BMJ Open

Study protocol for a national cohort of adults focused on respiratory health: the American Lung Association Lung Health Cohort (ALA-LHC) Study

Paul A Reyfman [ORCID],[1] Elizabeth Sugar,[2] Heather Hazucha,[3] Jenny Hixon,[1] Curt Reynolds,[2] Sonali Bose,[4] Mark T Dransfield,[5] MeiLan K Han,[6] Raul San Jose Estepar,[7,8] Mary B Rice,[9,10] George R Washko,[8,11] Mercedes Carnethon,[12] Ravi Kalhan [ORCID],[1,12] American Lung Association Airways Clinical Research Network

Correspondence to
Dr Ravi Kalhan;
r-kalhan@northwestern.edu

ABSTRACT

Introduction The current framework for investigating respiratory diseases is based on defining lung health as the absence of lung disease. In order to develop a comprehensive approach to prevent the development of lung disease, there is a need to evaluate the full spectrum of lung health spanning from ideal to impaired lung health. The American Lung Association (ALA) Lung Health Cohort is a new, population-based, cohort study focused primarily on characterising lung health in members of the millennial generation without diagnosed severe respiratory disease. Participants will be enrolled for the baseline study visit starting in 2021, and funding will be sought to support future study exams as part of a longitudinal cohort study. This study will be crucial for developing a novel paradigm of lung health throughout the adult life course.

Methods and analysis This study will leverage the existing infrastructure of the ALA Airways Clinical Research Centers network to enrol 4000 participants between ages 25 and 35 years old at 39 sites across the USA between April 2021 and December 2024. Study procedures will include physical assessment, spirometry, chest CT scan, accelerometry and collection of nasal epithelial lining fluid, nasal epithelial cells, blood and urine. Participants will complete questionnaires about their sociodemographic characteristics, home address histories and exposures, work history and exposure, medical histories, lung health and health behaviours and activity.

Ethics and dissemination The study was approved by the Johns Hopkins Medicine Institutional Review Board. Findings will be disseminated to the scientific community through peer-reviewed journals and at professional conferences. The lay public will receive scientific findings directly through the ALA infrastructure including the official public website. Deidentified datasets will be deposited to BioLINCC, and deidentified biospecimens may be made available to qualified investigators along with a limited-use datasets.

Strengths and limitations of this study

► This is the first large, prospective US cohort study of respiratory health over the adult life course.
► Lung health outcomes are defined broadly and include respiratory symptoms, physiological lung function and CT scan-defined lung and airways injury.
► A wide array of biospecimens including blood, urine and nasal samples will be collected at the baseline assessment, which will allow for discovery of biomarkers of impaired respiratory health.
► Individuals will provide extensive information on modifiable exposures and risk factors (eg, lifetime air pollution, potentially noxious inhalation exposure such as marijuana and e-cigarettes, reduced physical activity and workplace exposures), which will be updated throughout the course of follow-up.
► Current funding supports only a cross-sectional examination and remote follow-up contacts; additional funding will be required for future in-person study exams and the ascertainment of long-term lung health outcomes.

INTRODUCTION AND RATIONALE

Developing a robust approach to lung disease prevention requires a broad definition of lung health.[1] While the respiratory research community implicitly construes lung health as the absence of overt lung disease, the cardiovascular (CV) community has explicitly delineated a continuum ranging from ideal health to impaired health to established disease. According to this framework, CV health is composed of a set of factors, including several that directly occur along the causal pathway from health to disease. This paradigm of CV health serves as a foundation for developing preventive health measures by targeting risk factors for loss of health and recognition of measurable intermediate endotypes, such as heightened inflammation and hypercholesterolaemia.

The American Lung Association Lung Health Cohort (ALA-LHC) Study is motivated by the conviction that a large, population-based cohort of young adults focused on the respiratory system is needed to define the constructs of of 'ideal' versus 'impaired' lung health. Previous studies of subclinical lung disease have examined healthy controls in disease-focused studies or used limited information about respiratory symptoms, physiology or imaging derived from cohort studies focused on other chronic conditions. Existing prospective cohort studies within the respiratory field have a maximum follow-up of 10 years and are focused on specific respiratory diseases, such as asthma and chronic obstructive pulmonary disease (COPD). Life course studies of lung health are necessary to describe how an individual progresses from ideal lung health to an intermediate phenotype of impaired lung health and ultimately to chronic lung disease.

