (%)	32·5%	32·9%	0·314	36·9%	<0·0001
Current smoker (%)	11·5%	6·8%	<0·0001	18·0%	<0·0001
Pack/years Median (IQR) *	21 (11 – 33)	14 (8 – 24)	<0·0001	24 (13 – 37)	<0·0001
SOB walking on ground (%)	19·7%	8·6%	<0·0001	17·0%	<0·0001
Doctor diagnosed asthma	18·3%	11·0%	<0·0001	24·0%	<0·0001
Doctor diagnosed COPD	1·4%	0·3%	<0·0001	6·3%	<0·0001
Diabetes (%)	6·3%	2·7%	<0·0001	3·0%	<0·0001
Heart attack (%)	1·3%	0·5%	<0·0001	1·2%	0·21
Angina (%)	2·9%	1·3%	<0·0001	2·3%	0·0002
High blood pressure (%)	30·2%	21·7%	<0·0001	24·9%	<0·0001
Stroke (%)	1·6%	0·8%	<0·0001	1·4%	0·22

*Ex and current smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 2 and Table E11, there is no change interpretation of demographic differences between PRISm, control and stage I-IV obstruction when restricting analysis to women.

Table E12. Baseline demographics by PRISm trajectory, only men

Demographic at baseline	PRISm to PRISm N = 330	PRISm to Control Spirometry N = 474	P value*	PRISm to Stage I-IV Obstruction N = 101	P value†
Age (Years) Mean (SD)	53·7 (8)	54·0 (7)	0·77	57·0 (8)	0·0006
BMI (kg/m²) Mean (SD)	30·0 (5)	28·6 (4)	<0·0001	28 (4)	<0·0001

Never smoker (%)	51%	56%	0·23	52%	0·91
Ex-smoker (%)	37%	37%	0·89	34%	0·55
Current smoker (%)	11%	7%	0·061	14%	0·47
Pack years Median (IQR)**	22 (14 – 35)	19 (11 – 29)	0·0032	27 (17 – 36)	0·97
FEV$_1$ % predicted Median (IQR)	74 % (69 – 77)	76% (71 – 78)	0·0002	74% (69 – 77)	0·75
FVC % predicted Median (IQR)	75% (70 – 80)	77% (72 – 81)	0·0003	76% (71 – 81)	0·015
FEV$_1$/FVC Median (IQR)	76% (73 – 80)	76% (73 – 79)	0·35	73% (71 – 75)	<0·0001
SOB walking on ground (%)	5%	3%	0·28	9%	0·44
Doctor diagnosed asthma	16%	13%	0·47	27%	0·063
Doctor diagnosed COPD	3%	1%	0·21	2%	0·73

*P value comparing those that had PRISm at baseline and follow up with those that changed from PRISm at baseline to control at follow up. † P value comparing those with PRISm at baseline and follow up, with those that had PRISm at baseline and airflow obstruction at follow up. **Ex and never smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 2 and Table E12, there is no change in interpretation of differences in baseline demographics between PRISm trajectories when restricting analysis to only men.

Table E13. Baseline demographics by PRISm trajectory, only women

Demographic at baseline	PRISm to PRISm N = 415	PRISm to Normal Spirometry N = 513	P value*	PRISm to Stage I-IV Obstruction N = 140	P value†
Age (Years) Mean (SD)	54·0 (7)	53·7 (7)	0·63	56·8 (7)	<0·0001
BMI (kg/m^2) Mean (SD)	28·0 (5)	27·5 (5)	0·31	27·2 (5)	0·22
Never smoker (%)	62%	62%	0·88	52%	0·047
Ex-smoker (%)	32%	32%	0·98	36%	0·37
Current smoker (%)	6%	5%	0·71	11%	0·035
Pack years Median (IQR)**	19 (12 – 32)	14 (7 – 22)	0·0005	15 (7 – 29)	0·061
FEV$_1$ % predicted Median (IQR)	74% (70 – 77)	76% (71 – 78)	0·15	75% (70 – 78)	0·99
FVC % predicted Median (IQR)	77% (72 – 81)	78% (74 – 82)	0·30	81% (76 – 79)	0·0001
FEV$_1$/FVC Median (IQR)	76% (73 – 79)	76% (74 – 79)	0·65	72% (71 – 74)	<0·0001
SOB walking on ground (%)	13%	11%	0·65	22%	0·14
Doctor diagnosed asthma	20%	12%	0·012	24%	0·45
Doctor diagnosed COPD	0%	1%	0·71	0%	0·57

*P value comparing those that had PRISm at baseline and follow up with those that changed from PRISm at baseline to control at follow up. † P value comparing those with PRISm at baseline and follow up, with those that had PRISm at baseline and airflow obstruction at follow up. **Ex and never smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 2 and Table E13, there is no change in interpretation of differences in baseline demographics between PRISm trajectories when restricting analysis to only women.

