

VIEWPOINT: TURNING THE AIR BLUE

Is Achieving a Fully Race-Neutral Approach to Lung Function Classification Even Possible?

Amin Adibi¹, Mohsen Sadatsafavi¹, Emily P. Brigham², and Surya P. Bhatt³

¹Respiratory Evaluation Sciences Program, Faculty of Pharmaceutical Sciences, and ²Division of Respiratory Medicine, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada; and ³Division of Pulmonary, Allergy, and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, Alabama

ORCID IDs: 0000-0003-2748-4781 (A.A.); 0000-0002-0419-7862 (M.S.); 0000-0002-8449-5420 (E.P.B.); 0000-0002-8418-4497 (S.P.B.).

Airflow obstruction is assessed by estimating “normal” lung function using population-derived reference equations and measuring deviations from it using Z-scores (1). In 2023, the American Thoracic Society recommended replacing race-specific Global Lung Function Initiative (GLI)-2012 reference equations with race-weighted GLI-Global equations to address concerns about perpetuating race-based medicine and normalizing lower lung function in underserved populations (2).

Several studies refer to race-neutral Z-scores or percent predicted values (3–6). However, the definition of race neutrality is often unclear. We propose a more structured framework to define race neutrality and evaluate the statistical properties of spirometric values from different equations.

Race neutrality of spirometric values can be defined in at least three ways: 1) eliminating the collection of race information for spirometry (race-agnostic encounters), 2) ensuring that spirometric values have the same prognostic implications regardless of race (race-agnostic interpretability), and 3) ensuring that values do not vary statistically between racial groups and cannot predict race (race-unawareness). Criteria 2 and 3 align with notions of sufficiency and independence in algorithmic fairness literature, respectively (7). We demonstrate that no commonly used spirometry outputs can simultaneously satisfy all definitions of race neutrality and that the applicable definition depends on the context.

What Race Information Do Lung Function Metrics Contain?

We evaluated different definitions of race neutrality for lung function measures in the National Health and Nutrition Examination Survey 2007–2012. We included 10,410 self-identified Black and White participants with valid height and spirometry results meeting American Thoracic Society standards. To illustrate race-agnostic interpretability for a clinical outcome, we calculated marginal 10-year survival probabilities for Black and White 49-year-old females using a Cox proportional hazards model. For survival analysis, we included 7,867 participants with linked mortality follow-up through December 31, 2019. We compared race awareness of lung function metrics by examining how well they predicted self-identified race.

Figure 1 shows that none of the evaluated measures of lung function met all definitions of race neutrality. Black and White individuals with the same raw FEV₁ or the same FEV₁ Z-scores from GLI-Global equations had similar 10-year survival probabilities. FEV₁ Z-scores from GLI-Global could predict race with an area under the curve of 0.76. Z-scores from GLI-2012 had poor discriminative accuracy for predicting race, with an area under the curve of 0.53.

To summarize, FEV₁ and FEV₁ Z-scores from GLI-Global met the race-agnostic encounters and race-agnostic

interpretation criteria but not the race-unawareness criterion. In contrast, Z-scores from race-specific GLI-2012 only met the race-unawareness criterion. FVC Z-scores and FEV₁ and FVC percent predicted values exhibited patterns similar to FEV₁ Z-scores (results not shown).

Race-agnostic interpretability of raw lung function and Z-scores from GLI-Global is consistent with previous studies (8–10). Z-scores are generally unaware of any variable adjusted for in the reference equation. The decision to include a factor in reference equations should be based on the consequences of propagating those differences into decision making. For example, although, on average, females have lower lung function than males, these differences do not inherently represent differential lung health. Therefore, we ensure that Z-scores are sex unaware by adjusting for biological sex in reference equations.

Fairness Implications

Z-scores from GLI-Global achieve race-agnostic encounters and race-agnostic interpretability, whereas those from GLI-2012 meet the race-unawareness criterion. Achieving race-agnostic interpretability requires reference equations that do not adjust for race, which compromises race unawareness. This observation is related to the broader impossibility theorem, which proves that intuitive statistical fairness criteria cannot be concurrently satisfied

(Received in original form August 15, 2024; accepted in final form January 14, 2025)

Supported by the Canadian Institutes of Health Research with a team grant to the University of British Columbia (PHT 178432).

