A Framework for Considering the Value of Race and Ethnicity in Estimating Disease Risk

Madison Coots, MS; Soroush Saghafian, PhD, MS; David M. Kent, MD, MS; and Sharad Goel, PhD, MS

Background: Accounting for race and ethnicity in estimating disease risk may improve the accuracy of predictions but may also encourage a racialized view of medicine.

Objective: To present a decision analytic framework for considering the potential benefits of race-aware over race-unaware risk predictions, using cardiovascular disease, breast cancer, and lung cancer as case studies.

Design: Cross-sectional study.

Setting: NHANES (National Health and Nutrition Examination Survey), 2011 to 2018, and NLST (National Lung Screening Trial), 2002 to 2004.

Patients: U.S. adults.

Measurements: Starting with risk predictions from clinically recommended race-aware models, the researchers generated race-unaware predictions via statistical marginalization. They then estimated the utility gains of the race-aware over the race-unaware models, based on a simple utility function that assumes constant costs of screening and constant benefits of disease detection.

Results: The race-unaware predictions were substantially miscalibrated across racial and ethnic groups compared with the race-aware predictions as the benchmark. However, the clinical net benefit at the population level of race-aware predictions over race-unaware predictions was smaller than expected. This result stems from 2 empirical patterns: First, across all 3 diseases, 95% or more of individuals would receive the same decision regardless of whether race and ethnicity are included in risk models; second, for those who receive different decisions, the net benefit of screening or treatment is relatively small because these patients have disease risks close to the decision threshold (that is, the theoretical "point of indifference"). When used to inform rationing, race-aware models may have a more substantial net benefit.

Limitations: For illustrative purposes, the race-aware models were assumed to yield accurate estimates of risk given the input variables. The researchers used a simplified approach to generate race-unaware risk predictions from the race-aware models and a simple utility function to compare models.

Conclusion: The analysis highlights the importance of foregrounding changes in decisions and utility when evaluating the potential benefit of using race and ethnicity to estimate disease risk.

Primary Funding Source: The Greenwall Foundation.

Ann Intern Med. 2025;178:98-107. doi:10.7326/M23-3166
For author, article, and disclosure information, see end of text.
This article was published at Annals.org on 3 December 2024.

Statistical models are used to estimate individual risk for many of the most prevalent and deadly diseases faced by Americans. These risk estimates are often used to identify persons who would benefit from interventions, such as screening or prophylactic treatment (for example, pharmacotherapy or lifestyle changes), that would allow them to better manage their health and slow or halt the progression of their condition. However, screening and treatment come with potential harms and costs. Consequently, the medical community typically targets such interventions at those whose predicted risk for developing the disease is above a certain threshold. These thresholds are usually set at the level of risk above which the expected benefits exceed the expected harms and costs–that is, the "point of indifference."

See also:

Editorial comment. 132

Disease risk predictions are often produced using variables such as age, gender, relevant biomarkers, and lifestyle factors. There is debate over whether an individual's race and ethnicity should additionally be included to account for observed disparities in disease incidence and mortality rates across demographic subgroups in the United States (1, 2). Past work has demonstrated that including race and ethnicity improves the accuracy of clinical prediction models and that their omission could exacerbate disparities in health outcomes (3–10). Other work has argued that race and ethnicity can serve as a useful proxy for exposure to systemic racism, thereby offering a way to mitigate discrimination in health care (11). However, concern and criticism persist about the use of race and ethnicity in estimating disease risk (12–14). Their inclusion in predictive models may, for instance, inadvertently reinforce pernicious attitudes of biological determinism or

lead to greater stigmatization of already marginalized persons. In part for these reasons, race-aware estimates of glomerular filtration rate have largely been replaced by a "race-free" equation (15), both to avoid race-based predictions and to address concerns that a race-aware model may deprioritize Black patients for kidney transplantation (16–19). Similarly, the American Heart Association recently released race-unaware equations for predicting risk for cardiovascular disease events (PREVENT [Predicting Risk of cardiovascular disease EVENTs]), and researchers have released race-unaware calculators for estimating risk in other conditions (20–23).

In this work, we present a decision analytic framework for considering both the statistical and clinical utility of race and ethnicity in disease risk estimation. This approach considers not only improvements in accuracy from the use of race and ethnicity but also the extent to which those improvements affect decisions and utility. We apply this framework to cardiovascular disease, breast cancer, and lung cancer as illustrative case studies.

METHODS

Overview

To assess the value of race and ethnicity in estimating disease risk, we compare statistical predictions and clinical net benefit of race-unaware versus race-aware risk models for breast cancer, lung cancer, and cardiovascular disease. For each disease, we use a clinically recommended race-aware risk model to obtain risk estimates for a sample of individuals. We then convert the race-aware risk estimates to race-unaware risk estimates; we do so via statistical marginalization of race and ethnicity for simplicity. We then compare how clinical decisions would change using race-aware versus race-unaware risk estimates and quantify how these changed decisions translate into utility gains or losses for different racial and ethnic groups under a shared decision-making context and, separately, a rationing context. For illustrative purposes, we assume the race-aware models yield accurate estimates of risk given the input variables.

Data Sources

Our analysis of cardiovascular disease and breast cancer is based on publicly available data from NHANES (National Health and Nutrition Examination Survey), 2011 to 2018 (24), a cross-sectional survey representative of the community-dwelling U.S. population that combines interview responses with laboratory data to provide insight into health and nutrition. We restricted our samples to adults clinically eligible for each disease model (**Appendix**, available at Annals.org).

Our analysis of lung cancer is based on cross-sectional data from NLST (National Lung Screening Trial) (25), a randomized controlled trial done between August 2002 and April 2004 to assess whether low-dose computed tomography screening reduces lung cancer mortality relative to chest radiography among persons at high risk. Data on the approximately 54 000 participants included demographics, medical history, and lifestyle factors relevant to the development of lung cancer. Approximately 90% of NLST participants identified as non-Hispanic White, although these data have been used to investigate racial and ethnic disparities in lung cancer (26, 27). We reweighted participants to match the joint age, gender, and race distribution of Americans between age 40 and 80 years (28) (**Appendix**).

Risk Predictions

The **Table** describes the risk models we used, including covariates, risk thresholds, and the clinical decisions the models inform (**Appendix**).

The reference risk models are, by design, race-aware. In practice, the preferred approach to generate race-unaware models is to train new models that do not include race and ethnicity as inputs and to add other factors correlated with race and ethnicity that might improve the performance of race-unaware predictions (20, 21). However, because of data limitations, we could not retrain race-unaware models in this way. Instead, we estimated race-unaware models by taking a weighted average of the race-aware risk predictions, where the weights equal the population proportions of each group conditional on the nonrace risk factors. For example, to obtain the race-unaware lung cancer risk estimate for an individual, we used the Lung Cancer Risk Assessment Tool model to first produce 4 risk estimates, varying only race (that is, White, Black, Hispanic, or Asian) and holding all else constant. We then took a weighted average of these 4 race-aware predictions to obtain a race-unaware risk estimate, a well-established statistical technique for removing variables from risk models (Appendix). These race-unaware models are intended only to illustrate broad statistical principles and are not intended for clinical use.

The Utility Framework

To quantify the value of using race-aware risk scores to make screening and treatment recommendations, we adopt a utility framework where both costs and gains in health are expressed on a common scale. For each disease, we assume a constant cost of intervention (that is, screening or treatment) for those deemed high-risk. This cost encapsulates a wide range of monetary and nonmonetary considerations, such as the direct cost of screening or treatment and indirect costs like taking time off from work. We further assume a constant benefit for detecting disease across individuals, due to early detection and long-term treatment of the disease. This simplification assumes that the benefit of appropriate intervention (that is, intervening when the individual truly has the disease) does not vary by age, race, or other attributes and allows us to highlight broad patterns in a base case on which to expand our analysis.

Table. Risk Models, Inputs, Thresholds, and Decisions for Each Disease

Disease	Risk Model	Model Inputs	Risk Threshold	Decision Considered
Cardiovascular disease	U.S.-derived 2013 ASCVD pooled cohort equations (29, 30)	Sex, race and ethnicity, age, diabetes status, smoker status, untreated and treated systolic blood pressure, total cholesterol level, high-density lipoprotein cholesterol level	7.5% (31)	Recommendation of moderate-intensity statin therapy (31)
Breast cancer	Breast Cancer Risk Assessment Tool (32–37)	Age, race and ethnicity, number of first-degree relatives with breast cancer* (38), age at menarche, age at first live birth, history of breast biopsy*, and atypical hyperplasia*	1.67% (39, 40)	Recommendation of tamoxifen and raloxifene as chemoprevention (39, 40)
Lung cancer	Lung Cancer Risk Assessment Tool (41, 42)	Gender, race and ethnicity, age, smoking history, family history of lung cancer, body mass index, highest education level attained, and history of other diseases*	2.0% (41, 43)	Recommendation of CT lung cancer screening (41, 43)

ASCVD = atherosclerotic cardiovascular disease; CT = computed tomography.
* Further detail is the Appendix (available at Annals.org).

