Comparative Effectiveness Review
Number 268

Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

AHRQ
Agency for Healthcare
Research and Quality

Comparative Effectiveness Review

Number 268

Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. 75Q80120D00002

Prepared by:
ECRI-Penn Medicine Evidence-based Practice Center
Philadelphia, PA

Investigators:
Kelley Tipton, M.P.H., and Brian F. Leas, M.S., M.A. (Co-Leads)
Emilia Flores, Ph.D., R.N.
Christopher Jepson, Ph.D.
Jaya Aysola, M.D., M.P.H.
Jordana Cohen, M.D., M.S.C.E.
Michael Harhay, Ph.D.
Harald Schmidt, Ph.D.
Gary Weissman, M.D., M.S.H.P.
Jonathan Treadwell, Ph.D.
Nikhil K. Mull, M.D.
Shazia Mehmood Siddique, M.D., M.S.H.P.

AHRQ Publication No. 24-EHC004
December 2023

This report is based on research conducted by the ECRI-Penn Medicine Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 75Q80120D00002). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help healthcare decision makers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of healthcare services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

A representative from AHRQ served as a Contracting Officer's Representative and reviewed the contract deliverables for adherence to contract requirements and quality. AHRQ did not directly participate in the literature search, determination of study eligibility criteria, data analysis, interpretation of data, or preparation or drafting of this report.

AHRQ appreciates appropriate acknowledgment and citation of its work. Suggested language for acknowledgment: This work was based on an evidence report, Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare, by the Evidence-based Practice Center Program at the Agency for Healthcare Research and Quality (AHRQ).

Suggested citation: Tipton K, Leas BF, Flores E, Jepson C, Aysola J, Cohen J, Harhay M, Schmidt H, Weissman G, Treadwell J, Mull NK, Siddique SM. Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare. Comparative Effectiveness Review No. 268. (Prepared by the ECRI-Penn Medicine Evidence-based Practice Center under Contract

No. 75Q80120D00002.) AHRQ Publication No. 24-EHC004. Rockville, MD: Agency for Healthcare Research and Quality; December 2023. DOI: https://doi.org/10.23970/AHRQEPCCER268. Posted final reports are located on the Effective Health Care Program search page.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of healthcare in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new healthcare technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see https://effectivehealthcare.ahrq.gov/about/epc/evidence-synthesis.

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the healthcare system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the website (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an email list to learn about new program products and opportunities for input.

If you have comments on this systematic review, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Robert Otto Valdez, Ph.D., M.H.S.A.
Director
Agency for Healthcare Research and Quality

Craig A. Umscheid, M.D., M.S.
Director
Evidence-based Practice Center Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Arlene Bierman, M.D., M.S.
Director, Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Anjali Jain, M.D.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Christine S. Chang, M.D., M.P.H.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: Lindsey Miller (Project Manager); Kristina McShea (Medical Librarian); Meghan Lane-Fall, M.D., M.S.H.P. (University of Pennsylvania); Margie Beaudry, M.A., and Daria Turner, M.P.H. (Ripple Effect); and the reference management and editorial team: Katherine Donahue, Helen Dunn, Michael Phillips, and Britney Hall. We also thank AHRQ Task Order Officer Anjali Jain, M.D., and Christine Chang, M.D., M.P.H.

Key Informants

In designing the study questions, the EPC consulted several Key Informants who represent the end-users of research. The EPC sought the Key Informant input on the priority areas for research and synthesis. Key Informants are not involved in the analysis of the evidence or the writing of the report. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual Key Informants.

Key Informants must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any conflicts of interest.

The list of Key Informants who provided input to this report follows:

Yerado Abrahamian, M.H.S.
Team Lead, Clinical Practice Guidelines
Kaiser Permanente Southern California
Pasadena, CA

Shonta Chambers, M.S.W.
Executive Vice President of Health Equity Initiatives and Community Engagement
Patient Advocate Foundation
Atlanta, GA

Lauren Choi, J.D., M.A.
Managing Director, Health Data and Technology Policy
Blue Cross Blue Shield Association
Washington, DC

Malika Fair, M.D., M.P.H.
Senior Director, Equity and Social Accountability
Association of American Medical Colleges
Washington, DC

David Hunt, M.D.
Medical Director, Patient Safety and Health IT Adoption
Office of the National Coordinator for Health Information Technology

Washington, DC

LaShawn McIver, M.D., M.P.H.
Director, Office of Minority Health
Centers for Medicare & Medicaid ServicesBaltimore, MD

