

International Journal of Population Data Science

Journal Website: www.ijpds.org



Swansea University
Prifysgol Abertawe

Genetic risk score informed re-evaluation of phenotype quality control to maximise power in epidemiological studies: application to lung function

Jing Chen^{1,*}, Nick Shrine¹, Abril G Izquierdo¹, Anna Guyatt¹, Alex Williams¹, Henry Völzke², Stephanie London³, Ian P Hall⁴, Frank Dudbridge¹, SpiroMeta Consortium, CHARGE Consortium, Louise V Wain^{1,5}, Martin D Tobin^{1,5}, and Catherine John^{1,5*}

Submission History

Submitted:	15/04/2025
Accepted:	13/11/2025
Published:	09/02/2026

¹ Division of Public Health and Epidemiology, School of Medical Sciences, University of Leicester, Leicester, UK

² Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

³ Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC, USA

⁴ Division of Respiratory Medicine and NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, UK

⁵ National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, University of Leicester, Leicester, UK

* Joint corresponding authors

Abstract

Introduction

Clinical guidelines may reduce statistical power in epidemiological studies by discarding informative measures. Epidemiological studies of lung function may discard one-third to one-half of participants due to spirometry measures deemed “low quality” using criteria adapted from clinical practice.

Objectives

To optimise the signal-to-noise ratio in epidemiological studies of lung function, we aimed to develop a data-driven method to refine spirometry quality control (QC) criteria.

Methods

We proposed a genetic risk score (GRS) informed strategy to categorise spirometer blows by quality criteria. GRS was built using SNPs associated with lung function traits in non-UK Biobank cohorts. In the UK Biobank, we applied a step-wise testing of the GRS association across groups of spirometry blows stratified by acceptability flags to rank the blow quality. We reassessed QC criteria by comparing the genetic associations under different acceptability flags and repeatability thresholds to determine the trade-off between sample size and measurement error.

Results

We found that including blows previously excluded by strict QC criteria would maximise the statistical power for genome-wide association study and retain acceptable precision in the UK Biobank. This approach allowed the inclusion of 29% more participants compared to the strictest clinical guidelines and demonstrated genetic signals could be identified earlier.

Conclusions

Our GRS-based method offers an important framework to challenge prevailing practices that exclude informative measures and limit power in epidemiological studies.

Keywords

epidemiological methods; genome-wide association study; data-driven quality control; spirometry quality control; genetic/polygenic risk score

*Corresponding Authors:

Email Address: jc824@leicester.ac.uk (Jing Chen); cj153@leicester.ac.uk (Catherine John)

Introduction

Impairment of lung function, as measured by spirometry, is central to chronic respiratory diseases including chronic obstructive pulmonary disease (COPD), but also predicts mortality in the general population [1]. Genome-wide association studies (GWAS) have proven to be an effective tool for identifying genetic variants that are associated with complex diseases and traits, providing valuable insights into disease biology and informing development of diagnostic tools and potential treatments [2]. In the latest GWAS of lung function, associated genetic variants explained 33% of the heritability of the forced expiratory volume in one second to forced vital capacity ratio (FEV_1/FVC). Lower proportions were explained for forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC) [3], indicating that additional associations could be found in more powerful studies.

However, large-scale epidemiological studies face a fundamental trade-off between precision and inclusivity. GWAS of lung function may discard one-third to one-half of participants due to spirometry measures deemed “low quality”, substantially limiting the potential sample size [4]. Spirometry measures the flow and volume of air over time, typically in a forced expiratory manoeuvre, i.e. a vigorous, complete exhalation following maximal inhalation. Supplementary Figure 1 illustrates key spirometric parameters. Spirometry is effort and technique-dependent, and thorough quality control (QC) is considered essential to ensure measurements are unaffected by inadequate technique or artefacts (such as a cough). However, previous work has emphasised that spirometry QC criteria should be continuously evaluated and refined to accommodate advances in methodology and understanding [5, 6].

In clinical practice, QC procedures are applied at the level of individual assessment, with trained technicians ensuring patient cooperation and repeatability to support diagnostic decisions [7, 8]. In contrast, epidemiological studies often rely on automated QC algorithms and limited technician oversight due to scale, prioritising consistency and throughput over individualised accuracy. Moreover, QC for epidemiological studies may not require the same level of stringency as clinical practice, where results are used for diagnosis and management of individual patients. This fundamental difference means the QC criteria optimised for clinical practice may inadvertently exclude usable data in epidemiological studies. The variability in spirometer models further complicates QC. Each spirometer model employs unique software, firmware and diagnostic algorithms, leading to variations in how errors are coded and displayed. To mitigate unnecessary exclusion of valid data, Hankinson *et al.* have previously suggested visual inspection of blow curves by human reviewers in addition to computer assessment [9]. While this approach could enhance the inclusion of valid tests, it may become impractical for large-scale studies and could be subject to bias.

Genetic information can predict individual continuous traits such as lung function by tallying the number of risk alleles for each individual to give a genetic risk score (GRS) (sometimes called polygenic risk score, PRS, or polygenic score, PGS, Supplementary Figure 2) [10]. These may include hundreds to millions of genetic variants and

are typically weighted by the size of association of each allele with the trait of interest [10]. By checking concordance between genetically predicted lung function - adjusted for well-established demographic and environmental factors - and the actual measured lung function traits, we can reassess the value of spirometer blows that were previously deemed as “low quality”. Importantly, this approach provides a generalisable framework: GRS-informed QC could be applied to other traits in epidemiological studies. In this study, we aimed to propose a data-driven method to define new spirometry QC criteria informed by GRS to optimise the signal-to-noise ratio in epidemiological studies.

Methods

Spirometry quality control

We undertook analyses in the UK Biobank European population as illustration [11]. In UK Biobank, each individual was asked to perform up to three blows on a Vitalograph spirometer (Vitalograph Pneumotrac 6800). These blows were used to derive measures of lung function, including FEV_1 , FVC and FEV_1/FVC , from a spirogram (volume-time curve) recorded by the Vitalograph software. A blow received an automated error code from the spirometry software if: (1) there was hesitation (excessive extrapolated volume [12] at the start of the blow, coded “START”); (2) the time to peak flow was excessive (“EXPFLOW”); (3) a cough was detected during the manoeuvre (“COUGH”); (4) there was not an adequate plateau at the end of the blow (“END”) (see Table 1 for additional detail). It was also possible for the spirometer operator to explicitly reject the blow (“REJECT”). In previous GWAS of lung function, blows were deemed unacceptable and excluded from analysis if they had any of the above error codes [4].

Additionally, as in previous GWAS [3, 4], we checked each blow for inappropriate negative values which indicate a problem with the blow (for any of: forced expiratory flow at 25% of FVC; forced expiratory flow at 75% of FVC; average forced expiratory flow between 25% and 75% of the FVC; peak flow; extrapolated volume; volume at forced expiratory time) (coded “NEGATIVE”), the start of the blow underwent a further check for hesitation (“START2”), and consistency (<5% difference) was checked between the FEV_1 and FVC output from the spirometer and that rederived from the spirogram (“CONSISTENCY”). In this study, we identified blows that would previously have failed QC due to any of the above error codes or additional QC steps, and labelled these with corresponding “acceptability flags” (Table 1).

Testing the association between GRSs and lung function traits derived from spirometer blow measurements

We calculated the GRSs for European ancestry individuals in UK Biobank using European-specific weights trained from Shrine *et al.* [3] (Supplementary Methods). In the UK Biobank European population, we then conducted iterative testing of the GRS association within groups of spirometry blows stratified by acceptability flags, to identify and rank the flags most likely to cause failure of spirometry blow quality

Table 1: Reasons for spirometer blows failing quality control (QC)

