Annals of Internal Medicine:
Research and Reporting Methods

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[†]
Sharad Goel, PhD, MS[*]

[*] Harvard University, Cambridge, Massachusetts, USA
[†] Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies (ICRHPS), Tufts Medical Center/Tufts University School of Medicine, Boston, Massachusetts, USA

Running title: Race and Ethnicity in Estimating Disease Risk

Abstract word count: 283

Word count: 3497

References: 58

Tables and figures: 5

ABSTRACT

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: National Health and Nutrition Examination Survey, 2011-2018; National Lung Screening Trial, 2002-2004.

Patients (or Participants): U.S. adults

Measurements: Starting with risk predictions from clinically recommended race-aware models, we generated race-unaware predictions via statistical marginalization. We 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. 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 two empirical patterns: first, across all three diseases, 95% or more of individuals would receive the same decision regardless of whether race and ethnicity are included in risk models; and second, for those who receive different decisions, the net benefit of screening or treatment is relatively small, since these patients have disease risks close to the decision threshold (i.e., the theoretical "point of indifference"). When used to inform rationing, the net benefit of race-aware models may be more substantial.

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

Conclusion: Our 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: Dr. Kent is funded through a "Making a Difference" and Presidential Awards from The Greenwall Foundation.

INTRODUCTION

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 individuals who would benefit from interventions such as screening or prophylactic treatment—such as pharmacotherapy or lifestyle changes—that would allow individuals 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 with a predicted risk of developing the disease above a certain threshold. These thresholds are typically set at the level of risk above which the expected benefits exceed the expected harms and costs, i.e., above the "point of indifference".

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 healthcare (11). However, there is persistent concern and criticism over the use of race and ethnicity in estimating disease risk (12-14). Including race and ethnicity in predictive models may, for instance, inadvertently reinforce pernicious attitudes of biological determinism or lead to greater stigmatization of individuals who are already marginalized. In part for these reasons, race-aware glomerular filtration rate estimates 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 of CVD events (PREVENT), 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 the 2011-2018 National Health and Nutrition Examination Survey (NHANES) (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 (see Appendix for details).

Our analysis of lung cancer is based on cross-sectional data from the National Lung Screening Trial (NLST) (25), a randomized control trial conducted between August 2002 and April 2004 to assess whether low-dose CT screening reduces lung cancer mortality relative to chest radiography among high-risk individuals. 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, though

these data have been used to investigate racial and ethnic disparities in lung cancer (26, 27). We reweighted individuals to match the joint age, gender, and race distribution of Americans between 40 and 80 years old (28) (see Appendix for details).

RISK PREDICTIONS

Table 1 describes the risk models we used, including covariates, risk thresholds, and the clinical decisions the models inform (see Appendix for details).

Obtaining race-unaware risk estimates

The above 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 add other factors correlated with race and ethnicity that might improve the performance of race-unaware predictions (20, 21). However, due to data limitations, we were unable to 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 non-race risk factors. For example, to obtain the race-unaware lung cancer risk estimate for an individual, we used the LCRAT model to first produce four risk estimates, varying only race (i.e., White, Black, Hispanic, or Asian) and holding all else constant. We then took a weighted average of these four race-aware predictions to obtain a race-unaware risk estimate, a well-established statistical technique for removing variables from risk models (see Appendix). These race-unaware models are intended only to illustrate broad statistical principles and 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 (i.e., screening/treatment) for those deemed high risk. This cost encapsulates a wide range of monetary and non-monetary considerations, e.g., direct cost of screening or treatment and indirect costs such as taking time 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 the benefit of appropriate intervention (i.e., 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.

Figure 1 depicts the structure of the utility function. We set the utility of "no intervention" to 0, where individuals neither incur 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, since at this level of risk, the benefits of therapy are nullified by the costs, burdens, and harms of therapy.

Akin to 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 (see Appendix for details). 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.

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 row). Assuming the original, race-aware models yield accurate risk estimates, we find the race-unaware models underestimate risk of cardiovascular disease and lung cancer for Black individuals. In contrast, race-unaware models overestimate risk of breast and lung cancer for Asian individuals, and similarly overestimate risk of lung cancer for Hispanic individuals. For White individuals, the predicted risks were similar between the race-aware and race-unaware models for all three diseases. Miscalibrated predictions can result

in misclassifications that lead to inappropriately recommending screening/treatment in low risk individuals or failing to recommend screening/treatment for high-risk individuals. 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 fit 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, surprisingly, 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 (i.e., 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 Figure A1 in the Appendix for baseline utility results per disease by race and ethnicity.) For each disease, the subgroups that experience 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 individuals receive the same recommendation under a race-aware model as under a race-unaware model—as recommendations only change for the relatively few individuals close to the decision threshold. The fraction of individuals who receive