Neil R. Powe, M.D., M.P.H., M.B.A.*
Professor of Medicine
University of California-San Francisco
San Francisco, CA

Joan Reede, M.D., M.S., M.P.H., M.B.A.
Dean for Diversity and Community Partnership
Harvard University
Cambridge, MA

Lucy Savitz, Ph.D., M.B.A.
Professor of Health Policy and Management
University of Pittsburgh School of Public Health
Pittsburgh, PA

Denise Schofield
Executive Director, Blue Cross Blue Shield Association
Chicago, IL

Leith States, M.D., M.P.H., M.B.A.
Chief Medical Officer
Office of the Assistant Secretary for Health
Department of Health and Human Services
Washington, DC

Suzanne Tamang, Ph.D.
Senior Computer Scientist
U.S. Department of Veterans Affairs
Menlo Park, CA

Christo Wilson, Ph.D.
Associate Professor and Director, Cybersecurity Program
Northeastern University
Boston, MA

David Wong, M.D.
Chief Medical Officer
Office of Minority Health
Department of Health and Human Services
Washington, DC

*Provided input on Draft Report.

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who provided input to this report follows:

Suzanne Bakken, Ph.D., R.N.
Professor of Biomedical Informatics
Columbia University
New York, NY

Helen Burstin, M.D., M.P.H.
Chief Executive Officer
Council of Medical Specialty Societies
Washington, DC

Lee Fleisher, M.D.
Chief Medical Officer and Director
Center for Clinical Standards and Quality
Centers for Medicare & Medicaid Services
Baltimore, MD

Joseph Habboushe, M.D., M.B.A.
Chief Executive Officer, MDCalc
New York, NY

Andrew Hamilton, M.S.
Chief Informatics Officer, Alliance Chicago
Chicago, IL

Tina Hernandez-Boussard, Ph.D., M.S., M.P.H.*
Professor of Medicine, Stanford University
Stanford, CA

Maia Hightower, M.D., M.P.H., M.B.A.
Chief Executive Officer, Equality AI
Executive Vice President and Chief Digital and Technology Officer, University of Chicago Medicine
Chicago, IL

Vadim Khotilovich, Ph.D.
Director and Data Scientist, Cerner
Kansas City, MO

Kevin Larsen, M.D.*
Senior Vice President of Clinical Innovation, Optum
Washington, DC

Kathryn Marchesini, J.D.
Chief Privacy Officer
Office of the National Coordinator for Health Information Technology
Washington, DC

*Provided input on Draft Report.

Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report do not necessarily represent the views of individual reviewers. AHRQ may also seek comments from other Federal agencies when appropriate.

Peer Reviewers must disclose any financial conflicts of interest greater than $5,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

The list of Peer Reviewers follows:

Deborah Guadalupe Duran, Ph.D.
Senior Advisor to the Director, Data Science
National Institute on Minority Health and Health Disparities
Bethesda, MD

Aletha Maybank, M.D., M.P.H.
Senior Vice President and Chief Health Equity Officer
American Medical Association
Chicago, IL

Impact of Healthcare Algorithms on Racial and Ethnic Disparities in Health and Healthcare

Abstract

Objectives. To examine the evidence on whether and how healthcare algorithms (including algorithm-informed decision tools) exacerbate, perpetuate, or reduce racial and ethnic disparities in access to healthcare, quality of care, and health outcomes, and examine strategies that mitigate racial and ethnic bias in the development and use of algorithms.

Data sources. We searched published and grey literature for relevant studies published between January 2011 and February 2023. Based on expert guidance, we determined that earlier articles are unlikely to reflect current algorithms. We also hand-searched reference lists of relevant studies and reviewed suggestions from experts and stakeholders.

Review methods. Searches identified 11,500 unique records. Using predefined criteria and dual review, we screened and selected studies to assess one or both Key Questions (KQs): (1) the effect of algorithms on racial and ethnic disparities in health and healthcare outcomes and (2) the effect of strategies or approaches to mitigate racial and ethnic bias in the development, validation, dissemination, and implementation of algorithms. Outcomes of interest included access to healthcare, quality of care, and health outcomes. We assessed studies' methodologic risk of bias (ROB) using the ROBINS-I tool and piloted an appraisal supplement to assess racial and ethnic equity-related ROB. We completed a narrative synthesis and cataloged study characteristics and outcome data. We also examined four Contextual Questions (CQs) designed to explore the context and capture insights on practical aspects of potential algorithmic bias. CQ 1 examines the problem's scope within healthcare. CQ 2 describes recently emerging standards and guidance on how racial and ethnic bias can be prevented or mitigated during algorithm development and deployment. CQ 3 explores stakeholder awareness and perspectives about the interaction of algorithms and racial and ethnic disparities in health and healthcare. We addressed these CQs through supplemental literature reviews and conversations with experts and key stakeholders. For CQ 4, we conducted an in-depth analysis of a sample of six algorithms that have not been widely evaluated before in the published literature to better understand how their design and implementation might contribute to disparities.