Table E14. Baseline demographics, only never smokers

Demographic at baseline	PRISm N = 19777	Control Spirometry N = 146220	P value PRISm vs Control	Stage I-IV Obstruction N = 22699	P Value PRISM vs I-IV Obstruction
Age (Years) Mean (SD)	55·6 (8)	55·5 (8)	0·080	58 (8)	<0·0001
Female %	60·6%	59·2%	0·0002	48·8%	<0·0001
BMI (kg/m^2) Mean (SD)	28·1 (6)	27·0 (5)	<0·0001	26 (4)	<0·0001
FEV$_1$ % predicted Median (IQR)	74% (69 – 78)	98% (91 – 107%)	<0·0001	81% (85 – 91)	<0·0001
FVC % predicted Median (IQR)	76% (71 – 80)	99% (92 – 101)	<0·0001	97% (85 – 108)	<0·0001
FEV$_1$/FVC Median (IQR)	76% (73 – 79)	78 (75 – 81)	<0·0001	67% (63 – 69)	<0·0001
SOB walking on ground (%)	14·8%	6·4%	<0·0001	10·3%	<0·0001
Doctor diagnosed asthma %	18·2%	10·2%	<0·0001	31·5%	<0·0001
Doctor diagnosed COPD %	0·3%	0·2%	0·015	2·2%	<0·0001
Diabetes (%)	6·7%	3·1%	<0·0001	3·3%	<0·0001
Heart attack (%)	1·8%	1·0%	<0·0001	1·6%	0·054
Angina (%)	2·9%	1·7%	<0·0001	2·1%	<0·0001
High blood pressure (%)	30·9%	22·7%	<0·0001	25·0%	<0·0001
Stroke (%)	1·4%	0·9%	<0·0001	1·2%	0·035

P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 1 with Table E14, when restricting analysis to never smokers there is no change in demographic differences between PRISm, control and airflow obstruction.

Table E15. Baseline demographics, only ever smokers

Demographic at baseline	PRISm N = 18862	Control Spirometry N = 111423	P value PRISm vs Control	Stage I-IV Obstruction N = 32893	P Value PRISM vs I-IV Obstruction
Age (Years) Mean (SD)	57·2 (8)	57·1 (8)	<0·0001	60·1 (7)	<0·0001
Female %	49·9%	51·0%	0·0064	41·0%	<0·0001
BMI (kg/m^2) Mean (SD)	30 ·1 (6)	27·6 (4)	<0·0001	27·1 (5)	<0·0001
FEV$_1$ % predicted Median (IQR)	74% (69 – 78)	97% (90 – 106)	<0·0001	77% (64 – 88)	<0·0001
FVC % predicted Median (IQR)	77% (71 – 81)	99% (92 – 107)	<0·0001	93% (81 – 104)	<0·0001
FEV$_1$/FVC Median (IQR)	75% (72 – 78)	77% (75 – 80)	<0·0001	66% (61 – 68)	<0·0001
Ex-smokers (%)	74·6%	81·6%	<0·0001	32·4%	<0·0001
Current smokers (%)	25.4%	18·4%	<0·0001	67·6%	<0·0001
Pack/yrs*	23 (13 – 36)	16 (8 – 27)	<0·0001	27 (14 – 41)	<0·0001
SOB walking on ground (%)	20·8%	8·0%	<0·0001	18·6%	<0·0001
Doctor diagnosed asthma	15·3%	9·5%	<0·0001	23·7%	<0·0001
Doctor diagnosed COPD	3·6%	0·7%	<0·0001	11·0%	<0·0001
Diabetes (%)	10·8%	4·6%	<0·0001	5·8%	<0·0001
Heart attack (%)	5·0%	2·2%	<0·0001	4·6%	<0·0001
Angina (%)	6·5%	3·0%	<0·0001	5·2%	<0·0001
High blood pressure (%)	36·0%	26·5%	<0·0001	31·6%	<0·0001
Stroke (%)	2·6%	1·3%	<0·0001	2·5%	0·43

*Ex and current smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 1 with Table E15, when restricting analysis to never smokers there is no change in demographic differences between PRISm, control and airflow obstruction.