the same recommendation under both models is 98% for cardiovascular disease, 97% for breast cancer, and 95% for lung cancer. The vast majority of individuals 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 two 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 or not receiving the intervention—precisely since 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 for older people is the same as for younger people, even though treatment for younger people could lead to more life years gained. In theory, our race-unaware models could systematically under identify individuals who would benefit the most from treatment. However, in the appendix, we consider age-related heterogeneity in utility (Figure A2) 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—a tradeoff that can be better accommodated by race-aware decision-making. In the Appendix, we trace out these group-specific tradeoffs for each disease (Figure A3). If one believes there are sizable group-specific differences 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 healthcare 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 (e.g., 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 (e.g., pharmacotherapy for cardiovascular disease and breast cancer, and a CT 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 de-prioritize lower-risk Hispanic, Asian, and White patients for intervention. For example, if we imagine that only individuals with a CVD 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 further from their point of indifference and thus experience larger utility gains from reclassification. Additionally, given the distributions of risk in this example, more individuals receive different screening recommendations under 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 two patterns in the data: first, while the more accurate race-aware model changes predictions for all patients, decisions often change for only a small fraction of individuals; second, among those who do receive different decisions, the net value of intervention is relatively small since 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.

There are, however, circumstances 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. While 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 healthcare and prediction models are increasingly proposed to allocate resources. For example, during the COVID 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 three 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 demonstrates a need to additionally examine the gains in net utility from changed decisions. Lastly, our paper adds to previous work highlighting the important—and often overlooked—ethical distinctions in shared decision-making versus rationing, since 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 three 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 suffer from systematic inaccuracies (54), though 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 apparent disparity in disease risks is suspected to arise from label bias—e.g., 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 two or more thresholds might be appropriate for identifying low, medium, and high risk groups, 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). Additionally, there is also concern 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 U.S. have lower incidence rates of breast cancer compared to 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 a variety of contexts in medicine and beyond where using race and ethnicity in predictive models is contested. However, when used to inform rationing, the benefits of race-aware models may be much more substantial 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. But 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 trade-offs of using race and ethnicity when estimating risk.

ACKNOWLEDGEMENTS

Dr. Kent is funded through a “Making a Difference” and Presidential Awards from The Greenwall Foundation. Funders did not play a role in the design, conduct, or analysis of this study, nor in the decision to submit the manuscript for publication. The authors additionally thank the National Cancer Institute for access to NCI's data collected by the National Lung Screening Trial (NLST) (CDAS Project Number NLST-1121) The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

FINANCIAL SUPPORT

This study was partially supported by a "Making a Difference" and Presidential Awards from The Greenwall Foundation.

REPRODUCIBLE RESEARCH STATEMENT

Study protocol: Not available.

Statistical code: Available at https://github.com/madisoncoots/race-in-estimating-disease-risk

Data set: NHANES data is publicly available from the CDC. Processed versions of the NHANES datasets used for analysis are additionally included in the replication package posted at https://github.com/madisoncoots/race-in-estimating-disease-risk. NLST data is available by request from the National Cancer Institute.

CORRESPONDING AUTHOR

Madison Coots

John F. Kennedy School of Government, Harvard University

79 JFK St., Cambridge, MA 02138

mcoots@g.harvard.edu

REFERENCES

1. Chen T, Kharazmi E, Fallah M. Race and Ethnicity–Adjusted Age Recommendation for Initiating Breast Cancer Screening. JAMA Netw Open. 2023 Apr;6(4):e238893. 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 Sep;6(9):e2331155. doi: 10.1001/jamanetworkopen.2023.31155.

3. Manski CF. Patient‐centered appraisal of race‐free clinical risk assessment. Health Economics. 2022 Oct;31(10):2109-14.

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(35):1-e2303370120. 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. 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 Network Open. 2023 Jun 1;6(6):e2318495-.

7. Pandya A, Zhu J. Focusing on Decisions, Outcomes, and Value Judgments to Confront Algorithmic Bias. JAMA Network Open. 2023 Jun 1;6(6):e2318501.

8. Diao JA, He Y, Khazanchi R, et al. Implications of Race Adjustment in Lung-Function Equations. N Engl J Med. 2024 Jun 13;390(22):2083-2097. doi: 10.1056/NEJMsa2311809. Epub 2024 May 19. PMID: 38767252; PMCID: PMC11305821.

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. Published online July 29, 2024. doi:10.1001/jama.2024.12537

10. Zink, Anna, Ziad Obermeyer, and Emma Pierson. "Race adjustments in clinical algorithms can help correct for racial disparities in data quality." Proceedings of the National Academy of Sciences 121.34 (2024): e2402267121.

11. Lett E, Asabor E, Beltrán S, et al. Conceptualizing, Contextualizing, and Operationalizing Race in Quantitative Health Sciences Research. Ann Fam Med. 2022

Mar-Apr;20(2):157-163. doi: 10.1370/afm.2792. Epub 2022 Jan 19. PMID: 35045967; PMCID: PMC8959750.

12. Vyas DA, Eisenstein LG, Jones DS. Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms. The New England Journal of Medicine. 2020;383(9):874-82.

13. Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. Lancet (London, England). 2020;396(10257):1125-8.

14. Basu A. Use of race in clinical algorithms. Science Advances. 2023 May 26;9(21):eadd2704.

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(19):1737-1749. doi:10.1056/NEJMoa2102953.

