EBioMedicine 47 (2019) 58–64

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Determinants of impaired lung function and lung cancer prediction among never-smokers in the UK Biobank cohort

Matthew T. Warkentin [a,b], Stephen Lam [c,d], Rayjean J. Hung [a,b,*]

[a] Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health System, 60 Murray Street, M5T 3L9 Toronto, ON, Canada
[b] Dalla Lana School of Public Health, Department of Public Health Sciences, University of Toronto, 155 College Street, M5T 3M7 Toronto, ON, Canada
[c] Department of Respiratory Medicine, Department of Medicine, University of British Columbia, 2775 Laurel Street, 7th floor, Vancouver V5Z 1M9, British Columbia, Canada
[d] British Columbia Cancer Agency, 600 W 10th Ave, V5Z 4E6 Vancouver, British Columbia, Canada

ARTICLE INFO

Article history:
Received 29 May 2019
Received in revised form 23 August 2019
Accepted 24 August 2019
Available online 5 September 2019

ABSTRACT

Background: Impaired lung health represents a significant burden on global health, including chronic obstructive pulmonary disease (COPD) and lung cancer. Given its global health impact, it is important to understand the determinants of impaired lung function and its relation to lung cancer risk independent of smoking. However, to date, no study has evaluated determinants of impaired lung function in a cohort exclusively of never-smokers, who also represent a growing proportion of all lung cancers.

Methods: A total of 222,274 never-smokers with reproducible spirograms were identified in the UK Biobank population-based cohort and included in the analysis. Baseline volumetric measures of lung function, including forced expiratory volume in 1-s (FEV1) and forced vital capacity (FVC), were used to define lung function impairment. Determinants of impaired lung function were evaluated using Poisson regression with robust variance estimation. The added value of lung function in lung cancer prediction was evaluated using Fine and Gray regression accounting for the competing risk of all-cause mortality.

Findings: Lung function impairment was associated with low birthweight, ambient air pollution ($PM_{2.5}$ μg/mm^3), and overweight, after adjustment for other important risk factors. We observed modest improvement in discrimination by adding lung function to our lung cancer prediction model for never-smokers. The highest optimism-corrected AUC at 3 (0·700, 95% CI: 0·654–0·734) and 5 years (0·694, 95% CI: 0·658–0·736) included FEV1 (% of GLI predicted FEV1), while the highest AUC at 7 years was based on the inclusion of FEV1/FVC (0·722, 95% CI: 0·687–0·762).

Interpretation: We identified several modifiable risk factors associated with increased risk of lung function impairment among lifetime never-smokers in UKB. We achieved moderate discrimination for lung cancer risk-prediction for never-smokers, and found modest improvement with the inclusion of lung function.

Fund: This study was supported by a Canada Research Chair to RJH.

1. Introduction

Impaired lung health, including chronic obstructive pulmonary disease (COPD) and lung cancer, represents a significant burden on global health [1,2]. COPD has an estimated prevalence of 200 million persons worldwide, and is the 4th leading cause of death globally [1]. A recent comprehensive analysis of the Global Burden of Disease Study 2015 estimated that 3·2 million people died from COPD worldwide in 2015, a death rate per annum that has increased 12% since 1990 [2].

A recent study found that lung function indices were important predictors of all-cause mortality among several British cohorts of never-smokers, more so than systolic blood pressure and body mass index [3], highlighting the importance of studying lung function determinants for never-smokers. In addition, lung cancer is the leading cause of cancer-related mortality worldwide [4]. Given its significant global health impact, it is important to understand the determinants of impaired lung function, beyond the known demographic and anthropometric factors and tobacco smoking. While a number of studies have evaluated risk factors for lung impairment, most include a large proportion of ever-smokers, which may confound these findings. In addition, there is a dearth of studies that have looked at early life determinants.

Never smokers are of particular interest because they offer a clear insight into determinants of lung function without being confounded by tobacco smoking. While tobacco smoking is a major risk factor for COPD, an estimated 25–45% of all patients with COPD are never-

* Corresponding author at: 60 Murray St., Toronto, ON M5T 3L9, Canada.
E-mail address: Rayjean.Hung@lunenfeld.ca (R.J. Hung).

https://doi.org/10.1016/j.ebiom.2019.08.058

Research in context

Evidence before this study

Impaired lung health, including chronic obstructive pulmonary disease (COPD) and lung cancer, represent a significant burden on global health. Very little is known about the causes of impaired lung function beyond tobacco smoking, as previous studies often evaluated lung function determinants in cohorts including smokers, thus subject to confounding by smoking. Very few studies have evaluated the effect of early life factors on the risk of impaired lung function later in life. Furthermore, lung cancer among never-smokers is increasingly recognized as a separate disease entity, which represents a growing proportion of all lung cancers, especially in countries where smoking rates have been declining. While COPD is known to be associated with lung cancer, the lung cancer risk prediction model among never-smokers is underdeveloped. Therefore, we aim to investigate the determinants of impaired lung function specifically among never-smokers, and establish risk-predictions models for non-smoking lung cancer incorporating lung function measures.

Added value of this study

To our knowledge, this is the first study to comprehensively evaluate the role of lifestyle factors on impaired lung function exclusively in lifetime never-smokers. We identified several modifiable lifestyle factors which modified risk of lung function impairment, including air pollution in the form of particulate matter <2·5 μg, low birthweight, and overweight and obesity. Currently, never-smokers would not meet most lung screening enrollment criteria on the basis of absent smoking history. Therefore, we developed a lung cancer risk-prediction model that included lung function alongside lung cancer risk factors easily ascertainable at a physician visit. We were able to achieve moderate discrimination of lung cancers, and found modest improvement in discrimination when including lung function in the risk algorithm.

