Published in final edited form as:

N Engl J Med. 2024 June 13; 390(22): 2083–2097. doi:10.1056/NEJMsa2311809.

Implications of Race Adjustment in Lung-Function Equations

J.A. Diao,
Y. He,
R. Khazanchi,
M.J. Nguemeni Tiako,
J.I. Witonsky,
E. Pierson,
P. Rajpurkar,
J.R. Elhawary,
L. Melas-Kyriazi,
A. Yen,
A.R. Martin,
S. Levy,
C.J. Patel,
M. Farhat,
L.N. Borrell,
M.H. Cho,
E.K. Silverman,
E.G. Burchard,
A.K. Manrai

the Department of Biomedical Informatics, Harvard Medical School (J.A.D., P.R., L.M.-K., C.J.P., M.F., A.K.M.), the Computational Health Informatics Program, Boston Children's Hospital (J.A.D., A.K.M.), the Analytic and Translational Genetics Unit (Y.H., A.R.M.) and the Division of Pulmonary and Critical Care Medicine, Department of Medicine (M.F.), Massachusetts General Hospital, Harvard Internal Medicine–Pediatrics Combined Residency Program, Brigham and Women's Hospital, Boston Children's Hospital, and Boston Medical Center (R.K.), the François-Xavier Bagnoud Center for Health and Human Rights, Harvard University (R.K.), the Department of Medicine (M.J.N.T.) and the Channing Division of Network Medicine and the Division of Pulmonary and Critical Care Medicine, Department of Medicine (M.H.C., E.K.S.), Brigham and Women's Hospital, and the Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, Beth Israel Deaconess Medical Center (S.L.), Boston, and the Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge (Y.H., A.R.M.) — all in Massachusetts; the Departments of Pediatrics (J.I.W.), Medicine (J.R.E., E.G.B.),

Dr. Manrai can be contacted at arjun_manrai@hms.harvard.edu or at the Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck St., Boston, MA 02115.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

A data sharing statement provided by the authors is available with the full text of this article at NEJM.org.

and Bioengineering and Therapeutic Sciences (J.R.E., E.G.B.), University of California, San Francisco, San Francisco; the Department of Computer Science, Cornell University, Ithaca (E.P.), and the Department of Population Health Sciences, Weill Cornell Medical College (E.P.), and the Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, City University of New York (L.N.B.), New York — all in New York; the Department of Engineering Science, University of Oxford, Oxford, United Kingdom (L.M.-K.); and the Medical Scientist Training Program, University of Illinois at Chicago, Chicago (A.Y.).

Abstract

BACKGROUND—Adjustment for race is discouraged in lung-function testing, but the implications of adopting race-neutral equations have not been comprehensively quantified.

METHODS—We obtained longitudinal data from 369,077 participants in the National Health and Nutrition Examination Survey, U.K. Biobank, the Multi-Ethnic Study of Atherosclerosis, and the Organ Procurement and Transplantation Network. Using these data, we compared the race-based 2012 Global Lung Function Initiative (GLI-2012) equations with race-neutral equations introduced in 2022 (GLI-Global). Evaluated outcomes included national projections of clinical, occupational, and financial reclassifications; individual lung-allocation scores for transplantation priority; and concordance statistics (C statistics) for clinical prediction tasks.

RESULTS—Among the 249 million persons in the United States between 6 and 79 years of age who are able to produce high-quality spirometric results, the use of GLI-Global equations may reclassify ventilatory impairment for 12.5 million persons, medical impairment ratings for 8.16 million, occupational eligibility for 2.28 million, grading of chronic obstructive pulmonary disease for 2.05 million, and military disability compensation for 413,000. These potential changes differed according to race; for example, classifications of nonobstructive ventilatory impairment may change dramatically, increasing 141% (95% confidence interval [CI], 113 to 169) among Black persons and decreasing 69% (95% CI, 63 to 74) among White persons. Annual disability payments may increase by more than $1 billion among Black veterans and decrease by $0.5 billion among White veterans. GLI-2012 and GLI-Global equations had similar discriminative accuracy with regard to respiratory symptoms, health care utilization, new-onset disease, death from any cause, death related to respiratory disease, and death among persons on a transplant waiting list, with differences in C statistics ranging from −0.008 to 0.011.

CONCLUSIONS—The use of race-based and race-neutral equations generated similarly accurate predictions of respiratory outcomes but assigned different disease classifications, occupational eligibility, and disability compensation for millions of persons, with effects diverging according to race. (Funded by the National Heart Lung and Blood Institute and the National Institute of Environmental Health Sciences.)

Spirometry, a widely used test of lung function, is essential for the diagnosis, staging, and monitoring of lung disease. For more than a century, clinicians have interpreted spirometric measurements — including forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) — by means of comparison with a predicted normal range representing expected healthy values.[1] These norms are calculated on the basis of age, sex, height, and often race with the use of reference equations, which were designed to predict measured spirometric values in healthy nonsmokers. In clinical practice, these equations provide a

demographic-specific distribution of expected spirometric values against which measured spirometric values may be compared. For example, conditions such as chronic obstructive pulmonary disease (COPD) may be diagnosed by comparing measured spirometric values to the 5th-percentile lower limit of the normal range.[2] The degree of impairment (e.g., the COPD grade) may also be quantified by comparing reference-adjusted values against fixed thresholds that define mild, moderate, and severe disease[2,3]; such values include the percent of the predicted value (the ratio of the measured value to the predicted healthy value, with normal values typically considered to fall between 80% and 120% of the predicted healthy value) and z score (the number of standard deviations by which a measured value is above or below the predicted healthy value).

Adjustment for race in clinical algorithms has prompted controversy with regard to medicine generally[4] and lung function specifically[5,6] owing to its historical use in quantifying presumed deficiencies in Black persons and justifying their enslavement.[7] More recently, critical discussion has also emphasized outdated notions regarding racial essentialism and its effects on medical and economic inequalities.[8–12] In 2021, a technical standard from the European Respiratory Society (ERS) and the American Thoracic Society (ATS) stated that the "historical approach of fixed adjustment factors for race is not appropriate and is unequivocally discouraged."[2] In 2022, the Global Lung Function Initiative (GLI) sought to replace race-based GLI-2012 equations[13] with new race-neutral equations (GLI-Global equations) that do not include race or ethnic group as inputs.[14] GLI-Global equations were derived with the use of the same data and effectively constitute a weighted average across racial groups. As of early 2024, GLI-Global equations are the only lung-function reference equations officially endorsed by ATS and ERS.[15]

Although it is well-established that the choice of reference equation involves trade-offs,[14–17] the downstream consequences of including or removing race as an adjustment factor have not been comprehensively quantified. The consequences include potential changes to the predictive capacity of reference-adjusted lung-function indexes as well as the clinical, occupational, and financial outcomes that these indexes are used to determine. Using data from five cohorts, we quantified changes that are expected with widespread adoption of race-neutral equations, including U.S. national projections for reclassifications of lung disease, employment eligibility, and disability compensation; historical effects on lung-transplant priority; and discriminative accuracy of reference-adjusted lung-function indexes for respiratory symptoms, health care utilization, new-onset disease, and death.

METHODS

POPULATION CHARACTERISTICS

Data were obtained from five cohorts: 17,067 participants from the National Health and Nutrition Examination Survey (NHANES) 2007–2012 (NHANES IV),[18] 15,861 from NHANES 1988–1994 (NHANES III),[19] 290,136 from the U.K. Biobank,[20] 3262 from the Multi-Ethnic Study of Atherosclerosis (MESA),[21] and 42,751 from the Organ Procurement and Transplantation Network (OPTN)[22] (Table 1). Participants were selected on the basis of acceptable spirometric data and recorded age and height. Spirometry quality control was conducted in accordance with ATS–ERS standards, including quality grades of A or B (on

a scale of A to F, where A and B represent better-quality results) (see the Supplementary Methods section in the Supplementary Appendix, available with the full text of this article at NEJM.org).

