HHS Public Access

Author manuscript

Respirology. Author manuscript; available in PMC 2021 June 04.

Published in final edited form as:

Respirology. 2020 August ; 25(8): 855–862. doi:10.1111/resp.13783.

Reference values for high attenuation areas on chest CT in a healthy, never-smoker, multi-ethnic sample: The MESA study

Imaani EASTHAUSEN[1,2], **Anna PODOLANCZUK**[2], **Eric HOFFMAN**[3], **Steven KAWUT**[4], **Elizabeth OELSNER**[2], **John S. KIM**[5], **Ganesh RAGHU**[6], **Karen Hinckley STUKOVSKY**[7], **Susan REDLINE**[8], **Robyn L. MCCLELLAND**[7], **R Graham BARR**[2,9], **David J. LEDERER**[2,10]

[1]Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA;

[2]Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY, USA;

[3]Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA, USA;

[4]Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA;

[5]Department of Medicine, University of Virginia, Charlottesville, VA, USA;

[6]Department of Medicine, University of Washington, Seattle, WA, USA;

[7]Department of Biostatistics, University of Washington, Seattle, WA, USA;

[8]Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA;

[9]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA;

[10]Regeneron Pharmaceuticals, New York, NY, USA

Abstract

Correspondence: David J. Lederer, Regeneron Pharmaceuticals, 755 Old Saw Mill River Road, Rm 72-233, Tarrytown, NY 10591 USA. david.lederer@regeneron.com.

Author contributions: Conceptualization: I.E., A.P., S.K., J.S.K., G.R., R.G.B., D.J.L. Data curation: E.H., E.O., K.H.-S., R.L.M., R.G.B., D.J.L. Formal analysis: I.E., A.P., R.G.B., D.J.L. Methodology: I.E. Visualization: I.E., D.J.L. Funding acquisition: E.H., R.L.M., R.G.B., D.J.L. Investigation: E.H., E.O., K.H.-S., S.R., R.G.B., D.J.L. Methodology: A.P., E.H., S.K., E.O., J.S.K., K.H.-S., R.L.M., R.G.B., D.J.L. Project administration: K.H.-S., R.L.M., R.G.B., D.J.L. Resources: E.H., E.O., K.H.-S., R.L.M., R.G.B., D.J.L. Software: E.H., K.H.-S. Supervision: S.R., R.L.M., R.G.B., D.J.L. Writing—original draft: I.E., D.J.L. Writing—review and editing: I.E., A.P., E.H., S.K., E.O., J.S.K., G.R., K.H.-S., S.R., R.L.M., R.G.B., D.J.L.

Data availability statement: Anonymized data from the MESA study have been made publicly available at BioLincc (https://biolincc.nhlbi.nih.gov/home/) and/or dbGAP (https://www.ncbi.nlm.nih.gov/gap/).

Disclosure statement: E.H. is a founder and shareholder of VIDA Diagnostics, a company commercializing lung image analysis software developed, in part, at the University of Iowa.

Supplementary Information

Additional supplementary information can be accessed via the *html* version of this article at the publisher's website.

Background and objective: Normative values for HAA—a quantitative, CT-based measure of subclinical ILD—in healthy adults are needed to improve interpretability in clinical and research settings.

Methods: HAA was measured on full-lung CT in 3110 participants in the MESA study. Clinical prediction models were developed using a healthy never-smoker subset with normal spirometry (n = 696). RMSE on cross-validation was used as the primary criterion for model selection. Parametric and non-parametric methods were considered. z-Scores were calculated for the entire study sample. Associations between z-scores and several ILD features were estimated.

Results: In the healthy never-smoker subset, the mean age was 69 years with a range of 54–93 years. The median HAA was 4.3% with a range of 2.7–17.8%. Linear regression had better predictive performance than other methods. The final model included race, height, weight, age and sex. The standard error of the estimate was 1.62 with a cross-validated RMSE of 1.64 and an adjusted R^2 of 0.139. z-Scores were associated with several ILD outcomes in adjusted models, including ILA (OR: 1.40 per z-unit; 95% CI: 1.30, 1.52), exertional dyspnoea (OR: 1.08 per z-unit; 95% CI: 1.02, 1.15) and FVC (expected increase per z-unit: −2.49; 95% CI: −2.95, − 2.03).

Conclusion: We present a reference equation and z-scores to define expected values of HAA on full-lung CT to aid HAA interpretation in middle-aged and older adults.

Keywords

high attenuation area; pulmonary fibrosis; quantitative computed tomography; reference equations; subclinical interstitial lung disease

INTRODUCTION

Interstitial lung disease (ILD) refers to a family of closely related respiratory disorders that cause progressive fibrosis and inflammation in the lung.[1] Idiopathic pulmonary fibrosis (IPF) is the most common ILD. It affects nearly 1 in 200 older U.S. adults over the age of 65 years and carries a poor prognosis.[2,3] Two available therapies have been shown to slow disease progression in patients with mild to moderate IPF[4,5]; however, effective management depends on early disease detection.[6] There are no existing interventions that prevent, halt or reverse the development of fibrotic ILD.[7] Improved methods of identifying early stage ILD could allow for earlier intervention and may help reduce the public health burden of the disease.

High attenuation areas (HAA) are a quantitative, computed tomography (CT)-based measure of subclinical ILD.[8–12] HAA is associated with clinical respiratory outcomes, including ILD-specific hospitalization and death in community-dwelling adults.[9,10] These data support HAA as a biomarker of the earliest biological changes in the lung parenchyma leading to ILD. A better understanding of the distribution of HAA in a healthy sample is needed to improve interpretability in clinical and research settings.