Table E16. Baseline demographics by PRISm trajectory, only never smokers

Demographic at baseline	PRISm to PRISm N = 427	PRISm to Control Spirometry N = 585	P value*	PRISm to Stage I-IV Obstruction N = 126	P value†
Age (Years) Mean (SD)	52·5 (7)	53·2 (7)	0·13	56·1 (8)	<0·0001
BMI (kg/m^2) Mean (SD)	28·4 (5)	27·7 (5)	0·017	27.1 (4)	0·015
Female (%)	60%	55%	0·085	58%	0·685
FEV$_1$ % predicted Median (IQR)	74% (70 – 77)	76% (72 – 78)	0·103	75% (69 – 78)	0·49
FVC % predicted Median (IQR)	77% (71 – 80)	78% (73 – 81)	0·067	79 (74 – 83)	0·0046
FEV$_1$/FVC Median (IQR)	76% (74 – 80)	76% (74 – 79)	0·14	73% (71 – 75)	<0·0001
SOB walking on ground (%)	8%	6%	0·42	14%	0·26
Doctor diagnosed asthma	20%	12%	0·02	30%	0·061
Doctor diagnosed COPD	0%	1%	0·78	1%	0·35

*P value comparing those that had PRISm at baseline and follow up with those that changed from PRISm at baseline to control at follow up. † P value comparing those with PRISm at baseline and follow up, with those that had PRISm at baseline and airflow obstruction at follow up. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing Table 2 with Table E16, when restricting analysis to never smokers there is no change in demographic differences between PRISm trajectories.

Table E17. Baseline demographics by PRISm trajectory, only ever smokers

Demographic at baseline	PRISm to PRISm N = 318	PRISm to Normal Spirometry N = 402	P value*	PRISm to Stage I-IV Obstruction N = 115	P value†
Age (Years) Mean (SD)	55·0 (7)	54·6 (7)	0·48	57·6 (7)	0·0011

BMI (kg/m^2) Mean (SD)	29·3 (5)	28·5 (5)	0·035	28·0 (4)	0·002
Female (%)	50%	48%	0·64	58%	0·13
FEV$_1$ % predicted Median (IQR)	74% (69 – 77)	76% (71 – 78)	0·0004	75% (70 – 77)	0·57
FVC % predicted Median (IQR)	76% (71 – 80)	77% (73 – 81)	0·016	81% (73 – 83)	0·0035
FEV$_1$/FVC Median (IQR)	75% (73 – 78)	76% (73 – 79)	0·22	72% (71 – 74)	<0·0001
Ex-smokers (%)	80%	84%	0·18	74%	0·14
Current smokers (%)	20%	16%	0·18	26%	0·14
Pack years Median (IQR)	20 (13 – 34)	16 (8 – 26)	<0·0001	19 (9 – 34)	0·14
SOB walking on ground (%)	11%	8%	0·46	19%	0·21
Doctor diagnosed asthma	16%	13%	0·34	19%	0·58
Doctor diagnosed COPD	3%	1%	0·29	0%	0·20

*P value comparing those that had PRISm at baseline and follow up with those that changed from PRISm at baseline to control at follow up. † P value comparing those with PRISm at baseline and follow up, with those that had PRISm at baseline and airflow obstruction at follow up. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range.

Comparing Table 2 with Table E17, when restricting analysis to ever smokers there is no change in demographic differences between PRISm trajectories.

Table E18. Baseline demographics, only those BMI <25

Demographic at baseline	PRISm N = 8823	Control Spirometry N = 86053	P value PRISm vs Control	Stage I-IV Obstruction N = 21674	P Value PRISM vs I-IV Obstruction
Age (Years) Mean (SD)	55·2 (8)	55·1 (8)	0·16	58·5 (8)	<0·0001
Female %	80·0%	68%	<0·0001	54·3%	<0·0001
FEV$_1$ % predicted Median (IQR)	75% (70 – 78)	99% (91 – 107)	<0·0001	82% (70 – 92)	<0·0001
FVC % predicted Median (IQR)	78% (72 – 78)	101% (94 – 110)	<0·0001	98% (87 – 109)	<0·0001
FEV$_1$/FVC Median (IQR)	75% (72 – 78)	77% (74 – 80)	<0·0001	66% (62 – 68)	<0·0001
Never smoker (%)	58·3%	61·5%	<0·0001	45·3%	<0·0001
Ex-smoker (%)*	28·8%	30·7%	0·0002	32·2%	<0·0001
Current smoker (%)	13·0%	7·8%	<0·0001	22·5%	<0·0001
Pack/yrs Median (IQR)	18 (9 – 30)	13 (6 – 22)	<0·0001	24 (12 – 38)	<0·0001
SOB walking on ground (%)	6·8%	2·4%	<0·0001	9·4%	<0·0001
Doctor diagnosed asthma %	14·8%	8·9%	<0·0001	25·4%	<0·0001
Doctor diagnosed COPD %	0·93%	0·27%	<0·0001	5·2	<0·0001
Diabetes (%)	2·5%	1·3%	<0·0001	1·8%	<0·0001
Heart attack (%)	1·2%	0·7%	<0·0001	1·8%	<0·0001
Angina (%)	1·8%	1·0%	<0·0001	2·2%	0·064
High blood pressure (%)	16·6%	13·5%	<0·0001	18·6%	<0·0001
Stroke (%)	1·2%	0·7%	<0·0001	1·5%	0·059

*Ex and current smokers only. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Comparing table E18 with Table 1, although the mean age of those with PRISm is still higher than control, the evidence for a difference is statistically weak. In those with BMI <25 the prevalence of high blood pressure is higher in those with airflow obstruction than PRISm.