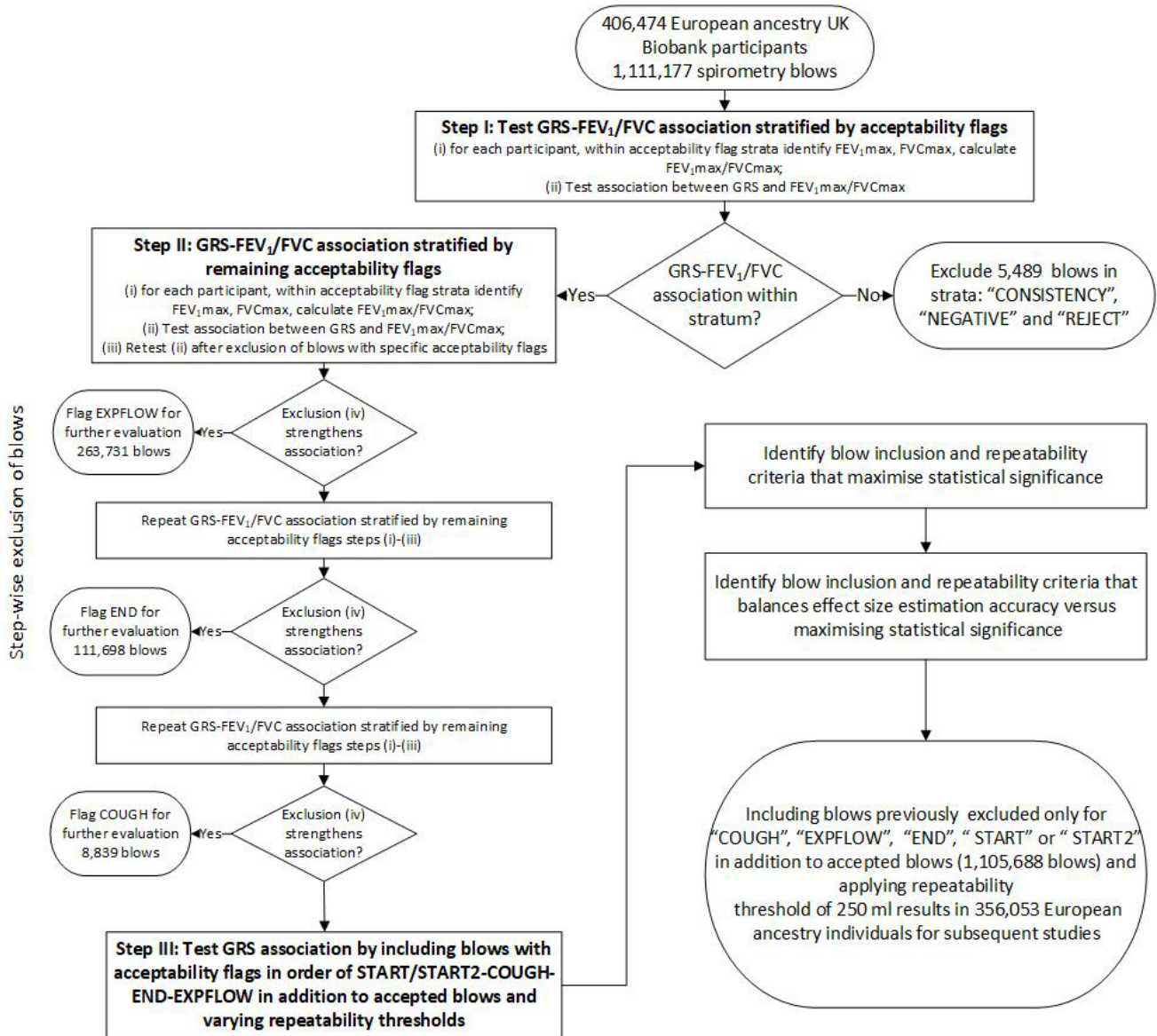
2019 ATS/ERS criterion [8]	Definition of ATS/ERS criterion (2019)	Acceptability flags in this study in this study	Acceptability flag definition in this study	Number of individuals with blows failing within acceptability flags stratum
Acceptable start of blow (i.e. an immediate, explosive start with no hesitation)	Extrapolated volume < 5% or 100ml (150ml in 2005 criteria)	START (spirometer output)	Present if the extrapolated volume at start of test is excessive.	54,922
		START2 (additional QC)	Present if the extrapolated volume is not less than 5% of FVC or 150 ml, whichever is greater.	53,840
		EXPFLOW (spirometer output)	Present if time to peak flow is excessive.	164,319
No cough in first second (for FEV1)	N/A	COUGH (spirometer output)	Present if a cough was detected during the manoeuvre.	11,857
Acceptable end-of-blow	Adequate terminal plateau (< 25ml flow in last 1s) or length of blow >=15s or FVC meets repeatability threshold	END (spirometer output)	Present if an adequate plateau at the end of the test does not exist.	103,092
N/A	N/A	REJECT (spirometer output)	Present if the spirometer operator explicitly rejects the blow.	1,610
N/A	N/A	NEGATIVE (additional QC)	Present if inappropriate negative values derived from the blow for any of: forced expiratory flow at 25% of FVC; forced expiratory flow at 75% of FVC; average forced expiratory flow between 25% and 75% of the FVC; peak flow; extrapolated volume; volume at forced expiratory time.	577
N/A	N/A	CONSISTENCY (additional QC)	Present if the difference between the FEV ₁ and FVC output from the spirometer and that rederived from the spirogram was greater than 5%.	3,183

In previous lung function GWAS and in our baseline analysis for this study, a repeatability threshold of 250ml was applied.

(Figure 1). All the GRS and lung function traits association tests were adjusted for age, age², height, smoking status, and top 10 genetic principal components. For each lung

function trait, we first stratified all the blows by acceptability flag (Figure 1, step I, Supplementary Methods). Where lung function measures within a stratum showed no significant

Figure 1: Flowchart for evaluating spirometer blow quality and spirometry QC criteria



Definitions of acceptability flags ("COUGH", "EXPFLOW", "END", "START", "START2", "CONSISTENCY", "NEGATIVE", "REJECT") were given in Table 1.

association with the GRS ($P > 0.05$), we considered that the corresponding acceptability flag to indicate unacceptable blow quality and thus a reason for exclusion. For the remaining blows, we applied an iterative selection process to identify the acceptability flags that were most likely to cause failure of spirometry blow quality as shown in Figure 1 (Step II). Based on the outcome of this, we ranked the remaining acceptability flags according to their impact on the spirometer blows in our new approach (see Results for details).

Re-evaluating spirometry QC criteria for association study

Based on the grouping of blows above, we then aimed to identify which acceptability and repeatability criteria would maximise the association of the lung function measures with GRS as measured by the level of statistical significance (Z-score for the association) (Figure 1, Step III). We varied our

acceptability criteria by including blows with acceptability flags in order of the ranking generated in the previous step, in addition to blows previously considered acceptable. We also varied our repeatability threshold from 150ml – the standard recommended in strict clinical practice by ATS/ERS [8] – up to 400ml in 50ml increments. Repeatability was based on the difference from any other blow, even if that blow was not accepted. Using these different criteria, we tested the association with the GRS. To ensure that effect size estimates retained acceptable precision while maximising statistical significance, we then tested genetic associations with the sentinel SNPs – that is, the SNPs showing the strongest statistical associations with previous lung function traits GWAS in genomic regions [3]. Using the different acceptability criteria and repeatability thresholds described above, we examined how these QC criteria influenced the estimated effect sizes and P values for the sentinels of lung function signals. All lung function traits were untransformed, adjusting for age,

age², height, smoking status, and relatedness (mixed models in BOLT-LMM [13]).

Assessing the credibility of GWAS findings

To illustrate the gain in power using our newly defined QC criteria in epidemiological studies, we conducted a GWAS in the UK Biobank European population as an example to compare GWAS findings with those found using the previous QC criteria. Credibility of newly identified signals was assessed using a Bayesian framework previously described by Okbay *et al* [14] and Turley *et al.* [15] (Supplementary methods).

Results

In UK Biobank, spirometry assessments were conducted by healthcare technicians or nurses certified to perform the assessments. In total, 445,541 individuals had at least two measures of FEV₁ and FVC, along with complete information for age, sex, standing height and derived smoking status (smoking status derived by Shrine *et al.* 2019 [4]). Of these individuals, 406,474 were assigned as European ancestry using k-means clustering and ADMIXTURE v1.3.0 [16] as described in Shrine *et al* [4]. We used FEV₁ and FVC measurements (UK Biobank Field IDs 3063 and 3062), along with the blow curve time series measurements (UK Biobank Field ID 20031) and the Vitalograph spirometer blow quality metrics (UK Biobank Field ID 20031).

Evaluating spirometer blows quality with different acceptability flags

Using the previous, more stringent approach to spirometry QC, we identified 713,885 spirometer blows as “accepted” for 354,746 European individuals. The best measures of FEV₁/FVC derived from “accepted” blows for these individuals were strongly associated with the GRS ($\beta_{SD_change_of_GRS} = 0.246$, 95% CI [0.243, 0.249], $P < 1.00E-300$) (Figure 2a).