In contrast to pulmonary function testing and quantitative measures of emphysema on CT,[13] a normative range for HAA has not been established. Here, we examine the natural variation of HAA in a healthy never-smoker sample of community-dwelling adults. We present an

age, sex and body size-specific prediction model that defines expected values of HAA and an upper limit of normal (ULN) on full-lung CT. We further demonstrate that adjusted z-scores generated by this model are associated with visually identifiable interstitial lung abnormalities (ILA) antecedent to clinical ILD and with other ILD features and outcomes.

METHODS

Study design and participants

The Multi-Ethnic Study of Atherosclerosis (MESA) is a National Heart, Lung, Blood Institute (NHLBI) funded prospective cohort study of 6814 community-dwelling adults without cardiovascular disease sampled from six communities across the United States between 2000 and 2002.[14] Of these, 3113 underwent full-lung CT imaging at the 10-year follow-up during 2010–2012. Three participants were excluded due to inadequate data. The present study is a cross-sectional analysis of 3110 MESA participants. All MESA participants provided informed consent, and the MESA study was approved by the institutional review boards at all centres.

The healthy never-smoker sample

In accordance with the NHANES III sampling criteria used to develop the spirometry reference equations[15] and validated in the MESA cohort,[16] healthy never-smokers were defined as all participants who denied all of the following: ever-smoking cigarettes, cigars or pipes; physician diagnosis of asthma, emphysema or lung cancer; and respiratory symptoms including chronic cough, chronic phlegm production, exertional dyspnoea and wheezing in chest. Further exclusions included abnormal spirometry, estimated glomerular filtration rate <30 mL/min/m^2 (eGFR), obesity (body mass index, BMI ≥ 35 kg/m^2), self-report of bronchitis or pneumonia in the 2 weeks prior to CT scan, and inadequate data for analysis (Fig. 1). We chose to include overweight participants (BMI between 25.0 and 34.9 kg/m^2) in the healthy never-smoker subset to make this cohort more representative of the general population in the United States where only one-third of adults over the age of 20 years have a BMI less than 25.[17] In total, 696 participants were included in the healthy never-smoker sample.

Respiratory symptoms

Cough, phlegm production and wheezing were assessed at 10-year follow-up during the years 2010–2012 with the following questions administered verbally: ‘Do you usually have a cough on most days for three or more months during the year?’; ‘Do you usually bring up phlegm from your chest on most days for three or more months during the year?’; and ‘In the last 12 months, have you had wheezing or whistling in your chest?’. Exertional dyspnoea was defined as affirmative response to any of the following questions administered verbally: ‘When walking on level ground, do you get more breathless than people your own age?’; ‘When walking up hills or stairs, do you get more breathless than people your own age?’; and ‘Do you ever have to stop walking because of breathlessness?’.

Lung function

Lung function was assessed by spirometry in accordance with the American Thoracic Society (ATS)/European Respiratory Society guidelines as previously described.[16,18]

HAA and interstitial lung abnormalities

Full-lung CT scans were obtained as previously described.[16] Image attenuation was measured using a modified version of the Pulmonary Analysis Software Suite (University of Iowa, Iowa City, Iowa, USA), and HAA were defined as the percent of imaged lung volume having CT attenuation between −600 and −250 Hounsfield Units (HU), as previously described.[8–10,12,19]

ILA on CT scans were defined as involvement of more than 5% of non-dependent lung by reticular abnormalities, ground-glass abnormalities, diffuse centrilobular nodularity, honeycombing, traction bronchiectasis and/or non-emphysematous cysts.[9,20,21]

Statistical analysis

We developed prediction models for full-lung HAA at 10-year follow-up using a healthy never-smoker sample. To optimize predictive performance, parametric and non-parametric methods were considered, and root of mean square error (RMSE) on 10-fold cross-validation was used as the primary criterion for model selection. Support vector machine (SVM) regression with a linear kernel, random forests, boosting, ordinary least squares (OLS) regression and elastic net regression were considered. Cross-validation to specify tuning parameters was nested inside of cross-validation to estimate model performance, and tuning parameter specification was performed on training sets as appropriate to the method.

To preserve comparability across models, cross-validation folds were defined prior to all analyses and the same folds were used for all models fitting (with the exception of tuning parameter specification).

Variable selection was performed by best subset selection for all methods. Candidate variables included race (as an indicator variable), waist circumference, hip circumference, height, weight, BMI, age, sex and a binary variable indicating if the subject received a low dose of radiation on CT scan due to having BMI below a certain threshold.

The presence of non-linear relationships and variable interactions were assessed on OLS models. For each continuous predictor, several non-linear terms, including a smoothing spline and polynomial terms of degree 2 through 6, were examined in the best subset setting. All possible subsets of all possible pairwise interaction terms were examined. Evidence of non-linearity or variable interactions was defined by reduction in RMSE of ≥0.1 when compared with the simplest model. Race/ethnicity- and sex-stratified OLS models were developed using the fitting process described above.

Predicted values of HAA were calculated for each study participant using the selected prediction model, and z-scores were calculated for each individual as described below:

$$\text{z-Score}_i = \frac{O_i - E_i}{\text{SEE}}$$

where O_i is the observed value of HAA for subject i, E_i is the expected value of HAA for subject i and SEE is the standard error of the estimate calculated in the healthy never-smoker subset. The reader should note that SEE is equivalent to the SD of the errors of prediction. Here, we refer to the SD of the errors of prediction as the SEE when it is estimated using the entire study sample and as the RMSE when estimated as a mean over cross-validation samples. We define elevated HAA as the upper fifth percentile of the healthy normal distribution. As the distribution of z-scores is right-skewed, we use the empirical 95th quantile (1.634) as the cut-off value for elevated HAA instead of the 95th quantile of a standard normal distribution.