We have developed a conceptual paradigm of lung health according to which there are two fundamental forms of impaired lung health: (1) impaired 'reserve' or lower peak lung function in young adult and (2) 'susceptibility' to impaired lung health or accelerated decline in lung function from young adulthood to middle age. The ALA-LHC will focus on modifiable exposures and clinical and behavioural risk factors in members of the millennial generation that we hypothesise are associated with impaired lung health reserve and on biomarkers of susceptibility to future lung disease. Identifying the factors associated with reserve and susceptibility will enhance the depiction of the continuum of lung health and will facilitate designing novel prevention strategies.

Specific aims

Our long-term goal is to determine factors evident in early adulthood that influence risk for future lung disease through detailed assessments of modifiable and non-modifiable environmental, clinical and behavioural risk factors and responses (physiological, imaging-based and biological including the nasal epithelial transcriptome). Our motivations were bolstered by a call from the National Heart, Lung, and Blood Institute (NHLBI) for the development of new epidemiology cohorts to address research hypothesis that were beyond the scope of existing cohorts.[2] We have observed that impaired lung health (lower lung function, self-report of respiratory symptoms and lung abnormalities on chest CT scan) precedes the development of chronic lung disease. Therefore, the focus of the initial contact will be to establish the cohort, characterise their current lung health and test hypotheses about the risk factors and manifestations of impaired lung health in young adults. There are three specific aims:

1. Determine whether modifiable exposures and risk factors (lifetime air pollution, potentially noxious inhalational exposures such as marijuana and e-cigarettes and reduced physical activity) are associated with lower lung function and greater burden of respiratory symptoms in young adults.

2. Determine whether CT measurements of small airway and parenchymal lung abnormalities are associated with lower lung function and greater burden of respiratory symptoms in young adults.

3. Collect and store biospecimens for future analyses to determine the association between the nasal respiratory epithelial transcriptome and selected blood biomarkers, CT measured small airway abnormality and parenchymal lung abnormalities in young adults.

Remote contacts will be used to keep the cohort engaged, update contact information and collect short-term follow-up information on the exposures and responses. We will endeavour to secure additional funding for future in-person exams and the evaluation of biospecimens.

METHODS AND ANALYSIS
Study design

The ALA-LHC is a longitudinal, multicentre cohort study that will enrol approximately 4000 participants aged 25–35 years old who do not have a diagnosis of severe lung disease recruited at participating research sites across the USA.

Patient and public involvement

The ALA-LHC was conceived and designed through close collaboration with the ALA, which regularly solicits input for its research programmes from participating patient advisory groups. Patients were not explicitly involved in the design of the ALA-LHC. The ALA will participate in the dissemination of results to the community of stakeholders interested in respiratory disease.

Eligibility criteria

Eligibility is deliberately broad and selected to be inclusive of young adults with a range of exposures that may influence lung health. However, individuals with established moderate-to-severe lung disease will not be enrolled. The detailed inclusion and exclusion criteria are provided in table 1.

Outcomes

The primary outcome measures are as follows:

▶ In-person spirometry: Prebronchodilator lung function (forced expiratory volume in 1 s (FEV_1), forced vital capacity (FVC) and peak flow) measured in accordance with the American Thoracic Society (ATS)/European Respiratory Society (ERS) standards and using United States National Health and Nutrition Examination Survey (NHANES) reference values.[3 4]

The primary outcome measures are as follows:

– Prebronchodilator FEV_1.
– Prebronchodilator FVC.

▶ ATS-Division of Lung Diseases (DLD) Respiratory Questionnaire: The ATS-DLD is a general respiratory symptom questionnaire. We will use the Lung Health

Table 1 Inclusion and exclusion criteria

Inclusion	Exclusion

Age 25–35 years old at the time of the enrolment	Severe asthma that is defined as any of the following:
	► Current Global Initiative for Asthma Step 4 or higher therapy[27]
	    OR
	► Three or more unscheduled healthcare visits (provider/urgent care/ER) for asthma in the past 12 months
	    OR
	► One asthma hospitalisation in the past 12 months

Able to read and understand English or Spanish	History of any chronic lung disease other than asthma including but not limited to COPD, cystic fibrosis, pulmonary fibrosis, pulmonary hypertension or congenital lung disease
Has and is willing to share a social security number	Current pregnancy
Resident (citizen or non-citizen) of the USA for at least 12 months prior to examination	History of cancer other than non-melanoma skin cancer
Willing to provide contact information for at least two proxies who are likely to know the whereabouts and vital status of the participant	Diagnosed cardiovascular disease (ie, congenital heart disease and coronary heart disease)
	History of cancer other than non-melanoma skin cancer
	Inability to comply with study procedures
	Any condition that puts the participant at risk by participating in the study as deemed by study investigators (eg, serious respiratory illness requiring antibiotics or steroids or severe fever at the time of the study visit)
	Institutionalisation

COPD, chronic obstructive pulmonary disease; ER, emergency room.