16. Powe NR. Black Kidney Function Matters: Use or Misuse of Race? JAMA. 2020;324(8):737-8.

17. Diao JA, Wu GJ, Taylor HA, et al. Clinical Implications of Removing Race From Estimates of Kidney Function. JAMA. 2021;325(2):184-6.

18. Eneanya ND, Yang W, Reese PP. Reconsidering the Consequences of Using Race to Estimate Kidney Function. JAMA. 2019;322(2):113-4.

19. Coots M, Linn KA, Goel S, et al. Racial Bias in Clinical and Population Health Algorithms: A Critical Review of Current Debates. Annual Review of Public Health. In press.

20. Khan SS, Coresh J, Pencina MJ, et al. 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. doi: 10.1161/CIR.0000000000001191.

21. Khan SS, Matsushita K, Sang Y, et al. Development and Validation of the American Heart Association Predicting Risk of Cardiovascular Disease EVENTs (PREVENTTM) Equations. Circulation 2023. DOI: 10.1161/CIRCULATIONAHA.123.067626(link opens in new window).

22. Grobman WA, Sandoval G, Rice MM, et al. Prediction of vaginal birth after cesarean delivery in term gestations: a calculator without race and ethnicity. Am J Obstet Gynecol.

2021 Dec;225(6):664.e1-664.e7. doi: 10.1016/j.ajog.2021.05.021. Epub 2021 May 24. PMID: 34043983; PMCID: PMC8611105.

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(6):569–575. doi:10.1001/jamapediatrics.2022.0700

24. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2023. https://www.cdc.gov/nchs/nhanes/index.html.

25. National Lung Screening Trial Research Team; Aberle DR, Berg CD, Black WC, et al. The National Lung Screening Trial: overview and study design. Radiology. 2011 Jan;258(1):243-53. doi: 10.1148/radiol.10091808. Epub 2010 Nov 2. PMID: 21045183; PMCID: PMC3009383.

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(8):e2119629. 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 Jul 15;192(2):200-8. doi: 10.1164/rccm.201502-0259OC. PMID: 25928649.

28. U.S. Census Bureau. National Population by Characteristics: 2020-2023. Accessed August 23, 2024. Available from: https://www.census.gov/data/tables/time-series/demo/popest/2020s-national-detail.html

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(24):3153-67. doi: 10.1016/j.jacc.2018.11.005.

30. Goff DC Jr, 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(25 Pt B):2935-2959. 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(24):e285-e350. 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. JNCI 2016; 109.

33. Matsuno RK, Costantino JP, Ziegler RG, et al. Projecting individualized absolute invasive breast cancer risk in asian and pacific islander american women. JNCI 103(12):951-61, 2011.

34. Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. JNCI 99(23):1782-92, 2007.

35. Costantino J, Gail MH, Pee D, et al. Validation studies for models to project the risk of invasive and total breast cancer. JNCI 91(18):1541-48, 1999.

36. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. JNCI 81(24): 1879-86, 1989.

37. Zhang F. BCRA: Breast Cancer Risk Assessment [R package version 2.1.2]. CRAN. 2020. Available from https://CRAN.R-project.org/package=BCRA

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 Dec 15;128(24):4232-4240. doi: 10.1002/cncr.34365. Epub 2022 Oct 19. PMID: 36262035; PMCID: PMC9712500.

39. Huilgol, Y.S., Keane, H., Shieh, Y. et al. Elevated risk thresholds predict endocrine risk-reducing medication use in the Athena screening registry. npj Breast Cancer 7, 102 (2021). https://doi.org/10.1038/s41523-021-00306-9

40. Waters, E. A., McNeel, T. S., McCaskill Stevens, W. & Freedman, A. N. Use of tamoxifen and raloxifene for breast cancer chemoprevention in 2010. Breast Cancer Res. Treat. 134, 875–880 (2012).

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(21):2300–2311. doi:10.1001/jama.2016.6255

42. Cheung, LC, Kovalchik, SA, Katki, HA. lcrisks: Lung Cancer Death Risk Predictor [R package version 4.1.1]. CRAN. 2023. Available from https://CRAN.R-project.org/package=lcrisks

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 Jul 3;169(1):10-19. doi: 10.7326/M17-2701. Epub 2018 May 15. PMID: 29800127; PMCID: PMC6557386.

44. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006 Nov-Dec;26(6):565-74. doi: 10.1177/0272989X06295361. PMID: 17099194; PMCID: PMC2577036.

45. Peirce CS. The numerical measure of the success of predictions. Science. 1884 Nov 14;4(93):453-4. doi: 10.1126/science.ns-4.93.453-a. PMID: 17795531.

46. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980 May 15;302(20):1109-17. doi: 10.1056/NEJM198005153022003. PMID: 7366635.

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. Annals of Internal Medicine. 2022 Jun;175(6):765-73.

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. Preventive medicine. 2021 Jul 1;148:106587.

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 Sep;33(9):1429-1430. doi: 10.1007/s11606-018-4475-x. PMID: 29766380; PMCID: PMC6109012.