From the association of GRS with the FEV₁/FVC derived from blows in each of our acceptability flag strata, blows with acceptability flags “CONSISTENCY”, “NEGATIVE” or “REJECT” were not associated with the GRS (Figure 2a), indicating that these acceptability flags represent unacceptable spirometry blow quality ($\beta_{SD_change_of_GRS} = -0.022$, 95% CI [-0.049, 0.006], $p = 0.1292$, Figure 2c, Group 3). To rank the remaining acceptability flags that cause failure of the spirometry blow quality in descending order of severity, we conducted an iterative selection process (Figure 2b). In the first round of selection (round 1), we identified excessive time to peak flow (“EXPFLOW”) as the next most likely cause of failure of spirometry quality. This was based on the observation that additionally removing blows with “EXPFLOW” flag led to an increase in the magnitude of the effect size estimate in the association results in all the remaining strata. Similarly, we identified an inadequate terminal plateau (“END”, round 2) and cough (“COUGH”, round 3) as the next most likely to cause failure of spirometry quality in the subsequent rounds of selection.

Collectively, blows from groups “EXPFLOW”, “END” and “COUGH” were associated with GRS but with

a smaller effect size than previously accepted blows ($\beta_{SD_change_of_GRS} = 0.149$, 95% CI [0.145, 0.153], $p < 1.00E-300$, Figure 2c, Group 2), after removing blows with the flags excluded in previous steps. For the remaining flags relating to hesitation (“START” and “START2”), the magnitude of the effect size of GRS association in the corresponding strata ($\beta_{SD_change_of_GRS} = 0.262$, 95% CI [0.240, 0.284], $P = 1.14E-118$, Figure 2c, Group 1) was similar to that from the acceptable blows ($\beta_{SD_change_of_GRS} = 0.246$, 95% CI [0.243, 0.249], $P < 1.00E-300$, Figure 2c, Accepted blows), after removing blows with all the other acceptability flags. For FEV₁ and FVC, we observed similar results to those for FEV₁/FVC above in ranking the spirometer blow quality but with the “COUGH” flag showing a similar effect size to accepted blows (Supplementary Figure 3 and Supplementary Figure 4). For FVC, “END” had a larger impact than “EXPFLOW”. Since the results suggest acceptability flags have greater impact on the measurements of FEV₁/FVC than FEV₁ and FVC, we based our subsequent analyses on FEV₁/FVC.

Re-evaluating spirometry QC criteria for association studies

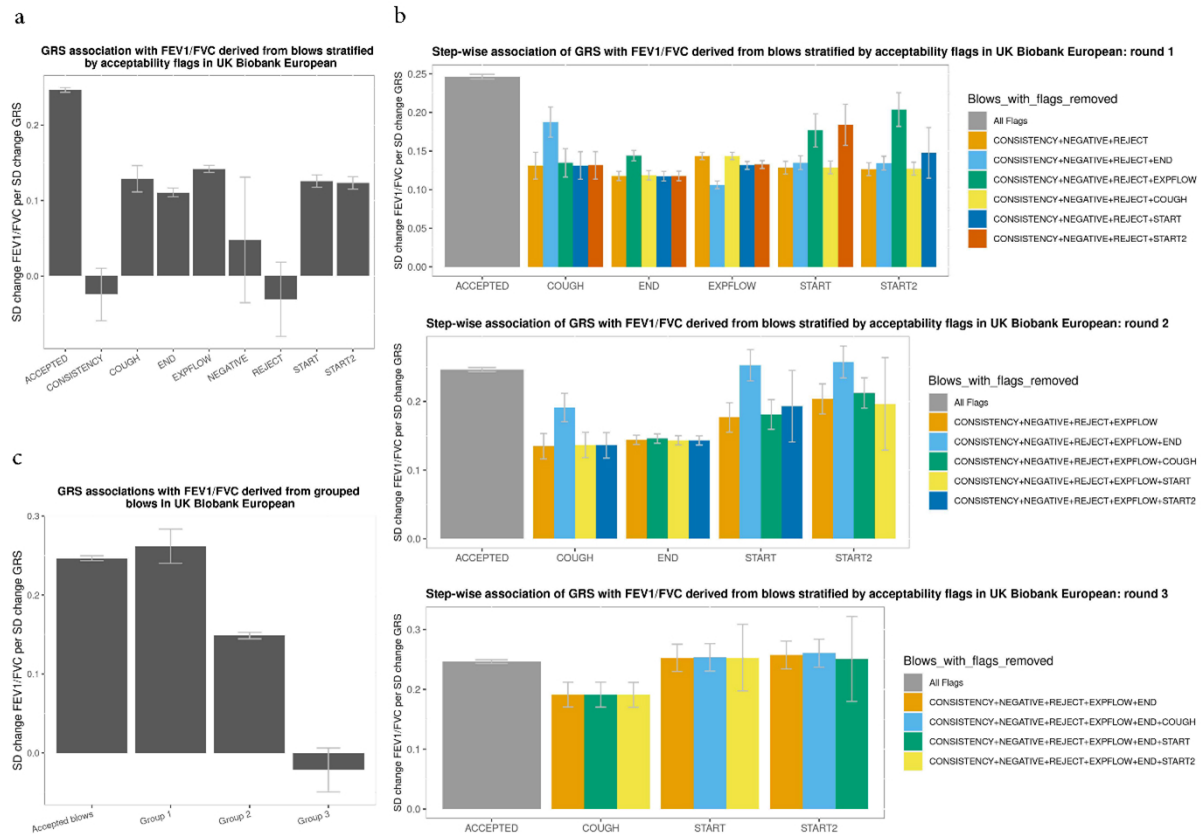
We investigated the trade-off between sample size and measurement error by re-evaluating various inclusion criteria for spirometry QC. To do this, we included blows with varying acceptability flags with their inclusion ordered according to the findings above, and applied different repeatability thresholds to assess their impact on the association between GRS and FEV₁/FVC. We found that including blows previously excluded for cough (COUGH), hesitation (START/START2), excessive time to peak flow (EXPFLOW) or lack of terminal plateau (END) in addition to accepted blows and applying a repeatability threshold of 350ml reached the maximum statistical significance in the association between GRS and FEV₁/FVC. However, as expected, the effect size in the association results attenuated toward zero as QC criteria were successively relaxed (Table 2).

To balance accuracy of effect size estimation versus maximising statistical significance (z-score) for GWAS, we examined the changes in the effect sizes and P values of the sentinels of FEV₁/FVC signals identified by Shrine *et al.* [3]. We found that additionally including blows previously excluded for cough (COUGH), hesitation (START/START2), excessive time to peak flow (EXPFLOW) or lack of terminal plateau (END), and applying a repeatability threshold of 250 ml optimised the signal-to-noise ratio in genetic association testing in UK Biobank (Figure 3). Based on this finding, we proposed a new spirometry QC strategy for epidemiological studies in UK Biobank, which retained 29% more participants using strictest ATS/ERS guidelines and increased the sample size from 275,084 to 356,053 individuals.

Illustrate the gain in power using newly defined QC criteria

Using the newly defined spirometry QC criteria, the UK Biobank sample size increased from 320,591 in the most recent GWAS of lung function to 356,053, an 11% gain in

Figure 2: Ranking the impact of acceptability flags on spirometer blows



a, GRS association with FEV₁/FVC derived from spirometer blows stratified by acceptability flags shown as the s.d. change in FEV₁/FVC per s.d. increase in GRS. N.B. A blow could be included in multiple acceptability flag strata if it carries multiple acceptability flags. b, Iterative selection process. c, GRS association with FEV₁/FVC derived from grouped spirometer blows shown as the s.d. change in FEV₁/FVC per s.d. increase in GRS. Group 1 represents blows with hesitation flags only ("START" or "START2"). Group 2 represents blows with cough, end-of-blow or time to peak flow flags ("COUGH", "END" or "EXPFLOW") but without flags of "REJECT", "CONSISTENCY" or "NEGATIVE". Group 3 represents blows with acceptability flags of "REJECT", "CONSISTENCY" or "NEGATIVE". The height of the bars shows the point estimate of the effect and whiskers show the 95% CI.

sample size, and the mean χ^2 statistic (i.e. squared z scores for SNP and FEV₁/FVC association) from GWAS increased from 1.27 to 1.29, leading to 15 additional sentinel SNPs associated with FEV₁/FVC (selected 2 Mb regions centred on the most significant variant for all regions containing a variant with $P < 5 \times 10^{-9}$), additional compared with analysis of UK Biobank alone using previous QC criteria [3]) (Supplementary Table 2). Of these, 7 were not identified in the largest GWAS of lung function to date. Examining the nearest genes to the sentinel SNPs, we found two novel genes in addition to 13 genes reported either in the most recent GWAS [3] or the EMBL-EBI GWAS Catalogue [17]. One of the novel genes, *FPR3*, encodes the formyl peptide receptor 3, a paralog of formyl peptide receptor 1. The functional role of *FPR3* is not fully understood. However, it is expressed in a range of immune cells, including macrophages and eosinophils, but not neutrophils, so has been hypothesised to play a role in allergic disease [18], and has also been associated with asthma and white blood cell counts in GWAS. We also found that enrichment of a previously implicated pathway [3], ESC pluripotency pathway, was strengthened by the newly identified gene *WNT16*.