ULN were defined by one-tailed 95% prediction intervals as follows

$$\text{ULN}_i = O_i + 1.634 \times \text{SEE}$$

where ULN_i is the ULN for subject i, and O_i and SEE are defined as described above. Elevated HAA is defined as $O_i \geq \text{ULN}_i$ or, equivalently, $\text{z-score}_i \geq 1.634$.

Similar methods were used to develop a reference equation, predicted values, ULN and z-scores for HAA measured from cardiac CT scans (Appendix S1, Table S1 in Supplementary Information).

z-Score associations with ILA, exertional dyspnoea and cough were estimated using logistic regression. z-Score associations with lung function measures and log-transformed pack-years were estimated using OLS regression. Associations were adjusted for study site, smoking status, pack-years, waist circumference, eGFR and educational attainment. The reader should note that adjusted z-scores function as a dimensionality reduction variable much like a principal component generated using principal component analysis. This means that the z-score contains *some* of the information present in the variables used to fit the prediction model, but not *all* of that information. Therefore, effect estimates were also examined in a model adjusted for all of the previously mentioned variables in addition to the predictor variables used to calculate the z-score (race, height, weight, age and sex). This model allows the reader to assess how effectively the dimensionality reduction variable (the z-score) is controlling for the confounding of the predictor variables used in the z-score.

The validity of all models used for inference was assessed by visual inspection of standard diagnostic plots and/or Hosmer–Lemeshow tests, as appropriate.

RESULTS

There were 3110 MESA participants with available data who underwent full-lung CT imaging. 52% of participants were women, 27% were Black, 39% were white, 21% were Hispanic and 13% were Chinese. About half (46%) of the participants were never-smokers,

47% were former smokers and 8% were current smokers. The mean ± SD age was 69 ± 6 years, the mean ± SD weight was 70.3 ± 13.8 kg and the mean ± SD height was 165.4 ± 9.9 cm. Mean ± SD percent of predicted forced vital capacity (FVC) was 97.1 ± 17.8%, mean ± SD percent of predicted forced expiratory volume in 1 s (FEV_1) was 94.9 ± 19.9% and mean ± SD FEV_1/FVC ratio was 74.0 ± 9.0% (Table 1).

Of these, 696 met criteria for inclusion in the healthy never-smoker subset (Fig. 1). The healthy never-smoker subset had a higher proportion of women (61.2%); a higher proportion of Chinese (26.1%); and lower proportions of Blacks (23.1%), whites (32.8%), and Hispanics (26.1%) when compared with the complete study sample. In the healthy never-smoker subset, the mean ± SD age (69 ± 9 years) and height (163.4 ± 9.4 cm) were similar to that of the complete study sample, while the mean ± SD weight (70.3 ± 13.8 kg) was lower in the healthy never-smoker subset. The healthy never-smoker subset also had better lung function (mean ± SD: FVC, 102.1 ± 14.7; FEV_1, 103.4 ± 15.1; FEV_1/FVC, 76.8 ± 6.0) when compared with the complete study sample (Table 1).

The distribution of HAA was right-skewed in both the healthy never-smoker subset and in the complete study sample (Fig. 2A). The mean ± SD value of HAA was 5.01 ± 2.38% in the complete study sample and was lower (4.76 ± 1.76%) in the healthy never-smoker subset (Table 1).

Prediction model for HAA using the healthy never-smoker subset

Linear regression with OLS estimation had improved predictive performance when compared with linear regression with an elastic net penalty implying good model stability despite covariate collinearity. OLS models also outperformed non-parametric methods, including random forests, boosting and support vector regression, suggesting good model specification with respect to parametric assumptions (Table 2).

The best-performing OLS model, given below, had an SEE of 1.62, an RMSE of 1.64 on 10-fold cross-validation and an adjusted coefficient of determination (adj. R^2) of 13.9%:

$$\begin{aligned} \mathrm{HAA}_{\mathrm{pred}} &= 10.953841 + 0.695687 \times I(\text{ race= Chinese }) \\ &+ 0.410185 \times I(\text{ race = Black }) + 0.964014 \\ &\times I(\text{ race = Hispanic }) - 0.065887 \times \text{ height} \\ &+ 0.036438 \times \text{ weight } + 0.022561 \times \text{ age} \\ &+ 0.043759 \times I(\text{ sex = male }) \end{aligned}$$

where $\mathrm{HAA}_{\mathrm{pred}}$ is the predicted value of HAA, height is measured in cm, weight is measured in kg and age is measured in years. The distribution of predicted values of HAA was normally distributed and symmetric (Fig. 2B).

The weighted average of the RMSE of the best-performing gender-stratified models was 1.65, suggesting no benefit to gender stratification. The weighted average RMSE for the race/ethnicity-stratified models was 1.56. This reduction in RMSE was not considered substantial enough to justify the increased complexity of stratified prediction models. No evidence of non-linearity or variable interactions was observed in stratified or unstratified models.

HAA z-score distribution

The distribution of z-scores in healthy never-smoker sample (n = 696) was right skewed (skewness: 2.95) with a median of −0.23. The interquartile range was from −0.57 to 0.26 and the range was from −1.44 to 7.45. By definition, 4.8% had elevated HAA (Fig. 2C).

The distribution of z-scores in the complete study sample (n = 3110) had a heavier right tail (skewness: 5.02) when compared with the distribution of z-scores in the healthy never-smoker subset. In the complete study population, the median z-score was −0.21, the mean was 0.07, the SD was 1.39, the interquartile range was from −0.58 to 0.29 and the range was from −2.1 to 23.42. The proportion of those with elevated HAA was higher in the complete study sample (6.2%) when compared with the healthy never-smoker subset.