Results. Fifty-eight studies met inclusion criteria, of which three were included for both KQs. One study was a randomized controlled trial, and all others used cohort, pre-post, or modeling approaches. The studies included numerous types of clinical assessments: need for intensive care or high-risk care management; measurement of kidney or lung function; suitability for kidney or lung transplant; risk of cardiovascular disease, stroke, lung cancer, prostate cancer, postpartum depression, or opioid misuse; and warfarin dosing. We found evidence suggesting that algorithms may: (a) reduce disparities (i.e., revised Kidney Allocation System, prostate cancer screening tools); (b) perpetuate or exacerbate disparities (e.g., estimated glomerular filtration rate [eGFR] for kidney function measurement, cardiovascular disease risk assessments); and/or (c) have no effect on racial or ethnic disparities. Algorithms for which mitigation strategies were identified are included in KQ 2. We identified six types of strategies often used to mitigate the

potential of algorithms to contribute to disparities: removing an input variable; replacing a variable; adding one or more variables; changing or diversifying the racial and ethnic composition of the patient population used to train or validate a model; creating separate algorithms or thresholds for different populations; and modifying the statistical or analytic techniques used by an algorithm. Most mitigation efforts improved proximal outcomes (e.g., algorithmic calibration) for targeted populations, but it is more challenging to infer or extrapolate effects on longer term outcomes, such as racial and ethnic disparities. The scope of racial and ethnic bias related to algorithms and their application is difficult to quantify, but it clearly extends across the spectrum of medicine. Regulatory, professional, and corporate stakeholders are undertaking numerous efforts to develop standards for algorithms, often emphasizing the need for transparency, accountability, and representativeness.

Conclusions. Algorithms have been shown to potentially perpetuate, exacerbate, and sometimes reduce racial and ethnic disparities. Disparities were reduced when race and ethnicity were incorporated into an algorithm to intentionally tackle known racial and ethnic disparities in resource allocation (e.g., kidney transplant allocation) or disparities in care (e.g., prostate cancer screening that historically led to Black men receiving more low-yield biopsies). It is important to note that in such cases the rationale for using race and ethnicity was clearly delineated and did not conflate race and ethnicity with ancestry and/or genetic predisposition. However, when algorithms include race and ethnicity without clear rationale, they may perpetuate the incorrect notion that race is a biologic construct and contribute to disparities. Finally, some algorithms may reduce or perpetuate disparities without containing race and ethnicity as an input. Several modeling studies showed that applying algorithms out of context of original development (e.g., illness severity scores used for crisis standards of care) could perpetuate or exacerbate disparities. On the other hand, algorithms may also reduce disparities by standardizing care and reducing opportunities for implicit bias (e.g., Lung Allocation Score for lung transplantation). Several mitigation strategies have been shown to potentially reduce the contribution of algorithms to racial and ethnic disparities. Results of mitigation efforts are highly context specific, relating to unique combinations of algorithm, clinical condition, population, setting, and outcomes. Important future steps include increasing transparency in algorithm development and implementation, increasing diversity of research and leadership teams, engaging diverse patient and community groups in the development to implementation lifecycle, promoting stakeholder awareness (including patients) of potential algorithmic risk, and investing in further research to assess the real-world effect of algorithms on racial and ethnic disparities before widespread implementation.