For the newly identified hits, we followed procedures previously described by Okbay *et al.* [14] to calculate the posterior probability of true association, which exceeded 99% for all 15 additional loci for any assumption about the prior probability of non-null SNPs in the range 1% to 99%. To test the sentinel SNPs for replication, we used meta-analysis results from 42 European ancestry cohorts (excluding participants from UK Biobank) of 249,114 individuals generated by Shrine *et al.* [3]. Due to the limited statistical power for replication of genome-wide significant association with a smaller sample size, we applied methods to assess the replication of the effect size of sentinel SNPs. For the set of 13 newly identified sentinel SNPs for FEV₁/FVC available in the meta-analysis results, we regressed the effect sizes in UK Biobank on the effect sizes in the meta-analysis results with intercept constrained to be zero, after correcting the UK Biobank effect size estimates for winner's curse bias using the method described in Turley *et al.* [15]. The regression slope was 0.75 (standard error = 0.1376), being statistically significantly greater than zero (one-sided $P = 1.474 \times 10^{-4}$) but not statistically distinguishable from one (one-sided $P = 0.0943$), suggesting that the newly identified sentinel SNPs were replicated in independent datasets.

Table 2: The GRS association with FEV₁/FVC derived from blows with varying inclusion criteria and repeatability threshold

Blow inclusion criteria	Repeatability threshold (ml)	Beta ^a (95% CI)	Z-score (95%CI)	Sample size
No acceptability flags	250	0.266 (0.263,0.269)	163.0 (161.2,164.8)	320591
No acceptability flags OR	150	0.270 (0.267,0.274)	154.5 (152.8,156.8)	275957
hesitation acceptability	200	0.268 (0.264,0.271)	160.7 (158.3,162.5)	304931
flags only ("START" or	250	0.266 (0.263,0.270)	164.1 (162.2,166.5)	321724
"START2")	300	0.264 (0.261,0.267)	164.8 (162.9,166.7)	331685
	350	0.262 (0.259,0.265)	164.9 (163.0,166.7)	337893
	400	0.260 (0.257,0.263)	164.7 (162.8,166.6)	342155
No acceptability flags OR	150	0.270 (0.267,0.274)	154.7 (153.0,157.0)	276696
hesitation or cough	200	0.268 (0.264,0.271)	160.8 (158.4,162.6)	305834
acceptability flags only	250	0.266 (0.263,0.269)	164.3 (162.4,166.1)	322734
("START", "START2",	300	0.264 (0.261,0.267)	165.0 (163.1,166.9)	332806
"COUGH")	350	0.262 (0.259,0.265)	165.1 (163.2,167.0)	339076
	400	0.260 (0.257,0.263)	165.0 (163.1,166.9)	343379
No acceptability flags OR	150	0.265 (0.262,0.269)	154.0 (152.3,156.4)	285804
hesitation, cough or	200	0.263 (0.259,0.266)	160.1 (157.7,162.0)	317024
end-of-blow acceptability	250	0.261 (0.258,0.264)	163.6 (161.7,165.4)	335310
flags only ("START",	300	0.258 (0.255,0.262)	164.3 (162.4,166.8)	346314
"START2", "COUGH",	350	0.265 (0.253,0.259)	164.4 (162.4,166.3)	353359
"END")	400	0.255 (0.252,0.258)	164.4 (162.4,166.3)	358243
No acceptability flags OR	150	0.265 (0.262,0.268)	157.5 (155.7,159.3)	299668
hesitation, cough,	200	0.262 (0.258,0.265)	164.0 (164.4,165.8)	334952
end-of-blow or time to	250	0.260 (0.257,0.263)	167.6 (165.7,169.5)	356163
peak flow acceptability	300	0.257 (0.254,0.260)	168.4 (166.5,170.4)	369039
flags only ("START",	350	0.255 (0.252,0.258)	168.7 (166.7,170.7)	377568
"START2", "COUGH",	400	0.253 (0.250,0.255)	168.5 (166.5,169.8)	383545
"END", "EXPFLOW")				

^a stands for standard deviation (SD) change of FEV₁/FVC per SD change of GRS.

Discussion

QC of spirometric measurements involves a range of metrics and criteria which are designed to ensure that clinical decision-making is based on accurate and reproducible measurements. Spirometry is also widely used in epidemiological research, including genetic association studies, where the aims and acceptable thresholds for data quality may differ from those in clinical settings. Leading experts in spirometry and its QC have previously highlighted uncertainty regarding the threshold at which test quality becomes insufficient for inclusion in research studies, and in particular have noted that end-of-test criteria may be applied too stringently [9]. The solution proposed of visual inspection of spirometry curves is not feasible in large-scale association studies, and in this context unnecessary exclusions may impact on power for novel discovery.

We propose a new GRS-based method to define a QC strategy which maximises power while maintaining acceptable precision, enabling an 29% increase in sample size using strictest ATS/ERS guidelines in UK Biobank. This identified 15 additional genetic loci not found in analysis of UK Biobank using the previous QC criteria, of which 7 were not identified in the largest GWAS of lung function to date, and two implicated novel genes highlighting new biology of interest. Eight were

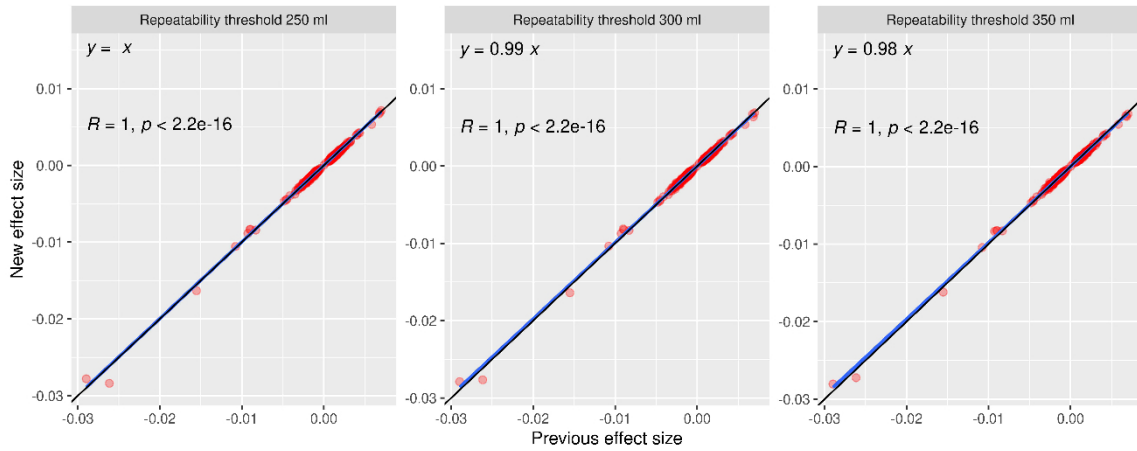
already identified in the largest consortium GWAS of lung function to date [3], but demonstrate that these discoveries could have been made earlier.

In this study, we introduce an iterative selection process to rank acceptability flags in descending order of their impact on spirometer blow quality failure. We then included blows with varying acceptability flags and repeatability threshold to refine the spirometry QC criteria for GWAS, with the aim to optimise the signal-to-noise ratio. Through the application of this strategy to lung function traits in UK Biobank, we demonstrated that the statistical power of GWAS can be increased by employing more inclusive spirometry QC criteria, as evidenced by substantial improvements in sample size, mean chi-square statistics and the identification of additional genetic association signals. These signals implicated two novel genes for lung function, one of which (*FPR3*) plays a role in innate immunity and has been implicated in asthma and allergic disease.