MESA, NHANES, and U.K. Biobank questionnaire data included participant-reported age, sex or gender, race and ethnic-group identification, medical conditions, and smoking behaviors; they also contained data from medical examinations, including height and spirometric measurements. OPTN data, which were reported by transplantation centers, included the same data fields as the MESA, NHANES, and U.K. Biobank questionnaires, with the addition of dates of referral, transplantation, or death as well as all data inputs required for calculating the lung-allocation score that was used in 2020. NHANES III and U.K. Biobank data also contained longitudinal outcomes with regard to new-onset disease and death. MESA and NHANES data were used to develop GLI-2012 and GLI-Global equations, but U.K. Biobank and OPTN data were not. NHANES IV was designed to provide a representative sample of the civilian noninstitutionalized U.S. population. OPTN data represented all persons on the 2020 U.S. lung-transplant waiting list.

Because changes resulting from including or excluding adjustment for race are expected to vary across groups defined according to the GLI racial taxonomy, we report outcomes stratified according to race or ethnic group for Black, Hispanic, and White participants. The remaining participants, who would be assigned the “Other” adjustment when GLI-2012 equations are used, were reported as “Asian or Other” to reflect that Asian persons made up the majority of that group. Additional details regarding the use of race and ethnic-group data in reporting outcomes are provided in the Supplementary Methods section.

OUTCOME DEFINITIONS

Spirometric and other criteria for the study outcomes are provided in Table 2. Obstructive ventilatory impairment, involving increased resistance to airflow, was defined as a ratio of FEV_1 to FVC that was below the 5th-percentile lower limit of the normal range.[2] Nonobstructive impairment, involving diverse intrapulmonary and extrapulmonary causes, was defined as either FEV_1 (preserved-ratio impaired spirometry[27]) or FVC (restrictive pattern) below their respective 5th-percentile lower limit of the normal ranges, with an FEV_1:FVC above the 5th-percentile lower limit of the normal range.[2] COPD was defined as an FEV_1:FVC of less than 0.70 and was assigned severity grades on the basis of the percent of the predicted FEV_1 and criteria defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD).[3]

We calculated medical impairment ratings among adult participants with work-related exposure to dust or fumes using the 2008 American Medical Association (AMA) Guides to the Evaluation of Permanent Impairment.[24] To assess occupational eligibility, we used the 2018 National Fire Protection Association Standard 1582 to identify adult participants whose lung function may disqualify them from firefighting occupations.[23] To quantify changes to compensation, we calculated Department of Veterans Affairs (VA) disability payments among adult veterans using the VA schedule for rating respiratory disabilities and the disability compensation rates for 2023.[25,26] Finally, we calculated the lung-allocation score, position on the waiting list, and expected wait time according to reference equation

for each of the 1399 persons listed on the 2020 U.S. transplant waiting list using scoring parameters and baseline survival data from 2020.[28]

STATISTICAL ANALYSIS

We used GLI-2012 and GLI-Global equations to calculate predicted normal FEV_1, FVC, and FEV_1:FVC values and the 5th-percentile lower limit of the normal range for all participants. We then used measured spirometry to derive percent of the predicted values and z scores. Using data from NHANES IV and applying appropriate survey weights, we calculated nationally representative projections for changes in clinical, occupational, and financial outcomes among the population of persons 6 to 79 years of age in the United States able to produce high-quality spirometric results (see the Supplementary Methods section). Movement on the lung-transplant waiting list was calculated on the basis of changes in lung-allocation score resulting from changes in percent of the predicted FVC. Following the 2020 policy of the United Network for Organ Sharing, we used spirometry inputs to calculate the lung-allocation score only among candidates assigned to the diagnosis group D (restrictive lung diseases).[28] Expected wait time was estimated from the initial position on the waiting list with the use of a linear equation derived from OPTN data (Fig. S1 in the Supplementary Appendix).

A key rationale for using reference-adjusted indexes (e.g., z scores or percent of the predicted values) rather than raw spirometric measurements is the improved ability to distinguish between states of health and disease.[29] To assess this ability, a statistical measure known as discriminative accuracy, we calculated Harrell's concordance statistics (C statistics) for the prediction of respiratory outcomes on the basis of spirometric z scores. For binary end points, the C statistic represents the probability that a random participant with a given clinical outcome (e.g., death) has lower lung function than a random participant without that clinical outcome; the C statistic is equivalent to the area under the receiver-operating-characteristic (ROC) curve. Notably, the C statistic measures an average performance across all possible lung-function thresholds and does not reference any specific threshold. A C statistic of 1.0 indicates perfect discrimination, whereas a C statistic of 0.5 indicates discriminative ability no better than random. In a secondary analysis, we calculated the sensitivity and specificity for predicting respiratory outcomes using a z-score threshold of −1.645, corresponding to the 5th-percentile lower limit of the normal range. Data regarding concurrent symptoms and recent health care utilization were derived from NHANES IV, data regarding new-onset asthma and COPD were derived from the U.K. Biobank, data regarding death from respiratory causes and death from any cause were derived from the 2019 Linked Mortality File for NHANES III, and data regarding deaths that occurred among persons on the lung-transplant waiting list were obtained from OPTN. Additional details regarding statistical analyses are provided in the Supplementary Methods section.

RESULTS

STUDY POPULATION

The 369,077 participants with acceptable results on spirometry are described in Table 1 and represented five demographically and socioeconomically diverse cohorts (Table S1). NHANES III and IV represented younger participants (median ages of 35 and 33, respectively) than MESA, the U.K. Biobank, and OPTN (median age range, 57 to 65). The U.K. Biobank and OPTN had a higher percentage of White participants (95.1% and 78.2%, respectively) than the other cohorts (percentages ranged from 35.2 to 39.6%). Median FVC across the cohorts ranged from 1.82 liters among lung-transplant candidates in OPTN to 3.62 liters among the NHANES IV population. When transplant data were omitted, the percentages of participants without respiratory symptoms, lung disease, or smoking history ranged from 31.6% in MESA to 56.2% in NHANES IV.

VENTILATORY IMPAIRMENT

Obstructive impairment is associated with disorders of airflow limitation (e.g., asthma and COPD). As compared with GLI-2012 equations, the use of GLI-Global equations with NHANES IV data resulted in increased findings of obstructive impairment among Black, Hispanic, and White participants and decreased findings among participants of Asian or other race or ethnic group (Fig. 1A). Scaled to the U.S. population, these changes in findings amount to 3.20 million reclassifications (95% confidence interval [CI], 2.63 million to 3.86 million): 2.64 million persons newly classified with obstruction and 565,000 no longer classified with obstruction (Table 3). Precise values for prevalence and total affected numbers are provided in Table S2, prevalence changes and relative changes are shown in Table S3, and reclassifications are shown in Table S4.

Nonobstructive impairment is a nonspecific finding that often involves follow-up testing for restrictive disease, early obstruction, muscle weakness, and other causes.[2] When GLI-Global equations were used, these findings more than doubled among Black persons and decreased by a factor of 3 to 4 among Hispanic and White persons (Fig. 1B). Scaled to the U.S. population, these changes amount to 2.34 million additional findings of nonobstructive impairment (95% CI, 1.93 million to 2.75 million) among Black persons, 1.37 million fewer findings (95% CI, 0.94 million to 1.80 million) among Hispanic persons, and 5.37 million fewer findings (95% CI, 4.19 million to 6.55 million) among White persons (Table 3). Similar relative changes were observed when GLI-Global equations were applied to NHANES III, MESA, and U.K. Biobank data (Fig. S2A and S2B). In total, 12.5 million persons may have reclassification of obstructive or nonobstructive ventilatory impairment.