Implications of all the available evidence

Several modifiable lifestyle risk factors were identified and may help to shape the trajectory of lung health and development of COPD later in life. Increasing our understanding of risk factors which contribute to lung morbidity among those without tobacco use as a primary factor may, in the future, help to guide prevention strategies and enhance our ability to identify never-smokers who could benefit from lung screening initiatives.

smokers [5]. Additionally, the proportion of lung cancers occurring among never-smokers is increasing [6,7], and it is increasingly recognized as a separate disease entity from tobacco-related lung cancer with different etiological factors and somatic mutation profiles [8]. If considered as a separate disease from tobacco smoking- related lung cancers, it ranks seventh among all cancers [8]. Currently never-smokers do not meet the guidelines (e.g. U.S. Preventive Services Task Force Recommendation Statement: Screening for Lung Cancer) [9], or risk-probability thresholds for screening eligibility [10]. Even though the causal role of impaired lung function in lung cancer risk remains unclear, impaired lung function may be leveraged as a useful risk predictor for identifying those at high risk who should undergo regular screening [11]. However, the added predictive value for lung function among never-smokers remains unknown.

To address these gaps in knowledge, we conducted this study to characterize the determinants of impaired lung function in lifetime never-smokers, particularly, the effect of exposure over the life- course from early life determinants to environmental and lifestyle exposures during adult life. Given the growing proportion of lung cancer in never-smokers and their ineligibility for CT screening under current guidelines, we also investigated the added value of lung function for lung cancer prediction.

2. Methods

2.1. Study participants: UK Biobank

This study is based on data collected in the UK Biobank cohort, the details of which have been described elsewhere [12]. In brief, this prospective study enrolled 502,616 participants aged 40–69 years between 2006 and 2010 at 22 assessment centres throughout the UK. The assessment centre visit comprised of questionnaires, physical and functional measures, and collection of biospecimens. Never-smokers are defined as self-reported non-current smokers with past smoking history reported as “never”, or with <100 lifetime cigarettes smoked for those with past history reported as “once or twice” or “occasional”; this was chosen to be consistent with the commonly used definition of never-smokers (<100 lifetime cigarettes).

Volumetric measures of lung function for this study were: forced-expiratory volume in 1-s (FEV1), forced vital capacity (FVC), and FEV1/FVC. We first excluded any participants without at least two spirograms with acceptable starts. For each participant, we then compared each FEV1 to their maximum FEV1, and spirograms were considered reproducible if they were within 250 mL of the maximum FEV1, based on standard spirometry guidelines [13]. Among reproducible spirograms, the maximum FEV1 and FVC for each participant were selected and used to derive FEV1/FVC. The Global Lung Initiative (GLI) 2012 equations were used to determine reference lung volumes for FEV1 to compute percent predicted FEV1 [14]. Impaired lung function was defined three ways: (1) FEV1 < 80% of the GLI predicted FEV1 reference value, (2) FEV1/FVC < 70%, or (3) both of these (i.e. moderate to severe COPD). Definition 2 and 3 align with COPD guidelines for airflow obstruction grading according to both Global Initiative for Chronic Obstructive Lung Disease (GOLD) and National Institute for Health and Care Excellence (NICE) [15,16]. As a sensitivity analysis, we also analyzed FEV1/FVC and FEV1 (% of GLI predicted FEV1) as continuous traits based on general linear models.

Air pollution was estimated as part of the European Study of Cohorts for Air Pollution Effects (ESCAPE) [17,18]. Physical activity was calculated as the sum of the duration of vigorous, moderate, and walking activities (minimum of 10 min and truncated at 300 min per day) multiplied by their metabolic equivalents (MET) values (vigorous = 8·0, moderate = 4·0, walking = 3·3), and then discretized into quartiles. Environmental tobacco smoke (ETS) exposure (yes vs. no) was based on any self-reported exposure to tobacco smoke at home or outside the home. Educational attainment was coded by mapping UKB qualifications to International Standard Classification of Education (ISCED) categories, as was done in previous studies in this cohort [19].

Linkages to National disease and death registries were used to identify incident lung cancers and mortality status. We excluded any never-smokers diagnosed with any respiratory cancer prior to recruitment (n = 94). Additionally, participants with lung cancers occurring <2 years from recruitment (n = 37) were excluded from all lung function analyses to minimize potential confounding due to subclinical lung cancer at baseline.

For lung cancer risk prediction modeling, we included only the first primary lung cancer occurrence. To exclude lung cancers that were potential metastases or recurrences, lung cancers occurring within 5 years after another primary cancer diagnoses were excluded. Deaths due to all-causes (excluding lung cancer-specific deaths) were considered a competing risk for lung cancer. Participants without primary lung cancer were observed for follow-up until death or September 30, 2014

(last date of complete cancer linkage), whichever came first. Due to incomplete cancer linkage beyond this date, deaths occurring after September 30, 2014 ($n = 1379$) were censored at this date, as a preceding lung cancer diagnosis could not be effectively ruled-out.

2.2. Statistical analysis

To evaluate lifestyle factors associations with impaired lung function, Poisson regression models with robust variance estimation were used to estimate incidence rate ratios (IRR) and 95% confidence intervals (CI) [20], excluding those with prevalent lung cancer, defined as lung cancer diagnosis before or up to 2 years after baseline. Potential non-linear relationships for continuous covariates with the outcomes were evaluated by adding polynomial terms to the regression models and improvements in model fit were determined using likelihood ratio tests. Stratified analysis by sex was performed and is presented in Supplementary Table 4. Given the mechanism of lung function impairment may differ in asthmatics vs. non-asthmatics, we also conducted stratified analysis by asthma status (Supplementary Table 5).