Correspondence and requests for reprints should be addressed to Amin Adibi, M.Sc., Respiratory Evaluation Sciences Program, Faculty of Pharmaceutical Sciences, The University of British Columbia, 4103B - 2405 Wesbrook Mall, Vancouver, BC, V6T 1Z3 Canada. E-mail: amin.adibi@ubc.ca.

Am J Respir Crit Care Med Vol 211, Iss 3, pp 432–435, Mar 2025

Copyright © 2025 by the American Thoracic Society

Originally Published in Press as 10.1164/rccm.202408-1599VP on January 21, 2025

Internet address: www.atsjournals.org

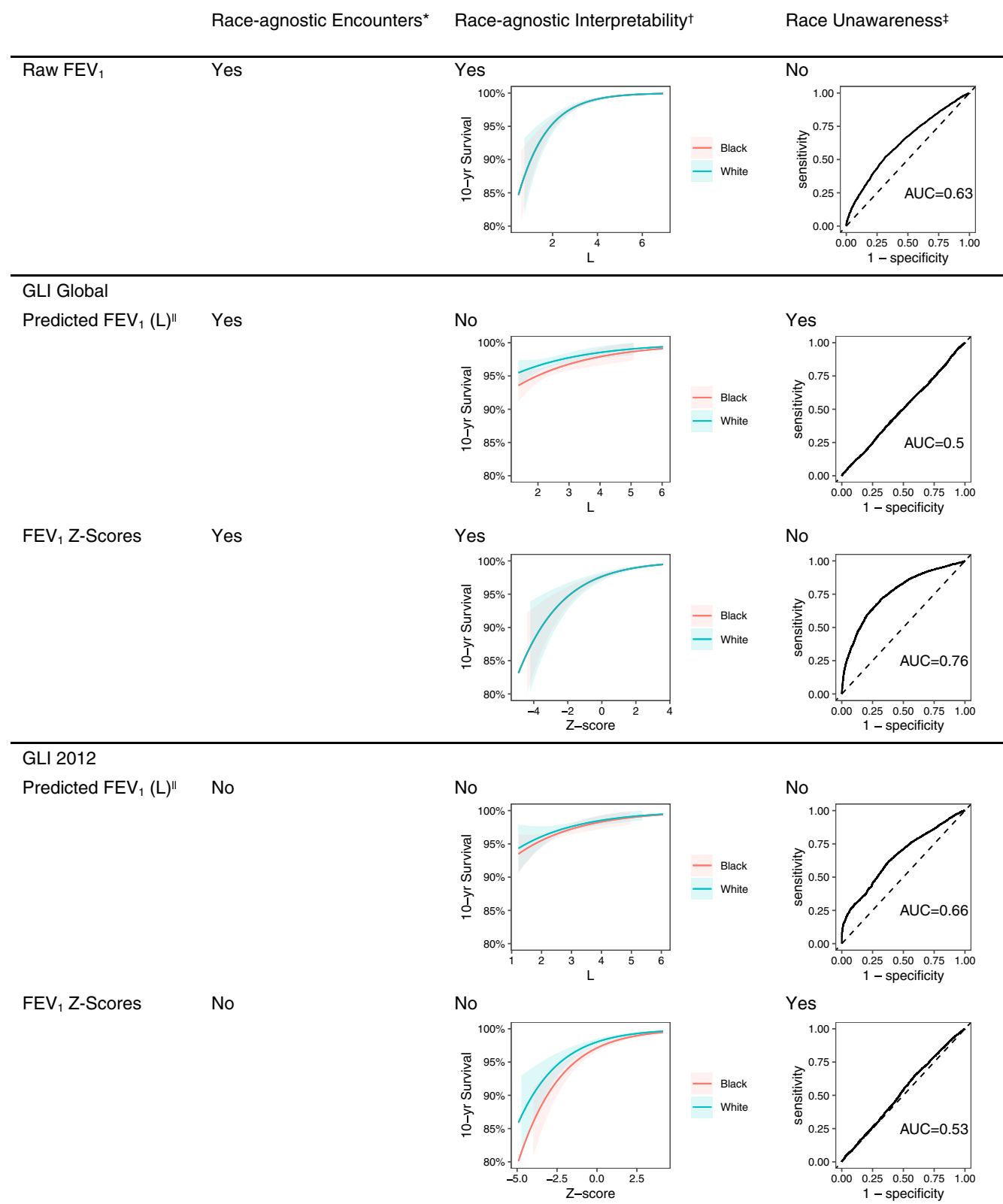


Figure 1. Race-neutrality of lung function measures among self-identified Black and White individuals.[§] *Race-agnostic encounters criterion means that obtaining the spirometric value does not require collecting the individual's self-identified race. †A metric meets race-agnostic interpretability if it carries the same predictive information for a clinically relevant outcome (10-yr survival in this example). Predicted survival probabilities are for hypothetical 49-year-old females with different degrees of observed lung function, calculated separately for self-identified

Figure 1. (Continued). Black and White persons. Survival was modeled in 7,867 participants eligible for mortality follow-up, using Cox proportional hazards with covariates for age, sex, race, and lung function as represented in each row. [‡]A metric is race unaware if its value cannot be used to predict race. AUCs shown in this column are for predicting binary race using the lung function metric. [§]All results were calculated using appropriate survey weights to account for complex survey design. Replacing FEV₁ with FVC, Z-scores with percent predicted values, or GLI-Global with GLI-2012-Other did not affect race-neutrality interpretations (results not shown). ^{||}Predicted FEV₁ is the immediate output of the reference equations and shows the predicted healthy lung function, expressed in liters. AUC = area under the curve; GLI = Global Lung Function Initiative.

outside special cases (7). As long as racial disparities in lung function exist, it is mathematically impossible to achieve race unawareness and race-agnostic interpretability of Z-scores simultaneously.

Which definition of race neutrality should be prioritized? In each context, one must choose either race-agnostic interpretability or race unawareness as the primary criterion. Race-agnostic encounters are always desirable because they avoid inadequate race categories (11) and mistrust concerns (12). However, race-agnostic encounters do not ensure fairness, because algorithms can still cause racially disproportionate harm without explicitly considering race (13, 14).