COPD SEVERITY GRADING

Spirometry is also used to grade COPD severity. When lung-function measurements for participants in NHANES IV were interpreted with the use of GLI-Global rather than GLI-2012 equations, classifications of moderate-to-severe COPD increased among Black participants and decreased among Hispanic and White participants (Fig. 1C). Scaled to the U.S. population, this amounts to 428,000 additional Black persons (95% CI, 300,000 to 556,000) and 1.10 million fewer White persons (95% CI, 0.72 million to 1.48 million) with

moderate-to-severe COPD (Table 3). The use of GLI-Global equations reclassified severity grades for 2.05 million persons with COPD (95% CI, 1.59 million to 2.51 million): 508,000 to more-severe grades and 1.54 million to less-severe grades (Table S4G). Similar relative changes were observed in other cohorts (Fig. S2C).

OCCUPATIONAL ELIGIBILITY

In some occupations, spirometric criteria are used to determine employment eligibility. When GLI-Global equations were used to assess adults with work-related exposures to dust or fumes, disqualifications from firefighting professions nearly doubled among Black adults and decreased by one fourth among White adults (Fig. 1D). This change amounts to 754,000 Black adults (95% CI, 540,000 to 969,000) who may no longer be eligible for firefighting jobs and 1.27 million White adults (95% CI, 0.81 million to 1.73 million) who may become newly eligible (Table 3). Overall, 2.28 million working-age U.S. adults (95% CI, 1.84 million to 2.78 million) may be subject to changes in firefighting eligibility. This estimate includes eligibility changes among the 1 million active firefighters in the United States[30] and the many applicants for firefighting jobs, but also includes applicants who would be excluded by other physical evaluations and persons not applying to firefighter jobs.

MEDICAL IMPAIRMENT RATINGS

Medical impairment ratings are assigned by clinicians to guide decisions regarding work eligibility and disability compensation. Among Black adults with work-related exposures to dust or fumes, classifications of moderate-to-severe medical impairment may more than double when GLI-Global equations are used (Fig. 1E). This amounts to 638,000 Black adults (95% CI, 478,000 to 797,000) who may receive increased payments for impairment-based compensation (Table 3). In contrast, moderate-to-severe impairment may decrease by one fourth among White adults, affecting 938,000 persons (95% CI, 570,000 to 1.31 million). The use of GLI-Global equations may reassign AMA impairment ratings for 8.16 million adults (95% CI, 6.93 million to 9.39 million): 2.68 million to more-severe impairment and 5.49 million to less-severe impairment (Table S4H).

DISABILITY COMPENSATION

The amounts of VA disability payments are determined in part on the basis of spirometric criteria. The use of GLI-Global equations to calculate compensation for respiratory impairment associated with military service among Black veterans may increase payments by 17.1% (95% CI, 8.5 to 25.8) (Fig. 1F), amounting to $1.10 billion (95% CI, 0.58 billion to 1.61 billion) annually (Table 3). Among the 216,000 Black veterans (9.5%) who stand to benefit, annual compensation could increase by $1,991 for 37.0% of that population, by $4,110 for 41.1%, by $9,740 for 19.0%, and by $27,600 for 2.9%. Conversely, compensation could decrease by 1.15% (95% CI, 0.29 to 2.00) (Fig. 1F) among White veterans, amounting to $0.52 billion (95% CI, 0.13 billion to 0.92 billion) annually (Table 3). Among the 150,000 White veterans (1.0%) who would be affected, annual compensation would decrease by $1,991 for 28.8% and $4,110 for the remaining 71.2%. In total, the use of GLI-Global equations may redistribute $1.94 billion (95% CI, 1.10 billion to 2.79 billion) in annual VA disability compensation among 413,000 veteran recipients (Table 3). The redistributed amount is less than 2% of the total VA disability compensation spending reported in 2022[31]

but probably represents a sizable proportion of the spending on respiratory conditions, which accounts for less than 5% of all service-connected disabilities among veterans.[31]

LUNG-TRANSPLANT PRIORITY

Until recently, spirometry was one of several measures used to determine lung-transplant priority. Of 1399 candidates on the 2020 U.S. lung-transplant waiting list, 1243 (88.8%) would undergo shifts in their position on the waiting list if priority were determined with GLI-Global equations instead of GLI-2012 equations (Fig. 2A). For the 632 candidates (45.2%) with restrictive lung disease, such shifts would result from changes to their percent of the predicted FVC and lung-allocation score (Fig. S3). Another 611 candidates (43.7%) would undergo shifts despite unchanged lung-allocation scores owing to rearrangement of the waiting list. If GLI-Global equations were used, Asian and Black candidates for transplantation would move forward (indicating higher priority) 21.2 positions on average, amounting to 4.3 fewer days of expected wait time. Hispanic and White transplant candidates would move back (indicating lower priority) 4.3 positions on average, amounting to an additional 1.1 days of expected wait time. The most advantaged patient would move forward 150 positions (10.7% of the waiting list), and the least advantaged patient would move backward 80 positions (5.7% of the waiting list), corresponding to expected changes in wait time of 4.6 fewer weeks and 2.5 additional weeks, respectively (Fig. 2B).

ASSOCIATIONS WITH RESPIRATORY OUTCOMES

Spirometric indexes adjusted with the use of GLI-2012 and GLI-Global equations had similar discriminative accuracy for prediction of dyspnea on exertion, wheezing that limits activity, lung or breathing problems that limit activity, medical visits for wheezing, overnight hospital admissions, new-onset asthma, death from chronic lower respiratory disease, death within 365 days while on a lung-transplant waiting list, and death from any cause (Table 4). Absolute differences were near or less than 1 percentage point, indicating few instances in which one equation outperformed the other. Secondary analyses of sensitivity and specificity, with a z-score threshold defined by the 5th-percentile lower limit of the normal range, showed that the use of GLI-Global equations increased sensitivity and decreased specificity for most outcomes among Black participants, with opposing effects among Hispanic and White participants (Tables S5 and S6). ROC curves showing sensitivity and specificity values across spirometric thresholds are provided in Figures S4, S5, and S6. Additional analyses comparing predicted normal spirometric results to measured spirometric results among healthy persons are provided in Tables S7, S8, and S9 and Figures S7 through S10.

DISCUSSION

In an official statement in 2023, the ATS recommended race-neutral interpretation of lung function[15] and called for investigation of “consequences for the yet-unquantified number of individuals with results near decision-making thresholds.” By comparing the results obtained with the use of race-stratified GLI-2012 equations with those obtained with race-neutral GLI-Global equations, our analyses showed that the choice of including or removing adjustment for race does not meaningfully change the discriminative accuracy of relevant

clinical outcomes but reclassifies lung diseases, occupational eligibility, and disability compensation for millions. These findings underscore the extent of medical decision making that is at stake with the use of race-based equations and warrant thoughtful consideration of the trade-offs involved.

The effect of including or removing adjustment for race or ethnic group for each person is expected to vary according to the race category to which the person was assigned in the GLI taxonomy. When race-neutral equations were used instead of race-based equations, Black participants in our study were classified as having greater ventilatory and medical impairment, more-severe COPD grades, more frequent occupational disqualifications, and higher amounts of disability payments, and Hispanic and White participants were classified as having opposing changes. These differences occurred because most outcomes were determined with the use of reference-adjusted lung-function values, which decreased among Black participants and increased among Hispanic and White participants. The only exception was an increased prevalence of obstructive impairment among Hispanic and White participants, which was the result of obstruction being determined on the basis of the FEV_1:FVC lower limit of the normal range, which increased for most race groups (Table S10).