50. "Race, Racism, Civil Rights Law, And The Equitable Allocation Of Scarce COVID-19 Treatments", Health Affairs Forefront, February 10, 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 News. April 4, 2022. Accessed August 26, 2024. https://www.statnews.com/2022/04/04/should-race-be-included-allocation-algorithms-covid-19-therapies/

52. Paulus JK, Kent DM. Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. npj Digit. Med. 3, 99 (2020). https://doi.org/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) Statement: Bias and Fairness in Clinical Prediction Models—Considerations when Race May Be a Prognostic Factor. Under revision at Nature Digital Medicine. 2024

54. Muntner P, Colantonio LD, Cushman M, et al. 2014. "Validation of the Atherosclerotic Cardiovascular Disease Pooled Cohort Risk Equations." JAMA 311 (14): 1406–15. https://doi.org/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 Mar 29;10(13):eadi8411. doi: 10.1126/sciadv.adi8411. Epub 2024 Mar 29. PMID: 38552013; PMCID: PMC10980258.

56. Boyd RW, Lindo EG, Weeks LD, McLemore MR. On Racism: A New Standard For Publishing on Racial Health Inequities. Health Affairs Blog, July 2, 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 Aug;90(8):1212-5. doi: 10.2105/ajph.90.8.1212. PMID: 10936998; PMCID: PMC1446334.

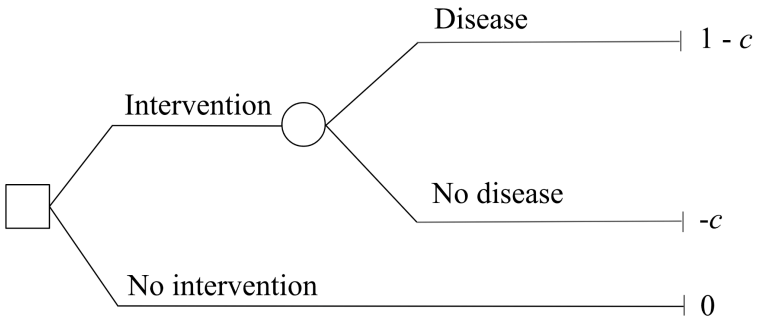
58. Giaquinto, AN, Sung, H, Miller, KD, et al. (2022), Breast Cancer Statistics, 2022. CA A Cancer J Clin, 72: 524-541. https://doi.org/10.3322/caac.21754

Table 1. Risk models, inputs, thresholds, and decisions for each disease.

Disease	**Risk Model**	**Model Inputs**	**Risk Threshold**	**Decision Considered**
Cardiovascular disease	US-derived 2013 ASCVD pooled cohort equations (PCE) (29, 30)	Sex, race and ethnicity, age, diabetes status, smoker status, untreated and treated systolic blood pressure, total cholesterol, high-density lipoprotein cholesterol	7.5% (31)	Recommendation of moderate-intensity statin therapy (31)
Breast cancer	Breast Cancer Risk Assessment Tool (BCRAT) (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 (LCRAT) (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 for CT lung cancer screening (41, 43)

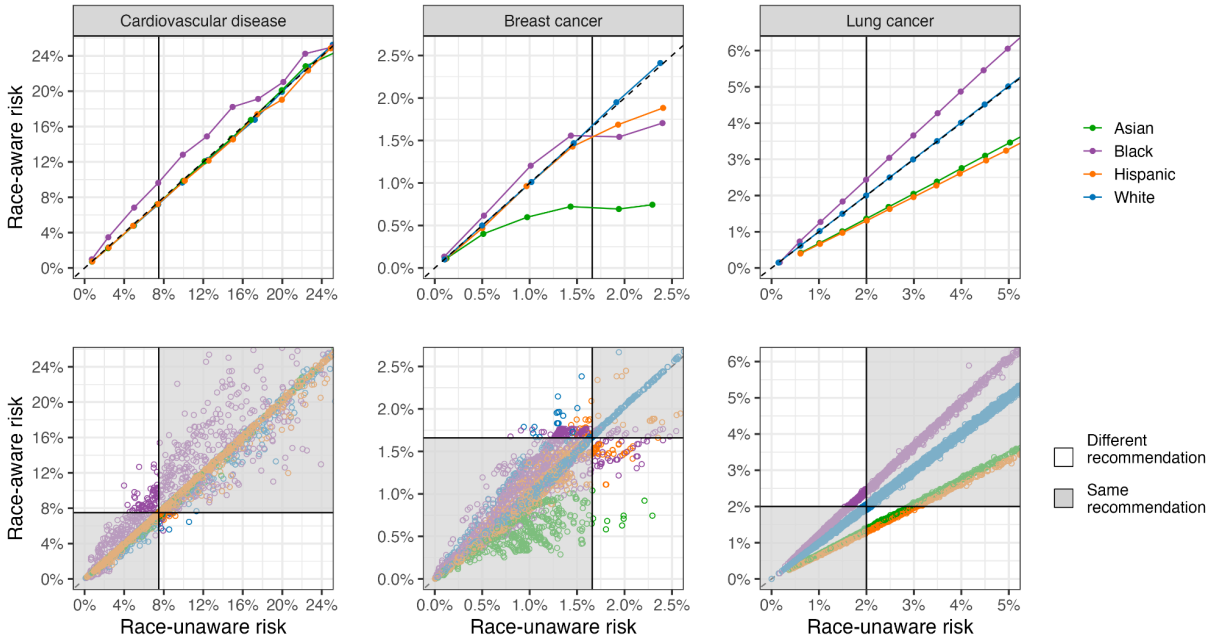
Asterisk (*) indicates that further detail is available in the Appendix.