In risk-based clinical decisions, such as in chronic obstructive pulmonary disease management, the goal is to achieve the most accurate assessment of a patient's risk, regardless of whether it stems from biological factors or complex exposures. Race-agnostic interpretable metrics such as Z-scores from GLI-Global might be preferable in this context, because they avoid normalizing racial disparities that impact health and have consistent prognostic implications. Race-agnostic interpretability is also desirable when determining the lowest FEV₁ thresholds for safe lung resection, ensuring equal risk across all patients.

Spirometry has other applications beyond clinical encounters, and a one-size-fits-all solution may not exist. For example, when determining occupational eligibility, different societal values may apply. To protect individuals with adverse exposures from additional occupational risks, race-aware Z-scores from equations such as GLI-Global that do not adjust for race may be suitable. However, this could inadvertently perpetuate racial discrimination in hiring (15).

Conversely, to prevent racial discrimination in hiring and life insurance premiums, race-unaware Z-scores may be preferred. In contexts involving access to limited resources or economic opportunities, race-unaware Z-scores help ensure that individuals affected by structural racism do not face further disadvantages because of it. Currently, this requires race-specific reference equations. Indeed, race adjustment in spirometry was initially mandated in the 1978 Cotton Dust Standard to prevent hiring discrimination (16). This standard successfully reduced exposure and byssinosis rates (17), although the differential impact of race-based spirometry standards on workers' health remains unclear. Because race is often a proxy for social determinants of health, one underexplored potential alternative is adjustment for deprivation indices, as

shown with PREVENT (predicting risk of cardiovascular disease events) equations (18). This approach achieves race-agnostic encounters in addition to race unawareness and indicates a focus on addressing structural inequalities. Although several countries have official deprivation indices (19), these scores must be scrutinized for their potential to introduce new biases.

Race unfortunately remains a determinant of health. As police brutality, apartheid, and ethnic cleansing remind us, racism has serious health consequences. Because racial inequality affects health in complex ways, fairness through unawareness can sometimes cause additional harm. What makes an algorithm fair is not whether it includes race but how it handles it. We echo calls for ongoing nuanced, careful, and context-specific assessments of the fairness consequences of different algorithms. Committing to fairness in clinical decision making requires continuous effort and a willingness to learn, relearn, and question long-standing and emerging orthodoxies. ■

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: The authors thank Dr. Stephen Pfohl, Dr. Abdollah Safari, and Dr. Anna Zink for their input.