Figure 1 shows the structure of the utility function. We set the utility of "no intervention" to 0, where individuals incur neither costs nor benefits of intervention. We normalize the benefit of appropriate intervention for each individual to 1 and assume a uniform cost *c* of intervening. Based on this framework, the optimal policy is to intervene if, and only if, a patient's predicted disease risk *r* exceeds a decision threshold *t* (the point of indifference), where the expected benefits of intervention equal the costs. This decision threshold implicitly accounts for the relative weights of a false-positive versus a false-negative prediction. For example, setting a risk threshold of 7.5% for cardiovascular disease treatment suggests that treating a patient with a statin whose risk is exactly 7.5% results in no utility gain because, at this level of risk, the benefits of therapy are nullified by the costs, burdens, and harms.

As in decision curve analysis (44–46), knowing the optimal threshold for a decision yields information on the relative costs and benefits of interventions. With the normalization in Figure 1, the implied value of *c* is precisely the threshold *t* (Appendix). For a given screening strategy, we call the resulting utility the "net benefit" of that strategy. To quantify the gains in net benefit of using a race-aware model over a race-unaware model to make a decision for an individual, we subtract the race-unaware utility from the race-aware utility.

Role of the Funding Source

Funders did not play a role in the design, conduct, or analysis of this study or in the decision to submit the manuscript for publication.

RESULTS

Miscalibration of Race-Unaware Risk Predictions

The race-unaware predictions that we developed exhibit substantial miscalibration across racial and ethnic groups (Figure 2, *top*). Assuming that the original, race-aware models yield accurate risk estimates, we find the race-unaware models underestimate risk for cardiovascular disease and lung cancer for Black individuals. In contrast, race-unaware models overestimate risk for breast and lung cancer for Asian individuals and similarly overestimate risk for lung cancer for Hispanic individuals. For White individuals, the predicted risks were similar between the race-aware and race-unaware models for all 3 diseases. Miscalibrated predictions can result in misclassifications that lead to inappropriately recommending screening or treatment of low-risk patients or failing to recommend screening or treatment of high-risk patients. The observed miscalibration of the marginalized race-unaware models we consider may not generalize to race-unaware models that are developed de novo, particularly if other covariates are included that correlate with race and ethnicity (20, 21). Nevertheless, similar patterns of miscalibration have been found previously for race-unaware disease models that were fitted directly (5).

Utility Gains From Race-Aware Predictions Assuming Constant Benefits

We start by considering the added value of race-aware predictions under our base-case utility model, where the benefit of appropriate intervention is constant across individuals. The overall clinical benefits of race-aware risk predictions in this base case were not as large as one might expect given the observed miscalibration of the race-unaware predictions. We find that the race-aware models yield an increase in net benefit of approximately 2.0 per 10 000 individuals for cardiovascular disease, 0.49 per 10 000 individuals for breast cancer, and 1.76 per 10 000 individuals for lung cancer. In Figure 3, we show the results of this analysis by race and ethnicity. To contextualize these results, the baseline utility (that is, the net benefit from using a race-unaware model relative to a policy of never intervening) is 388 per 10 000 individuals for cardiovascular disease, 11 per 10 000 for breast cancer, and 158 per 10 000 for lung cancer. (See Appendix

Figure 1, available at Annals.org, for baseline utility results per disease by race and ethnicity.) For each disease, the subgroups that have the largest gains in net benefit from race-aware risk estimates are those for whom the race-unaware miscalibration is worst. For breast cancer, Asian individuals benefit the most; for cardiovascular disease, Black individuals benefit the most; and for lung cancer, Hispanic individuals benefit the most. Across diseases and race subgroups, race-aware predictions lead to improvements in net benefit of at most 17 per 10 000 individuals.

Given that the race-unaware models are starkly miscalibrated, it is perhaps surprising that the race-aware models do not yield larger utility gains. Two factors help explain this phenomenon. First, as shown in the bottom rows of Figures 2 and 3, including race shifts predictions considerably for many patients, but most receive the same recommendation under a race-aware model as under a race-unaware model—because recommendations change only for the relatively few patients close to the decision threshold. The percentage who receive the same recommendation under both models is 98% for cardiovascular disease, 97% for breast cancer, and 95% for lung cancer. Most patients thus accrue no gains from using a race-aware model. Second, for the small number of patients near the threshold who do receive different recommendations under the 2 risk models, the utility gains are modest. To see this, note that those individuals with risk estimates equal to the decision threshold should, in theory, be completely indifferent between receiving and not receiving the intervention—precisely because the threshold was chosen to be the point of indifference. Similarly, those near the threshold should be largely indifferent between the alternatives.

Utility Gains From Race-Aware Predictions Assuming Heterogeneous Benefits

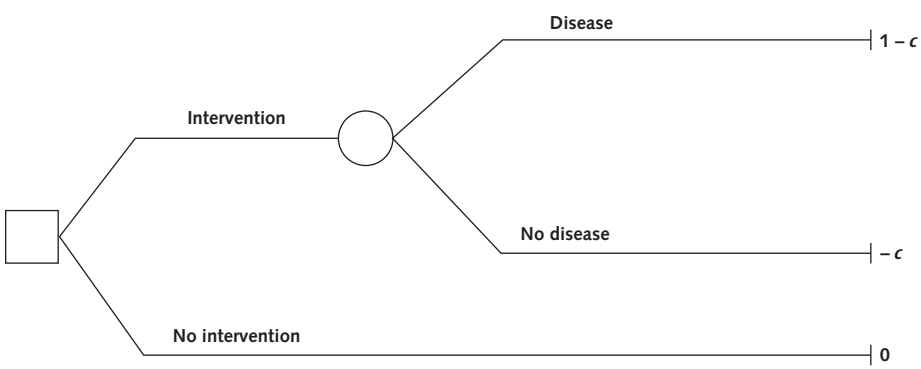
The results above assume, for simplicity, that the benefit of appropriate intervention is constant across individuals. For example, we have implicitly assumed that the value of appropriate treatment of older people is the same as that of younger people, even though treatment of younger people could lead to more life-years gained. In theory, our race-unaware models could systematically underidentify persons who would benefit the most from treatment. However, in the Appendix, we consider age-related heterogeneity in utility (Appendix Figure 2, available at Annals.org) and find qualitatively similar results to the base case.

There may also be heterogeneity in the utility function based on factors that relate to race and ethnicity. For example, certain racial or ethnic groups may exhibit lower responsiveness to treatment in later stages of disease or accrue greater utility from detecting or recommending prophylaxis before disease onset. As a result, there may be group-specific tradeoffs between costs of screening and benefits of detection—tradeoffs that can be better accommodated by race-aware decision making. In the Appendix, we trace out these group-specific tradeoffs for each disease (Appendix Figure 3, available at Annals.org). If sizable group-specific differences exist in the benefit of intervention, the value of a race-aware approach may be larger than what we find here (48).