Figure 1. The costs and benefits of intervention



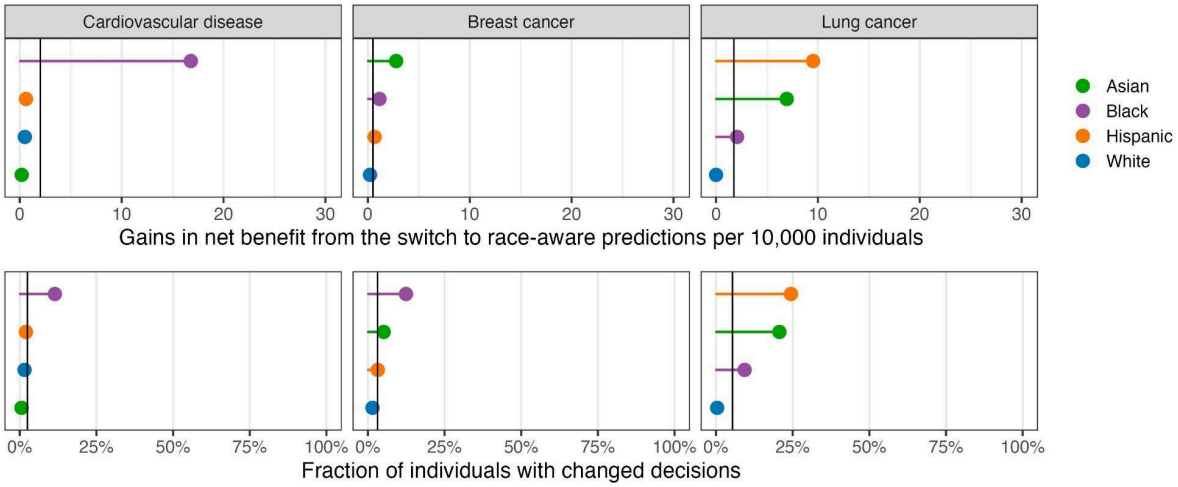
The figure shows the structure of the base case utility function used in the subsequent analysis. We normalize the benefit of appropriate intervention to equal 1 unit, with *c* denoting the cost of intervention.

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



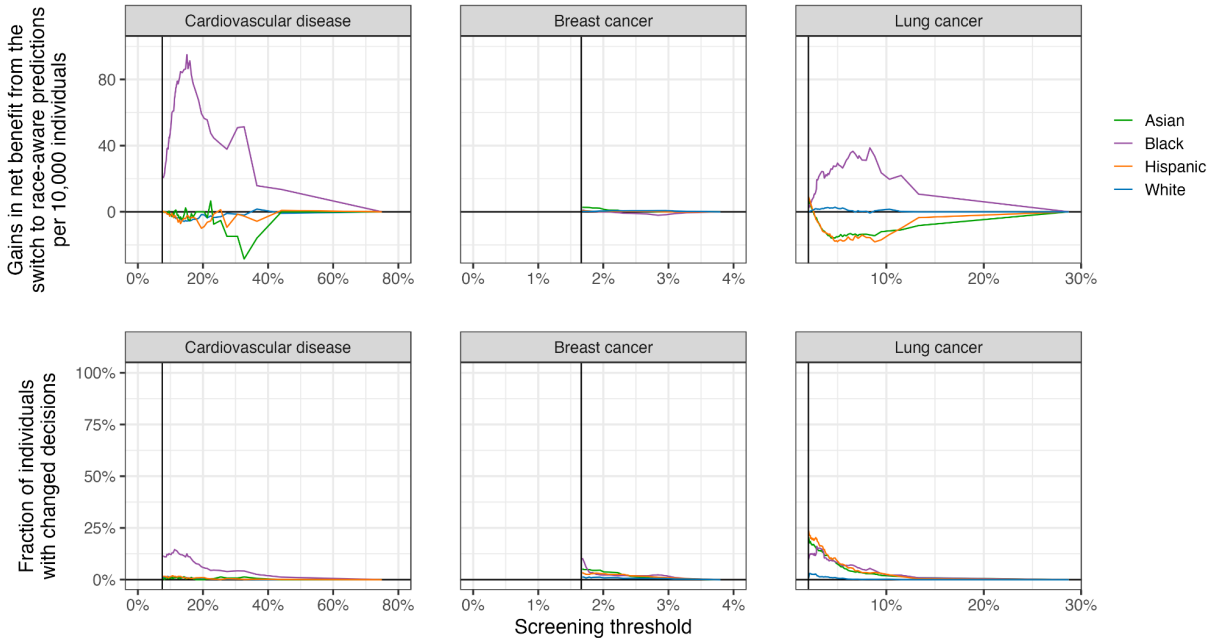
Top row: Calibration plots for cardiovascular disease, breast cancer, and lung cancer, showing race-unaware risk predictions plotted against race-aware risk predictions for each disease. The line $y = x$ denotes the line of perfect calibration (shown above by a dotted black line). The scales of the x- and y-axes differ across diseases due to differences in the risk distributions and thresholds for each disease. Across all three diseases, racial minorities experience more miscalibration in race-unaware predictions than White individuals. The solid black line marks the recommended screening or treatment threshold for each disease. Bottom row: Scatter plots showing race-unaware risk plotted against race-aware risk. Each dot represents an individual in the data. The unshaded regions contain 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 both models.