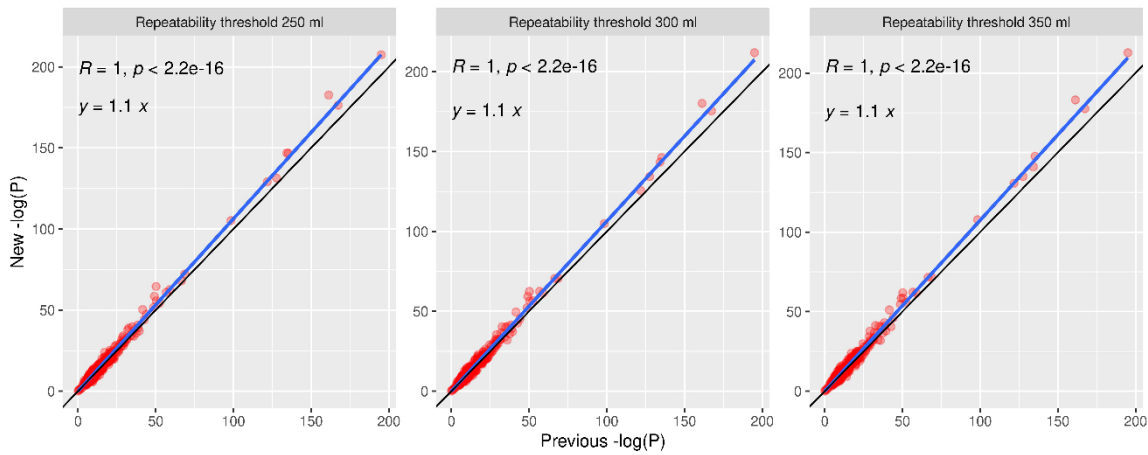
To date, although 1020 genetics signals have been discovered for lung function traits, much of the genetic contribution to lung function remains unexplained [3]. The problem of "missing heritability" could be explained by undetected common variant associations and rare variant associations. The approach we outline here will boost power for GWAS of common/ rare variants such as those now available

Figure 3: Examine the genetic association results of FEV₁/FVC signals estimated from relaxed blow inclusive criteria at varying repeatability threshold of 250 ml, 300 ml and 350 ml

a. Comparison of effect sizes



b. Comparison of P values



a, compare the effect sizes of FEV₁/FVC signals using relaxed spirometry QC (New effect size, including blows previously only excluded for cough, hesitation, excessive time to peak flow or lack of terminal plateau ("START", "START2", "COUGH", "END", "EXPFLOW")) to the result obtained from previous spirometry QC (Previous effect size, only including accepted blows). b, compare the p values of FEV₁/FVC signals using relaxed spirometry QC (New $-\log(P)$) to the result obtained from previous spirometry QC (Previous $-\log(P)$).

through the whole genome sequencing of UK Biobank [19], structural genomic variants as long read sequencing data become available. While the QC criteria established in UK Biobank may not be directly transferable to other studies, the same methodology can be applied. Our approach will be especially relevant where sample sizes are limited, such as in under-represented ancestries in genomic studies [20]. Furthermore, while genetic data are used to inform the spirometry QC, the increase in sample size and power from our approach could be applicable to a wide range of epidemiological research questions, such as assessing lung function associations with environmental factors or biomarker levels. Biomarker measures are becoming more available as biotechnology evolves. Moreover, this methodology could potentially be applicable to association studies of other complex traits such as thyroid stimulating hormone levels, where a recent genome-wide association study excluded 23.4% of measures outside the normal range [21].

Our study has several limitations that warrant consideration. We were not able to assess all ATS/ERS spirometry QC criteria [8]. Additionally, the new spirometry QC criteria derived from UK Biobank may not be directly generalisable to other cohorts, as different studies often employ spirometer models with distinct hardware, software and diagnostic algorithms. To fully meet 2019 ATS/ERS QC criteria, blows must also be free from glottic closure and from evidence of technical issues (faulty zero-flow setting, leak or obstruction). These are not generally included in automated spirometer output. Variations in error detections arise from differences in the software, firmware and diagnostic algorithms unique to each spirometer model. Consequently, defining customised QC criteria tailored to specific cohorts, as proposed in our methodology, is essential. Additionally, QC decisions informed by GRS may reinforce existing genetic correlations but, at the individual level, may not fully reflect physiological measures. Nevertheless, at the group level, we show that although the spirometry of some

participants groups does not meet standard QC criteria, their data show associations in our model that are comparable to those from “good” blows. Therefore, these data can be included in GWAS to increase statistical power. Our analysis could also benefit from a more powerful PRS, which would be expected to yield superior prediction performance.

Conclusion

In summary, our study highlights a useful application of GRS in epidemiological studies of lung function. GRS-informed QC boosts sample size and power for epidemiological studies, as illustrated by our discovery of new genetic associations for lung function. While the GRS-informed spirometry QC criteria may not directly translate to other epidemiological cohorts, the underlying methodology offers a framework for reassessing QC criteria that currently exclude informative measures and limit power in the epidemiological study of different traits.

Declarations

Acknowledgements

This research was supported by a Wellcome Discovery Award (WT 225221/Z/22/Z). The research was partially supported by the NIHR Leicester Biomedical Research Centre and through an NIHR Senior Investigator Award to MDT and IPH. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The research was conducted using UK Biobank, under applications 648. This research used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester.

Statement of conflict of interest

Catherine John, Martin D. Tobin and Anna Guyatt receive collaborative funding from Orion Pharma, unrelated to the submitted work. Louise V. Wain receive research funding from Orion, GSK and Genentech/Roche and has funded research collaborations with AstraZeneca, Nordic Bioscience and Sysmex (OGT), unrelated to the submitted work. Louise V. Wain has consultancy fees paid to institution from Galapagos, Boehringer Ingelheim and GSK. Louise V. Wain is on advisory board for Galapagos, associate editor for European Respiratory Journal and Medical Research Council Board member and Deputy Chair. Ian P. Hall is Vice Chair Trustees for Asthma + Lung UK (unpaid). The other authors declare no competing interests.

Ethics statement

The research was conducted using UK Biobank, under application 648. The study is carried out in UK Biobank, which has approval from the North West Multi-centre Research

Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval (21/NW/0157).

Data availability statement

UK Biobank data, including data on genomics and covariates used in our analyses, are available to approved bona fide researchers via application to UK Biobank. Weights for constructing the lung function traits genetic risk scores used in these analyses are publicly available from PGS Catalogue: <https://www.pgscatalog.org/publication/PGP000500/>. R scripts implementing this analysis are available from https://github.com/legeneipi/GRS_informed_QC.

References

1. Young, R.P., R. Hopkins, and T.E. Eaton, Forced expiratory volume in one second: not just a lung function test but a marker of premature death from all causes. *Eur Respir J*, 2007. 30(4): p. 616-22. <https://doi.org/10.1183/09031936.00021707>
2. Visscher, P.M., et al., 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, 2017. 101(1): p. 5-22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
3. Shrine, N., et al., Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk. *Nature Genetics*, 2023. 55(3): p. 410-422. <https://doi.org/10.1038/s41588-023-01314-0>
4. Shrine, N., et al., New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet*, 2019. 51(3): p. 481-493. <https://doi.org/10.1038/s41588-018-0321-7>
5. Graham, B.L., Pulmonary function standards: a work in progress. *Respir Care*, 2012. 57(7): p. 1199-200. <https://doi.org/10.4187/respcare.01903>
6. Haynes, J.M. and D.A. Kaminsky, The American Thoracic Society/European Respiratory Society Acceptability Criteria for Spirometry: Asking Too Much or Not Enough? *Respiratory Care*, 2015. 60(5): p. e113-e114. <https://doi.org/10.4187/respcare.04061>
7. Sylvester, K.P., et al., ARTP statement on pulmonary function testing 2020. *BMJ Open Respiratory Research*, 2020. 7(1): p. e000575. <https://doi.org/10.1136/bmjresp-2020-000575>
8. Graham, B.L., et al., Standardization of Spirometry 2019 Update. An Official American Thoracic Society and European Respiratory Society Technical Statement. *American Journal of Respiratory and Critical Care Medicine*, 2019. 200(8): p. e70-e88. <https://doi.org/10.1164/rccm.201908-1590ST>