Utility Gains From Race-Aware Predictions Under Conditions of Scarcity

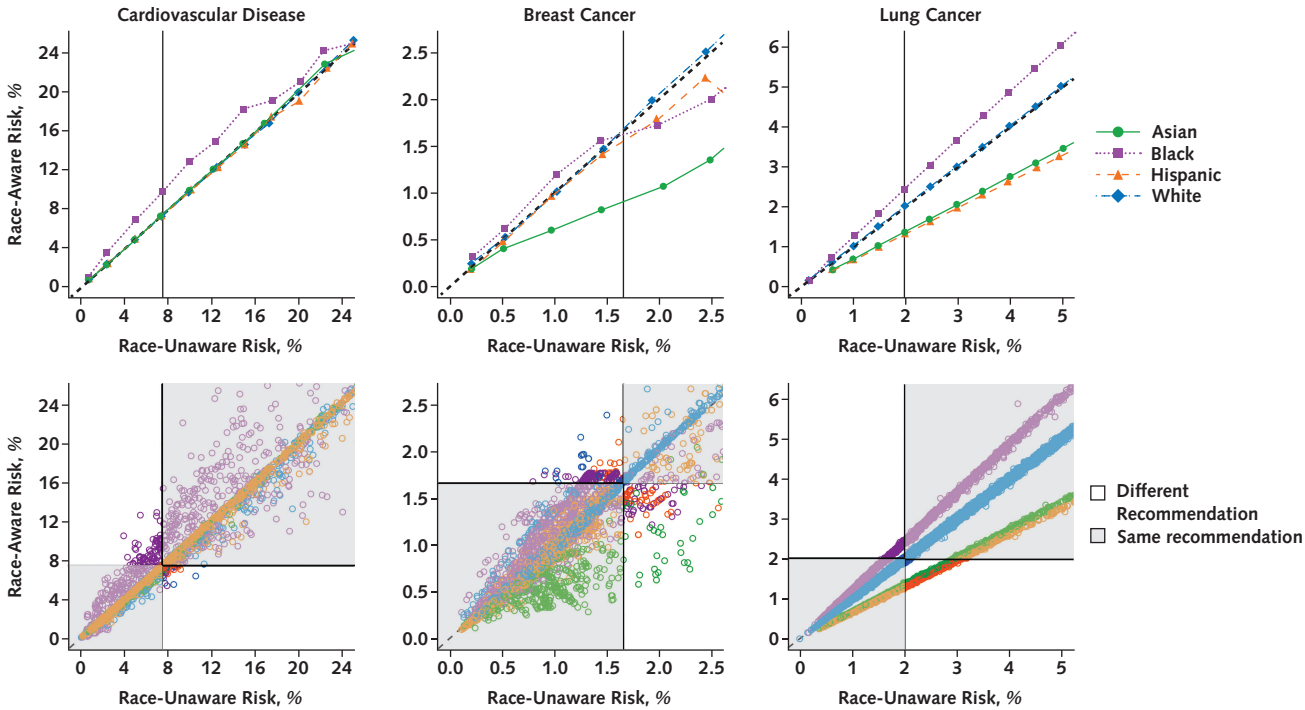
Finally, the empirical patterns discussed above may not hold under conditions of scarcity, where prediction models are used for efficient rationing of limited health care resources. In such circumstances, decision thresholds are determined not by the point of indifference but by capacity and may indeed be far from the point of indifference (for example, in organ transplantation). To demonstrate, we consider a hypothetical example where severe resource constraints mean that only individuals with risk scores above *K*% may receive the appropriate intervention for each disease (such as

Figure 1. The costs and benefits of intervention.



The figure shows the structure of the base-case utility function used in the subsequent analysis. We normalize the benefit of appropriate intervention to equal 1 unit, with c denoting the cost of intervention.

Figure 2. Assessing the effects of miscalibration of race-unaware risk estimates on screening and treatment recommendations.



Top. Calibration plots for cardiovascular disease, breast cancer, and lung cancer, showing race-unaware risk predictions plotted against race-aware risk predictions for each disease. The line $y = x$ denotes the line of perfect calibration (shown by a heavy dashed black line). The scales of the *x*- and *y*-axes differ across diseases because of differences in the risk distributions and thresholds for each disease. Across all 3 diseases, racial minorities have more miscalibration in race-unaware predictions than White individuals. The solid black line marks the recommended screening or treatment threshold for each disease. **Bottom.** Scatter plots showing race-unaware risk plotted against race-aware risk. Each point represents an individual in the data. The unshaded regions indicate individuals who would receive a different screening/treatment recommendation under the race-aware model than they would under the race-unaware model. Individuals in the shaded region receive the same recommendation under each model.

pharmacotherapy for cardiovascular disease and breast cancer, or a computed tomography scan for lung cancer), even though many individuals with risk scores below *K*% might benefit from the intervention. Using our base utility function assuming constant benefit, **Figure 4** shows the resulting group-specific gains in net benefit for various values of the screening threshold, demonstrating that gains from using race-aware predictions increase substantially under conditions of scarcity (unless *K* is very large).

In this hypothetical scenario, race-aware prediction models for cardiovascular disease and lung cancer would appropriately identify and prioritize higher-risk Black patients and deprioritize lower-risk Hispanic, Asian, and White patients for intervention. For example, if we imagine that only individuals with a cardiovascular disease risk score greater than 12% may receive statins—which corresponds to the riskiest 25% of individuals—using a race-aware risk model for Black patients would result in a net benefit of approximately 80 per 10 000 individuals, over a baseline net benefit of approximately 533 per 10 000 for the race-unaware model. This pattern is driven by the fact that, under scarcity, those who receive different recommendations are farther from their point of indifference and thus have larger utility gains from reclassification. In addition, given the distributions of risk in this example, more individuals receive screening recommendations that differ between the race-aware and race-unaware models.

DISCUSSION

In the shared decision-making context, our results suggest that race-aware risk models yield smaller gains in net benefit over race-unaware models than the improvement in predictions might suggest. This finding stems from 2 patterns in the data: First, although the race-aware model changes predictions for all patients, decisions often change for only a small fraction of patients; second, among those who do receive different decisions, the net value of intervention is relatively small because their disease risk is typically close to the decision threshold—which, in the shared decision-making case, is typically set at the theoretical point of indifference.

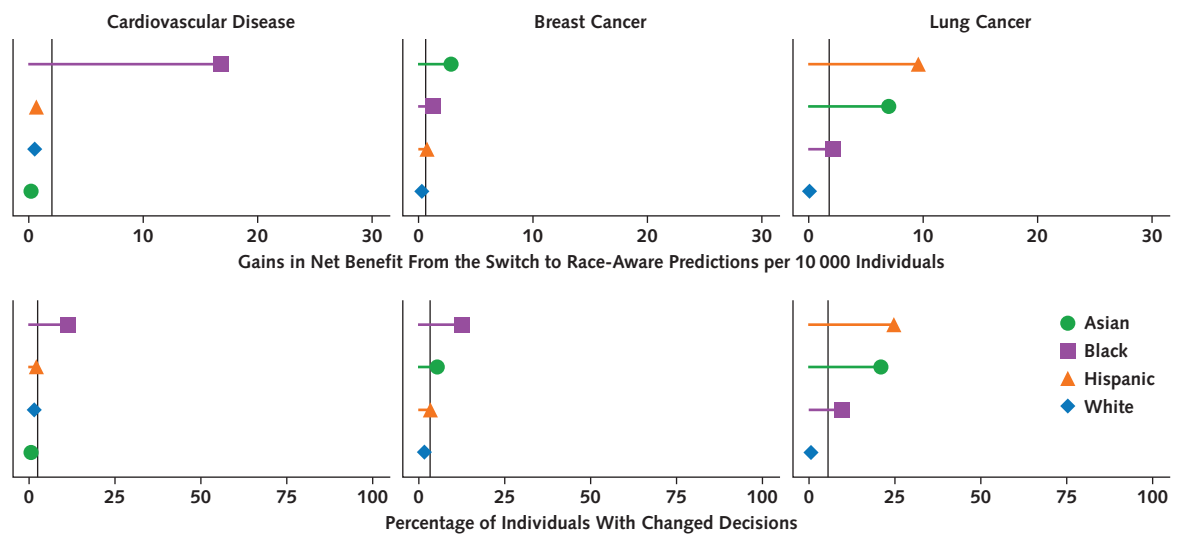
However, circumstances exist under which using race-aware models may yield greater net benefit, most notably in rationing contexts. Under rationing, the

decision threshold is not determined by the point of indifference but by capacity–and may be far from the point of indifference. Reclassifying patients across a decision threshold far from the point of indifference may be quite consequential. Moreover, the additional net benefit is preferentially directed to racial and ethnic minorities in our examples, such that using race-aware models is anticipated to reduce disparities. Although this pattern need not always hold, it would tend to when racial and ethnic subgroups are at higher risk for adverse outcomes, a common scenario in many clinical domains (49). The specific risk models we considered are intended for shared decision making, not rationing, but rationing is ubiquitous in health care and prediction models are increasingly proposed to allocate resources. For example, during the COVID-19 pandemic, many states developed race-aware algorithms to allocate scarce therapeutics, under the principle of "equal treatment for equal risk" (50, 51).

In evaluating the use of race and ethnicity in clinical risk algorithms, our work highlights the importance of foregrounding not just improvements in accuracy but changes in decisions and utility. Past work has largely focused on comparing the accuracy of race-unaware and race-aware models (10, 21–23, 47). However, as evidenced by our results with all 3 diseases, improvements in accuracy do not always translate to commensurately large changes in decisions and benefits. Other work that has measured the effects of race-aware predictions on decisions has stopped short of considering utility (2, 5, 6). Given the known costs of screening and treatment, our work shows a need to additionally examine the gains in net utility from changed decisions. Last, our research adds to previous work highlighting the important–and often overlooked–ethical distinctions in shared decision making versus rationing, because the latter gives rise to fairness concerns less relevant to the former (52). In particular, more care may be needed when omitting (or including) race and ethnicity for models used for rationing, given the larger consequences of risk reclassification in that context. Recent guidelines on model development further discuss these distinctions (53).