Study (LHS-III) modified version that includes information regarding occupational exposures and asthma:
- – Any symptoms
- – Individual symptoms
► CT measurements of small airway and parenchymal lung abnormalities will be made using both commercially available and open-source software (Chest Imaging Platform, an National Institutes of Health (NIH)-funded, open source platform, http://www.chestimagingplatform.org).[5] The primary outcome measurements are as follows:
- – Extent of parametric response map for functional small airways disease (PRM^{fSAD}).
- – Size of structural abnormalities (expressed as the percent of total lung with that classification (eg, 50% emphysema))

Recruitment

Participants will be recruited at each ALA Airways Clinical Research Centers (ACRC) site (box 1) through public advertising campaigns (digitally and through traditional media), public outreach, word of mouth and partnerships with community organisations. Information for potential and current participants will be posted on a publicly available study website hosted by the ALA, which will also organise a social media campaign.

The goal is to recruit a population that is balanced in terms of sex (target: 1/2 men and 1/2 women) and educational attainment (target: 1/3 high school education or less, 1/3 some college and 1/3 college degree or higher). Balancing recruitment by education as a proxy for socioeconomic status (SES) is based on data demonstrating individual behavioural factors (eg, occupation, smoking and physical activity) are associated with lung health and strongly correlated with SES.[6 7] Individuals with lower levels of education are more likely to reside in lower-income communities, which will also encourage a distribution of environmental exposures. Although the recruitment targets are based on SES rather than race, we do anticipate recruiting a diverse population, with an expected racial distribution of approximately 60% Caucasian, 30% African American, 7.5% Asian, 2.5% others and 25% Hispanic/Latinx.

Enrolment

To initiate an in-person enrolment, a site will schedule an initial, in-person visit with the potential participant. All participants will provide written informed consent. The procedures and questionnaires are expected to take approximately 4 hours. Young adults often have conflicts (eg, need for childcare and limited time off) that would make it difficult to schedule such a large block of time. In order to make the initial assessment more feasible, the in-person visit may be broken up into several shorter visits with emphasis on completing as many of the assessments remotely as possible to reduce the in-person time. The

Box 1 American Lung Association (ALA) Lung Health Cohort study sites and cores

Study sites
- *Baylor College of Medicine—Houston, Texas.*
- *Duke University consortium:*
  - Duke University Medical Center—Durham, North Carolina.
  - Johns Hopkins University.
  - University of North Carolina Hospital.
- *Michigan consortium:*
  - University of Michigan—Ann Arbor, Michigan.
  - University of Iowa.
- *Illinois consortium:*
  - Northwestern University.
  - Rush University.
  - University of Chicago.
  - University of Illinois at Chicago.
- *Pacific Northwest consortium:*
  VA Puget Sound Health Care System—Seattle, Washington.
- *Temple University consortium:*
  - Temple Lung Center—Philadelphia, Pennsylvania.
  - Ohio State University—Wexner Medical Center.
  - University of Pittsburgh—Emphysema COPD Research Center.
  - Medical University of South Carolina.
- *Nemours Children's Health consortium:*
  - Nemours Children's Health—Jacksonville, Florida.
  - University of Florida, Jacksonville.
- *National Jewish Health consortium:*
  - National Jewish Health—Denver, Colorado.
  - University of Colorado Anschutz Medical Campus.
  - Mount Sinai, New York City.
- *Northern New England consortium:*
  - University of Vermont—Burlington, Vermont.
  - University of Rochester.
- *New York consortium:*
  - Columbia University, New York, New York.
  - Brigham and Women's Hospital.
  - New York Medical College.
  - Presbyterian Brooklyn Methodist Hospital.
  - Cornell University.
  - New York University.
- *The UAB Lung Health Center—Birmingham, Alabama.*
- *University of Kansas consortium:*
  - University of Kansas Medical Center.
  - Wake Forest School of Medicine.
  - Vanderbilt University Medical Center.
  - University of Wisconsin Madison.
- *St Vincent Hospital and Health Care Center—Indianapolis, Indiana.*
- *University of California consortium:*
  - University of California, San Francisco.
  - University of California, Los Angeles.
- *University of Arizona—Tucson, Arizona.*

Study cores.
- Administrative Core at Northwestern University.
- Actigraph Reading Center at Northwestern University.
- Data Coordinating Center at Johns Hopkins University.
- Spirometry Reading Center at Johns Hopkins University.
- Imaging Core at Brigham and Women's Health and University of Michigan.

Continued

Box 1 Continued

- ALA Airways Clinical Research Centers (ACRC) Biorepository at Nemours Children's Hospital.
- Environmental Exposure Core at Beth Israel Deaconess Medical Center.

first part of the in-person visit will focus on evaluating eligibility and obtaining informed consent for participation in the ALA-LHC. The procedures and key questionnaires relating to the primary aims of the study must be completed in person. The remaining questionnaires may be completed remotely.