The population-level shifts arose from reversing the race-based calibration in GLI-2012 equations that normalized lower lung function among Asian and Black persons and higher lung function among Hispanic and White persons. This calibration assumes that healthy persons of different race groups have different lung functions. However, an imperfect selection filter for so-called healthy persons may reproduce demographic patterns of respiratory impairment in the development cohort. Adjustment for race would then appear to decrease model bias among this presumed healthy population while obscuring disparities in subclinical respiratory disease. Further study is needed to clarify whether new impairment findings constitute false positives or true undetected disease. Decreased impairment findings among healthy Hispanic and White persons prompt similar consideration.

When decision thresholds reflect compromises between risks and benefits, reclassifications will have dual effects. One recent study illustrates this trade-off: surgeons were less likely to recommend lung-cancer resection for Black patients when interpreting spirometric results using race-neutral equations.[17] This effect may limit potentially curative surgeries, but also may prevent surgical complications among patients who are at higher risk than previously recognized. Ultimately, the potential for benefit and harm depends on how accurately the equations in question can be used to classify disease states and forecast clinical outcomes. Our study showed that race-adjusted and race-neutral equations were similarly accurate in predicting the presence or occurrence of respiratory symptoms, health care utilization, new-onset disease, death from any cause, death from respiratory causes, and death while on a transplant waiting list. These findings expand on previous work that studied associations with patient-reported symptoms,[32–35] exercise tolerance,[32] emphysema on computed tomography,[32,35,36] hospitalization associated with chronic lower respiratory disease,[37] lung-transplant priority,[38,39] and death.[33,34,37,40,41]

Minor differences in discriminative accuracy may appear incongruent with our findings of substantial downstream implications. This discrepancy may be explained by two factors. First, the respiratory outcomes that were used to analyze predictive accuracy are distinct from the clinical, occupational, and financial outcomes that were used to analyze downstream implications. The former are measured independently of lung-function values, and the latter are directly defined with the use of lung-function thresholds. Second, the C statistic is a crude measure: it averages performance across the full range of lung-function thresholds, whereas clinical applications typically consider one or a few selected thresholds. Our secondary analysis using the threshold of 5th-percentile lower limit of the normal range to predict respiratory outcomes showed that removing adjustments for race by using GLI-Global equations increased both true and false positives among Black participants while decreasing true and false positives among Hispanic and White participants. Thus, inclusion or removal of race or ethnic group as an adjustment factor may produce reclassifications that exchange sensitivity and specificity while preserving discriminative accuracy overall.

The implications of adjustment for race extend beyond the outcomes evaluated in our study. For example, changes to COPD severity grades may determine eligibility for clinical trials[42] and influence treatment decisions for interventions that are approved on the basis of those trials.[43,44] Severity grades also affect insurance premiums, with cost multipliers for some conditions ranging from 50% to 175% greater than standard rates.[45] Changes in AMA impairment ratings affect payments from programs such as the Energy Employees Occupational Illness Compensation Program Act,[46] which provides $2,500 for each percentage point of impairment up to $250,000. In addition to firefighting, occupations in which workers are exposed to silica[47] and cotton dust[48] also determine occupational eligibility with preemployment lung-function testing. Further applications include fitness for lung-cancer resection,[17] indications for lung-transplantation referral,[49] and ventilatory support for patients with amyotrophic lateral sclerosis.[50]

Our study has limitations. First, spirometric classifications reflect physiological values and do not independently determine clinical diagnoses. Physical examination, imaging, diffusing-capacity testing, and functional testing frequently complement spirometry in assessments of respiratory impairment. Second, total reclassifications may be underestimated owing to the exclusion of persons with low-quality spirometric results or temporary contraindications for spirometry; reclassifications may also be overestimated owing to the inclusion of persons who would not be materially affected by reclassifications (e.g., disqualification from firefighting attributed to persons not considering the occupation). However, relative changes are less likely to be affected. Third, donor-lung allocation may be restricted by additional factors such as pediatric priority, blood type, and geographic distance, which were not modeled in our study. Our analysis of lung-transplantation outcomes is also specific to the 2020 waiting list — the lung-allocation score calculator that was updated in 2021 and the composite allocation score that was implemented in 2023 do not use spirometry to determine transplant priority. However, lung-function equations continue to affect candidacy for lung-transplant listings,[49] and previous outcomes may inform reparative policies.[51] Fourth, the five cohorts that we included in this study do not represent all populations globally, and our impact analyses are limited to the United States.

Beyond consideration of race, the practice of interpreting measured values relative to normal values deserves reconsideration. One alternative involves using fixed thresholds such as FEV_1:FVC of less than 0.70, which may be more accurate than the lower limit of the normal range in predicting COPD-related events.[52] The validity of fixed thresholds is debated,[2] but their use would decrease reliance on imprecise definitions of normal and align the interpretation of lung function with that of hypertension, obesity, diabetes, and other areas of medicine. Other approaches include personalized baselines that are derived from longitudinal assessments and consideration of more precise anthropometric, genetic,[53] socioeconomic,[54] and environmental[55] factors. Addressing these considerations will be essential for informing principled assessments of lung function in diverse populations and guiding interventions aimed at improving respiratory health.

Since the mid-19th century, spirometry has been used to support racial hierarchies that were based on assumptions of innate superiorities and deficiencies in lung function.[5,6] These distinctions obscure the continuum of human genetic and phenotypic variation and present additional challenges when a patient's race does not fit existing categories or is inappropriately assigned by clinicians. Race-neutral equations, although imperfect,[56] offer an opportunity to move beyond historical assumptions that group-level differences in lung function are natural and benign. However, race-neutral equations are not enough to rectify long-standing racial inequities, and their many trade-offs must be carefully considered. Responses to race-based equations for kidney function,[57] obstetrical risk,[58] and cognitive testing[59] may provide lessons; these include deliberation processes involving multiple stakeholders, unified recommendations, and interventions to redress quantified harms.[51,57,60]

Our study showed that the use of race in lung-function testing has broad clinical, occupational, and financial implications for millions of patients. We hope that data on the nature and extent of these implications may inform improvements to current reference equations and preparations for expected changes to care.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by awards (K01HL138259 and K23HL169911) from the National Heart, Lung, and Blood Institute and a grant (R01ES032470) from the National Institute of Environmental Health Sciences.

We thank Lundy Braun, Ph.D., Robert Brown, M.D., Diane Coyle, Ph.D., David C. Christiani, M.D., Michael A. LeNoir, M.D., and the staff of the African American Wellness Project for helpful discussions regarding the manuscript.

APPENDIX

The authors' full names and academic degrees are as follows: James A. Diao, M.Phil., Yixuan He, Ph.D., Rohan Khazanchi, M.D., M.P.H., Max Jordan Nguemeni Tiako, M.D., Jonathan I. Witonsky, M.D., Emma Pierson, Ph.D., Pranav Rajpurkar, Ph.D., Jennifer R. Elhawary, M.S., Luke Melas-Kyriazi, A.B., Albert Yen, B.S., Alicia R. Martin, Ph.D., Sean Levy, M.D., Chirag J. Patel, Ph.D., Maha Farhat, M.D., Luisa N. Borrell, D.D.S., Ph.D.,

Michael H. Cho, M.D., M.P.H., Edwin K. Silverman, M.D., Ph.D., Esteban G. Burchard, M.D., M.P.H., and Arjun K. Manrai, Ph.D.