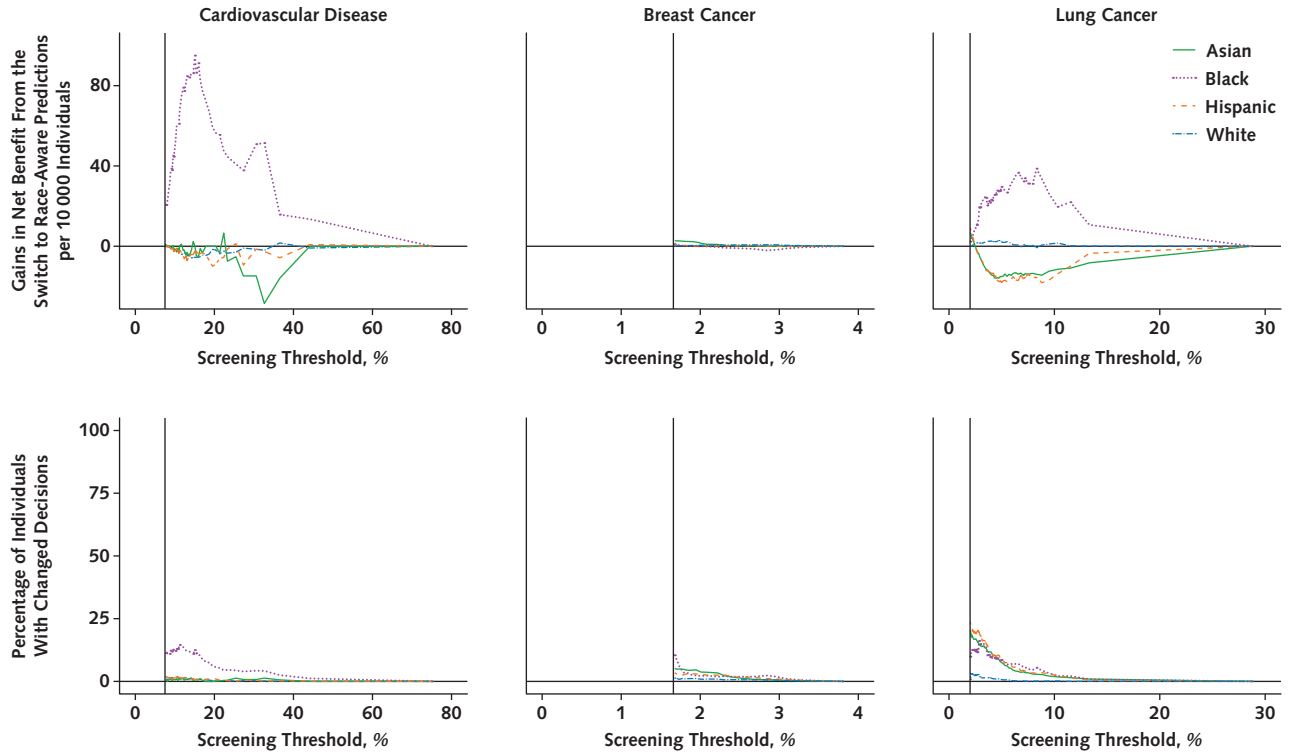
Our analyses are intended to illustrate general principles and should not be understood as specific recommendations for modeling risk in the 3 diseases examined. In particular, our work is subject to several important limitations. First, our analysis assumes that the clinically recommended race-aware models we consider yield accurate estimates of risk given the input variables–an assumption that lets us evaluate the relative performance of the derived race-unaware models. These race-aware models might have systematic inaccuracies (54), although we note that they were trained on widely used data with standard statistical methods. Moreover, our results showing statistical gains from using race-aware over race-unaware models are consistent with the general principle that adding prognostic information improves model performance (3, 4). Second, we have primarily considered per capita utility gains, but one could alternatively consider aggregate population-level utility benefits, which are considerably larger. Third, our results might not apply where the

Figure 3. The per capita utility gain–assuming constant benefit across individuals–of a race-aware risk model over a race-unaware model, disaggregated by race and ethnicity, and the percentage of individuals with different decisions under each model.



Top. The per capita utility gains by race and ethnicity group from using a race-aware model over a race-unaware model for making screening and treatment recommendations. The vertical line denotes the average gains across the entire population. Minority racial and ethnic groups consistently have the largest gains. This pattern is primarily driven by the fact that race-unaware risk predictions will most closely reflect the risk of the majority group, which in this case is White individuals. **Bottom.** The percentage of individuals within each race and ethnicity group that would receive different recommendations under race-aware and race-unaware models. In nearly every case, only a small proportion would receive different recommendations. The vertical line denotes the percentage of individuals with changed decisions across the entire population.

Figure 4. The effects of race-aware predictions under conditions of scarcity.



Top. As a function of the screening threshold *K*, the per capita utility gains from using race-aware risk predictions over race-unaware predictions, disaggregated by race and ethnicity. **Bottom.** As a function of the screening threshold *K*, the percentage of individuals within each race and ethnicity group who would receive different recommendations across the 2 models. This percentage includes individuals who would be recommended for an intervention under a race-aware model but not a race-unaware model, as well as the reverse. The solid black line denotes the standard recommended screening threshold for each disease. These results collectively show that Black individuals would have considerable gains in utility from using race-aware predictions under conditions of scarcity. In addition, we see that using a race-aware model for breast cancer would result in minimal gains in utility across groups. This result is primarily driven by relatively small gains in net benefit from using a race-aware model under normal circumstances.

apparent disparity in disease risks is suspected to arise from label bias–for example, arising from a difference in diagnostic labeling or outcome ascertainment rather than a true disparity in disease incidence or outcome (55). In the presence of label bias, using a race-aware model might in fact exacerbate statistical biases (55). Fourth, we obtained race-unaware risk estimates by taking a weighted average of race-aware risk estimates for an individual. In practice, race-unaware risk estimates would be obtained by training a separate race-unaware model. Finally, our analysis is contingent on the specific utility framework that we use to evaluate screening and treatment decisions. In particular, for simplicity we assumed a single decision threshold even though 2 or more thresholds might be appropriate for identifying groups at low, medium, and high risk, for example (46). Having more decision points might increase the number of patients reclassified.

At the heart of the debate over using race-unaware versus race-aware models to estimate disease risk is the goal of mitigating racial and ethnic disparities in health outcomes. Our work does not attempt to uncover the cause of such disparities in outcomes across racial and ethnic groups, but any efforts to do so should consider racism as a possible cause (56, 57). Concern also exists about the limited efficacy of risk predictions–either race-aware or race-unaware–for mitigating disparities in health outcomes apart from disease incidence. For example, Black women in the United States have lower incidence rates of breast cancer than White women but have a mortality rate 40% higher, highlighting the limitations of focusing solely on disease risk for mitigating disparities in outcomes that are downstream from screening, such as mortality rates (58).

Our main result–that large statistical gains from race-aware prediction may lead to only modest gains in utility–is based on broad principles, and so it likely extends to various contexts in medicine and beyond where using race and ethnicity in predictive models is contested. However, when used to inform rationing, race-aware models may be much more beneficial than when used in a shared decision-making context. We believe our work provides a widely adaptable framework for evaluating the consequences of including or excluding race and ethnicity from predictions. However, we also emphasize that the specifics in each case need

to be considered. We hope that our analytic framework helps researchers, practitioners, and policymakers better understand and balance the underlying tradeoffs of using race and ethnicity when estimating risk.

From Harvard University, Cambridge, Massachusetts (M.C., S.S., S.G.); and Predictive Analytics and Comparative Effectiveness Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center and Tufts University School of Medicine, Boston, Massachusetts (D.M.K.).

Disclaimer: The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by the National Cancer Institute.

Acknowledgment: The authors thank Johann Gaebler, Josh Grossman, Michael Zanger-Tishler, Alex Chohlas-Wood, Shira Gur-Arieh, Jacob Jameson, Max Spohn, Ravi Shroff, Ankur Pandya, and Julian Nyarko for helpful discussions. The authors additionally thank the National Cancer Institute for access to its data collected by NLST (Cancer Data Access System project NLST-1121).

Financial Support: In part by a "Making a Difference" grant and Presidential Awards from the Greenwall Foundation (Dr. Kent).

Disclosures: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M23-3166.

Reproducible Research Statement: *Study protocol:* Not available. *Statistical code:* Available at https://github.com/madisoncoots/race-in-estimating-disease-risk. *Data set:* NHANES data are publicly available from the Centers for Disease Control and Prevention. Processed versions of the NHANES data sets used for analysis are additionally included in the replication package posted at https://github.com/madisoncoots/race-in-estimating-disease-risk. NLST data are available on request from the National Cancer Institute.

Corresponding Author: Madison Coots, MS, John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138; e-mail, mcoots@g.harvard.edu.

Author contributions are available at Annals.org.