9. Hankinson, J.L., et al., Use of forced vital capacity and forced expiratory volume in 1 second quality criteria for determining a valid test. *Eur Respir J*, 2015. 45(5): p. 1283-92. <https://doi.org/10.1183/09031936.00116814>
10. Choi, S.W., T.S. Mak, and P.F. O'Reilly, Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*, 2020. 15(9): p. 2759-2772. <https://doi.org/10.1038/s41596-020-0353-1>
11. Sudlow, C., et al., UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*, 2015. 12(3): p. e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
12. Miller, M.R., et al., Standardisation of spirometry. *European Respiratory Journal*, 2005. 26(2): p. 319-338. <https://doi.org/10.1183/09031936.05.00034805>
13. Loh, P.-R., et al., Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nature Genetics*, 2015. 47(3): p. 284-290. <https://doi.org/10.1038/ng.3190>
14. Okbay, A., et al., Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*, 2016. 48(6): p. 624-33. <https://doi.org/10.1038/ng.3552>
15. Turley, P., et al., Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nature Genetics*, 2018. 50(2): p. 229-237. <https://doi.org/10.1038/s41588-017-0009-4>
16. Alexander, D.H., J. Novembre, and K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 2009. 19(9): p. 1655-64. <https://doi.org/10.1101/gr.094052.109>
17. Cerezo, M., et al., The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Res*, 2025. 53(D1): p. D998-d1005. <https://doi.org/10.1093/nar/gkae1070>
18. Dorward, D.A., et al., The role of formylated peptides and formyl peptide receptor 1 in governing neutrophil function during acute inflammation. *Am J Pathol*, 2015. 185(5): p. 1172-84. <https://doi.org/10.1016/j.ajpath.2015.01.020>
19. Whole-genome sequencing of 490,640 UK Biobank participants. *Nature*, 2025. 645(8081): p. 692-701. <https://doi.org/10.1038/s41586-025-09272-9>
20. Fatumo, S., et al., A roadmap to increase diversity in genomic studies. *Nat Med*, 2022. 28(2): p. 243-250. <https://doi.org/10.1038/s41591-021-01672-4>
21. Williams, A.T., et al., Genome-wide association study of thyroid-stimulating hormone highlights new genes, pathways and associations with thyroid disease. *Nat Commun*, 2023. 14(1): p. 6713. <https://doi.org/10.1038/s41467-023-42284-5>



Supplementary methods

Constructing GRS and testing its association with lung function traits

We built European-specific GRSs in a multi-ancestry study for FEV₁, FVC and FEV₁/FVC consisting of 425, 372 and 442 autosomal signals (genome-wide significant threshold of $P < 5 \times 10^{-9}$) associated with each trait respectively, using weights estimated from a multi-ancestry meta-regression in 244,472 individuals across 42 cohorts (Supplementary Table 1) that are independent from the UK Biobank [1]. We calculated the GRSs for 406,474 European-ancestry individuals in UK Biobank. The associations between GRSs and lung function traits were tested using a linear model, adjusted for age, age squared, sex, height, smoking status and ten principal components.

For the first step of GRS association with lung function traits stratified all the blows by acceptability flag, a blow could be included in multiple strata if it had multiple acceptability flags. Within each acceptability flag stratum, we identified the best lung function measures (i.e. the best measure was defined as the highest measure for FEV₁ and FVC, whilst FEV₁/FVC was derived from the selected FEV₁ and FVC) for each participant and tested the association of the GRS with the measured lung function trait to identify which strata were significantly associated with the GRS.

Bayesian framework to assess the credibility of GWAS findings

For genome-wide association testing under an additive model using BOLT-LMM [2], we used untransformed residuals from linear regression of lung function traits against age, age², sex,

height, smoking status. To assess the credibility of our findings from the genetic association study using the relaxed spirometry QC criteria, we used a standard Bayesian framework previously described by Okbay *et al* [3] and Turley *et al* [4]. Briefly, we used maximum likelihood to fit the SNP effect from the GWAS result using a mixture of a Gaussian prior with a point mass at zero. The prior for effect size corresponding to any given SNP j is

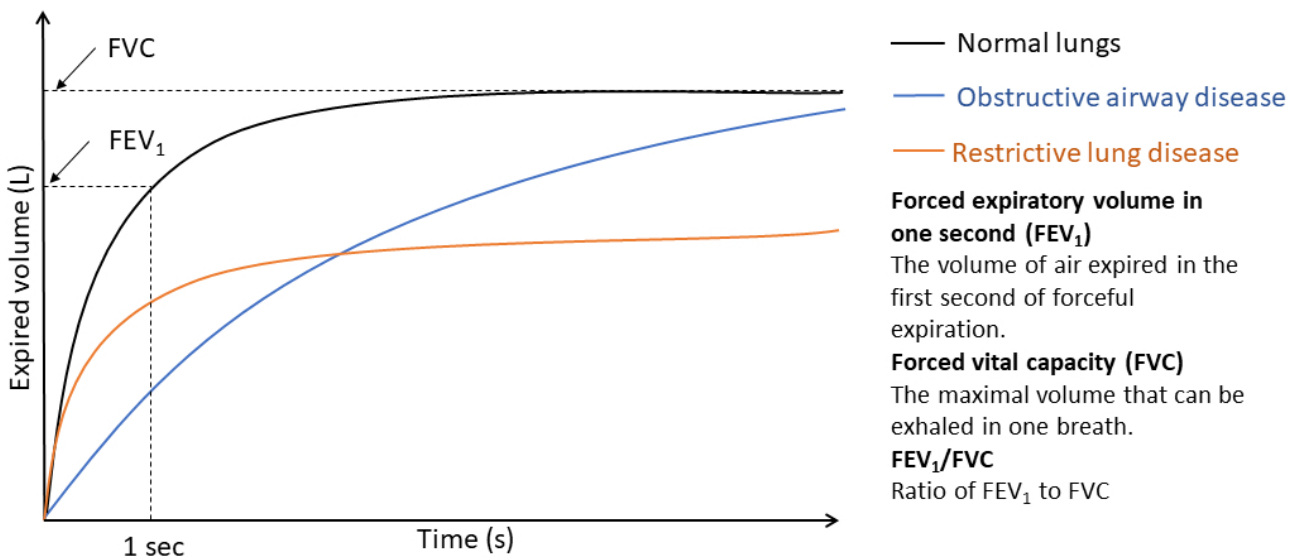
$$\beta_j \sim \begin{cases} N(0, \tau^2) & \text{with probability } \pi \\ 0 & \text{otherwise} \end{cases}$$

where τ^2 is the prior belief of the variance of effect size of non-null SNPs and π is the prior belief of the fraction of non-null SNPs. For each assumed prior π from 1% to 99%, we used the maximum likelihood estimates of parameters to calculate the posterior probability that a SNP is non-null given its estimated effect size and given that it is significant. To assess the replication of the effect size in independent study, the effect size for each SNP is corrected for winner's curse by

$$\hat{\beta}_{adj,j} = (1 - \pi_{posterior,j}) \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}_j^2} \hat{\beta}_j$$

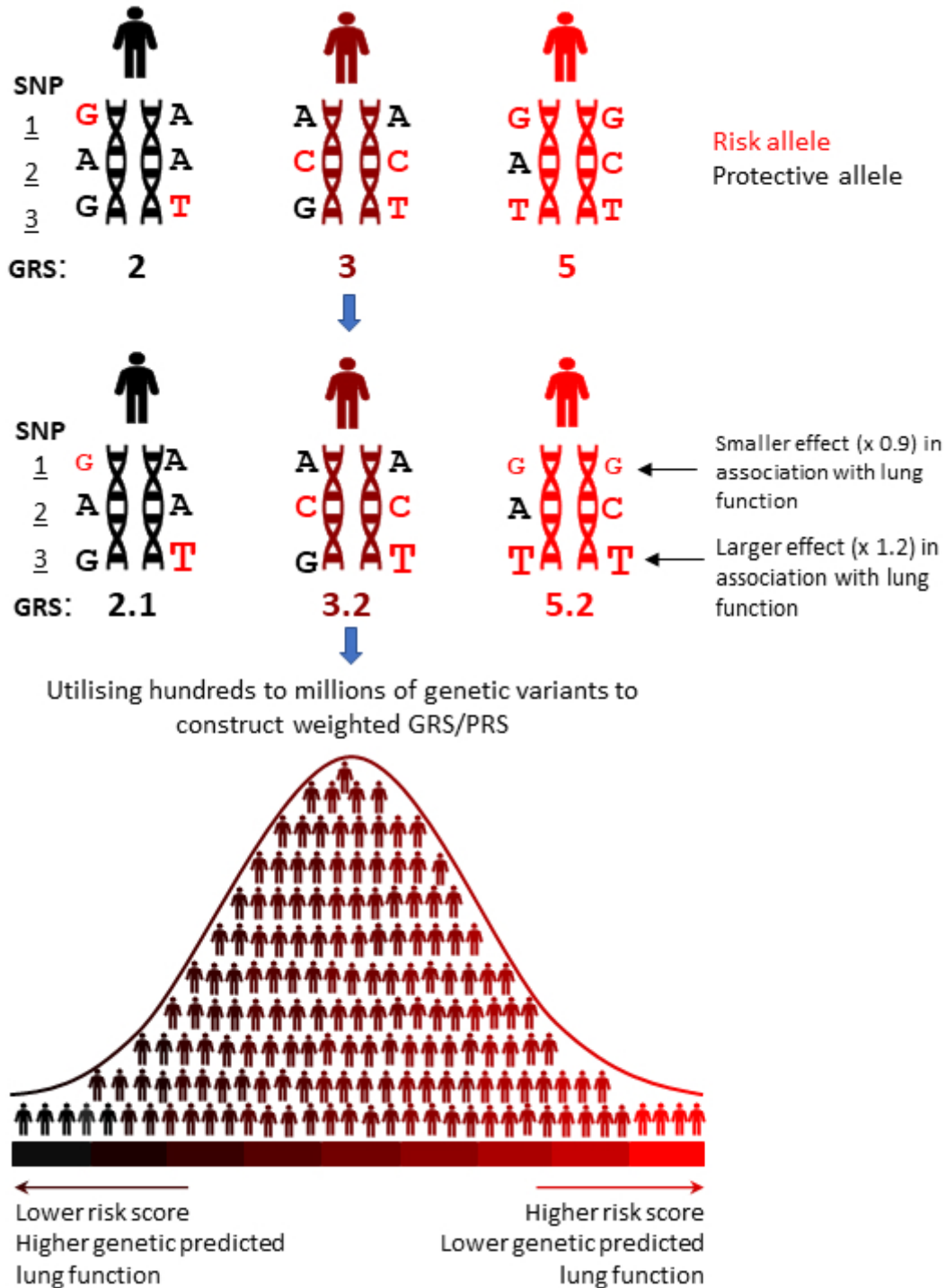
where $\pi_{posterior,j}$ is the posterior probability of the SNP j being non-null and $\hat{\sigma}_j^2$ is the squared standard error of the GWAS estimates for SNP j .