Study data collection
Baseline assessment

The baseline assessment includes an in-person evaluation, wearable device monitoring of activity and remote contacts. Box 2 lists the procedures and questionnaire data collection for the baseline assessment. The total time allocated for the completion of all procedures and questionnaires is 4 hours. The goal is to collect all assessments within 1 month (28 days of enrolment); however, due to potential barriers to that collection schedule related to the ongoing COVID-19 pandemic, the permitted time interval for evaluations following initial enrolment may be longer.

In-person activities

Individuals will be consented, tested for eligibility and enrolled at the beginning of the in-person evaluation. All procedures except for the actigraphy monitoring and home spirometry will be performed in the clinic. In addition, key questionnaires will be completed during the in-person visit.

In-home activities

Additional in-home activities are planned to collect information on lung function, activity and vaping after the in-person visit. Each participant will be asked to complete a home spirometry session on five separate days over a 14-day period within a month of the in-person visit with coaching from a site coordinator via a telemedicine visit. In addition, each participant will be asked to wear an activity monitor for 1 week and then mail it back to the site. Those individuals who indicated that they have used e-cigarettes and vaping or heat-not-burn (HNB)/heated tobacco devices will be asked to complete a 1-week daily diary recording the type and frequency of e-cigarette, vaping or HNB product use and any related symptoms.

Remote contact activities

Remote contacts will be used to facilitate the completion of the baseline assessment. Questionnaires that are not completed during the in-person visit will be completed remotely (eg, online by the participant using a link sent via email, text message, mail or phone).

Box 2 Activities and data collection

Consent

Eligibility evaluation

Enrolment

Procedures

► Physical assessment (eg, height, weight, resting heart rate and blood pressure).
► Prebronchodilator spirometry at study site.
► Home spirometry (prebronchodilator).
► Pregnancy testing in participants of childbearing potential (prior to chest CT).
► Chest CT scan.
► Nasal epithelial lining fluid collection (Nasosorption strips).
► Nasal epithelial curettage.
► Phlebotomy (complete blood count, plasma, serum, RNA, DNA and SARS-CoV-2 serology).
► Urine collection.
► Actigraphy.

Questionnaire domains

► Social security number.
► Contact information for participant and two proxies.
► Home address history.
► Demographics (eg, date of birth, age, race, ethnicity, sex and birth, gender identity, education, employment status, income and health insurance coverage).
► Lung health:
  – ATS-DLD Respiratory Questionnaire.
  – Modified Rhinitis Symptom Utility Index.[28]
► Substance use (eg, tobacco, e-cigarette/vaping/heat-not-burn, alcohol and marijuana).
► Environmental and occupational exposures (eg, secondhand smoke, pets, mould and fumes).
► Medication use.
► Medical history (eg, disease diagnoses, allergies and pregnancy/ early life factors).
► Activity:
  – International Physical Assessment Questionnaire (short form).[29]
► Sleep:
  – Pittsburgh Sleep Quality Instrument.[30]
► Social elements (eg, neighbourhood safety, social roles and activities and global health).
► Nutrition.
► COVID-19 history and associated health behaviours (tobacco, e-cigarettes, hand washing and social distancing).

Patient logs

► Activity log.
► Vaping diary.

Remote follow-up

During follow-up, participants will be contacted remotely (eg, via text message, email, mail or phone) every 3–6 months and asked to update their contact information and complete additional assessments related to the topics described above for the in-person visit (eg, lung health and behavioural habits) as well as health events (eg, urgent care visits, hospitalisation, vaccinations and diagnoses). In conjunction with each of the follow-up contacts, participants will be asked to schedule and complete a single, coached home spirometry session.

Study procedures

Questionnaires

Questionnaires may be completed in multiple ways. During the study visit, forms will be either directly data entered into the ALA-LHC REDCap data repository on tablets provided by the study or on paper forms, which will later be data entered by the site staff.

Environmental evaluation

The Environmental Exposure Core will use the lifetime address history of each participant to perform geocoding. The geocodes will be used to link relevant exposure data to each participant. Exposure data will include long-term averages of residential air pollution exposure (elemental carbon, organic carbon, nitrogen dioxide and sulfates), fine particulate matter ($PM_{2.5}$),[8–10] ozone, census-tract level indicators of neighbourhood socioeconomic position (median household income, median home value, measures of household crowding and distribution of education level) and weather data.