References

1. **Chen T, Kharazmi E, Fallah M.** Race and ethnicity-adjusted age recommendation for initiating breast cancer screening. JAMA Netw Open. 2023;6:e238893. [PMID: 37074714] doi:10.1001/jamanetworkopen.2023.8893
2. **Landy R, Gomez I, Caverly TJ, et al.** Methods for using race and ethnicity in prediction models for lung cancer screening eligibility. JAMA Netw Open. 2023;6:e2331155. [PMID: 37721755] doi:10.1001/jamanetworkopen.2023.31155
3. **Manski CF.** Patient-centered appraisal of race-free clinical risk assessment. Health Econ. 2022;31:2109-2114. [PMID: 35791466] doi:10.1002/hec.4569
4. **Manski CF, Mullahy J, Venkataramani AS.** Using measures of race to make clinical predictions: decision making, patient health, and fairness. Proc Natl Acad Sci U S A. 2023;120:e2303370120. [PMID: 37607231] doi:10.1073/pnas.2303370120
5. **Chohlas-Wood A, Coots M, Goel S, et al.** Designing equitable algorithms. Nat Comput Sci. 2023;3:601-610. [PMID: 38177749] doi:10.1038/s43588-023-00485-4
6. **Khor S, Haupt EC, Hahn EE, et al.** Racial and ethnic bias in risk prediction models for colorectal cancer recurrence when race and ethnicity are omitted as predictors. JAMA Netw Open. 2023;6:e2318495. [PMID: 37318804] doi:10.1001/jamanetworkopen.2023.18495
7. **Pandya A, Zhu J.** Focusing on decisions, outcomes, and value judgments to confront algorithmic bias. JAMA Netw Open. 2023;6:e2318501. [PMID: 37318809] doi:10.1001/jamanetworkopen.2023.18501
8. **Diao JA, He Y, Khazanchi R, et al.** Implications of race adjustment in lung-function equations. N Engl J Med. 2024;390:2083-2097. [PMID: 38767252] doi:10.1056/NEJMsa2311809
9. **Diao JA, Shi I, Murthy VL, et al.** Projected changes in statin and antihypertensive therapy eligibility with the AHA PREVENT cardiovascular risk equations. JAMA. 2024;332:989-1000. [PMID: 39073797] doi:10.1001/jama.2024.12537
10. **Zink A, Obermeyer Z, Pierson E.** Race adjustments in clinical algorithms can help correct for racial disparities in data quality. Proc Natl Acad Sci U S A. 2024;121:e2402267121. [PMID: 39136986] doi:10.1073/pnas.2402267121
11. **Lett E, Asabor E, Beltrán S, et al.** Conceptualizing, contextualizing, and operationalizing race in quantitative health sciences research. Ann Fam Med. 2022;20:157-163. [PMID: 35045967] doi:10.1370/afm.2792
12. **Vyas DA, Eisenstein LG, Jones DS.** Hidden in plain sight – reconsidering the use of race correction in clinical algorithms. N Engl J Med. 2020;383:874-882. [PMID: 32853499] doi:10.1056/NEJMms2004740
13. **Cerdeña JP, Plaisime MV, Tsai J.** From race-based to race-conscious medicine: how anti-racist uprisings call us to act. Lancet. 2020;396:1125-1128. [PMID: 33038972] doi:10.1016/S0140-6736(20)32076-6
14. **Basu A.** Use of race in clinical algorithms. Sci Adv. 2023;9:eadd2704. [PMID: 37235647] doi:10.1126/sciadv.add2704
15. **Inker LA, Eneanya ND, Coresh J, et al; Chronic Kidney Disease Epidemiology Collaboration.** New creatinine- and cystatin c-based equations to estimate GFR without race. N Engl J Med. 2021;385:1737-1749. [PMID: 34554658] doi:10.1056/NEJMoa2102953
16. **Powe NR.** Black kidney function matters: use or misuse of race? JAMA. 2020;324:737-738. [PMID: 32761164] doi:10.1001/jama.2020.13378
17. **Diao JA, Wu GJ, Taylor HA, et al.** Clinical implications of removing race from estimates of kidney function. JAMA. 2021;325:184-186. [PMID: 33263721] doi:10.1001/jama.2020.22124
18. **Eneanya ND, Yang W, Reese PP.** Reconsidering the consequences of using race to estimate kidney function. JAMA. 2019;322:113-114. [PMID: 31169890] doi:10.1001/jama.2019.5774
19. **Coots M, Linn KA, Goel S, et al.** Racial bias in clinical and population health algorithms: a critical review of current debates. Annu Rev Public Health. 2025. [Forthcoming].
20. **Khan SS, Coresh J, Pencina MJ, et al; American Heart Association.** Novel prediction equations for absolute risk assessment of total cardiovascular disease incorporating cardiovascular-kidney-metabolic health: a scientific statement from the American Heart Association. Circulation. 2023;148:1982-2004. [PMID: 37947094] doi:10.1161/CIR.0000000000001191
21. **Khan SS, Matsushita K, Sang Y, 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. [PMID: 37947085] doi:10.1161/CIRCULATIONAHA.123.067626
22. **Grobman WA, Sandoval G, Rice MM, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development**

Maternal-Fetal Medicine Units Network. Prediction of vaginal birth after cesarean delivery in term gestations: a calculator without race and ethnicity. Am J Obstet Gynecol. 2021;225:664.e1-664.e7. [PMID: 34043983] doi:10.1016/j.ajog.2021.05.021

23. **Shaikh N, Lee MC, Stokes LR, et al.** Reassessment of the role of race in calculating the risk for urinary tract infection: a systematic review and meta-analysis. JAMA Pediatr. 2022;176:569-575. [PMID: 35435935] doi:10.1001/jamapediatrics.2022.0700

24. **Centers for Disease Control and Prevention.** NHANES Questionnaires, Datasets, and Related Documentation. U.S. Department of Health and Human Services; 2023. Accessed at wwwn.cdc.gov/nchs/nhanes/Default.aspx on 5 November 2023.

25. **Aberle DR, Berg CD, Black WC, et al; National Lung Screening Trial Research Team.** The National Lung Screening Trial: overview and study design. Radiology. 2011;258:243-253. [PMID: 21045183] doi:10.1148/radiol.10091808

26. **Prosper AE, Inoue K, Brown K, et al.** Association of inclusion of more Black individuals in lung cancer screening with reduced mortality. JAMA Netw Open. 2021;4:e2119629. [PMID: 34427681] doi:10.1001/jamanetworkopen.2021.19629

27. **Tanner NT, Gebregziabher M, Hughes Halbert C, et al.** Racial differences in outcomes within the National Lung Screening Trial. Implications for widespread implementation. Am J Respir Crit Care Med. 2015;192:200-208. [PMID: 25928649] doi:10.1164/rccm.201502-0259OC

28. **U.S. Census Bureau.** National Population by Characteristics: 2020-2023. Updated 25 June 2024. Accessed at www.census.gov/data/tables/time-series/demo/popest/2020s-national-detail.html on 23 August 2024.

29. **Lloyd-Jones DM, Braun LT, Ndumele CE, et al.** Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: a special report from the American Heart Association and American College of Cardiology. J Am Coll Cardiol. 2019;73:3153-3167. [PMID: 30423392] doi:10.1016/j.jacc.2018.11.005

30. **Goff DC, Lloyd-Jones DM, Bennett G, et al.** 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. J Am Coll Cardiol. 2014;63:2935-2959. [PMID: 24239921] doi:10.1016/j.jacc.2013.11.005

31. **Grundy SM, Stone NJ, Bailey AL, et al.** 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol. 2019;73:e285-e350. [PMID: 30423393] doi:10.1016/j.jacc.2018.11.003

32. **Banegas MP, John EM, Slattery ML, et al.** Projecting individualized absolute invasive breast cancer risk in US Hispanic women. J Natl Cancer Inst. 2017;109 [PMID: 28003316] doi:10.1093/jnci/djw215

33. **Matsuno RK, Costantino JP, Ziegler RG, et al.** Projecting individualized absolute invasive breast cancer risk in Asian and Pacific Islander American women. J Natl Cancer Inst. 2011;103:951-961. [PMID: 21562243] doi:10.1093/jnci/djr154

34. **Gail MH, Costantino JP, Pee D, et al.** Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst. 2007;99:1782-1792. [PMID: 18042936] doi:10.1093/jnci/djm223

35. **Costantino JP, Gail MH, Pee D, et al.** Validation studies for models projecting the risk of invasive and total breast cancer incidence. J Natl Cancer Inst. 1999;91:1541-1548. [PMID: 10491430] doi:10.1093/jnci/91.18.1541

36. **Gail MH, Brinton LA, Byar DP, et al.** Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst. 1989;81:1879-1886. [PMID: 2593165] doi:10.1093/jnci/81.24.1879

37. **Zhang F.** BCRA: Breast Cancer Risk Assessment [R package version 2.1.2]. CRAN. 7 October 2020. Accessed at https://CRAN.R-project.org/package=BCRA on 1 April 2023.