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 fraction of individuals with different decisions under both models.



Top row: 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 experienced across the entire population. Minority racial and ethnic groups consistently experience 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 row: The fraction 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 fraction 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 fraction of individuals in the population who would receive different recommendations across the two models, disaggregated by race and ethnicity. This fraction 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. Collectively, these results show that Black individuals would experience considerable gains in utility from using race-aware predictions, under conditions of scarcity. Additionally, 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 minimal gains in net benefit from using a race-aware model under normal circumstances.

APPENDIX

Data samples for each disease

Cardiovascular disease

For cardiovascular disease, we restricted our data sample from NHANES to non-pregnant adults between the ages of 40 and 79 who had never taken statins or experienced any of the following cardiovascular events: congestive heart failure, coronary heart disease, angina, heart attack, or stroke. Lastly, we further restricted our sample to individuals with biomarkers in the appropriate range for use of the 2013 PCE (29, 30): high-density lipoprotein cholesterol greater than or equal to 20 mg/dl and less than or equal to 100 mg/dl, total cholesterol greater than or equal to 130 mg/dl and less than or equal to 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 35 years and older.

Lung Cancer

For lung cancer, we used the full data sample from the 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 US-derived 2013 race- and sex-specific pooled cohort equations (PCE) to estimate 10-year risk of atherosclerotic cardiovascular disease (ASCVD) events (29). The covariates used by the ASCVD PCE vary by race and ethnicity and sex due to there being four 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 (AHA), the equations for non-Hispanic White individuals may be used to estimate the risk of individuals of other ethnicities (30). The AHA guidelines recommend a risk threshold of 7.5% be used to identify individuals 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 their potential harms, burdens, and cost.

Breast Cancer

The National Cancer Institute maintains an online Breast Cancer Risk Assessment Tool (BCRAT) 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 Food and Drug Administration (FDA) has approved using a 1.67% threshold on 5-year BCRAT risk 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 dataset 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 to 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:

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

where

$$p_{White} = \frac{21{,}433+6{,}582}{235{,}629} \approx 0.12\ ,$$

$$p_{Black} = \frac{1{,}384+1{,}128}{27{,}179} \approx 0.09,$$

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

$$\text{and}\ \ p_{Hispanic} = \frac{546+256}{9{,}049} \approx 0.09.$$

These rates were taken from prior work on breast cancer incidence (specifically, Table 1 in (38)).

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

Lung cancer

For lung cancer, we computed 5-year risk predictions from the LCRAT model, which is the model used by the National Cancer Institute's Lung Cancer Risk Assessment Tool (LCRAT) (41). To compute risk estimates, we used the Lung Cancer Risk Models for 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 CT 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 LCRAT model is: the 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, prior 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 the patient has, in the past year, experienced 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 input as unknown values into the risk model.

Survey sample weights were not provided in the NLST dataset, and so we reweighted individuals to match the joint age, gender, and race distribution of Americans between 40 and 80 years old, mirroring the age range of individuals in the NLST data. To generate these weights, we used data on national population by characteristics (2020-2023) provided by the U.S. Census Bureau (28). Specifically, we used the projected monthly population estimate by age, sex, race and Hispanic

origin for the month of June 2024. The weight for an individual of age $A = a$, sex $S = s$, and race and ethnicity $R = r$ was computed as follows:

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

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

$$NLST\ proportion = \frac{\sum_{i=1}^{N_{NLST}} I(A_i{=}a, S_i{=}s, R_i{=}r)}{N_{NLST}}, \text{ and } I(\cdot) \text{ denotes the indicator function.}$$

Estimating race-unaware risk

To obtain race-unaware estimates of risk for each individual, we invoke the law of total probability as follows: $P(Disease \mid X) = \sum_{r} P(Disease \mid X,\ R = r) \cdot P(R = r \mid X)$, where X denotes the set of non-race covariates used to estimate risk for a given disease, and R denotes race and ethnicity. We estimate $P(R = r \mid X)$ 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.