The authors' affiliations are as follows: the Department of Biomedical Informatics, Harvard Medical School (J.A.D., P.R., L.M.-K., C.J.P., M.F., A.K.M.), the Computational Health Informatics Program, Boston Children's Hospital (J.A.D., A.K.M.), the Analytic and Translational Genetics Unit (Y.H., A.R.M.) and the Division of Pulmonary and Critical Care Medicine, Department of Medicine (M.F.), Massachusetts General Hospital, Harvard Internal Medicine–Pediatrics Combined Residency Program, Brigham and Women's Hospital, Boston Children's Hospital, and Boston Medical Center (R.K.), the François-Xavier Bagnoud Center for Health and Human Rights, Harvard University (R.K.), the Department of Medicine (M.J.N.T.) and the Channing Division of Network Medicine and the Division of Pulmonary and Critical Care Medicine, Department of Medicine (M.H.C., E.K.S.), Brigham and Women's Hospital, and the Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, Beth Israel Deaconess Medical Center (S.L.), Boston, and the Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge (Y.H., A.R.M.) — all in Massachusetts; the Departments of Pediatrics (J.I.W.), Medicine (J.R.E., E.G.B.), and Bioengineering and Therapeutic Sciences (J.R.E., E.G.B.), University of California, San Francisco, San Francisco; the Department of Computer Science, Cornell University, Ithaca (E.P.), and the Department of Population Health Sciences, Weill Cornell Medical College (E.P.), and the Department of Epidemiology and Biostatistics, Graduate School of Public Health and Health Policy, City University of New York (L.N.B.), New York — all in New York; the Department of Engineering Science, University of Oxford, Oxford, United Kingdom (L.M.-K.); and the Medical Scientist Training Program, University of Illinois at Chicago, Chicago (A.Y.).

REFERENCES

1. Kaminsky DA. Selecting reference values for pulmonary function tests. Philadelphia: UpToDate, 2023 (https://www.uptodate.com/contents/selecting-reference-values-for-pulmonary-function-tests).
2. Stanojevic S, Kaminsky DA, Miller MR, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. Eur Respir J 2022; 60: 2101499.
3. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the prevention, diagnosis and management of COPD: 2023 report. 2023 (https://goldcopd.org/2023-gold-report-2/).
4. Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight — reconsidering the use of race correction in clinical algorithms. N Engl J Med 2020; 383: 874–82. [PubMed: 32853499]
5. Braun L Breathing race into the machine: the surprising career of the spirometer from plantation to genetics. Minneapolis: University of Minnesota Press, 2014.
6. Race Braun L., ethnicity and lung function: a brief history. Can J Respir Ther 2015; 51: 99–101. [PubMed: 26566381]
7. Elliott EN. Cotton is king, and pro-slavery arguments: comprising the writings of Hammond, Harper, Christy, Stringfellow, Hodge, Bledsoe, and Cartwright, on this important subject. Augusta, GA: Pritchard, Abbott & Loomis, 1860.
8. Moffett AT, Eneanya ND, Halpern SD, Weissman GE. The impact of race correction on the interpretation of pulmonary function testing among black patients. Am J Respir Crit Care Med 2021; 203: A1030. abstract (10.1164/ajrccm-conference.2021.203.1_MeetingAbstracts.A1030).
9. McClure ES, Vasudevan P, Bailey Z, Patel S, Robinson WR. Racial capitalism within public health-how occupational settings drive COVID-19 disparities. Am J Epidemiol 2020; 189: 1244–53. [PubMed: 32619007]

10. Anderson MA, Malhotra A, Non AL. Could routine race-adjustment of spirometers exacerbate racial disparities in COVID-19 recovery? Lancet Respir Med 2021;9:124–5. [PubMed: 33308418]
11. Bhakta NR, Kaminsky DA, Bime C, et al. Addressing race in pulmonary function testing by aligning intent and evidence with practice and perception. Chest 2022 161 288–97. [PubMed: 34437887]
12. American Medical Association. Redressing the harms of misusing race in medicine: H-65.943. 2023 (https://policy-search.ama-assn.org/policyfinder/detail/Redressing%20the%20Harms%20of%20Misusing%20Race%20in%20Medicine%20H-65.943?uri=%2FAMADoc%2FHOD.xml-H-65.943.xml).
13. Quanjer PH, Stanojevic S, Cole TJ, 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]
14. Bowerman C, Bhakta NR, Brazzale D, et al. A race-neutral approach to the interpretation of lung function measurements. Am J Respir Crit Care Med 2023;207:768–74. [PubMed: 36383197]
15. Bhakta NR, Bime C, Kaminsky DA, et al. Race and ethnicity in pulmonary function test interpretation: an official American Thoracic Society statement. Am J Respir Crit Care Med 2023;207:978–95. [PubMed: 36973004]
16. Moffett AT, Bowerman C, Stanojevic S, Eneanya ND, Halpern SD, Weissman GE. Global, race-neutral reference equations and pulmonary function test interpretation. JAMA Netw Open 2023;6(6):e2316174. [PubMed: 37261830]
17. Bonner SN, Lagisetty K, Reddy RM, Engeda Y, Griggs JJ, Valley TS. Clinical implications of removing race-corrected pulmonary function tests for African American patients requiring surgery for lung cancer. JAMA Surg 2023;158:1061–8. [PubMed: 37585181]
18. Curtin LR, Mohadjer LK, Dohrmann SM, et al. National Health and Nutrition Examination Survey: sample design, 2007–2010. Vital Health Stat 2 2013;(160):1–23.
19. National Center for Health Statistics. Plan and operation of the third National Health and Nutrition Examination Survey, 1988–94. Hyattsville, MD: Public Health Service, 1994.
20. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203–9. [PubMed: 30305743]
21. 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 multiethnic study of atherosclerosis (MESA) lung study. Chest 2010;137:138–45. [PubMed: 19741060]
22. OPTN/SRTR 2020 annual data report: preface. Am J Transplant 2022;22:Suppl 2 1–10.
23. National Fire Protection Association. NFPA 1582: standard on comprehensive occupational medical program for fire departments. 2023 (https://link.nfpa.org/all-publications/1582/2007).
24. American Medical Association. Guides to the evaluation of permanent impairment. 2008 (https://ama-guides.ama-assn.org/books/book/3/AMA-Guides-to-the-Evaluation-of-Permanent).
25. Legal Information Institute. 38 CFR § 4.97, schedule of ratings — respiratory system (https://www.law.cornell.edu/cfr/text/38/4.97).
26. Department of Veterans Affairs. 2024 Veterans disability compensation rates. 2022 (https://www.va.gov/disability/compensation-rates/veteran-rates/).
27. Wan ES, Castaldi PJ, Cho MH, et al. Epidemiology, genetics, and subtyping of preserved ratio impaired spirometry (PRISm) in COPDGene. Respir Res 2014;15:89. [PubMed: 25096860]
28. United Network for Organ Sharing. A guide to calculating the lung allocation score. 2020 (https://unos.org/wp-content/uploads/unos/lung-allocation-score.pdf).
29. Stanojevic S, Wade A, Stocks J. Reference values for lung function: past, present and future. Eur Respir J 2010;36:12–9. [PubMed: 20595163]
30. Fahy R, Evarts B, Stein GP. US fire department profile 2020. Quincy, MA: National Fire Protection Association, September 2022 (https://npr.brightspotcdn.com/03/bf/7bb8a5f04651b98de071b3464196/osfdprofile.pdf).
31. Veterans Benefits Administration. The fiscal year 2022 annual benefits report. 2022 (https://www.benefits.va.gov/REPORTS/abr/docs/2022-abr.pdf).