HAA z-scores and ILA

Among those with elevated HAA (n = 127), 38.6% had ILA compared to 11.0% among those with normal HAA (n = 2279). Associations between z-scores and ILA were observed in the unadjusted model (OR: 1.39 per z-unit; 95% CI: 1.29, 1.51), the minimally adjusted model (OR: 1.39 per z-unit; 95% CI: 1.29, 1.51) and the fully adjusted model (OR per z-unit: 1.40; 95% CI: 1.30, 1.52). Similar results were observed for the association between (binary) elevated HAA and ILA (Table 3). These results are consistent with our previously published work.[9]

The 95th percentile threshold for elevated HAA (z-score = 1.634) was associated with a sensitivity of 0.97 for the detection of ILA and a specificity of 0.06. Alternative thresholds for elevated HAA were examined using receiver operating curve (ROC) analysis. The area under the curve (AUC) was 0.67 (95% CI: 0.64, 0.70). The optimal threshold for ILA detection identified by Youden's J-statistic was −0.02 (sensitivity: 0.61; specificity: 0.66; Fig. 2D).

HAA z-scores and clinical features

Higher HAA z-score was associated with lower FVC in the unadjusted model (mean difference per z-unit: −2.61; 95% CI: −3.10, −2.13), the minimally adjusted model (mean difference per z-unit: −2.49; 95% CI: −2.96, −2.03) and the fully adjusted model (mean difference per z-unit: −2.49; 95% CI: −2.95, − 2.03). Higher HAA z-score was also associated with lower FEV_1 in unadjusted (mean difference per z-unit: −2.00; 95% CI: −2.55, − 1.45), minimally adjusted (mean difference per z-unit: −1.89; 95% CI: −2.42, − 1.36) and fully adjusted (mean difference per z-unit: −1.80; 95% CI: −2.33, − 1.28) models. Higher FEV_1/FVC ratio was also observed to be associated with increasing HAA z-score in unadjusted (mean difference per z-unit: 0.58; 95% CI: 0.25, 0.90), minimally adjusted (mean difference per z-unit: 0.57; 95% CI: 0.25, 0.89) and fully adjusted (mean difference per z-unit increase: 0.65; 95% CI: 0.33, 0.97) models. Similar results were observed for the association between these lung function measures and binary elevated HAA (Table 4).

Higher HAA z-score was associated with dyspnoea in unadjusted (OR: 1.09 per z-unit; 95% CI: 1.03, 1.15), minimally adjusted (OR: 1.08 per z-unit; 95% CI: 1.02, 1.14) and fully adjusted (OR: 1.08 per z-unit; 95% CI: 1.02, 1.15) models (Table 3). An association was

also observed between binary elevated HAA and exertional dyspnoea in the unadjusted (OR: 1.75 per z-unit; 95% CI: 1.28, 2.37) and minimally adjusted models (OR: 1.52 per z-unit; 95% CI: 1.10, 2.09), but the estimate was unstable in the fully adjusted model (OR: 1.37 per z-unit; 95% CI: 0.98, 1.91) (Table 4).

No meaningful associations were observed between z-scores and cough (Table 3) or smoking pack-years (Table 4).

DISCUSSION

We examined several prediction modelling methods in order to develop optimized HAA reference equations and z-scores using a multi-ethnic, healthy, never-smoker sample of older adults. We demonstrate the validity of HAA z-scores as a measure of disease risk by showing that z-scores are associated with ILD features, such as smoking, and respiratory outcomes, including ILA, lung function and exertional dyspnoea.

Spirometric measures of lung function are known to have wide margins of variation in healthy samples.[16,22,23] Clinical interpretations of spirometric measurements require reference to ranges standardized by height, weight, age, sex and race/ethnicity.[16,23] More recently, CT-based measures of emphysematous lung have been shown to vary substantially by these same factors in a healthy non-smoking sample suggesting that demographic and body size characteristics play key roles in the natural variation of lung health measures.[13] Importantly, models containing these same variables were shown to be optimal for predicting HAA out of several variables considered.

z-Scores measure the difference between an individual's observed measurement and what would be expected for that individual based on his or her demographic and anthropometric characteristics. This makes z-scores ideal for clinical decision-making because they adjust for variables that may confound interpretations of 'normal'.[24,25] As z-scores also function as a dimensionality reduction variable, they may be used in place of observed values of HAA in small-sample research settings when the number of predictors is large relative to the sample size, impeding appropriate adjustment. However, we caution that as dimensionality reduction reduces the total information present in confounding variables, they may be underpowered to adjust for confounding.

Our methodology follows the latest recommendations for clinical prediction modelling[26,27] and considers both parametric and non-parametric modelling approaches. In this sample, linear regression with OLS outperformed non-parametric machine learning approaches and the elastic net fit. Parametric methods may outperform non-parametric approaches for small data when the model is well specified with respect to parametric assumptions. OLS regression may also outperform an elastic net penalty when predictor collinearity does not affect coefficient stability and the model is at low risk for overfitting. OLS regression has the added benefit of providing unbiased coefficient estimates.

Some limitations of this study include the sample size of the healthy never-smoker subset used to develop the prediction models and the lack of an external validation set. It is also likely that additional variables not considered here play a role in the natural variation of

HAA. Furthermore, this cross-sectional analysis of the MESA cohort is left truncated on 10-year survivorship.

In conclusion, this study fills an important knowledge gap by establishing a normative range for HAA and by presenting reference equations to define expected values with respect to key variables associated with natural variation in HAA. These tools will aid in the interpretability of HAA in future studies and help to move HAA into the clinical sphere.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

The authors thank the other investigators, staff and the participants of the MESA study for their valuable contributions. A complete list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. This research was supported by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, R01-HL103676, R01-HL-103676-S1, K24-HL-131937, R01 HL077612, K24 HL103844 and R01 HL093081 from the National Heart, Lung, and Blood Institute, and by grants UL1-TR-000040, UL1-TR-001079 and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS).