Spirometry at the study site

Prebronchodilator spirometry will be performed at the in-person baseline visit if local conditions and policies permit it. Each site will follow their institutional guidelines for conduct of research and COVID-19-related precautions and procedures and, if required at their local institution, institutional review board (IRB) approval.

Spirometry measurements taken at the site will adhere to ATS/ERS standards.[3] Quality control, training and interpretation of results will be led by the Spirometry Reading Center at Johns Hopkins University. Spirometry measurements will include the FVC, FEV_1 and $FEV_1/$FVC ratio. Only prebronchodilator spirometry will be performed (participants will not receive albuterol). Spirometry reference values will be those of Hankinson *et al* from NHANES.[4] Spirometry sessions will be performed by a certified technician.

Home spirometry

All participants will be provided with a Spirobank Smart spirometer (MIR USA International Research), which is an Food and Drug Administration 510(k)-cleared device (K072979) for home use. After the in-person visit, participants will be asked to perform home spirometry on five separate days over a 14-day period within 1 month of the in-person visit. Therefore, participants will be asked to complete quarterly home spirometry sessions in conjunction with the quarterly contacts. The remote spirometry sessions will be coached by a certified technician using a video conference link.

Chest CT acquisition

Each participant will undergo CT imaging. CT scan will be acquired with multidetector CT scanners (64 detector channels) at full inspiration and expiratory effort using a

low-dose protocol with dose modulation. Imaging reconstruction will use submillimetre slice thickness with four reconstructions per scan.

Standardised parameters will be developed for the CT scanner to be used at each site to ensure consistency of results. CT scanners at each site will be calibrated using a standard lung phantom prior to enrolling participants. All participants will be imaged in the supine body posture both at full inflation (approximating total lung capacity) and at the end of forcible exhalation (approximating residual volume).

Study measurements obtained on the CT scan will include features based on analyses of the lung parenchyma through the local histogram to determine lung inflammation, oedema, emphysema and interstitial changes as well as the PRM^{fSAD} as has been used previously.[11–16]

Nasal epithelial curettage
Scraping of the nasal epithelium will be performed using five passes with the disposable plastic ASI Rhino-Pro nasal cytology curette from the inferior nasal turbine. Samples will be immediately frozen in RNA lysis buffer and stored at −80°C until shipped to the repository for future processing.

Nasal epithelial lining fluid
For collection of nasal epithelial lining fluid, a thin absorbent, fibrous matrix designed to fit within the nasal passages (Mucosal Diagnostics Nasosorption FX·i) will be inserted into each nostril. Participants will be asked to pinch their nose closed for 1 min prior to removal of each strip. The two fibrous strips will be placed in separate tubes, capped and immediately frozen and stored at −80°C until shipped to the repository.

Phlebotomy
Approximately 50 mL of blood will be collected by venipuncture from the arm or the back of the hand. One 4 mL tube will be transmitted to the local hospital laboratory without processing for measurement of complete blood count and chemistries. The remaining blood will be divided into 83 250 uL aliquots of serum and plasma for freezing and bulk storage pending shipping. One 2.5 mL PAXgene RNA tube will be used to collect blood for RNA analysis. Also, one lavender tube containing a remaining pellet after aliquoting plasma will be saved for DNA analysis. Both these tubes will be stored at −70 or −80°C until shipped to the repository.

Urine collection
Participants will be instructed on how to obtain a clean catch urine specimen. A minimum of 20 mL will be collected from each participant. Urine will be aliquotted into 10 1 mL non-preserved aliquoting vials and placed at −80°C for storage until shipment.

Accelerometry
At the conclusion of the in-person visit, participants will be given the ActiGraph GT3X. At the same time, they will be given a paper log to record when the device is removed and put back on and the intervals in which they were sleeping. Participants will be asked to wear the device on their non-dominant arm continuously for 24 hours a day over 7 days after leaving the site and to mail the device back to the clinic.

COVID-19 exposure and history
Participants will be given a questionnaire to assess self-reported history of COVID-19 and factors related to severity of disease. Exposure to SARS-CoV-2 will be assessed using serology testing.

Data management
Data collection and transfer
Data will be collected, stored and transferred in compliance with Health Insurance Portability and Accountability Act (HIPAA) Privacy and Security rules.

The primary mechanism for data collection and management in the ALA-LHC will be REDCap, a secure, web-based application hosted by the Data Coordinating Center (DCC) at Johns Hopkins University that can support direct electronic data entry or forms-based data capture in compliance with HIPAA Security Rule.[17 18] The ALA-LHS REDCap Project will be managed by the DCC. The majority of enrolment and study data for the study visit will be directly entered to REDCap. Tablets will be provided to each site to facilitate real-time capture of data elements. However, paper versions of all instruments will be available as an alternative in case of connectivity issues, participant preference (in the completion of self-reported forms) or other barriers to direct data entry. In addition, each participant will be provided with a tablet to take home that can be used to complete forms remotely.