We constructed prediction models for lung cancer risk among never-smokers using competing-risk regression based on Fine and Gray models. Absolute risks of lung cancer (cumulative incidences) were calculated using Breslow-type estimates of the cumulative baseline subhazard and the linear predictor from the Fine-Gray models (see Supplementary Methods for more details). We included age, sex, personal cancer history, family history of lung cancer, and lung function in risk models based on their consistent associations with lung cancer. Model calibration was assessed by plotting observed and predicted risks across quantiles of predicted risks. Model discrimination was evaluated using time-dependent area under the receiver operating characteristic curves (AUC). We computed optimism-corrected AUCs using 100 bootstrap replicates [21]. We tested the improvement in model fit when including lung function in models using likelihood ratio tests, which has been recommended over directly testing differences in AUCs [22].

Missing data were generally minimal (<10%) for the potential determinants of impaired lung function, such as BMI, asthma, family history of lung cancer, previous cancer history, ETS exposures, alcohol drinking, and air pollution. Since we aim to estimate the association of each predictor simultaneously, only participants with complete data for all predictors were included in ($n = 181,805$) analyses. Given that birth weight and physical activity were only available in a subset of the population (58% and 78%, respectively), we conducted the analysis for these two determinants in separate models adjusting for the rest of the determinants as covariates based on the subset with complete data ($n = 88,874$).

To investigate whether any associations in complete-case analyses were biased due to missing data, we compared the study characteristics between those that were included in the analysis versus total population, and the frequency distributions were similar (Supplementary Table 2). In addition, we performed sensitivity analysis using multiple imputed data and results were materially similar (Supplementary Table 3). We therefore report the results based on the complete-case analysis herein. All data preparation and analyses were done using Stata 14 (Stata Corporation, College Station, Texas) and R version 3.5.3 [23].

2.3. Role of the funding source

This study was support by a Canada Research Chair to RJH. The funding source had no role in any aspects of the conception, design, implementation, analysis, interpretation of this study, or the decision to submit the paper for publication.

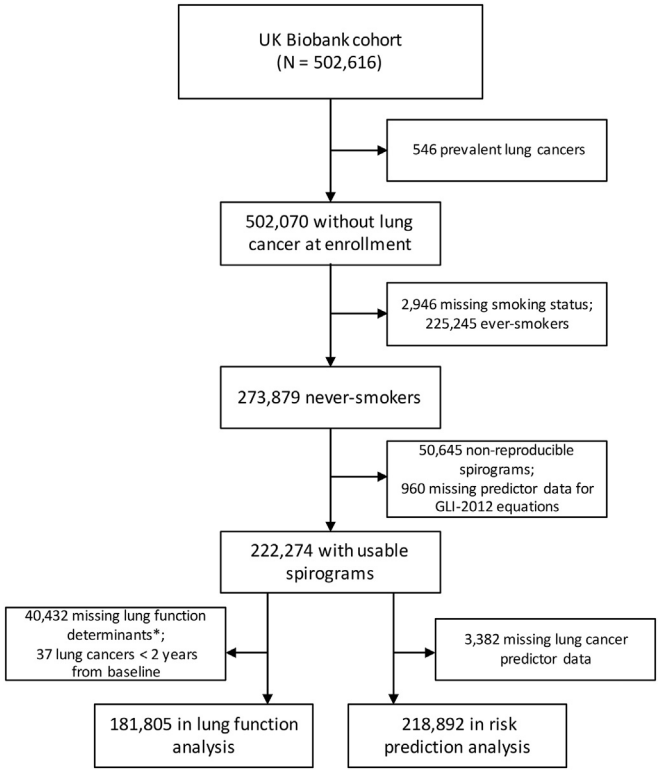

Fig. 1. Study flow of participants used in lung function and lung cancer risk-prediction analyses. Additional lung function analyses were performed including physical activity and birthweight as risk factors, which included a total of 88,874 participants after removing additional missing data for these variables.

3. Results

In the UK Biobank study, there were 222,274 never-smokers with reproducible spirograms (see Fig. 1). A total of 218,892 were included for risk prediction modeling after excluding those with missing predictor data. Based on 218,892 participants and 1·23 million person-years of follow-up, we identified 165 incident lung cancers among never-smokers through cancer registry linkage with a median follow-up of 5·6 years. Among never-smokers, the incidence rate of lung cancer was 13·4 per 100,000 person-years (95% CI: 11·5–15·6). Never-smokers accounted for 13% of the total lung cancers in UKB. In terms of competing risks, there were 2662 deaths due to all-causes (other than lung cancer-related mortality) and the mortality rate was 216·1 per 100,000 person-years (95% CI: 208·0–224·4). Study characteristics are summarized in Table 1 according to lung cancer status. Study characteristics by lung function impairment status are summarized in Supplementary Table 1. In general, never-smokers with lung cancer were older at baseline and more likely to be female, and had generally poorer lung function. Lung impairment was nearly 1·5 to 2-fold more common in those who developed lung cancer (FEV1 < 80%: 22·6% vs. 15·9%, $P_{\chi^2} = 0·018$; FEV1/FVC < 70%: 17·9% vs. 11·4%, $P_{\chi^2} = 0·008$; Both: 10·7% vs. 5·1%, $P_{\chi^2} = 0·019$).

A total of 181,805 never-smoking participants were included in the multivariable regression to identify determinants of impaired lung function, after the exclusions described in the Methods. The association between putative risk factors and lung function impairment, estimated as incidence rate ratios (IRR) based on multivariable-adjusted associations from Poisson models are shown in Table 2. As expected, there is a strong association between asthma history and impaired lung function, with IRR of 2·25 (95% CI: 2·20–2·30), 3·17 (95% CI: 3·09–3·25), 5·17 (95% CI: 4·97–5·38) for FEV1, FEV1/FVC, and both criteria, respectively. This association was stronger in males than in females in stratified analysis (Supplementary Table 4). Exposure to ambient air

Table 1
Characteristics of never-smokers with reproducible spirograms according to lung cancer status in the UK Biobank cohort ($N = 222,274$).