Abbreviations:

AUC	area under the curve
CT	computed tomography
eGFR	estimated glomerular filtration rate
FEV_1	forced expiratory volume in 1 s
FVC	forced vital capacity
HAA	high attenuation area
ILA	interstitial lung abnormality
ILD	interstitial lung disease
IPF	idiopathic pulmonary fibrosis
MESA	Multi-Ethnic Study of Atherosclerosis
OLS	ordinary least square
RMSE	root of mean square error
ROC	receiver operating curve
ULN	upper limit of normal

REFERENCES

1. Rosas IO, Dellaripa PF, Lederer DJ, Khanna D, Young LR, Martinez FJ. Interstitial lung disease: NHLBI workshop on the primary prevention of chronic lung diseases. Ann. Am. Thorac. Soc 2014; 11(Suppl. 3): S169–77. [PubMed: 24754826]
2. Raghu G, Chen SY, Yeh WS, Maroni B, Li Q, Lee YC, Collard HR. Idiopathic pulmonary fibrosis in US Medicare beneficiaries aged 65 years and older: incidence, prevalence, and survival, 2001–11. Lancet. Respir. Med 2014; 2: 566–72. [PubMed: 24875841]
3. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA et al. An official ATS/-ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am. J. Respir. Crit. Care Med 2011; 183: 788–824. [PubMed: 21471066]
4. King TE Jr, Bradford WZ, Castro-Bernardini S, Fagan EA, Glaspole I, Glassberg MK, Gorina E, Hopkins PM, Kardatzke D, Lancaster L et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N. Engl. J. Med 2014; 370: 2083–92. [PubMed: 24836312]
5. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, Cottin V, Flaherty KR, Hansell DM, Inoue Y et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N. Engl. J. Med 2014; 370: 2071–82. [PubMed: 24836310]
6. Caminati A, Cassandro R, Torre O, Harari S. Severe idiopathic pulmonary fibrosis: what can be done? Eur. Respir. Rev 2017; 26: 170047. [PubMed: 28954763]
7. Soo E, Adamali H, Edey AJ. Idiopathic pulmonary fibrosis: current and future directions. Clin. Radiol 2017; 72: 343–55. [PubMed: 28118995]
8. Lederer DJ, Enright PL, Kawut SM, Hoffman EA, Hunninghake G, van Beek EJ, Austin JH, Jiang R, Lovasi GS, Barr RG. Cigarette smoking is associated with subclinical parenchymal lung disease: the Multi-Ethnic Study of Atherosclerosis (MESA)-lung study. Am. J. Respir. Crit. Care Med 2009; 180: 407–14. [PubMed: 19542480]
9. Podolanczuk AJ, Oelsner EC, Barr RG, Hoffman EA, Armstrong HF, Austin JH, Basner RC, Bartels MN, Christie JD, Enright PL et al. High attenuation areas on chest computed tomography in community-dwelling adults: the MESA study. Eur. Respir. J 2016; 48: 1442–52. [PubMed: 27471206]
10. Podolanczuk AJ, Oelsner EC, Barr RG, Bernstein EJ, Hoffman EA, Easthausen IJ, Stukovsky KH, RoyChoudhury A, Michos ED, Raghu G et al. High-attenuation areas on chest computed tomography and clinical respiratory outcomes in community-dwelling adults. Am. J. Respir. Crit. Care Med 2017; 196: 1434–42. [PubMed: 28613921]
11. Cronkhite JT, Xing C, Raghu G, Chin KM, Torres F, Rosenblatt RL, Garcia CK. Telomere shortening in familial and sporadic pulmonary fibrosis. Am. J. Respir. Crit. Care Med 2008; 178: 729–37. [PubMed: 18635888]
12. Bernstein EJ, Barr RG, Austin JHM, Kawut SM, Raghu G, Sell JL, Hoffman EA, Newell JD Jr, Watts JR Jr, Nath PH et al. Rheumatoid arthritis-associated autoantibodies and subclinical interstitial lung disease: the multi-ethnic study of atherosclerosis. Thorax 2016; 71: 1082–90. [PubMed: 27609750]
13. Hoffman EA, Ahmed FS, Baumhauer H, Budoff M, Carr JJ, Kronmal R, Reddy S, Barr RG. Variation in the percent of emphysema-like lung in a healthy, nonsmoking multiethnic sample. The MESA lung study. Ann. Am. Thorac. Soc 2014; 11: 898–907. [PubMed: 24983825]
14. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K et al. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am. J. Epidemiol 2002; 156: 871–81. [PubMed: 12397006]
15. 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. [PubMed: 9872837]
16. Hankinson JL, Kawut SM, Shahar E, Smith LJ, Stukovsky KH, Barr RG. Performance of American Thoracic Society-recommended spirometry reference values in a multiethnic sample of adults: the Multi-Ethnic Study of Atherosclerosis (MESA) lung study. Chest 2010; 137: 138–45. [PubMed: 19741060]

17. Fryer CD, Carroll MD, Ogden CL. Prevalence of overweight, obesity, and severe obesity among adults aged 20 and over: United States, 1960–1962 through 2015–2016. NCHS Health E-Stats. National Center for Health Statistics, Centers for Disease Control and Prevention, US Department of Health and Human Services, 2018. [Accessed 7 Feb 2020.] Available at https://www.cdc.gov/nchs/data/hestat/obesity_adult_15_16/obesity_adult_15_16.pdf.

18. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, Crapo R, Enright P, van der Grinten CP, Gustafsson P et al. Standardisation of spirometry. Eur. Respir. J 2005; 26: 319–38. [PubMed: 16055882]

19. Kim JS, Podolanczuk AJ, Borker P, Kawut SM, Raghu G, Kaufman JD, Stukovsky KDH, Hoffman EA, Barr RG, Gottlieb DJ et al. Obstructive sleep apnea and subclinical interstitial lung disease in the Multi-Ethnic Study of Atherosclerosis (MESA). Ann. Am. Thorac. Soc 2017; 14: 1786–95. [PubMed: 28613935]

20. Washko GR, Hunninghake GM, Fernandez IE, Nishino M, Okajima Y, Yamashiro T, Ross JC, Estepar RS, Lynch DA, Brehm JM et al. Lung volumes and emphysema in smokers with interstitial lung abnormalities. N. Engl. J. Med 2011; 364: 897–906. [PubMed: 21388308]

21. Washko GR, Lynch DA, Matsuoka S, Ross JC, Umeoka S, Diaz A, Sciurba FC, Hunninghake GM, San Jose Estepar R, Silverman EK et al. Identification of early interstitial lung disease in smokers from the COPDGene study. Acad. Radiol 2010; 17: 48–53. [PubMed: 19781963]

22. Morris JF, Koski A, Johnson LC. Spirometric standards for healthy nonsmoking adults. Am. Rev. Respir. Dis 1971; 103: 57–67. [PubMed: 5540840]

23. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MS, Zheng J et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. Eur. Respir. J 2012; 40: 1324–43. [PubMed: 22743675]

24. Curtis AE, Smith TA, Ziganshin BA, Elefteriades JA. The mystery of the Z-score. Aorta (Stamford) 2016; 4: 124–30. [PubMed: 28097194]

25. Colan SD. The why and how of Z scores. J. Am. Soc. Echocardiogr 2013; 26: 38–40. [PubMed: 23261367]

26. Steyerberg E Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York, Springer Science and Business Media, 2008.

27. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur. Heart J 2014; 35: 1925–31. [PubMed: 24898551]

SUMMARY AT A GLANCE

To better understand the natural variation of HAA (a novel quantitative CT-based measure of subclinical ILD), we developed HAA reference equations and z-scores to define expected values of HAA with adjustment for key demographic and anthropometric variables, and we demonstrated that HAA z-scores correlate with several ILD features.

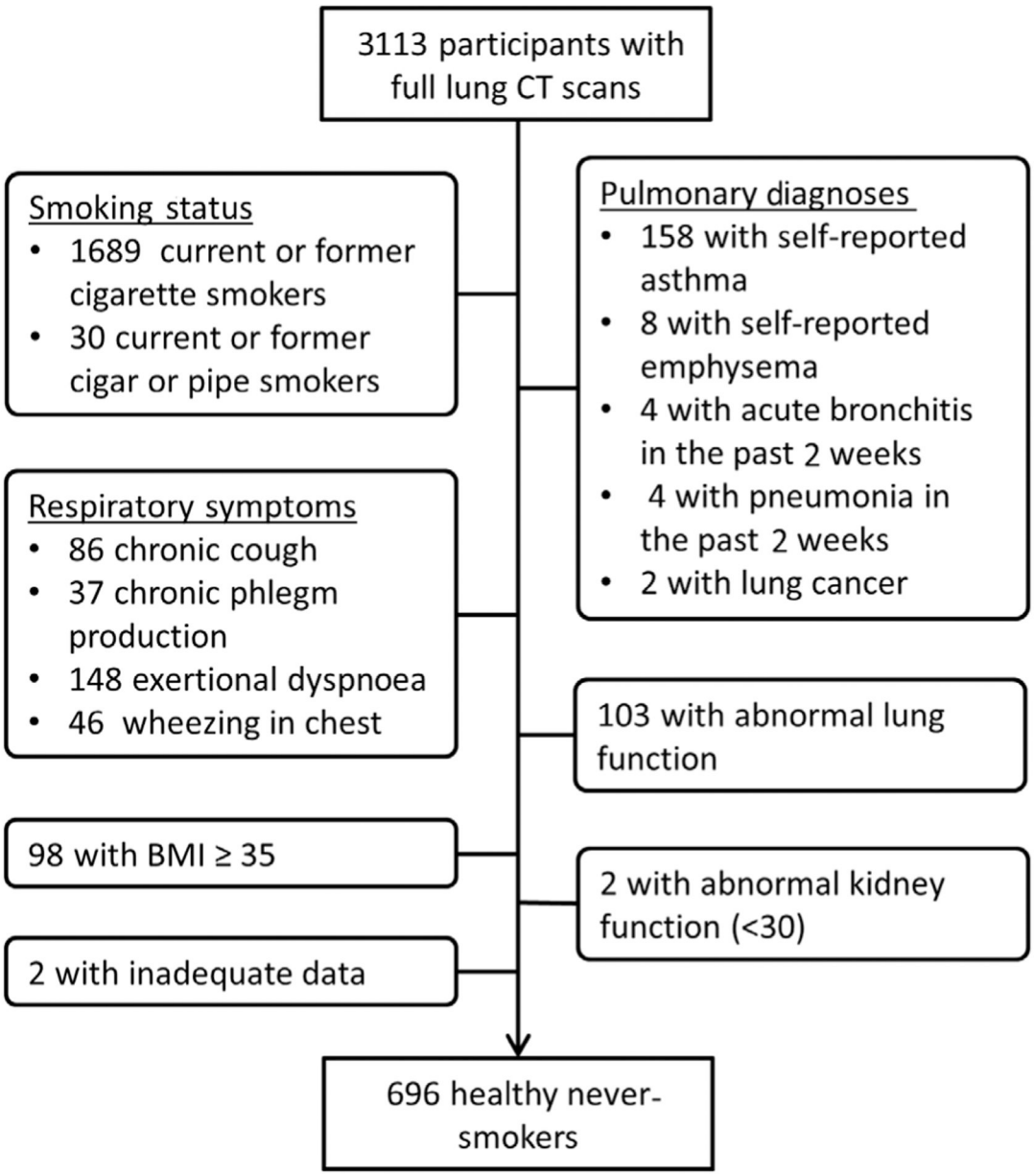

Figure 1.
Flow diagram of MESA participants included in the healthy never-smoker subset. BMI, body mass index; CT, computed tomography; MESA, Multi-Ethnic Study of Atherosclerosis.