38. **Durham DD, Abraham LA, Roberts MC, et al.** Breast cancer incidence among women with a family history of breast cancer by relative's age at diagnosis. Cancer. 2022;128:4232-4240. [PMID: 36262035] doi:10.1002/cncr.34365

39. **Huilgol YS, Keane H, Shieh Y, et al; Athena Breast Health Network Investigators and Advocate Partners.** Elevated risk thresholds predict endocrine risk-reducing medication use in the Athena screening registry. NPJ Breast Cancer. 2021;7:102. [PMID: 34344894] doi:10.1038/s41523-021-00306-9

40. **Waters EA, McNeel TS, McCaskill Stevens W, et al.** Use of tamoxifen and raloxifene for breast cancer chemoprevention in 2010. Breast Cancer Res Treat. 2012;134:875-880. [PMID: 22622807] doi:10.1007/s10549-012-2089-2

41. **Katki HA, Kovalchik SA, Berg CD, et al.** Development and validation of risk models to select ever-smokers for CT lung cancer screening. JAMA. 2016;315:2300-2311. [PMID: 27179989] doi:10.1001/jama.2016.6255

42. **Cheung LC, Kovalchik SA, Katki HA.** Lung cancer risk models for screening (R package: lcrisks). 2023. Accessed at https://dceg.cancer.gov/tools/risk-assessment/lcrisks on 2 November 2023.

43. **Katki HA, Kovalchik SA, Petito LC, et al.** Implications of nine risk prediction models for selecting ever-smokers for computed tomography lung cancer screening. Ann Intern Med. 2018;169:10-19. [PMID: 29800127] doi:10.7326/M17-2701

44. **Vickers AJ, Elkin EB.** Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26:565-574. [PMID: 17099194] doi:10.1177/0272989X06295361

45. **Peirce CS.** The numerical measure of the success of predictions. Science. 1884;4:453-454. [PMID: 17795531] doi:10.1126/science.ns-4.93.453-a

46. **Pauker SG, Kassirer JP.** The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109-1117. [PMID: 7366635] doi:10.1056/NEJM198005153022003

47. **Aggarwal R, Bibbins-Domingo K, Yeh RW, et al.** Diabetes screening by race and ethnicity in the United States: equivalent body mass index and age thresholds. Ann Intern Med. 2022;175:765-773. [PMID: 35533384] doi:10.7326/M20-8079

48. **Thomas TW, Golin C, Samuel-Hodge CD, et al.** Race and gender differences in abnormal blood glucose screening and clinician response to prediabetes: a mixed-methods assessment. Prev Med. 2021;148:106587. [PMID: 33930437] doi:10.1016/j.ypmed.2021.106587

49. **Paulus JK, Wessler BS, Lundquist CM, et al.** Effects of race are rarely included in clinical prediction models for cardiovascular disease. J Gen Intern Med. 2018;33:1429-1430. [PMID: 29766380] doi:10.1007/s11606-018-4475-x

50. **Khazanchi R, Marcelin J, Abdul-Mutakabbir J, et al.** Race, racism, civil rights law, and the equitable allocation of scarce COVID-19 treatments. Health Affairs Forefront. 10 February 2022. doi:10.1377/forefront.20220208.453850

51. **Kent DM, Ladin K, Duru OK.** Equal treatment for equal risk: should race be included in allocation algorithms for Covid-19 therapies? STAT. 4 April 2022. Accessed at www.statnews.com/2022/04/04/should-race-be-included-allocation-algorithms-covid-19-therapies on 26 August 2024.

52. **Paulus JK, Kent DM.** Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. NPJ Digit Med. 2020;3:99. [PMID: 32821854] doi:10.1038/s41746-020-0304-9

53. **Ladin K, Cuddeback J, Duru OK, et al.** Guidance for unbiased predictive information for healthcare decision-making and equity (GUIDE): considerations when race may be a prognostic factor. NPJ

Digit Med. 2024;7:290. [PMID: 39427028] doi:10.1038/s41746-024-01245-y
54. **Muntner P, Colantonio LD, Cushman M, et al.** Validation of the atherosclerotic cardiovascular disease pooled cohort risk equations. JAMA. 2014;311:1406-1415. [PMID: 24682252] doi:10.1001/jama.2014.2630
55. **Zanger-Tishler M, Nyarko J, Goel S.** Risk scores, label bias, and everything but the kitchen sink. Sci Adv. 2024;10:eadi8411. [PMID: 38552013] doi:10.1126/sciadv.adi8411
56. **Boyd RW, Lindo EG, Weeks LD, et al.** On racism: a new standard for publishing on racial health inequities. Health Affairs Blog. 2 July 2020. doi:10.1377/hblog20200630.939347
57. **Jones CP.** Levels of racism: a theoretic framework and a gardener's tale. Am J Public Health. 2000;90:1212-1215. [PMID: 10936998] doi:10.2105/ajph.90.8.1212
58. **Giaquinto AN, Sung H, Miller KD, et al.** Breast cancer statistics, 2022. CA Cancer J Clin. 2022;72:524-541. [PMID: 36190501] doi:10.3322/caac.21754

Author Contributions: Conception and design: M. Coots, S. Goel, S. Saghafian.
Analysis and interpretation of the data: M. Coots, S. Goel, D.M. Kent, S. Saghafian.
Drafting of the article: M. Coots, S. Goel, D.M. Kent, S. Saghafian.
Critical revision for important intellectual content: M. Coots, S. Goel, D.M. Kent, S. Saghafian.
Final approval of the article: M. Coots, S. Goel, D.M. Kent, S. Saghafian.
Statistical expertise: M. Coots, S. Goel, S. Saghafian.
Obtaining of funding: D.M. Kent.
Collection and assembly of data: M. Coots.

Appendix: Data, Risk Predictions, and Utility Analyses

Data Samples for Each Disease

Cardiovascular Disease

For cardiovascular disease, we restricted our data sample from NHANES to nonpregnant adults between the ages of 40 and 79 years who had never taken statins or had any of the following cardiovascular events: congestive heart failure, coronary heart disease, angina, heart attack, or stroke. We further restricted our sample to participants with biomarkers in the appropriate range for use of the 2013 pooled cohort equations (29, 30): high-density lipoprotein cholesterol level greater than or equal to 0.52 mmol/L (20 mg/dL) and less than or equal to 2.59 mmol/L (100 mg/dL), total cholesterol level greater than or equal to 3.37 mmol/L (130 mg/dL) and less than or equal to 8.28 mmol/L (320 mg/dL), and systolic blood pressure greater than or equal to 90 mm Hg and less than or equal to 200 mm Hg.

Breast Cancer

For breast cancer, we restricted our data sample from NHANES to women aged 35 years or older.

Lung Cancer

For lung cancer, we used the full data sample from NLST.

Risk Models for Each Disease

Cardiovascular Disease

For cardiovascular disease, the 2018 Cholesterol Clinical Practice Guidelines and the 2017 Hypertension Clinical Practice Guidelines recommend using the U.S.-derived, race- and sex-specific pooled cohort equations from 2013 to estimate 10-year risk for atherosclerotic cardiovascular disease events (29). The covariates used by the atherosclerotic cardiovascular disease pooled cohort equations vary by race and ethnicity and sex because there are 4 separate equations based on sex (male or female) and race and ethnicity (non-Hispanic Black or non-Hispanic White). According to the 2013 guidelines from the American Heart Association, the equations for non-Hispanic White persons may be used to estimate the risk for persons of other ethnicities (30). The American Heart Association guidelines recommend that a risk threshold of 7.5% be used to identify persons who would benefit from starting a moderate-intensity statin therapy (31). This is the decision threshold we considered for the cardiovascular disease analysis, above which the benefits of statin therapy are generally considered to outweigh its potential harms, burdens, and cost.

Breast Cancer

The National Cancer Institute maintains an online Breast Cancer Risk Assessment Tool that estimates an individual's 5-year risk based on the Gail model, which we likewise use to compute breast cancer risk estimates (32–36). To do so, we used the BCRA R package (version 2.1.2) published by the National Cancer Institute (37). The U.S. Food and Drug Administration has approved using a 1.67% threshold on 5-year risk as determined by the Breast Cancer Risk Assessment Tool for prescribing tamoxifen and raloxifene as chemoprevention for breast cancer (39, 40). This is the decision threshold we considered for our breast cancer analysis.