32. Erratum: Reconsidering the utility of race-specific lung function prediction equations. Am J Respir Crit Care Med 2022;206:230. [PubMed: 35838593]
33. Ekström M, Mannino D. Research race-specific reference values and lung function impairment, breathlessness and prognosis: analysis of NHANES 2007–2012. Respir Res 2022;23:271. [PubMed: 36182912]
34. Ekström M, Backman H, Mannino D. Clinical implications of the global lung function initiative race-neutral spirometry reference equations in terms of breathlessness and mortality. Am J Respir Crit Care Med 2024;209:104–6. [PubMed: 37187171]
35. Non AL, Bailey B, Bhatt SP, et al. Race-specific spirometry equations do not improve models of dyspnea and quantitative chest CT phenotypes. Chest 2023;164:1492–504. [PubMed: 37507005]
36. Liu GY, Khan SS, Colangelo LA, et al. Comparing racial differences in emphysema prevalence among adults with normal spirometry: a secondary data analysis of the CARDIA lung study. Ann Intern Med 2022;175:1118–25. [PubMed: 35849828]
37. Elmaleh-Sachs A, Balte P, Oelsner EC, et al. Race/ethnicity, spirometry reference equations, and prediction of incident clinical events: the Multi-Ethnic Study of Atherosclerosis (MESA) lung study. Am J Respir Crit Care Med 2022;205:700–10. [PubMed: 34913853]
38. Brems JH, Balasubramanian A, Psoter KJ, et al. Race-specific interpretation of spirometry: impact on the lung allocation score. Ann Am Thorac Soc 2023;20:1408–15. [PubMed: 37315331]
39. Colon Hidalgo D, Ramos KJ, Harlan EA, et al. Historic use of race-based spirometry values lowered transplant priority for black patients. Chest 2024;165:381–8. [PubMed: 37832783]
40. Burney PGJ, Hooper RL. The use of ethnically specific norms for ventilatory function in African-American and white populations. Int J Epidemiol 2012 41:782–90. [PubMed: 22434864]
41. McCormack MC, Balasubramanian A, Matsui EC, Peng RD, Wise RA, Keet CA. Race, lung function, and long-term mortality in the National Health and Nutrition Examination Survey III. Am J Respir Crit Care Med 2022;205:723–4. [PubMed: 34597248]
42. Food and Drug Administration. Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials. Draft guidance for industry. April 2022 (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations).
43. Decramer ML, Chapman KR, Dahl R, et al. Once-daily indacaterol versus tiotropium for patients with severe chronic obstructive pulmonary disease (INVIGORATE): a randomised, blinded, parallel-group study. Lancet Respir Med 2013;1:524–33. [PubMed: 24461613]
44. Donohue JF, Fogarty C, Lötvall J, et al. Once-daily bronchodilators for chronic obstructive pulmonary disease: indacaterol versus tiotropium. Am J Respir Crit Care Med 2010;182:155–62. [PubMed: 20463178]
45. Cornelius N Life insurance and COPD: buyer's guide. Quotacy. September 30, 2020 (https://www.quotacy.com/life-insurance-chronic-obstructive-pulmonary-disease-buyers-guide/).
46. Energy Employees Occupational Illness Compensation Program Act. Department of Labor, Employment Standards Administration, Office of Workers' Compensation Program. 2003.
47. Department of Labor. App B — medical surveillance guidelines. Standard no. 1926.1153 (https://www.osha.ov/laws-regs/regulations/standardnumber/1926/1926.1153AppB).
48. Townsend MC, Cowl CT. U.S. occupational historical perspective on race and lung function. Am J Respir Crit Care Med 2022;206:789–90. [PubMed: 35503517]
49. Shweish O, Dronavalli G. Indications for lung transplant referral and listing. J Thorac Dis 2019;11:Suppl 14:S1708–S1720. [PubMed: 31632748]
50. Miller RG, Jackson CE, Kasarskis EJ, et al. Practice parameter update: the care of the patient with amyotrophic lateral sclerosis: drug, nutritional, and respiratory therapies (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2009;73:1218–26. [PubMed: 19822872]
51. Mohottige D, Purnell TS, Boulware LE. Redressing the harms of race-based kidney function estimation. JAMA 2023;329:881–2. [PubMed: 36848168]
52. Bhatt SP, Balte PP, Schwartz JE, et al. Discriminative accuracy of FEV1:FVC thresholds for COPD-related hospitalization and mortality. JAMA 2019;321:2438–47. [PubMed: 31237643]

53. Kumar R, Seibold MA, Aldrich MC, et al. Genetic ancestry in lung-function predictions. N Engl J Med 2010;363:321–30. [PubMed: 20647190]

54. Hegewald MJ, Crapo RO. Socioeconomic status and lung function. Chest 2007;132:1608–14. [PubMed: 17998360]

55. Gauderman WJ, Avol E, Gilliland F, et al. The effect of air pollution on lung development from 10 to 18 years of age. N Engl J Med 2004;351:1057–67. [PubMed: 15356303]

56. Wang RJ. Beyond race-specific spirometry reference equations: what comes next? Am J Respir Crit Care Med 2024;209:117–8. [PubMed: 37595271]

57. Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease. J Am Soc Nephrol 2021;32:2994–3015. [PubMed: 34556489]

58. Palmer K Changing the equation: researchers remove race from a calculator for childbirth. STAT. June 3, 2021 (https://www.statnews.com/2021/06/03/vbac-calculator-birth-cesarean/).

59. Possin KL, Tsoy E, Windon CC. Perils of race-based norms in cognitive testing: the case of former NFL players. JAMA Neurol 2021;78:377–8. [PubMed: 33346785]

60. Khazanchi R, Morse M. NYC Coalition to End Racism in Clinical Algorithms: inaugural report. New York: New York City Department of Health and Mental Hygiene, September 2022 (https://www.nyc.gov/assets/doh/downloads/pdf/cmo/cerca-report.pdf).

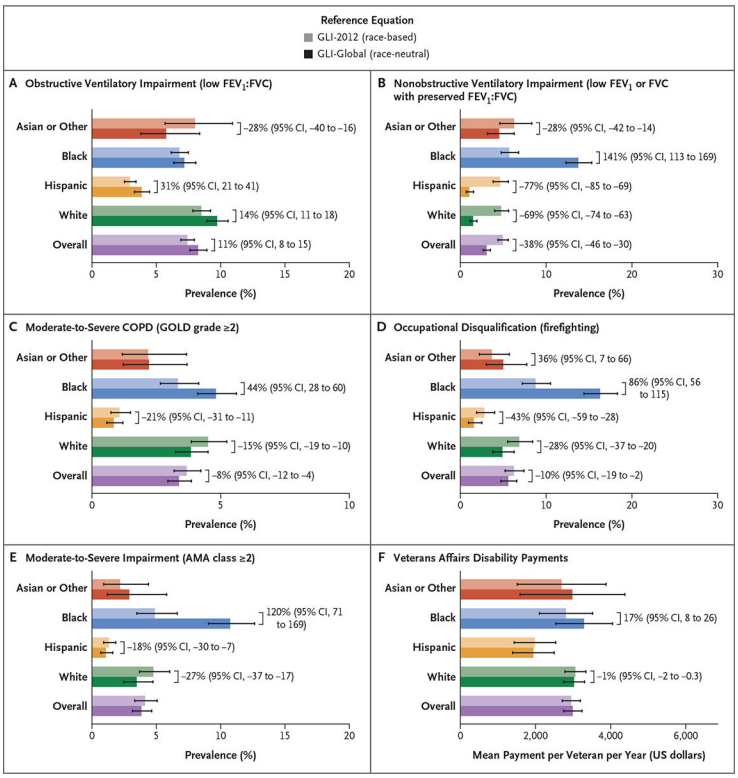
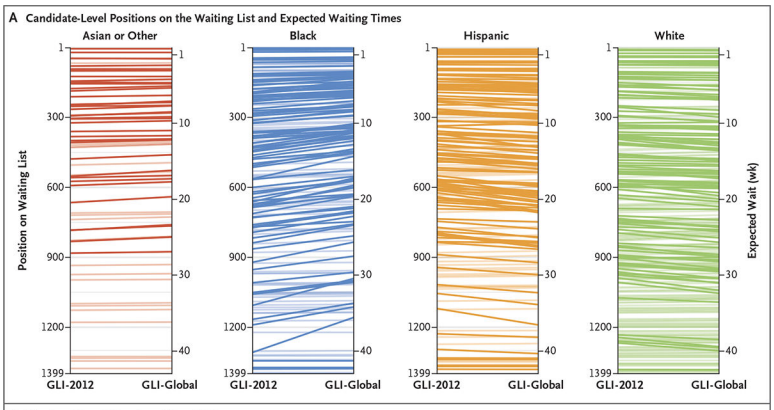


Figure 1. Clinical, Occupational, and Financial Outcomes in the United States Calculated with Race-Based versus Race-Neutral Lung-Function Equations.