Data sharing
The ALA-ACRC DCC has developed and tested procedures for deidentification of personal health information (PHI) that meet the deidentification standard established by Section 164.514 of the HIPAA Privacy Rule. Deidentified datasets will be deposited to BioLINCC at scheduled milestones (eg, following completion of the baseline exam by all participants). Deidentified (coded) biospecimens collected in the study that are not analysed may be made available to qualified investigators along with a limited-use dataset in line with Office for Human Research Protections (OHRP) guidelines with a material transfer agreement and/or data-use agreement as applicable.

Analyses
Hypotheses
There are three primary hypotheses:
► Hypothesis 1: Air pollution and other inhalational exposures are associated with lower pulmonary function and more frequent respiratory symptoms, whereas physical activity is associated with higher function and less frequent symptoms.
► Hypothesis 2: CT-measured small airways and parenchymal lung abnormalities are associated with lower

pulmonary function and more frequent respiratory symptoms in young adults.
► Hypothesis 3:
  – The pattern of nasal epithelial gene expression is different in participants with CT-measured small airways and parenchymal lung abnormalities.
  – Blood biomarkers previously associated with distinct chronic lung diseases such as Interstitial Lung Disease (ILD) and COPD (fibrinogen, C reactive protein, soluble receptor for advanced glycosylation end product, surfactant protein D, club cell protein-16 and matrix metalloproteinases 7 and 9) are also associated with different CT patterns of lung abnormalities in young adults.

Hypotheses 1 and 2 will be addressed using data collected in the main ALA-LHC Study. Specimens will be collected to enable evaluation of Hypothesis 3 but will be deferred until additional funding is obtained.

Sample size
Detectable effect sizes were calculated assuming a sample size of 4000 (assuming 235 per site) and two-sided type I error rate of 5% (unless otherwise specified). Simulations incorporating a random site effect produced similar results for variables used in prior ALA-ACRC studies due to low intraclass correlation coefficient among participants seen at the same site. Based on prior studies,[19–21] we conservatively estimated that the SD of percentage of predicted FEV_1 is to be 15. Data from the Framingham Study suggest that the SD of long $PM_{2.5}$ is 1.24 µg/m^3 with an IQR width (IQRw) of approximately 1.87 (ie, the SD for $PM_{2.5}$ rescaled in units of IQRw is 0.66). We have 80% and 90% power to detect 0.98% and 1.13% decreases in percentage of predicted FEV_1 of per IQRw increase in $PM_{2.5}$ exposure, a clinically meaningful change as decreases of 1%–1.5% are associated with all cause mortality.[22] In the Coronary Artery Risk Development in Young Adults Study, a cohort of 18–30 year olds enrolled in 1985; baseline symptom prevalence ranged from 10% (bronchitis) to 45% (wheeze).[23] Assuming reference prevalence of 10%, 15%, 30% and 45%, we have 90% power to detect an increase in the risk of symptoms to 12%, 17%, 33% and 48% (ie, OR: 1.19, 1.15, 1.12 and 1.11) per IQRw increase in $PM_{2.5}$.

Primary analyses
The cross-sectional associations between lung health and modifiable risk factors (Hypothesis 1) and CT measurements (Hypothesis 2), as well as the association between CT measurements and biological markers (Hypothesis 3), will be assessed using mixed-effect models with a random intercept for study site. Robust SEs will be computed using statistical programme-based approaches when available and the bootstrap otherwise (eg, absolute risk differences). During model selection and formal testing, graphical (eg, residual plots and splines to assess linearity of the effect of continuous exposures) and analytical (eg,

tests of normality) methods will be used to detect violations of modelling assumptions.

To evaluate Hypothesis 1, we will examine the cross-sectional associations of modifiable risk factors (air pollution exposure, inhalational exposures and physical activity) with continuous (FEV_1 and FVC) and binary (respiratory symptoms) measurements of lung health. In addition to the stratification variables, we will adjust all models for race/ethnicity; body mass index (BMI) (or height and weight); and, when appropriate, additional confounders. For environmental exposure, a centred exposure variable and a centre-specific mean exposure variable will be used to model within-centre and between-centre contrasts, or alternatively, centre will be included as a categorical fixed effect to ensure that there is no residual confounding.

To evaluate Hypothesis 2, we will examine the cross-sectional associations of continuous CT measurements of lung abnormalities (extent of PRM^{fSAD} and percent lung abnormalities, log transformed if necessary) with continuous (FEV_1 and FVC) and categorical (symptoms) measurements of lung health. In addition to the stratification variables, all models will include age, smoking and BMI (or height and weight).