NHANES does not contain information on the number of first-degree relatives with breast cancer—one of the inputs of the Gail model. To account for this gap, for each individual in our data set we randomly generated a value for the number of their first-degree relatives with breast cancer, based on national race- and ethnicity-specific statistics (38). Specifically, for each individual in our data sample, we sampled a binary value as an approximation of the number of first-degree relatives with breast cancer using a Bernoulli distribution parameterized by race- and ethnicity-specific probabilities of having a first-degree relative with breast cancer for women in the United States:

$$\textit{Number of relatives} \sim \textit{Bernoulli}\,(p = p_{race})$$

where

$$p_{White} = \frac{21\,433 + 6582}{235\,629} \approx 0.12$$

$$p_{Black} = \frac{1384 + 1128}{27\,179} \approx 0.09$$

$$p_{Asian} = \frac{514 + 377}{11\,780} \approx 0.08$$

and

$$p_{Hispanic} = \frac{546 + 256}{9049} \approx 0.09$$

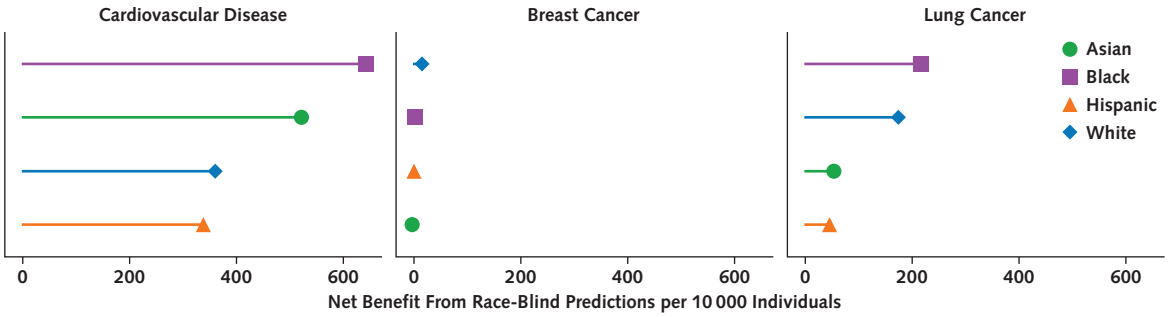
These rates were taken from prior work on breast cancer incidence (**Table 1** in Durham and colleagues [38]).

NHANES also does not contain information on the history of breast biopsy or atypical hyperplasia. Therefore, histories of breast biopsy and atypical hyperplasia were inputted as unknown values into the risk model.

Lung Cancer

For lung cancer, we computed 5-year risk predictions from the National Cancer Institute's Lung Cancer Risk Assessment Tool model (41). To compute risk estimates, we used the Lung Cancer Risk Models for

Appendix Figure 1. Baseline utility results per disease, by race and ethnicity.



Screening (lcrisks) R package (version 4.1.1) published by the National Cancer Institute (42). Past work has assessed and recommended a risk threshold of approximately 2.0% for selecting ever-smokers for computed tomography lung cancer screening (41, 43), which is likewise the decision threshold we considered for our lung cancer analysis.

The full set of covariates used by the Lung Cancer Risk Assessment Tool model is as follows: year of assessment; age; gender; number of years smoked; number of years quit; number of cigarettes per day; race and ethnicity; whether the individual has any health problems requiring special equipment; whether the individual has chronic obstructive pulmonary disease or emphysema; number of parents with lung cancer; body mass index; highest education level attained; history of cancer; indicator variables for hypertension, coronary heart disease, angina pectoris, heart attack, other heart disease, stroke, and diabetes; and indicator variables for whether in the past year the patient has had chronic bronchitis, weak or failing kidneys, or a liver condition (42). The NLST data do not contain information on other heart disease, kidney issues, liver issues, angina, or the presence of conditions that require special medical equipment; therefore, these were inputted as unknown values into the risk model.

Survey sample weights were not provided in the NLST data set, so we reweighted participants to match the joint age, gender, and race distribution of Americans between ages 40 and 80 years, mirroring the age range of participants in the NLST data. To generate these weights, we used data on the national population by characteristics (2020 to 2023) provided by the U.S. Census Bureau (28). Specifically, we used the projected monthly population estimate by age, sex, race, and Hispanic origin for June 2024. The weight for an individual of age $A=a$, sex $S=s$, and race and ethnicity $R=r$ was computed as follows:

$$Weight = \frac{U.S.\,population\,proportion}{NLST\,proportion}$$

where

$$U.S.\,population\,proportion = \frac{\sum_{i=1}^{N_{Census}} I(A_i = a, S_i = s, R_i = r)}{N_{Census}}$$

and

$$NLST\,proportion = \frac{\sum_{i=1}^{N_{NLST}} I(A_i = a, S_i = s, R_i = r)}{N_{NLST}}$$

and $I(\cdot)$ denotes the indicator function.

Estimating Race-Unaware Risk

To obtain race-unaware estimates of risk for each individual, we invoke the law of total probability as follows:

$$P(Disease|X) = \sum_r P(Disease\,|X, R = r)\ \cdot P(R = r\,|X)$$

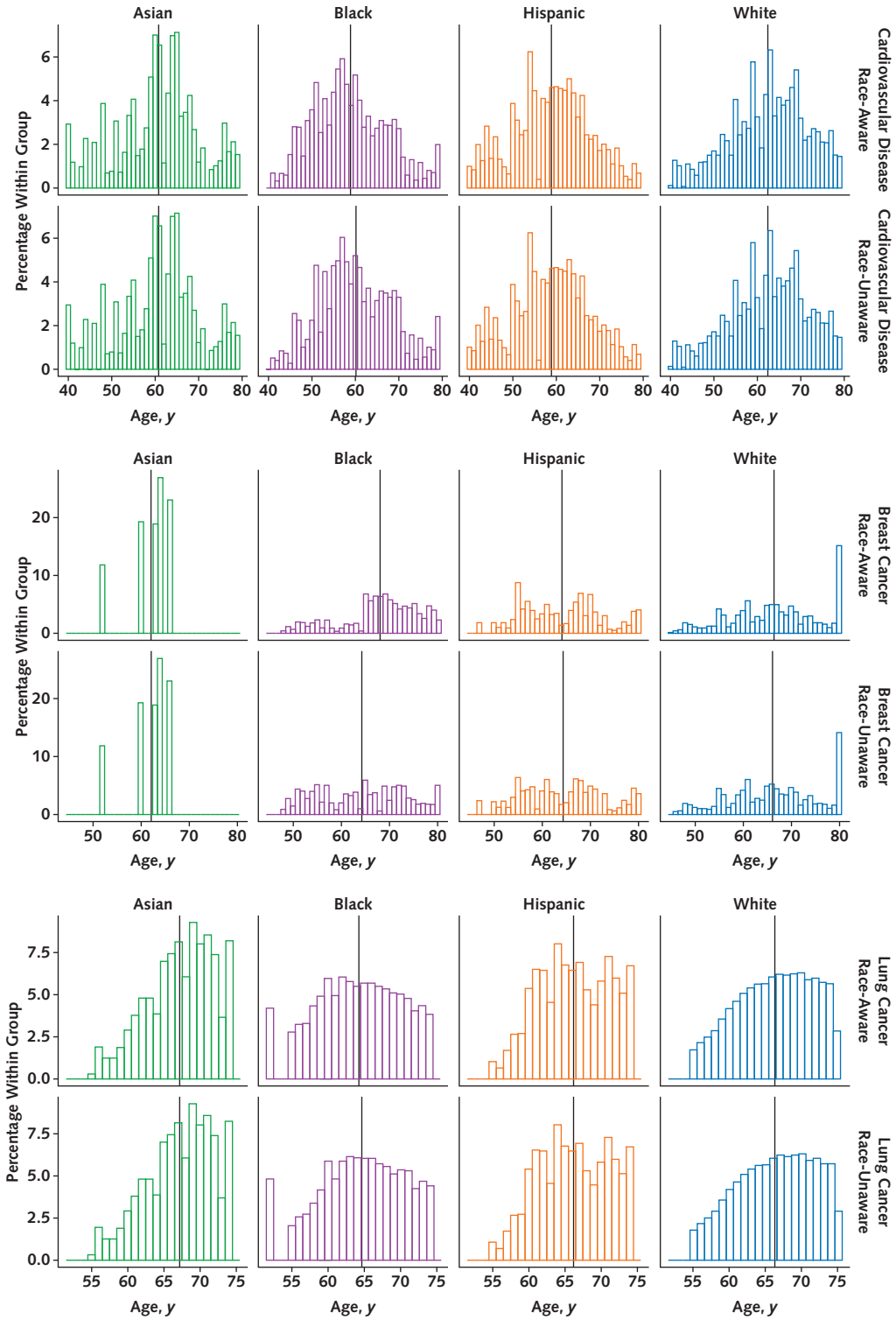
where X denotes the set of nonrace covariates used to estimate risk for a given disease and R denotes race and ethnicity. We estimate $P(R=r|X)$ using a multinomial regression model that predicts race using the covariates used in the risk model for each disease. The right-hand side of the equation is equal to the weighted sum of the risk estimates obtained from the race-aware risk model. The above equation then amounts to marginalizing out race and ethnicity from the risk prediction.