Variables	Healthy ($n = 222,106$)	Lung cancer ($n = 168$)
Age at recruitment (y, mean [SD])	55·6 [8·1]	60·7 [6·5]
Male sex (%)	85,530 (38·5%)	50 (29·8%)
Body mass index (kg/m^2, mean [SD])	27·0 [4·8]	26·6 [4·1]
Ethnicity (%) [a]		
White	208,236 (93·8%)	161 (95·8%)
Black	4123 (1·8%)	2 (1·2%)
Northeast (NE) Asian	992 (0·4%)	0 (0·0%)
Southeast (SE) Asian	5557 (2·5%)	3 (1·8%)
Other/mixed	3198 (1·4%)	2 (1·2%)
University education (%)	81,412 (37·0%)	61 (37·0%)
Positive family history of lung cancer (%)	25,950 (11·9%)	19 (11·5%)
Positive personal history of cancer (%)	15,761 (7·1%)	21 (12·5%)
Asthma, yes (%)	25,730 (11·6%)	17 (10·2%)
Birth weight		
Low (< 2500 g)	13,729 (10·6%)	14 (15·4%)
Normal (2500–4200 g)	107,178 (82·9%)	69 (75·8%)
High (> 4200 g)	8438 (6·5%)	8 (8·8%)
$PM_{2·5}$ (mcg/m^2, mean [SD])	1·0 [0·1]	1·0 [0·1]
ETS exposure, yes (%)	39,892 (19·8%)	28 (18·3%)
FEV1 maximum (L, mean [SD])	2·8 [0·8]	2·5 [0·7]
FVC maximum (L, mean [SD])	3·7 [1·0]	3·4 [0·9]
FEV1/FVC (%, mean [SD])	76·8 [6·4]	75·3 [7·2]
GLI percent predicted FEV1 (%, mean [SD]) [b]	94·5 [15·9]	91·5 [19·2]
Lung impairment (%)		
FEV1 < 80% reference FEV1	35,263 (15·9%)	38 (22·6%)
FEV1/FVC < 70%	25,346 (11·4%)	30 (17·9%)
Both (moderate-to-severe COPD)	11,290 (5·1%)	18 (10·7%)

Abbreviations: COPD, chronic obstructive pulmonary disease; ETS, environmental tobacco smoke; FEV1, forced expiratory volume, 1-s; FVC, force vital capacity; GLI, Global Lung Initiative; PM, particulate matter; SD, standard deviation.

[a] Ethnicities recorded for UK Biobank were grouped into GLI ethnic groups (White, Black, NE Asian, SE Asian, Other/Mixed) in order to use of the GLI-2012 equations to predicted reference FEV1 values.

[b] Reference FEV1 values were derived using the GLI-2012 equations (Quanjer et al. [14]) based on age, sex, height, and ethnicity.

pollution in the form of $PM_{2·5}$ micrograms per cubic meter was associated with increased risk of impaired lung function across all 3 criteria definitions, with IRRs ranging from 1·29–1·85 (per 10 micrograms per cubic meter). $PM_{2·5}$ was a stronger association among non-asthmatics, across all of the definitions of lung function impairment (Supplementary Table 5).

In terms of alcohol consumption, each category of drinkers had lower risk of lung impairment when compared to abstainers, according to impaired FEV1, FEV1/FVC, or both. For FEV1 lung impairment (FEV1 < 80% predicted), the largest difference in risk was between abstainers and daily consumers of alcohol (0·83 [95% CI: 0·79–0·87]). Increased physical activity was found to be associated with decreased risk based on FEV1 (per quartile, IRR = 0·94, 95% CI: 0·93–0·95), but not based on FEV1/FVC or both combined. For body mass index, a trend was observed for increased risk of impaired FEV1 with obesity, but the reverse trend was shown for FEV1/FVC. ETS was low variance and lacked detailed dosage data, but were suggestive of a modest increase in risk for lung impairment. No consistent relationship was found for a previous cancer (excluding lung cancers) history, or for family history of lung cancer.

In terms of early life determinants, an individuals with low birth weight (< 2500 g) had an increased risk of lung function impairment with IRR of 1·19 (95% CI: 1·13–1·25), 1·06 (95% CI: 1·00–1·13), and 1·21 (95% CI: 1·10–1·33), based on impaired FEV1, FEV1/FVC, or both, respectively. No significant associations were observed for maternal smoking status (data not shown). The results of analyzing lung function as continuous traits are shown in Supplementary Table 6, where we observed similar direction of the associations based on the general linear estimates.

For never-smoker lung cancer risk-prediction, age was the strongest single predictor among the set of variables we evaluated. The model without lung function included the predictors age, sex, personal history of cancer (yes vs. no), and family history of lung cancer (yes vs. no). The model with lung function included all of these predictors plus FEV1/FVC. Exponentiated model coefficients for the Fine and Gray model (sub-distribution hazard ratios) are shown in Table 3. We computed absolute

Table 2
Poisson regression (with robust variance estimation) estimates for the adjusted associations of risk factors on impaired lung function according to FEV1 (<80% GLI predicted FEV1), FEV1/FVC <70%, or both criteria, among lifetime never-smokers in the UK Biobank cohort.