Appendix 9. Demographic tables, only those BMI <25

Table E19. Baseline demographics of participants by PRISm trajectory at follow up, only those BMI <25

Demographic at baseline	PRISm to PRISm N = 194	PRISm to Control Spirometry N = 271	P value*	PRISm to airflow obstruction N = 82	P value†
Age (Years) Mean (SD)	52·2 (7)	52·9 (8)	0·34	55·4 (8)	0·0011
Female (%)	74%	66%	0·044	63%	0·057
Never smoker (%)	66%	67%	0·79	55%	0·081
Ex-smoker (%)	26%	27%	0·65	35%	0·107
Current smoker (%)	8%	5%	0·18	10%	0·68
Pack years Median (IQR)	15 (8 – 30)	12 (5 – 20)	0·020	14 (6 – 20)	0·11

FEV$_1$ % predicted Median (IQR)	75% (70 – 78)	75% (72 – 78)	0·024	75% (73 – 77)	0·34
FVC % predicted Median (IQR)	78% (73 – 81)	79% (74 – 83)	0·108	82% (78 – 84)	0·0011
FEV$_1$/FVC Median (IQR)	75% (73 – 78)	76% (71 – 78)	0·78	72% (71 – 74)	<0·0001
SOB walking on ground (%)	4%	1%	0·35	3%	0·089
Doctor diagnosed asthma	14%	10%	0·34	18%	0·53
Doctor diagnosed COPD	1%	1%	0·86	0%	0·51

FEV$_1$ % predicted Median (IQR) | 75% (70 – 78) | 75% (72 – 78) | 0·024 | 75% (73 – 77) | 0·34
FVC % predicted Median (IQR) | 78% (73 – 81) | 79% (74 – 83) | 0·108 | 82% (78 – 84) | 0·0011
FEV$_1$/FVC Median (IQR) | 75% (73 – 78) | 76% (71 – 78) | 0·78 | 72% (71 – 74) | <0·0001

Appendix 10. PRISm trajectories across all subgroup analysis

Table E20. PRISm trajectories across all subgroup analyses

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%

Appendix 11. Cross sectional and longitudinal associations of PRISm in subgroup analysis

Table E21. Multivariable logistic regression of demographics associated with PRISm vs control at baseline

Demographic	Main analysis	LLN criteria	Obstruction as GOLD II-IV	Men only	Women only	Never smokers	Ever smokers	Non-doctor diagnosed asthma	Not overweight
Age	1·00 (1·00 – 1·01) pval 0.17	0·99 (0·98 – 0·99) pval <0·0001*	1·00 (1·00 - 1·01) pval 0·17	1·00 (0·99 – 1·00) pval 0·76	1·00 (1·00 – 1·01) pval 0·37	1·00 (0·99 – 1·00) pval 0·67	1·01 (1·00 – 1·01) pval 0·013	1·00 (1·00 – 1·00. pval 0·49)	1·00 (0·99 - 1·00. pval 0·57)
Sex (Female)	1·08 (1·03 – 1·13) pval <0·0010	1·00 (0·95 – 1·05) pval 0·90	1·08 (1·03 - 1·13) pval 0·0010	NA	NA	1·11 (1·05 – 1·18) pval 0·0007	1·04 (0·97 – 1·12) pval <0·24	1·10 (1·05 – 1·16. pval 0·0001)	1·27 (1·15 - 1·39. Pval <0.0001)
Overweight (BMI ≥25 and <30)	1·30 (1·23 – 1·37) pval <0·0001	1·30 (1·22 – 1·38) pval <0·0001	1·30 (1·23 - 1·37) pval <0·0001	1·41 (1·28 – 1·55) pval <0·0001	1·27 (1·19 – 1·37) pval <0·0001	1·31 (1·22 – 1·41) pval <0·0001	1·29 (1·18 – 1·41) pval <0·0001	1·30 (1·22 – 1·38. pval <0.0001)	NA
Obese (BMI ≥30)	2·40 (2·26 – 2·55) pval <0·0001	2·37 (2·22 – 2·53) pval <0·0001	2·40 (2·26 - 2·55) pval <0·0001	3·00 (2·70 – 3·30) pval <0·0001	2·09 (1·94 – 2·26) pval <0·0001	2·33 (2·15 – 2·51) pval <0·0001	2·47 (2·25 – 2·70) pval <0·0001	2·43 (2·28 – 2·60. pval <0·0001)	NA
Ex-smoker	1·00 (0·95 – 1·05) pval 0·95	1·01 (0·95 – 1·06) pval 0·82	1·00 (0·95 - 1·05) pval 0·95	1·05 (0·97 – 1·13) pval 0·25	0·95 (0·89 – 1·02) pval 0·17	NA	NA	1·02 (0·97 – 1·08. pval 0·76)	0·94 (0·86 - 1·04. pval 0·24)
Current smoker	1·48 (1·36 – 1·62) pval <0·0001	1·67 (1·53 – 1·83) pval <0·0001	1·48 (1·03 - 1·13) pval <0·0001	1·40 (1·23 – 1·60) pval <0·0001	1·57 (1·40 – 1·78) pval <0.0001	NA	NA	1·49 (1·36 – 1·64. pval <0·0001)	1·65 (1·40 - 1·94. pval <0·0001)
Doctor diagnosis of asthma	1·76 (1·66 – 1·88) pval <0·0001	1·95 (1·83 – 2·09) pval <0·0001	1·76 (1·66 - 1·18) pval <0·0001	1·90 (1·70 – 2·09) pval <0·0001	1·71 (1·58 – 1·85) pval <0·0001	1·83 (1·68 – 1·96) pval <0·0001	1·65 (1·49 – 1·82) pval <0·0001	NA	1·78 (1·57 - 2·01. pval <0·0001)