Shown are outcomes for study participants with regard to obstructive ventilatory impairment (Panel A), nonobstructive ventilatory impairment (Panel B), chronic obstructive pulmonary disease (COPD) of grade 2 or higher on the Global Initiative for Chronic Obstructive Lung Disease (GOLD) scale (Panel C), disqualification from firefighting occupations (Panel D), American Medical Association (AMA) impairment classifications (Panel E), and Department of Veterans Affairs disability payments (Panel E) when lung function was interpreted with the use of race-based Global Lung Function Initiative 2012 (GLI-2012) equations and with race-neutral GLI-Global equations. Data from the National Health and Nutrition Examination Survey (NHANES) 2007–2012 were survey-adjusted to be representative of the U.S. population (see the Supplementary Methods section). Error bars represent 95% confidence intervals for outcome values. Numeric labels represent relative changes between the outcomes projected on the basis of GLI-2012 equations (lighter) and GLI-Global equations (darker); confidence intervals overlapping 0 were not labeled. Because outcome values calculated with the use of GLI-2012 and GLI-Global equations are highly correlated, uncertainties in adjacent bars cannot be used to approximate the uncertainty in their difference.



B Most and Least Advantaged Candidates

Advantage	Race or Ethnic Group	Gender	Age	FVC	Percent of Predicted FVC		Lung-Allocation Score		Position on Waiting List		Expected Wait	
					Original	Change	Original	Change	Original	Change	Original	Change
			yr	*liter*							*wk*	
Highest												
First	Black	Female	72	1.96	75.0	−10.0	32.3	0.66	1308	−150	40.2	−4.6
Second	Black	Female	53	2.38	74.4	−7.5	33.1	0.48	1106	−115	34.0	−3.5
Third	Black	Female	70	1.28	56.8	−6.0	36.4	0.91	662	−102	20.3	−3.1
Lowest												
First	White	Male	61	2.78	67.6	5.3	33.5	−0.35	1029	80	31.6	2.5
Second	White	Female	62	1.94	69.6	5.3	33.4	−0.31	1030	75	31.6	2.3
Third	Hispanic	Male	59	2.72	62.6	4.8	33.1	−0.22	1120	70	34.4	2.1

Figure 2. Implications of Race-Based Lung-Function Equations for the 2020 U.S. Lung-Transplant Waiting List.

Panel A shows the position on the U.S. lung-transplant waiting list and expected waiting time for 1399 candidates on January 1, 2020. The percent of the predicted forced vital capacity (FVC) was calculated with the use of race-based GLI-2012 equations or race-neutral GLI-Global equations. Positions on the waiting list are ordered according to decreasing lung-allocation score. A lower position on the waiting list and higher lung-allocation score indicate higher priority, with ties broken by accrued wait time. This retrospective analysis is specific to the 2020 waiting list; newer allocation scores do not use spirometry to determine transplant priority. Dark-colored lines indicate candidates who had changes in both waiting-list position and lung-allocation score. Light-colored lines indicate candidates who had changes in waiting-list position but not in lung-allocation score. Gray lines indicate candidates who had no changes in either waiting-list position or lung-allocation score. White candidates were downsampled by 70% to aid visualization. Expected wait time is a linear function of the initial position on the waiting list, allowing dual-axis plotting (Fig. S1). Data are from the Organ Procurement and Transplantation Network (OPTN). Panel B shows demographic, clinical, and waiting-list characteristics of the candidates who were most and least advantaged by the use of GLI-Global equations rather than GLI-2012 equations, with advantage measured as change in expected wait time. OPTN data in the Gender column represent patient-reported gender identification.

Table 1.

Demographic Characteristics, Spirometric Measurements, and Respiratory Conditions.*

Characteristic	NHANES III (N = 31,311)	NHANES IV (N = 30,442)	MESA (N = 6814)	U.K. Biobank (N = 501,723)	OPTN (N = 42,751)
Data-collection period	1988–1994	2007–2012	2005–2007	2006–2010	2005–2023
Participants included — no. (%)†	15,861 (50.7)	17,067 (56.1)	3262 (47.8)	290,136 (57.8)	42,751 (100)
Female sex or gender — %‡	51.2	50.0	53.1	57.2	43.5
Race or ethnic group — %§					
Asian	—	4.4	16.5	2.0	2.6
Black	28.7	21.9	25.0	1.3	9.8
Hispanic	28.5	29.6	11.3	—	8.8
White	38.9	39.6	35.2	95.1	78.2
Multiracial, other race, or unknown	3.9	4.5	—	1.1	0.7
Median age (IQR) — yr	35 (20–54)	33 (16–52)	65 (57–73)	57 (50–63)	59 (49–65)
Median height (IQR) — cm	165 (157–173)	165 (156–173)	165 (158–173)	167 (161–175)	170 (161–177)
Median spirometric values (IQR)					
FEV_1 — liters	2.91 (2.29–3.56)	2.92 (2.23–3.61)	2.30 (1.87–2.83)	2.71 (2.26–3.26)	—
FVC — liters	3.60 (2.90–4.42)	3.62 (2.81–4.48)	3.06 (2.48–3.81)	3.55 (2.99–4.29)	1.82 (1.37–2.38)
FEV_1:FVC	0.81 (0.76–0.86)	0.82 (0.77–0.86)	0.76 (0.71–0.80)	0.77 (0.73–0.80)	—
Respiratory factors — %¶					
Smoking history	41.4	33.9	49.8	21.1	44.1
Respiratory symptoms	24.0	15.5	14.3	22.0	100
Respiratory disease	20.1	16.1	32.1	23.3	100

*FEV_1 denotes forced expiratory volume in 1 second, FVC forced vital capacity, IQR interquartile range, MESA Multi-Ethnic Study of Atherosclerosis, NHANES National Health and Nutrition Examination Survey, and OPTN Organ Procurement and Transplantation Network.

†Each sample includes participants 3 to 95 years of age with acceptable spirometric data and recorded age and height. Additional details regarding acceptability criteria are provided in the Supplementary Methods section.

‡Participants included in the NHANES IV, NHANES III, and OPTN cohorts reported gender, and those in the MESA and U.K. Biobank cohorts reported sex.

§Race and ethnic group were reported by the participants for all cohorts except OPTN, in which data were reported by transplantation centers. Additional details on the use of race and ethnic-group data are provided in the Supplementary Methods section.

¶Respiratory symptoms include chronic wheezing, coughing, or phlegm on most days. Respiratory disease includes asthma, chronic bronchitis, chronic obstructive pulmonary disease, and lung cancer. All lung-transplant candidates on the OPTN waiting list were presumed to have symptomatic respiratory disease.

Table 2.