	FEV1 < 80%	FEV1/FVC < 70%	Both
	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)
(a) Based on complete-case analysis (N = 181,805)			
Asthma (yes vs. no)	2·25 (2·20–2·30)	3·17 (3·09–3·25)	5·17 (4·97–5·38)
$PM_{2·5}$ (per 10 micrograms)	1·29 (1·16–1·43)	1·85 (1·64–2·09)	1·80 (1·49–2·17)
BMI groups			
Normal (<25 kg/m^2)	1·0 (ref.)	1·0 (ref.)	1·0 (ref.)
Overweight (25–30 kg/m^2)	1·11 (1·08–1·13)	0·73 (0·72–0·76)	0·88 (0·84–0·91)
Obese (≥30 kg/m^2)	1·52 (1·48–1·56)	0·58 (0·56–0·60)	0·83 (0·79–0·88)
Family history of lung cancer (yes vs. no)	0·96 (0·93–1·00)	0·98 (0·94–1·01)	0·95 (0·90–1·01)
Previous cancer (yes vs. no)	1·08 (1·04–1·12)	1·02 (0·97–1·06)	1·03 (0·96–1·11)
Alcohol status			
Never	1·0 (ref.)	1·0 (ref.)	1·0 (ref.)
Rarely	0·89 (0·86–0·92)	0·90 (0·85–0·94)	0·87 (0·81–0·94)
Monthly/weekly	0·84 (0·81–0·87)	0·95 (0·91–0·99)	0·89 (0·83–0·96)
Daily	0·83 (0·79–0·87)	0·95 (0·90–1·00)	0·91 (0·84–0·99)
ETS exposure (yes vs. no)	1·07 (1·05–1·10)	1·02 (0·99–1·05)	1·04 (0·99–1·10)
(b) Based on complete-case analysis in the subset ($N = 81,874$)			
Birth weight groups			
Low (<2500 g)	1·19 (1·13–1·25)	1·06 (1·00–1·13)	1·21 (1·10–1·33)
Normal (2500–4200 g)	1·0 (ref.)	1·0 (ref.)	1·0 (ref.)
High (>4200 g)	0·90 (0·85–0·97)	0·99 (0·92–1·06)	0·95 (0·85–1·07)
MET quartiles (per quartile) [a]	0·94 (0·93–0·95)	1·01 (0·99–1·03)	0·98 (0·95–1·00)

Abbreviations: BMI, body mass index; ETS, environmental tobacco smoke; FEV1, forced expiratory volume, 1-s; IRR, incidence rate ratio; MET, metabolic equivalents; PM, particulate matter.

Note: (a) Multivariable regression models including all determinants listed under a; (b) multivariable regression model including all variables listed under both (a) and (b).

Note: All models are adjusted for age, sex, ethnicity, height, education, and household income. Education (qualifications) in the UKB were mapped to their nearest equivalent according to the International Standard Classification of Education as has been done in previous studies using this cohort.

[a] The boundaries used to discretize metabolic equivalents into quartiles were: <824, 824–1806, 1806–3653, >3653.

Table 3
Subdistribution hazard ratios (SHR) and beta coefficients (log-SHR) from multivariable Fine & Gray competing risk regression models for the risk of lung cancer, with the competing risk of all-cause mortality, among never-smokers in the UK Biobank cohort.

	With lung function		Without lung function[a]	
	Beta coefficients	SHR (95% CI)	Beta coefficients	SHR (95% CI)
Age (per year) [b]	0·083100	1·09 (1·06–1·11)	0·085519	1·09 (1·07–1·11)
Sex (male vs. female)	−0·399,467	0·67 (0·48–0·94)	−0·387,205	0·68 (0·48–0·95)
Family history of lung cancer (yes vs. no)	−0·156,943	0·85 (0·53–1·38)	−0·158,382	0·85 (0·53–1·38)
Previous cancer (yes vs. no)	0·331,328	1·39 (0·87–2·24)	0·334,233	1·40 (0·87–2·24)
FEV1/FVC (per 5% increase) [b]	−0·071250	0·93 (0·84–1·04)	*NA*	*NA*

Abbreviations: CI, confidence interval; FEV1, forced expiratory volume, 1 s; FVC, forced vital capacity; NA, not applicable; SHR, subhazard ratio.

Note: Absolute risk predictions were made by computing the cumulative incidence function (CIF) for cause k as $\hat{I}_k(t) = 1 - \exp(-\hat{H}_k(t))$, where k is lung cancer, t is the time horizon, and $\hat{H}_k$ is the cumulative subdistribution hazard for lung cancer for a particular covariate pattern using a Breslow-type estimator of the baseline cumulative subdistribution hazard. More details are provided in the Supplementary Methods.

[a] The P-value for the likelihood ratio test (LRT) for the improvement in model fit when including lung function was 0·216.

[b] Age was centered to 55 years. FEV1/FVC was centered to 75% and then scaled to a 5% change in FEV1/FVC. Therefore, the cumulative baseline subdistribution hazard used for computing absolute risks represents the baseline risk for a 55-year old, female, no family history of lung cancer, no personal history of other cancers, with a FEV1/FVC of 75%.

risks of lung cancer at 3, 5, and 7-years to assess time-dependent model calibration and discrimination (AUC). Model calibration for the Fine-Gray model including lung function at fixed-year time horizons are presented in Supplementary Fig. 1, and the model-predicted absolute risks and observed risks showed moderate calibration across model-predicted risk quantiles.

Due to the relatively rare occurrence of lung cancers among never-smokers in this population-based cohort, we decided against using split-sample methods for model training and validation. Instead, model discrimination was internally validated by computing optimism-corrected AUCs using bootstrap methods with 100 bootstrap replicates. The 3, 5, and 7-year optimism-corrected AUCs and bootstrap quantile-based confidence intervals for the models without lung function were 0·686 (95% CI: 0·642–0·733), 0·687 (95% CI: 0·647–0·729), and 0·719 (95% CI: 0·688–0·763), respectively. The optimism-corrected time-dependent AUCs for the full model (incl. FEV1/FVC) at 3, 5, and 7 years were 0·686 (95% CI: 0·638–0·733), 0·688 (95% CI: 0·647–0·730), and 0·722 (95% CI: 0·687–0·762), respectively. Minimal improvement in model fit was observed when including FEV1/FVC ($P_{LRT} = 0·22$, Table 3). We note that FEV1 (% of GLI predicted FEV1) was modestly informative as a lung function predictor at 3- and 5-year time horizons, conferring a 1·4% and 0·7% AUC gain, respectively (Supplementary Table 7).