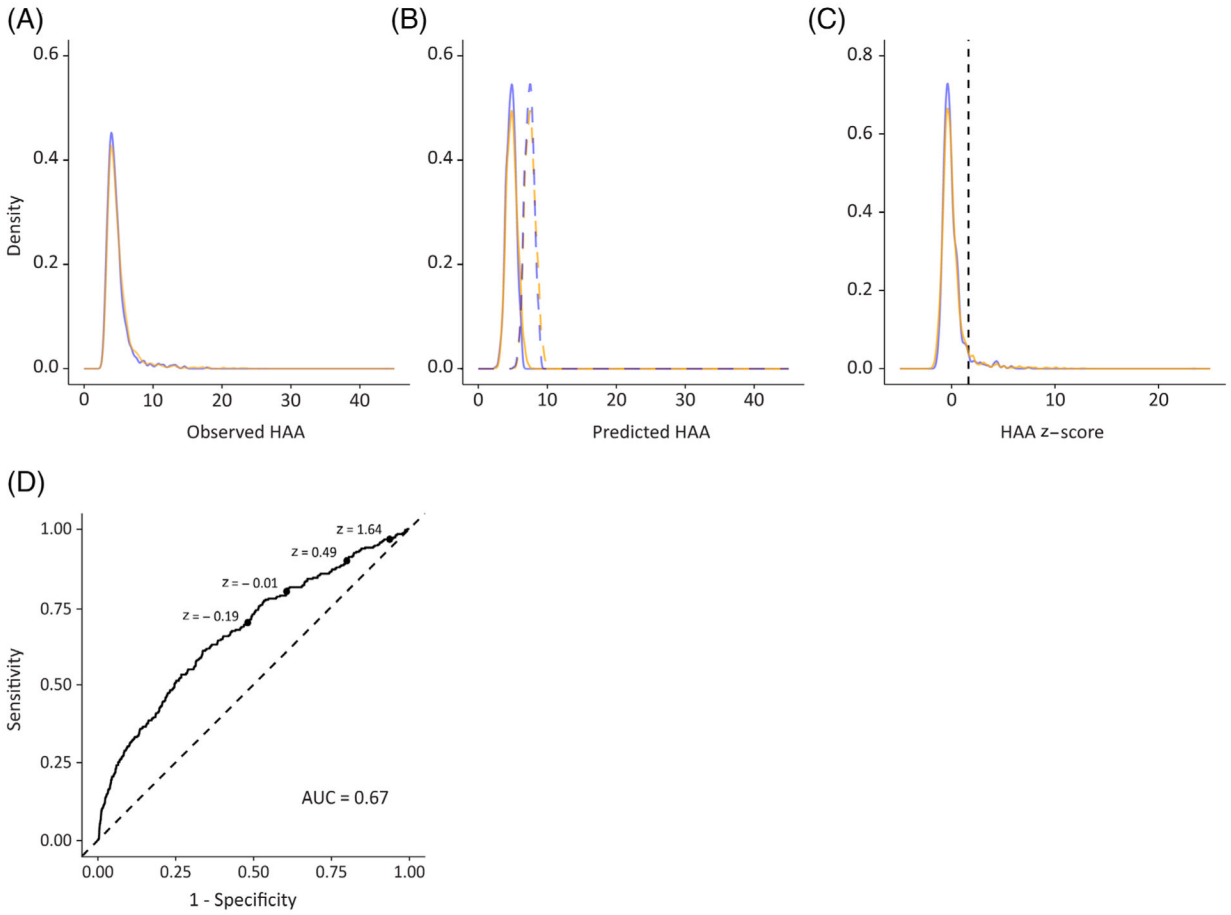

Figure 2.

Density plots showing the distribution of: (A) observed HAA values, (B) predicted HAA values (solid) and ULN (dashed) and (C) z-scores (solid) and ULN (z-score = 1.634). Blue denotes the healthy never-smoker subset and orange denotes the complete study sample. (D) ROC curve showing sensitivity versus false positive rate (1 − specificity) for ILA detection for varying elevated HAA threshold values. AUC, area under the curve; HAA, high attenuation area; ILA, interstitial lung abnormality; ROC, receiver operating curve; ULN, upper limit of normal.

Table 1

Characteristics of study participants

	Complete study population	Healthy never-smoker subset
Participants, *n*	3110	696
HAA (%)	5.01 ± 2.38	4.76 ± 1.76
Age (years)	69 ± 9	69 ± 9
Sex		
Male	1489 (47.9%)	270 (38.8%)
Female	1621 (52.1%)	426 (61.2%)
Race/ethnicity		
Black	839 (27.0%)	161 (23.1%)
White	1202 (38.6%)	228 (32.8%)
Hispanic	658 (21.2%)	125 (18.0%)
Chinese	411 (13.2%)	182 (26.1%)
Weight (kg)	78.3 ± 17.5	70.3 ± 13.8
Height (cm)	165.4 ± 9.9	163.4 ± 9.4
Smoking history		
Never-smokers	1417 (45.6%)	696 (100%)
Former smokers	1459 (46.9%)	0 (0%)
Current smokers	234 (7.5%)	0 (0%)
Cigarette pack-years[†]	20 ± 25	0 ± 0
Lung function		
FVC (percent predicted)	97.1 ± 17.8	102.1 ± 14.7
FEV_1 (percent predicted)	94.9 ± 19.9	103.4 ± 15.1
FEV_1/FVC ratio	74.0 ± 9.0	76.8 ± 6.0

Data are presented as mean ± SD or *n* (%). All variables were measured at MESA 10-year follow-up examinations during the years 2010–2012.