Criteria for Assessing Clinical, Occupational, and Financial Outcomes.*

Outcome	Spirometric Criteria†	Age Criteria	Other Criteria	Source
Ventilatory impairment		6–79 yr	None	ERS-ATS (2021)[2]
Obstructive	FEV_1:FVC <LLN			
Nonobstructive	FEV_1 or FVC <LLN and FEV_1:FVC >LLN			
COPD severity		6–79 yr	None	GOLD (2023)[3]
Grade 1 (least severe)	FEV_1:FVC <0.70 and FEV_1 ≥80% of predicted			
Grade 2	FEV_1:FVC <0.70 and FEV_1 50–79% of predicted			
Grade 3	FEV_1:FVC <0.70 and FEV_1 30–49% of predicted			
Grade 4 (most severe)	FEV_1:FVC <0.70 and FEV_1 <30% of predicted			
Occupational disqualification from firefighting	FEV_1 or FVC <70% of predicted, *or* FEV_1 or FVC <80% of predicted and FEV_1:FVC <0.75, *or*	18–65 yr	Work exposure to dust or fumes	NFPA (2007)[23]
	FEV_1 or FVC <90% of predicted and previous diagnosis of asthma			
Medical impairment ratings		18–79 yr	Work exposure to dust or fumes	AMA (2008)[24]
Class 1 (least severe)	FEV_1 65–79% of predicted or FVC 70–79% of predicted			
Class 2	FEV_1 55–64% of predicted or FVC 60–69% of predicted			
Class 3	FEV_1 45–54% of predicted or FVC 50–59% of predicted			
Class 4 (most severe)	FEV_1 <45% of predicted or FVC <50% of predicted			
VA disability ratings		18–79 yr	Served in the U.S. Armed Forces	VA (2023)[25,26]
10% (least severe)	FEV_1 71–80% of predicted or FEV1:FVC 0.71–0.80			
30%	FEV_1 56–70% of predicted or FEV1:FVC 0.56–0.70			
60%	FEV_1 40–55% of predicted or FEV1:FVC 0.40–0.55			
100% (most severe)	FEV_1 <40% of predicted or FEV1:FVC <0.40			

*AMA denotes American Medical Association, ATS American Thoracic Society, COPD chronic obstructive pulmonary disease, ERS European Respiratory Society, GLI Global Lung Function Initiative, GOLD Global Initiative for Chronic Obstructive Lung Disease, LLN 5th-percentile lower limit of normal range, NFPA National Fire Protection Association, and VA Department of Veterans Affairs.

†The percent of the predicted value is the ratio of the measured value to the predicted healthy value, with normal values typically considered to fall between 80% and 120% of the predicted healthy value. Percent of the predicted and LLN values were determined with the use of race-based GLI-2012 equations or race-neutral GLI-Global equations. When multiple spirometric values were assessed, the value resulting in the more severe rating was used. Additional details regarding assessed and nonassessed criteria are provided in the Supplementary Methods section.

N Engl J Med. Author manuscript; available in PMC 2024 August 07.

Table 3.

Extrapolated Changes in Outcomes Calculated from Race-Based and Race-Neutral Lung-Function Equations, According to Persons Affected and Disability Payments in the United States.*

Outcome	Net Change					Total Change		
	Overall	Asian or Other	Black	Hispanic	White	Overall	Newly Classified	No Longer Classified
Persons affected — no., in thousands (95% CI)								
Obstructive impairment	2070 (1330 to 2820)	−392 (−608 to −177)	110 (−17 to 237)	349 (231 to 466)	2010 (1360 to 2650)	3200 (2630 to 3860)	2640 (2080 to 3290)	565 (338 to 884)
Nonobstructive impairment	−4700 (−6140 to −3270)	−304 (−498 to −110)	2340 (1930 to 2750)	−1370 (−1800 to −942)	−5370 (−6550 to −4190)	9620 (8570 to 10,800)	2460 (1960 to 3040)	7160 (6130 to 8320)
Moderate-to-severe COPD: GOLD grade ≥2	−749 (−1180 to −316)	6 (−15 to 27)	428 (300 to 556)	−85 (−139 to −32)	−1100 (−1480 to −715)	1660 (1300 to 2070)	453 (324 to 617)	1200 (857 to 1640)
Occupational disqualification from firefighting	−624 (−1200 to −48)	72 (10 to 135)	754 (540 to 969)	−181 (−262 to −100)	−1270 (−1730 to −809)	2280 (1840 to 2780)	826 (601 to 1110)	1450 (1030 to 1980)
Moderate-to-severe impairment: AMA class ≥2	−297 (−725 to 131)	41 (−9 to 92)	638 (478 to 797)	−38 (−59 to −16)	−938 (−1310 to −570)	1580 (1250 to 1960)	646 (481 to 849)	929 (632 to 1320)
Annual VA disability payments — U.S. $, in millions (95% CI)	806 (−42 to 1,650)	279 (−268 to 825)	1,100 (585 to 1,610)	46 (−99 to 7)	−524 (−917 to −131)	1,940 (1,100 to 2,790)	1,380 (627 to 2,120)	570 (173 to 966)

* Differences are shown for the total number of affected persons or disability payment dollars when criteria are calculated with the use of race-neutral GLI-Global equations as compared with race-based GLI-2012 equations. Data from NHANES IV were survey-adjusted to be representative of the U.S. population (see the Supplementary Methods section).

Table 4.

Accuracy of Reference-Adjusted Spirometry for Discriminating Respiratory Symptoms, Health Care Utilization, New-Onset Disease, and Death.*

Outcome and Cohort	Best Spirometric Predictor	Discriminative Accuracy (95% CI)†		
		GLI-2012	GLI-Global	Difference
		C statistic		
Concurrent respiratory symptoms, NHANES IV				
Dyspnea on exertion	FEV_1 z score	0.634 (0.619 to 0.649)	0.632 (0.616 to 0.647)	−0.002 (−0.009 to 0.004)
Wheezing that limits activity	FEV_1 z score	0.685 (0.655 to 0.714)	0.689 (0.661 to 0.718)	0.005 (−0.008 to 0.017)
Lung or breathing problem that limits activity	FEV_1 z score	0.737 (0.695 to 0.780)	0.746 (0.705 to 0.787)	0.009 (−0.008 to 0.025)
Recent health care utilization, NHANES IV				
Medical visit for wheezing in past yr	FEV_1 z score	0.676 (0.644 to 0.708)	0.676 (0.644 to 0.707)	−0.001 (−0.013 to 0.012)
Overnight hospital admission in past yr	FEV_1 z score	0.573 (0.548 to 0.598)	0.584 (0.559 to 0.609)	0.011 (0.001 to 0.021)
New-onset respiratory disease, U.K. Biobank				
Asthma ≤10 yr	FEV_1:FVC z score	0.587 (0.559 to 0.616)	0.588 (0.559 to 0.617)	0.001 (−0.002 to 0.003)
COPD ≤10 yr	FEV_1:FVC z score	0.786 (0.750 to 0.823)	0.792 (0.755 to 0.828)	0.005 (0.002 to 0.008)
Death, NHANES III				
30-yr incidence from chronic lower respiratory disease	FEV_1:FVC z score	0.838 (0.628 to 0.981)	0.833 (0.601 to 0.981)	−0.004 (−0.037 to 0.013)
10-yr incidence from any cause	FEV_1 z score	0.620 (0.530 to 0.705)	0.620 (0.528 to 0.706)	−0.001 (−0.022 to 0.022)
Death while on transplant waiting list, OPTN				
45-day incidence	FVC z score	0.573 (0.545 to 0.598)	0.564 (0.538 to 0.590)	−0.008 (−0.013 to −0.003)
365-day incidence	FVC z score	0.573 (0.547 to 0.599)	0.568 (0.542 to 0.594)	−0.005 (−0.011 to 0.001)

*Predictors included spirometric z scores, indicating the number of standard deviations by which a measured value is above or below the predicted normal value. Results are provided for the spirometric z score with the highest mean concordance statistic (C statistic) among FEV_1, FVC, and the ratio of FEV_1 to FVC.

†Harrell's C statistics were calculated with the use of z scores derived from race-based GLI-2012 equations or race-neutral GLI-Global equations to predict respiratory outcomes.