4. Discussion

Based on one of the largest analyses of never-smokers conducted to date, we identified an association between low birthweight, ambient air pollution ($PM_{2·5}$), obesity, alcohol intake and the risk of impaired lung function. However, we found limited improvement in predictive performance when incorporating lung function into the risk prediction model for lung cancer among never-smokers.

4.1. Impaired lung function

Our study found that low birthweight is a risk factor for impaired lung function or COPD in adulthood among a large cohort of never-smokers with no confounding by smoking. Previous studies with much smaller sample sizes found that those with low birthweight have poorer lung function in adulthood, but these studies contain large proportions of smokers [24,25]. To our knowledge, this is the very first time the association between low birthweight and impaired lung function has been reported among otherwise healthy never-smokers.

Our finding that increased exposure to ambient air pollution in the form of PM2·5 increased lung impairment is consistent with previous studies [26–29]. However, many previous studies contained considerable proportions of former and current smokers, and were limited by their cross-sectional nature, smaller sample sizes, and self-report exposures. Particulate matter are considered Group I carcinogens according to the International Agency for Research on Cancer (IARC) and has been consistently associated with increase lung cancer risk [30]. Similarly, the finding that increased physical activity is associated with better lung function in otherwise healthy adults is consistent with previous studies in the literature [31,32]. The possibility of reverse-causation cannot be ruled out in our study, and it is likely that the relationship between physical activity and lung function is bi-directional.

The association of alcohol consumption with lung function impairment and COPD is contentious. Early studies of alcohol consumption on COPD included notable proportions of heavy-drinkers (including alcoholics) and smokers that may partially confound the results [33]. In the early 2000's, two epidemiologic studies from Europe reported a U-shaped dose-response relationship between alcohol and COPD, similar to those reported for cardiovascular mortality [34,35]. The hypothetical mechanisms are that low alcohol exposure may enhance mucociliary clearance, stimulate bronchodilation, and may attenuate airway inflammation and injury found in those with COPD, while chronic and high-dose alcohol exposure likely worsens lung function. In our study, we observed higher risk of impaired lung function among abstainers. This is likely to be due to the difference in general health background of the complete abstainers, rather than a true causal relationship. However, the alcohol consumption in this healthy population-based cohort is unlikely to be extreme enough to observe the deleterious effects expected at chronically-high doses.

We observed generally null associations for ETS exposure and lung impairment. This was likely due, at least in part, to the sparsity of ETS data collected, and the low-dose and low-frequency SHS exposure among these never-smokers. The effect of ETS on pulmonary function and COPD has been well-described in the literature and several synthesizing studies have been conducted and evidence suggests SHS increases COPD. Although, weaknesses in existing studies, such as potential confounding by smoking need to be addressed to provide more definitive evidence [36,37].

We observed opposite effects for the association of BMI with FEV1 and FEV1/FVC ratio, with overweight being associated with increased risk of impairment according to FEV1, but a decreased risk for a reduced FEV1/FVC ratio (<70%). We also observed that compared to normal weight, overweight and obesity were associated with a decrease in absolute FEV1 and FVC in adjusted analyses, but an increase in the FEV1/FVC ratio. This association remained when body fat percentage was tested in place of BMI (results not shown), which suggests this finding is unlikely to be residually confounded by height. Previous studies have similarly observed a decrease in FEV1, FVC, and total lung capacity with increasing BMI [32,38–41]. One potential explanation for this finding is that being overweight can result in the diaphragm being pushed upward, effectively reducing the FVC and can artificially increase the

FEV1/FVC ratio. Overall, these findings suggest the effect of obesity is primarily on lung volume and not airway obstruction [40].

For the sex-stratified analysis, there were minor differences in effect estimates by sex, and no clear differential pattern was observed. Asthma was a strong risk factor for impairment in males, compared to females. We observed that particulate matter ($PM_{2.5}$) was a more pronounced risk factor in non-asthmatics, and the association between low birth weight and lung function impairment was slightly stronger among non-asthmatics. The stratified analysis by asthma status provided additional insight to the determinants of lung function impairment, where we observed a stronger association with air pollution and low birthweight among the non-asthmatics. This is potentially related to the different mechanisms of lung function impairment between asthmatics, where lung function impairment is reversible, versus non-asthmatics, where lung impairment is irreversible [42].

4.2. Lung cancer risk prediction

The mechanistic relationship between COPD and lung cancer has been suggested to operate through many pathways including genetic susceptibility, DNA damage and repair, epigenetics, downregulation of specific microRNA, expression of pro-inflammatory genes, and immune adaptive responses [43]. Several studies have shown that COPD confers a 2- to 4-fold increase in lung cancer risk [43–46]. An increase in lung cancer risk has been shown for as little as a 10% reduction in FEV1 from predicted reference values [44].

Lung cancer in never-smokers is a growing issue and an increasingly larger proportion of lung cancers are being diagnosed in lifetime never-smokers [6]. We found that using only a simple set of predictors (age, sex, family history of lung cancer, and personal non-lung cancer history), we were able to achieve moderate discrimination of lung cancers after correcting for model over-optimism. At 3 and 5 years time horizons, the continuous measurement of FEV1 (% of GLI predicted FEV1) added modest predictive value to the lung cancer prediction model. This aligns with the fact that low FEV1 values reflect obstruction of the airway and it is an indicator of the impairment severity. All of these measures can be collected from a routine physician visit. Despite the modest improvement in model fit and discrimination when including lung function, the overall absolute risk of lung cancer remains low in this population-based cohort, consisting of healthy recruits at baseline.