[†]Among current and former smokers.

FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; HAA, high attenuation area; MESA, Multi-Ethnic Study of Atherosclerosis.

Table 2

Predictive performance of parametric and non-parametric modelling approaches for prediction of HAA measured from full lung CT scans

Modelling method	Variables included in the best subset	RMSE
Linear regression with least squares estimation	Race, height, weight, age, sex	1.64
Linear regression with elastic net penalty	Race, height, weight, age	1.72
Random forests	Race, weight, age, sex	1.68
Boosting	Race, sex	1.70
Support vector regression	Weight	1.81

RMSE is calculated on 10-fold cross-validation.

CT, computed tomography; HAA, high attenuation area; RMSE, root of mean square error.

Table 3

Associations between HAA z-scores and ILA, dyspnoea and cough

	ILA				Exertional dyspnoea				Cough			
	n	**Events**	**OR (95% CI)**	***P*-value**	***n***	**Events**	**OR (95% CI)**	***P*-value**	***n***	**Events**	**OR (95% CI)**	***P*-value**
HAA z-score												
Unadjusted	2406	299	1.39 (1.29, 1.51)	<0.001	3085	793	1.09 (1.03, 1.15)	0.002	3104	307	1.01 (0.92, 1.09)	0.81
Model 1	2352	285	1.39 (1.29, 1.51)	<0.001	3009	779	1.08 (1.02, 1.14)	0.01	3028	303	1.01 (0.93, 1.09)	0.77
Model 2	2352	285	1.40 (1.30, 1.53)	<0.001	3009	779	1.08 (1.02, 1.15)	0.01	3028	303	1.02 (0.93, 1.10)	0.68
Elevated HAA												
Unadjusted	2406	299	5.10 (3.47, 7.44)	<0.001	3085	793	1.75 (1.28, 2.37)	<0.001	3104	307	1.07 (0.64, 1.69)	0.78
Model 1	2352	285	5.42 (3.65, 8.00)	<0.001	3009	779	1.52 (1.10, 2.09)	0.01	3028	303	1.06 (0.64, 1.69)	0.80
Model 2	2352	285	5.68 (3.74, 8.60)	<0.001	3009	779	1.37 (0.98, 1.91)	0.06	3028	303	1.11 (0.66, 1.77)	0.68

OR are estimated by logistic regression, *P*-values are calculated from two-sided Wald tests and error rate is calculated by 10-fold cross-validation. Elevated HAA is defined as the upper fifth percentile of the population (z-score ≥ 1.645). Model 1 is adjusted for study site, smoking status, pack-years, waist circumference, estimated glomerular filtration rate and educational attainment. Model 2 is adjusted for all terms included in model 1 and race, height, BMI, age and sex.

BMI, body mass index; HAA, high attenuation area; ILA, interstitial lung abnormality; OR, odds ratio per z-unit increase or for elevated HAA compared to non-elevated HAA.

Table 4

Associations between HAA z-scores and lung function and pack-years

	FVC			FEV_1			FEV_1/FVC			Pack-years†		
	n	**Expected rate of change (95% CI)**	***P*-value**	***n***	**Expected rate of change (95% CI)**	***P*-value**	***n***	**Expected rate of change (95% CI)**	***P*-value**	***n***	**Expected rate of change (95% CI)**	***P*-value**
HAA z-score												
Unadjusted	2740	−2.61 (−3.10, −2.13)	<0.001	2735	−2.00 (−2.55, −1.45)	<0.001	2789	0.58 (0.25, 0.90)	0.001	1425	0.06 (0.00, 0.11)	0.06
Model 1	2674	− 2.49 (− 2.96, −2.03)	<0.001	2669	−1.89 (−2.42, −1.36)	<0.001	2669	0.57 (0.25, 0.89)	<0.001	1412	0.06 (0.00, 0.11)	0.06
Model 2	2674	−2.49 (−2.95, −2.03)	<0.001	2669	−1.80 (−2.33, −1.28)	<0.001	2669	0.65 (0.33, 0.97)	<0.001	1412	0.05 (−0.01, 0.10)	0.10
Elevated HAA												
Unadjusted	2740	−8.05 (−10.91, −5.18)	<0.001	2735	−5.12 (−8.36, −1.88)	0.002	2789	2.86 (0.95, 4.77)	0.003	1425	0.18 (−0.14, 0.49)	0.28
Model 1	2674	−6.61 (−9.39, − 3.83)	<0.001	2669	−3.83 (−6.97, −0.68)	0.02	2669	2.74 (0.86, 4.61)	0.004	1412	0.15 (−0.16, 0.47)	0.34
Model 2	2674	−7.57 (−10.31, −4.84)	<0.001	2669	−5.03 (−8.13, −1.93)	0.001	2669	2.32 (0.46, 4.18)	0.01	1412	0.21 (−0.10, 0.52)	0.19

Associations are estimated by ordinary least squares regressions, *P*-values are calculated from two-sided t-tests and RMSE is calculated on 10-fold cross-validation. Elevated HAA is defined as the upper fifth percentile of the population (z-score ≥ 1.645). Model 1 is adjusted for study site, smoking status, pack-years, waist circumference, estimated glomerular filtration rate and educational attainment. Model 2 is adjusted for all terms included in model 1 and race, height, BMI, age and sex.

†Analysis is conducted among current and former smokers, and models are not adjusted for smoking status or pack-years.

BMI, body mass index; FEV_1, forced expiratory volume in 1 s; FVC, forced vital capacity; HAA, high attenuation area; RMSE, root of mean square error.