Supplementary Figure 1: Volume-time curve



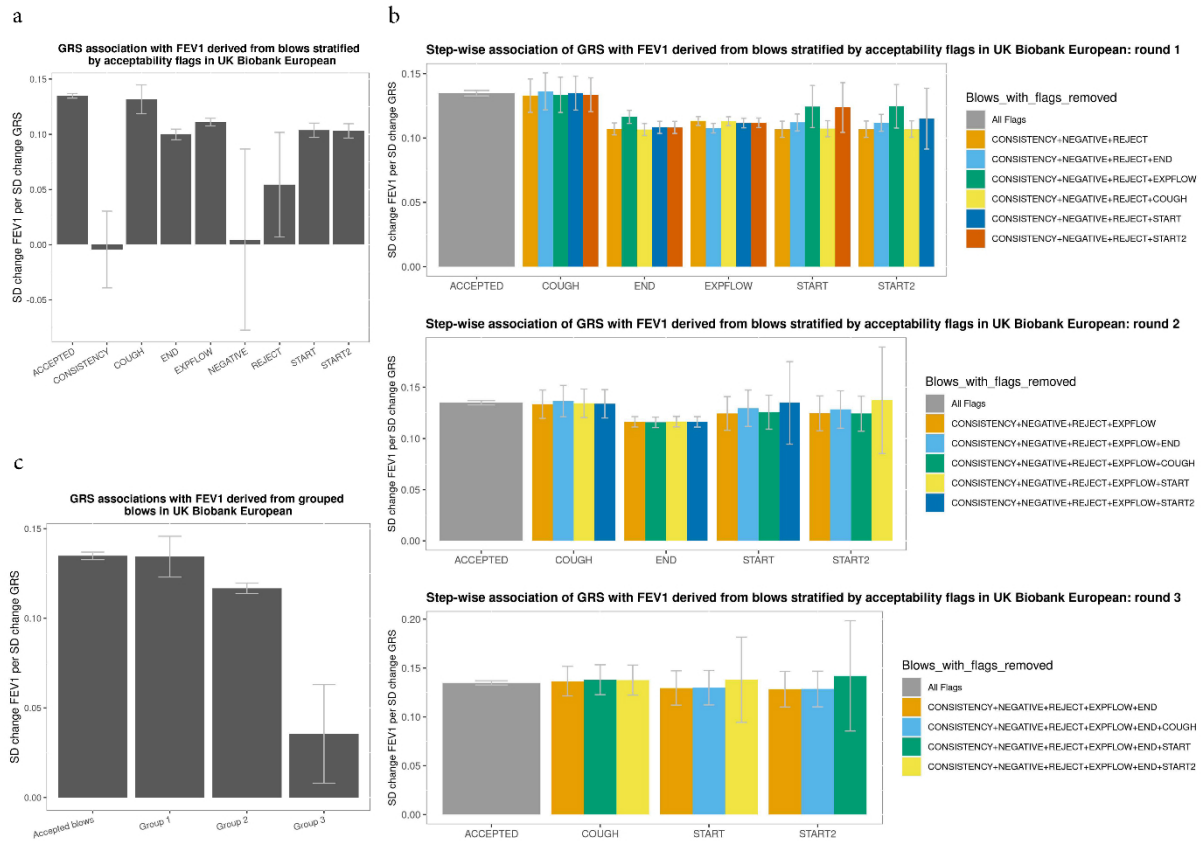
The curve plots the total volume of air expired by time, from full inspiration until full expiration. Normal blow shows rapid increase in volume of air expired initially, then curve forms a plateau; Obstructive blow shows prolonged increase but ends the same point; Restrictive blow shows rapid increase as normal, but curve forms a plateau much sooner.

Supplementary Figure 2: Basic principle of constructing GRS/PRS



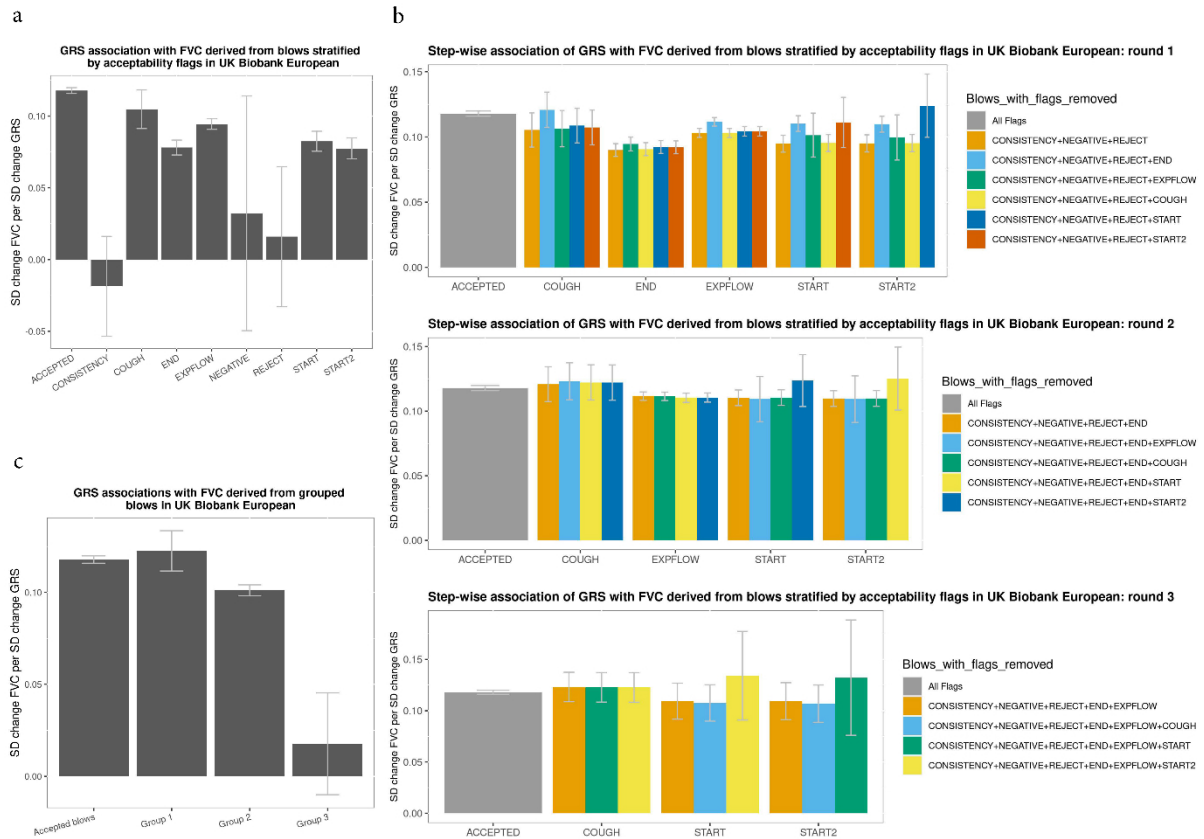
Schematic showing the principles of constructing a GRS. Top section (a) shows a theoretical example of an unweighted risk score is calculated at three SNPs, assuming an additive model; middle section (b) shows a weighted risk score for the same alleles, whereby the score for each allele is weighted according to its association with lung function; bottom section (c) shows a normal distribution of individual scores for a GRS/PRS using hundreds to millions of genetic variants, where the score at each variant is weighted according to its association with the trait of interest (lung function), ideally in an independent population.