We note that, in practice, if one wanted to generate race-unaware risk models for any of the diseases considered, the preferred way to do so would be to train a new model without using race and ethnicity as features, perhaps also including social and biological determinants correlated with race and ethnicity that might improve the performance of race-unaware predictions.

Deriving the Value of Screening or Treatment Cost

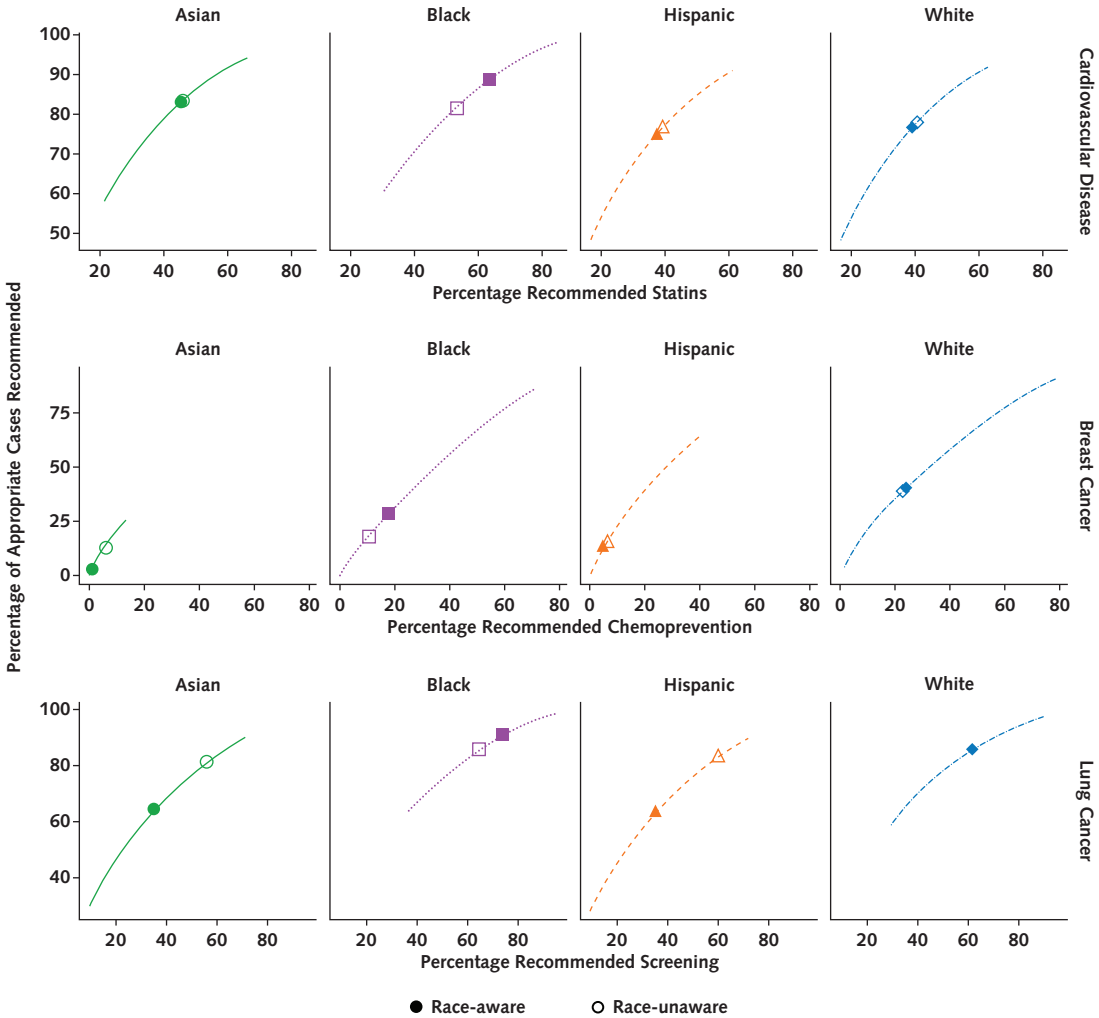
For each disease, we assume that the decision threshold is set at the point of indifference–that is, where the expected benefits of an intervention (either screening or treatment) equal the expected costs. Based on the tree structure depicted in **Figure 1**, we use this fact to derive

Appendix Figure 2. Subgroup analysis of the age distribution of individuals appropriately recommended for screening or treatment under race-aware and race-unaware models.



For each disease, we show the age distribution of individuals who are appropriately recommended for screening or treatment by a race-aware model and by a race-unaware model. The group-level mean age is shown by the black vertical line. For each subgroup within each disease, the race-aware and race-unaware models recommend similar groups of individuals for screening or treatment. These results suggest that an age-based utility function (e.g., based on quality-adjusted life-years) would yield results similar to those in our base-case analysis, where we assumed uniform benefit across individuals.

Appendix Figure 3. Sensitivity analysis for when the utility benefit varies across groups.



Across racial and ethnic subgroups for cardiovascular disease, breast cancer, and lung cancer, the frontier of the fraction of appropriate cases that are recommended for intervention (sensitivity or true-positive rate) as a function of the fraction of the subgroup population that is recommended for intervention. Each plot represents a subgroup and a specific disease, showing the relationship between the fraction of the population recommended for intervention (*x*-axis) and the actual fraction of appropriate cases identified (*y*-axis). Solid points indicate the tradeoff obtained using a race-aware model with the recommended decision threshold, and hollow points show the tradeoff obtained using a race-unaware model with the recommended decision threshold.

the value of the intervention cost c in terms of the threshold t. In particular, for an individual on the threshold (that is, with $P(Disease) = t$) we have:

$$\begin{aligned} 0 &= E[Intervention\ Utility] \\ &= P(Disease) \cdot (1 - c) - \Big(1 - P(Disease)\Big) \cdot c \\ &= t \cdot (1 - c) - (1 - t) \cdot c \\ &= t - tc - c + tc \\ &= t - c \end{aligned}$$

As a result, $c = t$.

Heterogeneous Utility

We start by considering age-related heterogeneity in utility. To do so, we examine the age of persons recommended for screening or treatment by race-aware versus race-unaware models. For all 3 diseases we study, we find that the average age of persons appropriately recommended for intervention is nearly identical between the race-aware and race-unaware models. Further, not just the means but the full distributions of age are likewise nearly identical across the race-aware and race-unaware models, as shown in Appendix Figure 2. As a result, a heterogeneous form of the utility function based solely on age (for example, one based on quality-adjusted life-years) would produce qualitatively similar results to what we find in the base case.

In using statistical risk models to inform intervention decisions, there may be group-specific tradeoffs between the costs and benefits of intervention–a tradeoff that we shed greater light on in this analysis. To do so, we trace out the frontier of the fraction of appropriate cases that are recommended for intervention (sensitivity or true-positive rate) as a function of the fraction of the subgroup population that is recommended for intervention. For each

Appendix Table. Verification of Optimal Race-Aware and Race-Unaware Risk Thresholds*

Disease	Recommended Threshold, %	Optimized Race-Aware Threshold, %	Optimized Race-Unaware Threshold, %
Cardiovascular disease	7.5	7.51	7.49
Breast cancer	1.67	1.66	1.66
Lung cancer	2.0	2.0	2.1

* Optimal thresholds under race-aware and race-unaware models are approximately equal.

disease and race group, we do this by starting with the race-aware risk model and then, for each (race-specific) decision threshold, plotting the resulting intervention rate and true-positive rate obtained with that threshold. We show the results in Appendix Figure 3, where each point on the curve is an outcome that is theoretically achievable with a race-aware decision tool.

The solid point in each subplot in Appendix Figure 3 corresponds to the tradeoff between intervention and detection using the recommended decision threshold with a race-aware model. The hollow point in each subplot corresponds to the tradeoff using the recommended decision threshold with a race-unaware model. For several subgroups and diseases, the race-aware and race-unaware models yield different tradeoffs. If one believes there are group-specific differences in the costs and benefits of appropriate intervention, then a policymaker may want to recommend intervention for a particular subgroup at higher rates than for the general population to obtain a higher fraction of cases that are detected or appropriately treated within that subgroup. In such a scenario, the relative utility of a race-aware risk assessment tool over a race-unaware tool could be greater than suggested by our base-case analysis.

Optimal Screening Thresholds

We verified that the optimal threshold under a race-aware model is approximately equal to the optimal threshold under a race-unaware model. The results of this analysis are shown in the Appendix Table. To verify optimal threshold values, we computed the total population utility (under either a race-unaware model or a race-aware model) as a function of threshold values t in the range (0,1). We then determined the value of t that maximized the total population utility under each model.