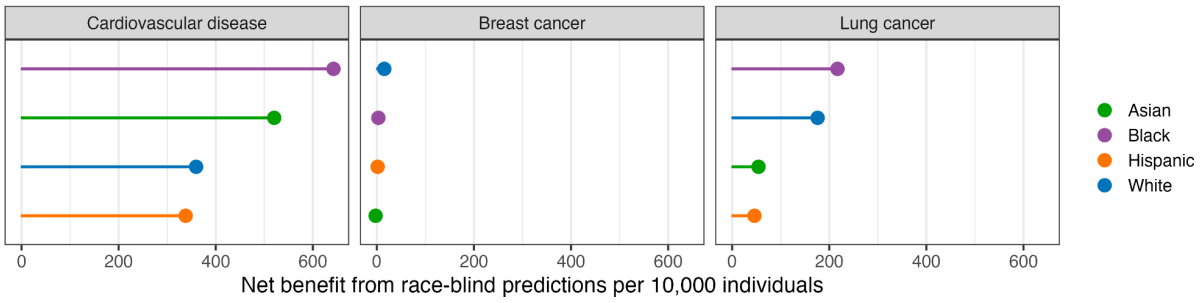
Deriving the value of screening or treatment cost

For each disease, we assume that the decision threshold is set at the point of indifference, i.e., 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 the value of the intervention cost c in terms of the threshold t. In particular, for an individual on the threshold, i.e., with $P(Disease) = t$, we have

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

As a result, $c = t$.

Figure A1. Baseline utility results per disease by race and ethnicity



Heterogeneous utility

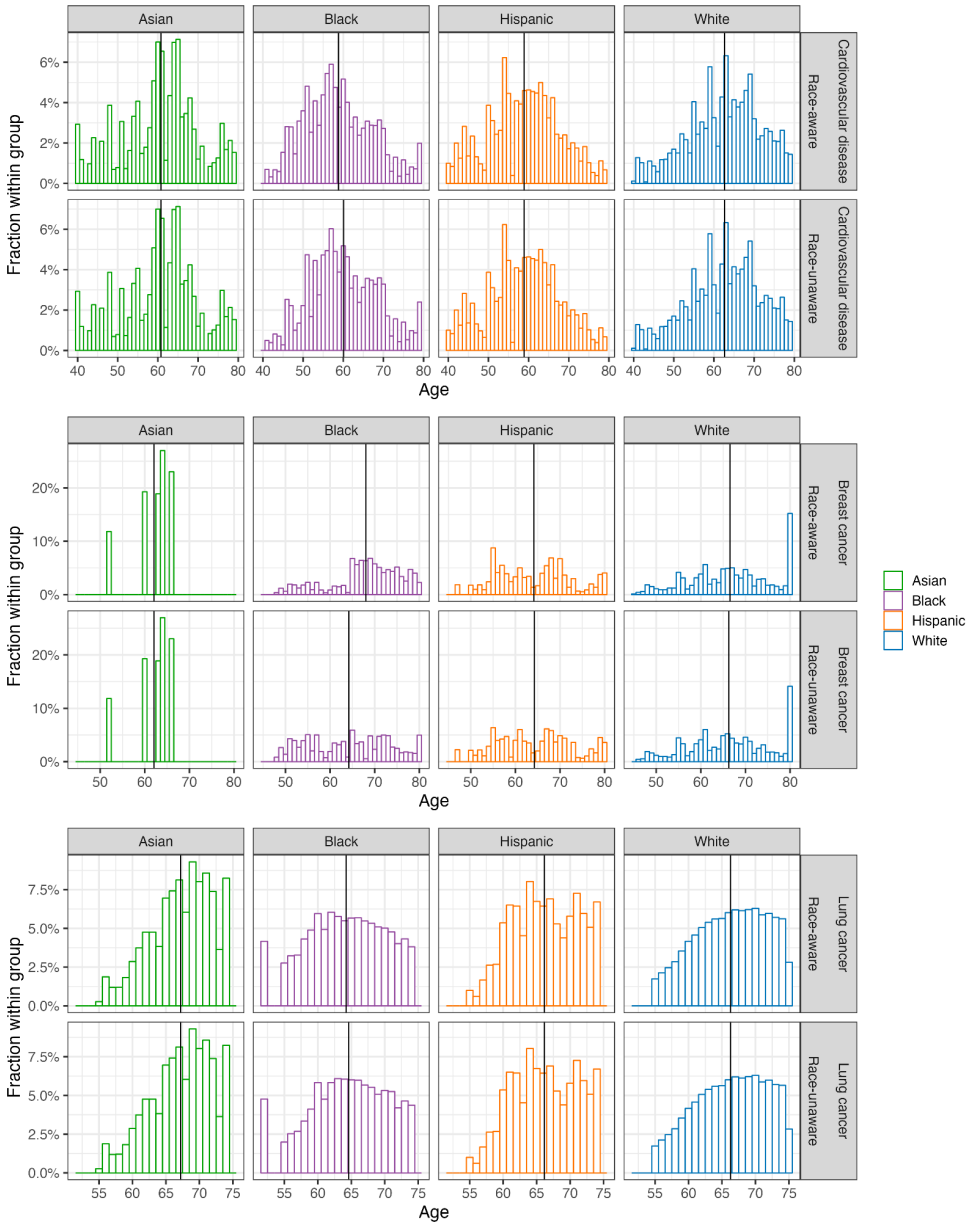
We start by considering age-related heterogeneity in utility. To do so, we examine the age of individuals recommended for screening or treatment by race-aware versus race-unaware models. For all three diseases we study, we find that the average age of individuals appropriately recommended for intervention 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 Figure A2. As a result, a heterogeneous form of the utility function based solely on age (e.g., 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 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 Figure A3, where each point on the curve is an outcome that is theoretically achievable with a race-aware decision tool.