Values are reported as Odds Ratio (95% Confidence intervals).

*In this model age did not have a linear effect, thereby violating linear assumption of logistic regression. We have reported in this format for continuity, and this is an as isolated result in sensitivity analysis.

Table E22. Multivariable multinomial logistic regression of baseline demographic, PRISm changing to control trajectory vs. persistent PRISm

Demographic	Main analysis	LLN criteria	Obstruction as GOLD II-IV	Men only	Women only	Never smokers	Ever smokers	Non-doctor diagnosed asthma	Not overweight (BMI <25)
Age	1·01 (0·99 – 1·03) pval 0·46	1·03 (1·00 – 1·05) pval 0·016	1·01 (0·99 – 1·03) pval 0·43	1·01 (0·98 – 1·03) pval 0·71	1·01 (0·99 – 1·04) pval 0·44	1·02 (1·00 – 1·04) pval 0·17	0·99 (0·96 – 1·02) pval 0·63	1·01 (0·99 – 1·03) pval 0·30	1·03 (0·99 - 1·06) pval 0·16

	Main analysis	LLN	Obstruction as GOLD II-IV	Men only	Women only	Never smokers	Ever smokers	Non-doctor diagnosed asthma	Not overweight (BMI <25)
Sex (Female)	0·88 (0·68 – 1·14) pval 0·34	0·95 (0·70 – 1·29) pval 0·74	0·88 (0·68 – 1·14) pval 0·33	NA	NA	0·90 (0·64 – 1·27) pval 0·54	0·82 (0·53 – 1·23) pval 0·33	0·94 (0·95 – 1·91) pval 0·09	0·97 (0·57 - 1·66) pval 0·91
Overweight (BMI ≥25 and <30)	1·42 (1·02 – 1·96) pval 0·037	1·48 (1·00 – 2·17) pval 0·48	1·41 (1·02 – 1·96) pval 0·039	1·57 (0·92 – 2·70) pval 0·097	1·29 (0·84 – 1·97) pval 0·23	1·56 (1·03 – 2·35) pval 0·032	1·19 (0·68 – 2·08) pval 0·53	1·35 (0·95 – 1·92) pval 0·24	NA
Obese (BMI ≥30)	0·74 (0·53 – 1·04) pval 0·081	0·85 (0·58 - 1·25) pval 0·41	0·74 (0·53 – 1·04) pval 0·079	0·73 (0·62 - 1·48) pval 0·27	0·75 (0·49 - 1·17) pval 0·21	0·79 (0·51 - 1·21) pval 0·28	0·65 (0·37 - 1·15) pval 0·14	0·80 (0·95 - 1·92) pval 0·24	NA
Ex-smoker	0·90 (0·68 – 1·19) pval 0.46	0·91 (0·65 - 1·27) pval 0·58	0·90 (0·68 – 1·20) pval 0.50	0·96 (0·62 - 1·47) pval 0·84	0·85 (0·59 - 1·25) pval 0·41	NA	NA	0·87 (0·64 - 1·18) pval 0·36	1·06 (0·59 - 1·90) pval 0·84
Current smoker	0·79 (0·46 – 1·35) pval 0·39	1·12 (0·65 - 1·27) pval 0·56	0·79 (0·47 - 1·35) pval 0·40	0·76 (0·37 - 1·56) pval 0·46	0·88 (0·39 - 1·98) pval 0·76	NA	NA	0·73 (0·41 - 1·30) pval 0·28	0·87 (0·29 - 2·50)
Doctor diagnosis of asthma	0·67 (0·47 – 0·96) pval 0·030	0·71 (0·48 – 1·03) pval 0·073	0·68 (0·47 – 0·97) pval 0·034	0·85 (0·49 – 1·50) pval 0·59	0·56 (0·35 – 0·90) pval 0·017	0·61 (0·39 - 0·97) pval 0·037	0·82 (0·43 - 1·37) pval 0·38	NA	0·73 (0·34 - 1·56) pval 0·41