To evaluate Hypothesis 3, we will model the association between blood biomarkers and continuous measures of lung abnormalities evaluated by CT (extent of PRM^{fSAD} and percent lung abnormality) using the mixed-effect models described above. In addition to the stratification variables, all models will adjust for age, race/ethnicity, BMI (or height and weight) and smoking status.

Subgroups
Following the model of prior population studies, we will stratify recruitment to include an approximately equal number of men and women. This design enables us to test hypothesised risk factors separately in women and men. We anticipate that 40% of our sample will be non-white race and that 25% will be Hispanic/Latinx. Although we are not adequately powered to test for race/ethnicity-based interactions for most of our risk factors of interest, we will do stratified analyses by race/ethnicity that can be informative for describing whether the patterns of association within subgroups are consistent with those observed in the full population.

Secondary analyses
The analyses of continuous and binary secondary cross-sectional outcomes will use the methods and modelling techniques described for the primary outcomes. Models will include adjustment for potential confounders, which may vary depending on the outcome of interest.

In addition to the cross-sectional data collected at the in-person visit, longitudinal data will be collected through the remote contacts. Continuous and binary outcomes will be modelled as described for the primary analyses with the addition of random effects and fixed effects (if appropriate) for time to account for repeated measurements within the

same individual. Events will be evaluated in two ways. Time to first event will be graphically explored using Kaplan-Meier curves, and comparisons will be made using Cox proportional hazards models for individual outcomes (eg, time to hospitalisation). Event rates and comparisons for repeated events (eg, rates of influenza vaccination) will be modelled using negative binomial regression to account for overdispersion, and the model will include a random effect to account for within-site correlation.[24]

Missing data

A variety of sensitivity analyses will be performed to determine the potential impact of missing data on our conclusions. 'Best' and 'worst' case single imputation techniques will be implemented to define the range of impact. In addition, the effect size that would need to be observed in the missing data in order to change inference will be computed. For a more sophisticated approach, a variety of tools including but not limited to multiple imputation and inverse probability weighting will be used.[25][26]

Ethics and dissemination
Observational Study Monitoring Board (OSMB)
The NHLBI appointed an OSMB who will review the charter, protocol and consent prior to the study initiation. Thereafter, the OSMB will meet annually with the research study team to review data regarding recruitment, data completeness, data quality and safety with additional meetings as needed. The OSMB will also review ancillary study proposals to evaluate the impact on participant burden and safety and the ability to achieve the ALA-LHC's objectives.

Institutional review board

IRB approval was obtained through a single IRB (Johns Hopkins Medicine). Participating study sites are expected to cede approval to the single IRB as per current NIH policy.

Dissemination of data and biospecimens

The ALA-ACRC DCC has developed and tested procedures for deidentification of PHI that meet the deidentification standard established by Section 164.514 of the HIPAA Privacy Rule. Deidentified datasets will be deposited to BioLINCC at scheduled milestones (eg, following completion of the baseline exam by all participants). Deidentified (coded) biospecimens collected in the study that are not analysed may be made available to qualified investigators along with a limited-use dataset in line with OHRP guidelines with a material transfer agreement and/or data-use agreement as applicable. The ALA-ACRC steering committee will review and, in consultation with LHC multiple principal investigator (mPI) team, approve requests for the use of biospecimens. A similar model will be used for sharing images. The ALA-ACRC DCC has extensive experience in preparation of limited-use datasets and maintains policies and model agreements for data-use agreements and material transfer agreements.

Author affiliations
[1]Pulmonary and Critical Care Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA
[2]Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[3]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[4]Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
[5]Pulmonary, Allergy and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA
[6]Pulmonary and Critical Care Medicine, University of Michigan Michigan Medicine, Ann Arbor, Michigan, USA
[7]Radiology, Brigham and Women's Hospital, Boston, Massachusetts, USA
[8]Harvard Medical School, Boston, Massachusetts, USA
[9]Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[10]Environmental Health, Harvard University T H Chan School of Public Health, Boston, Massachusetts, USA
[11]Pulmonary and Critical Care, Brigham and Women's Hospital, Boston, Massachusetts, USA
[12]Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

ORCID iDs
Paul A Reyfman http://orcid.org/0000-0002-6435-6001
Ravi Kalhan http://orcid.org/0000-0003-2443-0876

REFERENCES

1 Reyfman PA, Washko GR, Dransfield MT, *et al*. Defining impaired respiratory health. A paradigm shift for pulmonary medicine. *Am J Respir Crit Care Med* 2018;198:440–6.