To our knowledge, there are only two studies on lung cancer risk prediction specifically in never-smokers [10,47]: $PLCO_{M2014}$, which was developed in ever-smokers [48], and was later adapted and applied to never-smokers with an AUC of 0·662 in never-smokers [10]; the risk model developed by Wu et al, in a Taiwanese cohort that reached an AUC of 0·806 (95% CI: 0·790–0·819) using demographics, family history, alpha-fetoprotein, maximum mid-expiratory flow, and carcinoembryonic antigen as predictors [47]. The results of the Taiwanese study suggest that further incremental gain in prediction for never-smokers can be achieved through the inclusion of lung function and the key biomarkers. Comparing to a previous study that assessed the role of lung function in lung cancer risk prediction based on the overall population, where lung cancer patients are predominantly smokers [49], our approach substantially differs from this previous work because we focused solely on building a never-smoker model, and modeled risk at several fixed time-points of lung cancer (e.g. 5-year risk, a commonly used time horizon for screening program enrollment).

Our study has several strengths. We were able to conduct this study in one of the largest cohorts of never-smokers with an extensive collection of lifestyle, demographic, and physical measurements combined with National disease and death registries. We were able to investigate the determinants of lung function using spirometry measures in a cohort exclusively of never-smokers, whose lung function would not be confounded due to primary smoking habits. The large cohort size and length of follow-up allowed us to exclude never-smokers with potentially subclinical lung cancer at baseline in order minimize the potential confounding effect on lung function associations and provided sufficient accrual of never-smoking lung cancer cases for risk analysis.

A limitation of this study is that more detailed ETS history was not available. A small set of questions with substantial missing data hindered the ability to sufficiently evaluate the role of ETS on lung impairment or lung cancer risk. There is potential for selection bias for the participants included in the complete case analysis, especially for birth weight and physical activity where data is only available in a subset of the participants. The subset with available data contains more males and whites and are less overweight compared to the total population. However, sex and ethnicity are adjusted for in the models and having a larger proportion of participants with normal BMI is likely to bias the results toward the null, and thus would be unlikely to have created the observed associations. To utilize the GLI-2012 equations, we grouped the ethnicities in UKB cohort into the 5 GLI ethnic groups (such as Caucasian, Black, Northeastern Asian, Southeastern Asian, and Other). Some of the UKB self-reported ethnicities do not have a direct correspondence to a GLI ethnic group, therefore we have aligned them to the closest GLI ethnic subgroup. For example participants with Indian, Pakistani, and Bangladeshi ancestry were aligned with the GLI-SE Asian group, as the closest proxy. The GLI-SE Asian equations have been shown to be applicable to Indian children, although the validity in adults is less clear [50]. To assess whether this would affect our results, we conducted a sensitivity analysis by excluding Indian, Pakistani, and Bangladesh participants ($n = 3417$) and the results were very similar to the results based on the full never smoker population presented in Table 2.

In conclusion, low birthweight, ambient air pollution, asthma, and obesity were found be associated with impaired lung function in lifetime never-smokers. We achieved moderate discrimination of lung cancers among never-smokers when using a small set of predictors that can be routinely collected during physician visits. Lung function demonstrated a limited increase in model predictive performance based on the small number of lung cancer cases available. Future investigation of these models based on a larger sample size of never-smoking lung cancers may help to determine whether lung function can improve risk-stratification among never-smokers and identify those who would potentially benefit from lung cancer screening.

Contributors

MTW and RJH contributed equally to the design and analysis of the study. All authors contributed to the interpretation of the results, initial manuscript draft, manuscript revisions, and approval of the final version for submission.

Declaration of Competing Interest

None of the authors have any financial or personal conflicts of interest to disclose.

Acknowledgments

This study was support by a Canada Research Chair awarded to Rayjean J. Hung. The funding source had no role in any aspect of the conception, design, implementation, analysis, interpretation, or decision to publish this study. This research has been conducted using the UK Biobank Resource under Application Number 23261.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.08.058.