Values are reported as Relative Risk Ratios (95% Confidence intervals)

Table E23. Multivariable multinomial logistic regression of baseline demographic, PRISm changing to airflow obstruction trajectory vs. persistent PRISm

Demographic	Main analysis	LLN	Obstruction as GOLD II-IV	Men only	Women only	Never smokers	Ever smokers	Non-doctor diagnosed asthma	Not overweight (BMI <25)
Age	1·07 (1·04 – 1·10) pval <0·0001	1·08 (1·04 – 1·14) pval 0·00097	1·06 (1·02 – 1·10) pval 0·0002	1·07 (1·03 – 1·13) pval 0·0013	1·06 (1·02 – 1·10) pval 0·0016	1·06 (1·02 – 1·10) pval 0·0028	1·09 (0·96 – 1·02) pval 0·63	1·08 (1·04 – 1·12) pval <0·0001	1·04 (1·00 - 1·10) pval 0·072
Sex (Female)	1·00 (0·66 – 1·52) pval 0·99	1·03 (0·50 – 2·08) pval 0·94	1·02 (0·64 – 1·65) pval 0·91	NA	NA	0·73 (0·42 – 1·27) pval 0·27	0·85 (0·54 – 1·23) pval 0·53	1·19 (0·74 – 1·93) pval 0·48	0·62 (0·31 - 1·27) pval 0·20
Overweight (BMI ≥ 25 and <30)	0·73 (0·45 – 1·17) pval 0·19	0·80 (0·35 – 1·83) pval 0·60	0·79 (0·46 – 1·36) pval 0·40	0·52 (0·24 - 1·12) pval 0·096	0·89 (0·48 – 1·65) pval 0·72	1·01 (0·54 – 1·90) pval 0·96	1·19 (0·69 – 2·08) pval 0·53	0·60 (0·35 – 1·04) pval 0·067	NA
Obese (BMI ≥30)	0·28 (0·16 - 0·49) pval <0·0001	0·37 (0·15 - 0·94) pval 0·036)	0·34 (0·18 - 0·64) pval 0·0008)	0·17 (0·07 - 0·42) pval 0·0001	0·40 (0·19 - 0·82) pval 0·012	0·42 (0·20 - 0·88) pval 0·021	0·16 (0·07 - 0·40) pval <0·0001	0·30 (0·16 - 0·57) pval 0002	NA
Ex-smoker	0·94 (0·59 – 1·47) pval 0·78	1·27 (0·56 – 2·76) pval 0·59	1·02 (0·61 – 1·70. pval 0·93)	0·61 (0·29 – 1·28) pval 0·19	1·25 (0·71 – 2·22) pval 0·44	NA	NA	0·98 (0·58 – 1·63) pval 0·93	1·84 (0·86 - 3·95) pval 0·11
Current smoker	1·92 (0·94 - 3·91) pval 0·074	5·11 (1·86 - 14·0) pval 0·0016	2·55 (1·21 - 5·36) pval 0·014	1·44 (0·47 - 4·47) pval 0·52	2·80 (1·08 - 7·27) pval 0·035	NA	NA	2·32 (1·06 - 5·09) pval 0·036	3·22 (0·96 - 10·8) pval 0·057
Doctor diagnosis of asthma	1·91 (1·17 – 3·13) pval 0·0010	1·18 (0·52 - 2·69) pval 0·69	1·71 (0·98 - 2·99) pval 0·060	3·19 (1·43 – 7·13) pval 0·0045	1·39 (0·74 – 2·65) pval 0·30	2·08 (1·12 – 3·89) pval 0·021	1·61 (1·00 – 4·80) pval 0·94	NA	1·54 (0·62 - 3·88) pval 0·35

Values are reported as Relative Risk Ratios (95% Confidence intervals. P-Value)

Note. As this was multinomial logistic regression factors associated with PRISm to control and PRISm to COPD (vs PRISm to PRISm) were modelled simultaneously. However, the results are presented in two separate tables for ease of interpretation.

Appendix 11. Trajectories of normal spirometry and airflow obstruction at baseline

Compared to those with PRISm at baseline, those with control and airflow obstruction at baseline had much more stable spirometry. 62% of those with PRISm at baseline changed spirometric classification at follow up, compared to 12% of those with control and 34% of those with airflow obstruction.