Supplementary Figure 3: Ranking the impact of acceptability flags on spirometer blows



a, GRS association with FEV₁ derived from spirometer blows stratified by acceptability flags shown as the s.d. change in FEV₁ per s.d. increase in GRS. N.B. A blow could be included in multiple acceptability flag strata if it carries multiple acceptability flags. **b**, Iterative selection process. **c**, GRS association with FEV₁ derived from grouped spirometer blows shown as the s.d. change in FEV₁ per s.d. increase in GRS. Group 1 represents blows with hesitation flags only ("START" or "START2"). Group 2 represents blows with cough, end-of-blow or time to peak flow flags ("COUGH", "END" or "EXPFLOW") but without flags of "REJECT", "CONSISTENCY" or "NEGATIVE". Group 3 represents blows with acceptability flags of "REJECT", "CONSISTENCY" or "NEGATIVE". The height of the bars shows the point estimate of the effect and whiskers show the 95% CI.

Supplementary Figure 4: Ranking the impact of acceptability flags on spirometer blows



a, GRS association with FVC derived from spirometer blows stratified by acceptability flags shown as the s.d. change in FVC per s.d. increase in GRS. N.B. A blow could be included in multiple acceptability flag strata if it carries multiple acceptability flags. **b**, Iterative selection process. **c**, GRS association with FVC derived from grouped spirometer blows shown as the s.d. change in FVC per s.d. increase in GRS. Group 1 represents blows with hesitation flags only ("START" or "START2"). Group 2 represents blows with cough, end-of-blow or time to peak flow flags ("COUGH", "END" or "EXPFLOW") but without flags of "REJECT", "CONSISTENCY" or "NEGATIVE". Group 3 represents blows with acceptability flags of "REJECT", "CONSISTENCY" or "NEGATIVE". The height of the bars shows the point estimate of the effect and whiskers show the 95% CI.

Supplementary Table 1: Sample size in 42 non-UKB cohorts

Cohort	Ancestry	N Total	N Males	N Females
ALHS	EUR	2,848	1,447	1,401
ALSPAC	EUR	4,857	2,444	2,413
ARIC-EA	EUR	9,275	4,376	4,899
B58C	EUR	5,788	2,862	2,926
BHS	EUR	4,272	1,884	2,388
CHS-EA	EUR	3,220	1,259	1,961
Croatia-Korcula	EUR	2,520	929	1,591
Croatia-Split	EUR	961	379	582
Croatia-Vis	EUR	925	390	535
EPIC-Norfolk	EUR	20,771	9,664	11,107
EXCEED	EUR	1,334	818	516
FinnTwin	EUR	539	91	448
FHS	EUR	7,905	3,656	4,249
GS	EUR	16,048	6,633	10,415
H2000	EUR	3,808	1,655	2,153
HUNT	EUR	16,485	7,648	8,837
KORA-F4	EUR	1,474	717	757
KORA-S3	EUR	1,147	551	596
LBC1936	EUR	1,002	509	493
MESA-CAU	EUR	1,397	707	690
NFBC1966	EUR	5,078	2,417	2,661
NFBC1986	EUR	3,210	1,516	1,694
SAPALDIA-Edinburgh	EUR	2,584	1,338	1,246
SAPALDIA-Gabriel	EUR	1,355	656	699
Orcades	EUR	1,821	730	1,091
PIVUS	EUR	806	395	411
Raine Study	EUR	1,213	637	576
RS1	EUR	1,240	526	714
RS2	EUR	1,143	532	611
RS3	EUR	1,431	624	807
SHIP-START	EUR	1,759	860	899
SHIP-TREND	EUR	2,519	1,265	1,254
TwinsUK	EUR	4,227	380	3,847
UKHLS	EUR	7,442	3,290	4,152
VIKING	EUR	1,701	672	1,029
YFS	EUR	419	198	221
ARIC-AA	AFR	2,820	1,044	1,776
CHS-AA	AFR	635	238	397
MESA-AFR	AFR	908	439	469
HCHS-SOL	AMR	10,965	4,558	6,407
MESA-HIS	AMR	905	427	478
CKB	EAS	83,715	36,399	47,316



Supplementary Table 2: Additional signals for FEV₁/FVC using relaxed spirometry QC criteria

Nearest gene	SNP	CHR	Genomic position (GRCh37)	Effect allele	Other allele	Effect allele frequency	Beta	Se	P values	Annotation	Novel SNP *	Gene reported previously**	Involved in pathway implicated in Shrine et al. 2023
GNL2	rs2294875	1	38052445	C	T	0.39	0.0009	0.0001	2.50E-09	intronic	Yes	GWAS Catalog	No
LOC100422212	1:163587821_AG_A	1	163587821	AG	A	0.56	0.0008	0.0001	3.00E-09	intergenic	No	Shrine et al. 2023	No
ZAK	rs13004345	2	174037347	C	T	0.35	-0.0010	0.0002	1.20E-10	intronic	No	Shrine et al. 2023	No
SPATS2L	rs985256	2	201208692	A	C	0.22	0.0011	0.0002	1.30E-10	intronic	No	GWAS Catalog	No
LOC646736	rs758076580	2	227174417	C	CAA	0.87	-0.0013	0.0002	1.60E-09	intergenic	Yes	No	No
TRAIP	rs112908707	3	49865628	G	GA	0.84	0.0014	0.0002	8.90E-12	downstream	Yes	Shrine et al. 2023	No
LOC101929123	rs10022937	4	17062269	T	C	0.87	0.0013	0.0002	1.30E-09	intergenic	Yes	Shrine et al. 2023	No
CPVL	rs1861116	7	29130953	A	C	0.94	-0.0018	0.0003	7.60E-10	intronic	No	Shrine et al. 2023	No
WNT16	rs2536189	7	120973621	C	G	0.55	0.0009	0.0001	6.30E-10	intronic	Yes	GWAS Catalog	ESC Pluripotency Pathways
CCDC171	rs6474931	9	15617677	G	T	0.55	0.0009	0.0001	1.50E-09	intronic	No	Shrine et al. 2023	No
ATP6V0A2	rs2333838	12	124220354	A	C	0.41	0.0009	0.0001	7.70E-10	intronic	No	Shrine et al. 2023	No
SPATA13	rs12430586	13	24603722	C	T	0.62	0.0009	0.0001	2.60E-09	intronic	No	Shrine et al. 2023	No
NID2	rs2516599	14	52491754	T	C	0.34	-0.0009	0.0002	1.90E-09	intronic	No	GWAS Catalog	No
REC114	rs7176074	15	73833600	G	T	0.95	-0.0020	0.0003	1.50E-09	intronic	Yes	GWAS Catalog	No
FPR3	19:52300562_CTT_C	19	52300562	CTT	C	0.78	0.0010	0.0002	9.40E-10	intronic	Yes	No	No

*It was defined as novel if the sentinel SNP was not within 1 MB distance to the sentinels reported in Shrine *et al.* 2023 or if the sentinel SNP was not in LD with the sentinels reported in Shrine *et al.* 2023 [1] ($R^2 < 0.2$).

**This column indicates whether the nearest gene of the sentinel SNP was reported either in Shrine *et al.* 2023[1] or other studies in GWAS Catalog [5].



References

1. Shrine, N., et al., *Multi-ancestry genome-wide association analyses improve resolution of genes and pathways influencing lung function and chronic obstructive pulmonary disease risk*. Nature Genetics, 2023. **55**(3): p. 410-422.
2. Loh, P.-R., et al., *Efficient Bayesian mixed-model analysis increases association power in large cohorts*. Nature Genetics, 2015. **47**(3): p. 284-290.
3. Okbay, A., et al., *Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses*. Nat Genet, 2016. **48**(6): p. 624-33.
4. Turley, P., et al., *Multi-trait analysis of genome-wide association summary statistics using MTAG*. Nature Genetics, 2018. **50**(2): p. 229-237.
5. Cerezo, M., et al., *The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity*. Nucleic Acids Research, 2024. **53**(D1): p. D998-D1005.