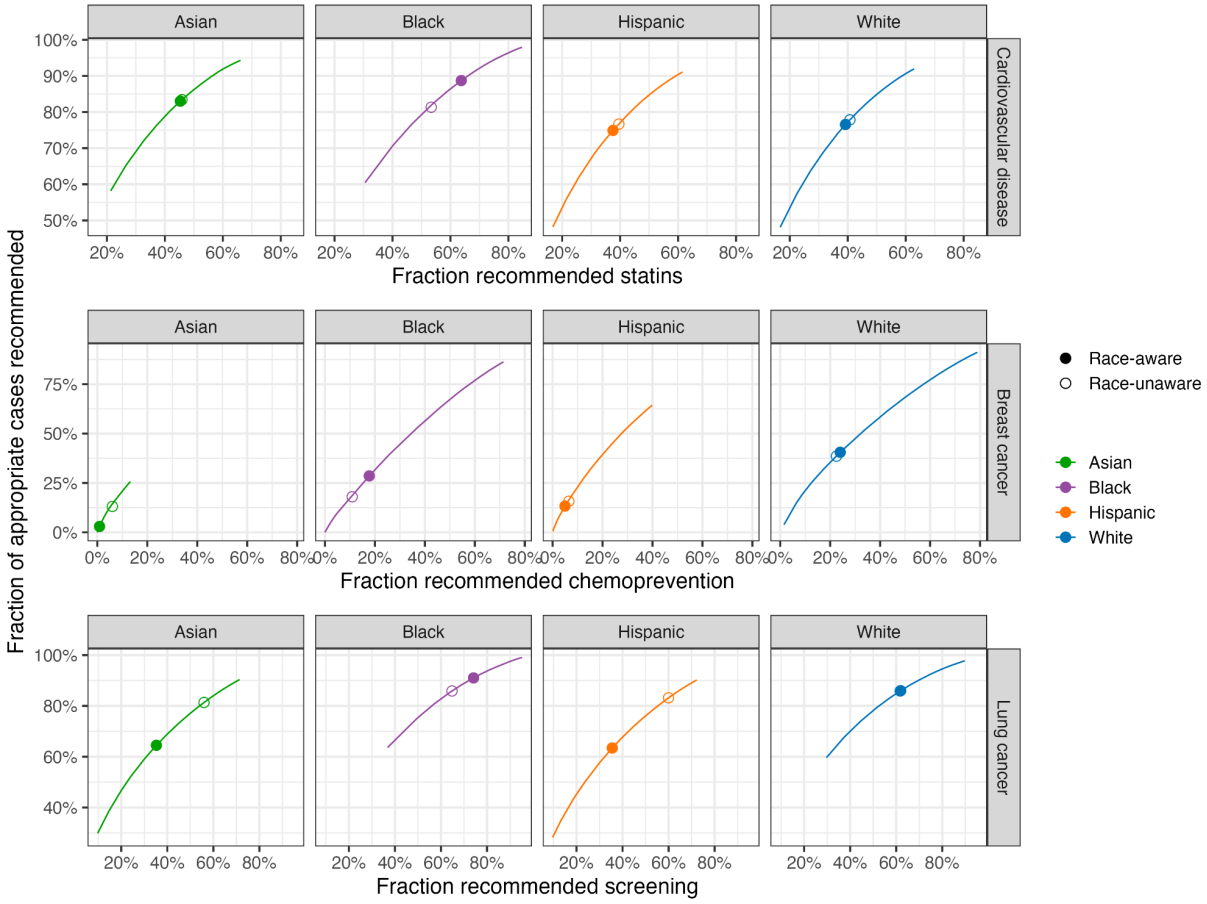
The solid dot in each subplot in Figure A3 corresponds to the trade-off between intervention and detection using the recommended decision threshold with a race-aware model. The hollow dot in each subplot corresponds to the trade-off using the recommended decision threshold with a race-unaware model. For several subgroups and diseases, the race-aware and race-unaware models yield different trade-offs. 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 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.

Figure A2. 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 that 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.

Figure A3. 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 dots indicate the trade-off obtained using a race-aware model with the recommended decision threshold, and hollow dots show the trade-off obtained using a race-unaware model with the recommended decision threshold.

Table A1. 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%

We verified that the optimal threshold under a race-aware model is approximately equal to the optimal threshold under a race-unaware model. To do so, 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.