References

1. Stanojevic S, Kaminsky DA, Miller MR, Thompson B, Aliverti A, Barjaktarevic I, et al. ERS/ATS technical standard on interpretive strategies for routine lung function tests. *Eur Respir J* 2022;60: 2101499.
2. Bhakta NR, Bime C, Kaminsky DA, McCormack MC, Thakur N, Stanojevic S, et al. Race and ethnicity in pulmonary function test interpretation: an official American Thoracic Society statement. *Am J Respir Crit Care Med* 2023; 207:978–995.
3. Kanj AN, Niven AS. Race-neutral Z-score classification of airflow obstruction: a measured step forward. *Am J Respir Crit Care Med* 2024; 210:1287–1289.
4. Liu GY, Khan SS, Colangelo LA, Meza D, Washko GR, Sporn PHS, 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;175:1118–1125.
5. Cannon MF, Goldfarb DG, Zeig-Owens RA, Hall CB, Choi J, Cohen HW, et al. Normal lung function and mortality in World Trade Center responders and NHANES III participants. *Am J Respir Crit Care Med* 2024;209:1229–1237.
6. Sheshadri A, Rajaram R, Baugh A, Castro M, Correa AM, Soto F, et al. Association of preoperative lung function with complications after lobectomy using race-neutral and race-specific normative equations. *Ann Am Thorac Soc* 2024;21:38–46.
7. Barocas S, Hardt M, Narayanan A. Fairness and machine learning: limitations and opportunities. Cambridge, MA: MIT Press; 2023.
8. 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;205:723–724.
9. Gaffney AW, McCormick D, Woolhandler S, Christiani DC, Himmelstein DU. Prognostic implications of differences in forced vital capacity in black and white US adults: findings from NHANES III with long-term mortality follow-up. *EClinicalMedicine* 2021;39:101073.
10. Burney P, Hooper R. The use of ethnically specific norms for ventilatory function in African-American and white populations. *Int J Epidemiol* 2012;41:782–790.

11. Movva R, Shanmugam D, Hou K, Pathak P, Gutttag J, Garg N, *et al.* Coarse race data conceals disparities in clinical risk score performance. *Proc 8th Mach Learn Healthcare Conf PMLR* 2023;219:443–472.
12. Schmidt IM, Shohet M, Serrano M, Yadati P, Menn-Josephy H, Ilori T, *et al.* Patients' perspectives on race and the use of race-based algorithms in clinical decision-making: a qualitative study. *J Gen Intern Med* 2023;38:2045–2051.
13. Siddique SM, Tipton K, Leas B, Jepson C, Aysola J, Cohen JB, *et al.* The impact of health care algorithms on racial and ethnic disparities. *Ann Intern Med* 2024;177:484–496.
14. Cary MP, Zink A, Wei S, Olson A, Yan M, Senior R, *et al.* Mitigating racial and ethnic bias and advancing health equity in clinical algorithms: a scoping review. *Health Aff (Millwood)* 2023;42:1359–1368.
15. Diao JA, He Y, Khazanchi R, Nguemeni Tiako MJ, Witonsky JI, Pierson E, *et al.* Implications of race adjustment in lung-function equations. *N Engl J Med* 2024;390:2083–2097.
16. Townsend MC, Cowl CT. US occupational historical perspective on race and lung function. *Am J Respir Crit Care Med* 2022;206:789–790.
17. Office of Program Evaluation, Occupational Safety and Health Administration. Regulatory Review of OSHA's Cotton Dust Standard, 29 C.F.R. § 1910.1043; 2000 [accessed 2025 Jan 17]. Available from: https://www.osha.gov/sites/default/files/cottondust_final2000.pdf.
18. Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, *et al.*; Chronic Kidney Disease Prognosis Consortium and the American Heart Association Cardiovascular-Kidney-Metabolic Science Advisory Group. Development and validation of the American Heart Association's PREVENT Equations. *Circulation* 2024;149:430–449.
19. Phillips RL, Liaw W, Crampton P, Exeter DJ, Bazemore A, Vickery KD, *et al.* How other countries use deprivation indices—and why the United States desperately needs one. *Health Aff (Millwood)* 2016;35:1991–1998.