2 National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services. New epidemiology cohort studies in heart, lung, blood, and sleep diseases and disorders (u01 - clinical trial not allowed), 2018. Available: https://grants.nih.gov/grants/guide/pa-files/PAR-18-577.htm

3 Miller MR, Hankinson J, Brusasco V, *et al*. Standardisation of spirometry. *Eur Respir J* 2005;26:319–38.

4 Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–87.

5 Fedorov A, Beichel R, Kalpathy-Cramer J, *et al*. 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 2012;30:1323–41.

6 Cheng TL, Goodman E. Committee on pediatric R. race, ethnicity, and socioeconomic status in research on child health. *Pediatrics* 2015;135:e225–37.

7 Polak M, Szafraniec K, Kozela M, *et al*. Socioeconomic status and pulmonary function, transition from childhood to adulthood: cross-sectional results from the Polish part of the HAPIEE study. *BMJ Open* 2019;9:e022638.

8 Di Q, Dominici F, Schwartz JD. Air pollution and mortality in the Medicare population. *N Engl J Med* 2017;377:2513–22.

9 Di Q, Rowland S, Koutrakis P, *et al*. A hybrid model for spatially and temporally resolved ozone exposures in the continental United States. *J Air Waste Manag Assoc* 2017;67:39–52.

10 Kloog I, Chudnovsky AA, Just AC, *et al*. A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmos Environ* 2014;95:581–90.

11 Ash SY, Harmouche R, Ross JC, *et al*. The objective identification and quantification of interstitial lung abnormalities in smokers. *Acad Radiol* 2017;24:941–6.

12 Ash SY, Harmouche R, Ross JC, *et al*. Interstitial features at chest CT enhance the deleterious effects of emphysema in the COPDGene cohort. *Radiology* 2018;288:600–9.

13 Ash SY, Harmouche R, Vallejo DLL, *et al*. Densitometric and local histogram based analysis of computed tomography images in patients with idiopathic pulmonary fibrosis. *Respir Res* 2017;18:45.

14 Bhatt SP, Soler X, Wang X, *et al*. Association between functional small airway disease and FEV1 decline in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2016;194:178–84.

15 Galbán CJ, Han MK, Boes JL, *et al*. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med* 2012;18:1711–5.

16 Kalhan R, Dransfield MT, Colangelo LA, *et al*. Respiratory symptoms in young adults and future lung disease. The cardia lung study. *Am J Respir Crit Care Med* 2018;197:1616–24.

17 Harris PA, Taylor R, Minor BL, *et al*. The REDCap Consortium: building an international community of software platform partners. *J Biomed Inform* 2019;95:103208.

18 Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009;42:377–81.

19 Kalhan R, Arynchyn A, Colangelo LA, *et al*. Lung function in young adults predicts airflow obstruction 20 years later. *Am J Med* 2010;123:468.e1–468.e7.

20 Smith LJ, Kalhan R, Wise RA, *et al*. Effect of a soy isoflavone supplement on lung function and clinical outcomes in patients with poorly controlled asthma. *JAMA* 2015;313:2033–43.

21 American Lung Association Asthma Clinical Research Centers, Peters SP, Anthonisen N, *et al*. Randomized comparison of strategies for reducing treatment in mild persistent asthma. *N Engl J Med* 2007;356:2027–39.

22 Sin DD, Wu L, Man SFP. The relationship between reduced lung function and cardiovascular mortality: a population-based study and a systematic review of the literature. *Chest* 2005;127:1952–9.

23 Kalhan R, Dransfield MT, Colangelo LA, *et al*. Respiratory symptoms in young adults and future lung disease. The cardia lung study. *Am J Respir Crit Care Med* 2018;197:1616–24.

24 Keene ON, Jones MRK, Lane PW, *et al*. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat* 2007;6:89–97.

25 Daniels MH JW. *Missing data in longitudinal studies; strategies for Bayesian modeling and sensitivity analysis*. Boca Raton: Chapman & Hall/CRC Press, 2008.

26 Little RJA, Rubin DB. *Statistical analysis with missing data. 2nd ed*. Hoboken, N.J: Wiley, 2002.

27 Global Initiative for Asthma. Global strategy for asthma management and prevention, 2021. Available: www.ginasthma.org

28 Togias A, Gergen PJ, Hu JW, *et al*. Rhinitis in children and adolescents with asthma: ubiquitous, difficult to control, and associated with asthma outcomes. *J Allergy Clin Immunol* 2019;143:1003–11.

29 Chu AHY, Ng SHX, Koh D, *et al*. Reliability and validity of the self- and Interviewer-Administered versions of the global physical activity questionnaire (GPAQ). *PLoS One* 2015;10:e0136944.

30 Buysse DJ, Reynolds CF, Monk TH, *et al*. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193–213.