4·2% of those with control spirometry at baseline changed to PRISm at follow up. At baseline those that changed to PRISm had worse lung function, higher rates of smoking and were more short of breath (Table E24). Multivariable multinomial logistic regression analysis showed that female sex, being overweight, obesity and current smoking were all strongly associated with this transition, whereas doctor diagnosis of asthma was not. However, it was strongly associated with those that change from control to airflow obstruction (Table E25).

5·7% of those with airflow obstruction at baseline changed to PRISm at follow up. Those that transitioned to PRISm were younger, female predominant, with better lung function at baseline, less smoking history, and lower rates of asthma. (Table E26) Multivariable multinomial logistic regression analysis showed current smoking and doctor diagnosis asthma were strongly associated with transition to PRISm (Table E27).

Table E24. Demographics at baseline by control spirometry trajectory

Demographic at baseline	Control to Control N = 16967	Control to PRISm N = 815	P value*	Control to Stage I-IV Obstruction N = 1413	P value†
Age (Years) Mean (SD)	54·0 (7)	54·0 (7)	0·89	56·7 (7)	<0·0001
BMI (kg/m^2) Mean (SD)	26·7 (4)	27·9 (5)	<0·0001	25·8 (4)	<0·0001
Female (%)	52·6%	56·0%	0·052	51·2%	0·326
FEV$_1$ % predicted Median (IQR)	99% (92 – 108)	86% (83 – 92)	<0·0001	97% (89 – 98)	<0·0001
FVC % predicted Median (IQR)	101% (93 – 109%	90% (85 – 96)	<0·0001	100% (94 – 109)	0·061
FEV$_1$/FVC Median (IQR)	78% (75 – 81)	77% (74 – 80%)	<0·0001	74% (72 – 78)	<0·0001
Never smoker (%)	62·0%	60·4%	0·35	55·7%	<0·0001
Ex-smoker (%)	33·0%	29·0%	0·017	36·2%	0·015
Current smoker (%)	5·1%	10·7%	<0·0001	8·1%	<0·0001
Pack/years Median (IQR) **	13 (7 – 23)	20 (11 – 29)	<0·0001	16 (8 – 27)	<0·0001
SOB walking on ground (%)	4·1%	9·0%	0·0001	2·7%	0·11
Doctor diagnosed asthma	10·0%	12·6%	0·084	12·1%	0·070
Doctor diagnosed COPD	0·2%	1·1%	0·0001	0·5%	0·093

*P value comparing those that had control spirometry at baseline and follow up with those that changed from control at baseline to PRISm at follow up. † P value comparing those with control at baseline and follow up, with those that had control at baseline and airflow obstruction at follow up. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Table E25. Multivariable multinomial logistic regression of baseline demographics association with control to prism and control to airflow obstruction vs. persistent control trajectory

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.

Table E26. Demographics at baseline by airflow obstruction trajectory

Demographic at baseline	COPD to COPD N = 2138	COPD to PRISm N = 839	P value*	COPD to Control N = 259	P value†
Age (Years) Mean (SD)	60·0 (7)	56·3 (7)	0·013	54·7 (7)	<0·0001
BMI (kg/m^2) Mean (SD)	25·8 (4)	26·0 (4)	0·061	30·0 (4)	<0·0001
Female (%)	37·6%	41·2%	0·73	43·6%	0·60
FEV$_1$ % predicted Median (IQR)	82% (71 – 92)	89% (83 – 98)	<0·0001	74% (66 – 79)	<0·0001
FVC % predicted Median (IQR)	98% (87 – 109)	104% (96 – 114)	<0·0001	90% (83 – 96)	<0·0001
FEV$_1$/FVC Median (IQR)	66% (62 – 68)	68% (68 – 69)	0·0009	67% (60 – 68)	0·0001
Never smoker (%)	49·6%	55·5%	0·0004	56·0%	0·053
Ex-smoker (%)	38·6%	37·5%	0·59	35·9%	0·40
Current smoker (%)	11·8%	6·9%	<0·0001	8·1%	0·078
Pack/years Median (IQR) **	21 (10 – 34)	18 (10 – 30)	0·30	17 (10 – 21)	0·062
SOB walking on ground (%)	6·9%	4·1%	0·12	11·4%	0·15
Doctor diagnosed asthma	32·5%	19·7%	<0·0001	20·5%	0·0054
Doctor diagnosed COPD	6·2%	1·3%	<0·0001	2·4%	0·081

*P value comparing those that had airflow obstruction at baseline and follow up with those that changed from airflow obstruction at baseline to PRISm at follow up. † P value comparing those with airflow obstruction at baseline and follow up, with those that had airflow obstruction at baseline and control spirometry at follow up. P-values calculated using Z-score for continuous outcomes, Pearson's chi-squared for categorical outcomes. SD – standard deviation. IQR – Inter quartile range