References

[1] Ferkol T, Schraufnagel D. The global burden of respiratory disease. Ann Am Thorac Soc 2014;11(3):404–6.
[2] Soriano JB, Abajobir AA, Abate KH, Abera SF, Agrawal A, Ahmed MB, et al. Global, regional, and national deaths, prevalence, disability-adjusted life years, and years lived with disability for chronic obstructive pulmonary disease and asthma, 1990–2015: a systematic analysis for the global burden of disease study 2015. Lancet Respir Med 2017;5(9):691–706.
[3] Gupta RP, Strachan DP. Ventilatory function as a predictor of mortality in lifelong non-smokers: evidence from large british cohort studies. BMJ Open 2017;7(7):e015381.
[4] Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. JAMA Oncol 2017;3(4):524–48.
[5] Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. Lancet 2009;374(9691):733–43.
[6] Pallis AG, Syrigos KN. Lung cancer in never smokers: disease characteristics and risk factors. Crit Rev Oncol Hematol 2013;88(3):494–503.
[7] Planchard D, Besse B. Lung cancer in never-smokers; 2015.
[8] Sun S, Schiller J, Gazdar A. Lung cancer in never smokers–a different disease. Nat Rev Cancer 2007;7(10):778–90.
[9] Moyer VA. Screening for lung cancer: US preventive services task force recommendation statement. Ann Intern Med 2014;160(5):330–8.
[10] Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the lung cancer risks at which to screen ever-and never-smokers: screening rules applied to the plco and nlst cohorts. PLoS Med 2014;11(12):e1001764.
[11] Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental value of pulmonary function and sputum dna image cytometry in lung cancer risk prediction. Cancer Prev Res 2011;4(4):552–61.
[12] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, 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):e1001779.
[13] Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J 2005;26(2):319–38.
[14] Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations; 2012.
[15] Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease 2019 report. https://goldcopd.org/; 2019.
[16] National Clinical Guideline Centre. Chronic obstructive pulmonary disease: management of chronic obstructive pulmonary disease in adults in primary and secondary care. London: National Clinical Guideline Centre; 2010.
[17] Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, et al. Development of no2 and nox land use regression models for estimating air pollution exposure in 36 study areas in europe–the escape project. Atmos Environ 2013;72:10–23.
[18] Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M, et al. Development of land use regression models for pm2. 5, pm2. 5 absorbance, pm10 and pmcoarse in 20 european study areas; results of the escape project. Environ Sci Technol 2012;46(20):11195–205.
[19] Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a 1.1-million-person gwas of educational attainment. Nat Genet 2018;50(8):1112.
[20] Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159(7):702–6.
[21] Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361–87.
[22] Harrell Jr FE. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer; 2015.
[23] R Core Team. R: a language and environment for statistical computing [internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019 Available from https://www.R-project.org/.
[24] Saad NJ, Patel J, Burney P, Minelli C. Birth weight and lung function in adulthood: a systematic review and meta-analysis. Ann Am Thorac Soc 2017;14(6):994–1004.
[25] Saarenpää H-K, Tikanmäki M, Sipola-Leppänen M, Hovi P, Wehkalampi K, Siltanen M, et al. Lung function in very low birth weight adults. Pediatrics 2015;136(4):642–50.
[26] Adam M, Schikowski T, Carsin AE, Cai Y, Jacquemin B, Sanchez M, et al. Adult lung function and long-term air pollution exposure. ESCAPE: a multicentre cohort study and meta-analysis. Eur Respir J 2015;45(1):38–50.
[27] Liu Q, Xu C, Ji G, Liu H, Shao W, Zhang C, et al. Effect of exposure to ambient pm2. 5 pollution on the risk of respiratory tract diseases: a meta-analysis of cohort studies. J Biomed Res 2017;31(2):130.
[28] Paulin L, Hansel N. Particulate air pollution and impaired lung function. F1000Research 2016;5.
[29] Xing Y-F, Xu Y-H, Shi M-H, Lian Y-X. The impact of pm2. 5 on the human respiratory system. J Thorac Dis 2016;8(1):E69.
[30] Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, et al. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. Environ Health Perspect 2014;122(9):906–11.
[31] Fuertes E, Carsin A-E, Antó JM, Bono R, Corsico AG, Demoly P, et al. Leisure-time vigorous physical activity is associated with better lung function: the prospective ecrhs study. Thorax 2018;73(4):376–84.
[32] Sin DD, Jones RL, Mannino DM, Man SP. Forced expiratory volume in 1 second and physical activity in the general population. Am J Med 2004;117(4):270–3.
[33] Sisson JH. Alcohol and airways function in health and disease. Alcohol 2007;41(5):293–307.
[34] Tabak C, Smit HA, Räsänen L, Fidanza F, Menotti A, Nissinen A, et al. Alcohol consumption in relation to 20-year copd mortality and pulmonary function in middle-aged men from three european countries. Epidemiology 2001;12(2):239–45.
[35] Tabak C, Smit H, Heederik D, Ocke M, Kromhout D. Diet and chronic obstructive pulmonary disease: independent beneficial effects of fruits, whole grains, and alcohol (the morgen study). Clin Exp Allergy 2001;31(5):747–55.
[36] Fischer F, Kraemer A. Meta-analysis of the association between second-hand smoke exposure and ischaemic heart diseases, copd and stroke. BMC Public Health 2015;15(1):1202.
[37] Lee PN, Forey BA, Coombs KJ, Hamling JS, Thornton AJ. Epidemiological evidence relating environmental smoke to copd in lifelong non-smokers: a systematic review. F1000Research 2018;7.
[38] Melo LC, da Silva MAM, do Calles ACN. Obesity and lung function: a systematic review. Einstein (Sao Paulo) 2014;12(1):120–5.
[39] Salome CM, King GG, Berend N. Physiology of obesity and effects on lung function. J Appl Physiol 2009;108(1):206–11.
[40] Spelta F, Pasini AF, Cazzoletti L, Ferrari M. Body weight and mortality in copd: focus on the obesity paradox. Eating and weight disorders-studies on anorexia, bulimia and obesity, 23(1). ; 2018. p. 15–22.
[41] Watson R, Pride NB, Thomas EL, Fitzpatrick J, Durighel G, McCarthy J, et al. Reduction of total lung capacity in obese men: comparison of total intrathoracic and gas volumes. J Appl Physiol 2010;108(6):1605–12.
[42] Rogliani P, Ora J, Puxeddu E, Cazzola M. Airflow obstruction: is it asthma or is it copd? Int J Chron Obstruct Pulmon Dis 2016;11:3007–13.
[43] Gonzalez J, Marín M, Sánchez-Salcedo P, Zulueta JJ. Lung cancer screening in patients with chronic obstructive pulmonary disease. Ann Transl Med 2016;4(8).
[44] Calabrò E, Randi G, La Vecchia C, Sverzellati N, Marchianò A, Villani M, et al. Lung function predicts lung cancer risk in smokers: a tool for targeting screening programmes. Eur Respir J 2010;35(1):146–51.
[45] Henschke CI, Yip R, Boffetta P, Markowitz S, Miller A, Hanaoka T, et al. CT screening for lung cancer: importance of emphysema for never smokers and smokers. Lung Cancer 2015;88(1):42–7.
[46] Wasswa-Kintu S, Gan W, Man S, Pare P, Sin D. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: a systematic review and meta-analysis. Thorax 2005;60(7):570–5.
[47] Wu X, Wen CP, Ye Y, Tsai M, Wen C, Roth JA, et al. Personalized risk assessment in never, light, and heavy smokers in a prospective cohort in Taiwan. Sci Rep 2016;6:36482.
[48] Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection criteria for lung-cancer screening. N Engl J Med 2013;368(8):728–36.
[49] Muller DC, Johansson M, Brennan P. Lung cancer risk prediction model incorporating lung function: development and validation in the Uk biobank prospective cohort study. J Clin Oncol 2017;35(8):861–9.
[50] Kirkby J, Stocks J, Lum S, Rao P, Sonnappa S. Global lung function initiative (GLI) spirometry equations: comparison of lung function between indigenous Indian and UK-Indian children. Eur Respir Soc 2013;42.