Table E27. Multivariable multinomial logistic regression of baseline demographics association with airflow obstruction to prism and airflow obstruction to control vs persistent airflow obstruction trajectory

Demographic	RRR (95%CI) Airflow obstruction to PRISm	RRR (95%CI) Airflow obstruction to Control
Age	0·99 (0·98 – 1·01) pval 0·18	0·97 (0·95 - 1·00) pval 0·042
Sex	1·22 (0·89 – 1·45) pval 0·31	1·44 (0·98 - 2·21) pval 0·061
Overweight (BMI ≥25 and <30)	1·13 (0·89 - 1·45) pval 0·31	1·68 (1·10 - 2·56) pval 0·017
Obese (BMI ≥30)	1·28 (0·91 - 1·80) pval 0·16	2·15 (1·26 - 3·71) pval 0·0054
Ex-smoker	0·92 (0·73 - 1·17) pval 0·50	1·20 (0·29 - 0·73) pval 0·37
Current smoker	0·40 (0·35 - 0·60) pval <0·0001	0·75 (0·38 - 1·48) pval 0·409
Doctor diagnosis of asthma	0·46 (0·35 – 0·60) pval <0·0001	0·46 (0·29 - 0·73) pval 0·104

RRR – relative risk ratio

Appendix 13. FEV$_1$ decline in persistent phenotypes

Persistent PRISm participants had lower mean FEV$_1$ annual decline than those with persistently normal spirometry (-23mls vs -28mls, p-value <0·001) and those with persistent airflow obstruction (-30mls, p-value <0·001). However, these are highly selected populations.

Appendix 14. Directed acyclic graphs of analysis plan, confounding and collider bias

Figure E2. Example DAG

Figure E3. DAG of analysis plan

As seen in Figure E2, there are a number of confounding factors when examining any extra-pulmonary associations of PRISm/COPD. These need to be adjusted to gain more accurate estimates of association. In our analysis we adjusted for age, smoking, gender, BMI, and diabetes to reduce the effect of these measured confounding factors. In some analysis diabetes was the outcome so was not adjusted.

Figure E4. DAG demonstrating possible collider bias

Collider bias occurs when two variables independently influence a third variable, and that third variable is conditioned upon. [1] PRISm/COPD/Control are conditional phenotypes influenced by both measured and unmeasured factors, making them colliders. By selecting on these phenotypes it could introduce collider bias, meaning the lung function phenotypes are inherently associated with the factors influencing them. Traditional observational epidemiology is not able to account for this. As per our discussion, this problem can be surmounted by using Mendelian Randomization with the latest techniques i.e. slope hunter, when GWAS of PRISm become available.[2,3]

References
1.   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.
2.   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-.
3.   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.
4.   Schwartz A, Arnold N, Skinner B, et al. Preserved Ratio Impaired Spirometry in a Spirometry Database. *Respiratory care* 2021; **66**(1): 58-65.
5.   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.
6.   WHO T. https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/. 2020.
7.   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.
8.   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.

9.   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).

10.   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.

11.   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.

12.   Barnes PJ, Vestbo J, Calverley PM. The Pressing Need to Redefine "COPD". *Chronic Obstr Pulm Dis* 2019; **6**(5): 380-3.

13.   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.

14.   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.

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

16.   Lytras T. 2020. https://github.com/thlytras/rspiro2020).

17.   Global Iniative for Chronic Obstructive Pulmonary Lung Disease, 2019.

18.   STATACorp. Stata Statistical Software: Release 15. 2017.

19.   Linden A. RTMCI: Stata module to estimate regression to the mean effects with confidence intervals. In: S457757 SSC, editor.; 2013.

20.   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.

21.   Thomas ET, Guppy M, Straus SE, Bell KJL, Glasziou P. Rate of normal lung function decline in ageing adults: a systematic review of prospective cohort studies. *BMJ Open* 2019; **9**(6): e028150.

22.   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.

23.   Maclay JD, MacNee W. Cardiovascular disease in COPD: mechanisms. *Chest* 2013; **143**(3): 798-807.

24.   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.

25.   Burney PGJ, Hooper R. Forced vital capacity, airway obstruction and survival in a general population sample from the USA. *Thorax* 2011; **66**(1): 49.

26.   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.

27.   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.

28.   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.

29.   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.

30.   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.

31.	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.

32.	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.

33.	Mahmoud O, Dudbridge F, Smith GD, Munafo M, Tilling K. Slope-Hunter: A robust method for index-event bias correction in genome-wide association studies of subsequent traits. *bioRxiv* 2020: 2020.01.31.928077.

34.	Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLoS genetics* 2017; **13**(10): e1006944-e.