Contents

Tables

Figures

Appendixes

Executive Summary

Main Points

- We examined two Key Questions (KQs). KQ 1 explored the effect of healthcare algorithms on racial and ethnic disparities in access to care, quality of care, and health outcomes. KQ 2 identified strategies to mitigate racial and ethnic bias associated with algorithms.
- For KQ 1, we identified 17 studies examining the effect of 18 algorithms on racial and ethnic disparities in health and healthcare. Four of the 18 algorithms included race or ethnicity as an input variable. The most frequently examined algorithms are used (or, in a few instances, are suggested for use) to inform resource allocation in a crisis setting (e.g., crisis standards of care) (4 studies), guide emergency department (ED) care decisions (3 studies), determine eligibility for lung cancer screening (3 studies), and determine eligibility for prostate cancer screening (2 studies). None was a randomized controlled trial (RCT).
- KQ 1 studies found that algorithms may reduce racial and ethnic disparities (5 studies), perpetuate or exacerbate disparities (11 studies; 3 of 11 included an examination of methods to mitigate these disparities and thus addressed both KQ 1 and 2), or have no effect on disparities (1 study). Three of the four studies examining algorithms that included race and ethnicity as an input variable found that these algorithms actually or potentially reduced disparities. In some algorithms (e.g., revised Kidney Allocation System), race and ethnicity input variables were included specifically to address existing racial and ethnic disparities.
- Studies addressing KQ 2 examined strategies for mitigating racial and ethnic disparities associated with algorithms. We included 44 studies across a range of clinical applications, including measurement of kidney function and lung function; risk assessment for cardiovascular disease, stroke, lung cancer, opioid misuse, and postpartum depression; suitability for kidney and liver transplant; and anticoagulation titration.
- Six types of mitigation strategies were identified: removing a race or ethnicity input variable from the algorithm (24 studies); replacing race or another input variable with a different measure (5 studies); adding an input variable (9 studies); recalibrating the algorithm with a more representative patient population (4 studies); stratifying algorithms to assess Black and White patients separately (2 studies); and using different statistical techniques within algorithms (3 studies).
- Most studies that examined the impact of removing a race coefficient from a common kidney function measure (eGFR) found an increase in diagnoses of chronic and severe kidney disease in Black patients. This may lead to a decrease in disparities in both early nephrology referrals for chronic kidney disease and referrals for kidney transplant. Conversely, some studies demonstrated a potentially negative impact of removing the race coefficient on health and healthcare outcomes other than transplant eligibility among Black patients when removing the race coefficient (e.g., patients reclassified as having more severe kidney disease could be deemed ineligible, possibly inappropriately, for medication or enrollment in clinical trials).

- Algorithms are often developed by electronic health record vendors, payers, and health systems. Due to the propriety nature of these algorithms, little is known about the development approach and potential impact on racial and ethnic disparities.
- Awareness is low among patients, healthcare providers, payers, and policymakers of the potential for algorithms to affect racial and ethnic disparities.

Background and Purpose

Healthcare algorithms are frequently used to guide clinical decision making at the point of care and as part of resource allocation and healthcare management. Race and ethnicity are often used as input variables in these algorithms.[1-3] However, because race and ethnicity are socially constructed and thereby poor proxies for biological markers or genetic predisposition, when used in algorithms to guide clinical care their inclusion may introduce or exacerbate inappropriate, unequal treatment (healthcare disparities) and thereby contribute to or exacerbate unequal health outcomes (health disparities).[4-6] In September 2020, the Agency for Healthcare Research and Quality received a request from Congress to review the evidence on the use of race and ethnicity in clinical algorithms and the potential of algorithms to contribute to disparities in healthcare. This review responds to that request by exploring two KQs addressing how healthcare algorithms affect racial and ethnic disparities in access to care, quality of care, and health outcomes. KQ 1 asks: What is the effect of healthcare algorithms on racial and ethnic differences in access to care, quality of care, and health outcomes? KQ 2 focuses on potential solutions: What is the effect of interventions, models of interventions, or other approaches to mitigate racial and ethnic bias in the development, validation, dissemination, and implementation of healthcare algorithms?

Four Contextual Questions (CQs) designed to capture insights on practical aspects of potential racial and ethnic bias were also examined. CQ 1 examines the scope of healthcare algorithms that explicitly include race and ethnicity as an input variable; CQ 2 summarizes recently emerging standards and guidance on how racial and ethnic bias can be prevented or mitigated during algorithm development and deployment; and CQ 3 explores various stakeholders' awareness of and their perspectives on associations between algorithms, race and ethnicity, and healthcare. To respond to CQ 4, we conducted an in-depth evaluation of a sample of six healthcare algorithms, not previously evaluated in the published literature, to better understand how their design and implementation might contribute to racial and ethnic disparities.

Methods

We searched electronic databases (Embase®, MEDLINE®, PubMed®, and the Cochrane Library) from January 1, 2011, to February 7, 2023. Using predefined criteria and dual review, we screened all records for KQ 1 and KQ 2 and selected eligible studies that assessed one or both KQs. We included studies that examined actual outcomes among patients managed using algorithms as well as those that modeled potential effects of the use of algorithms in both real-world and synthetic datasets. Eligible studies were required to report on at least one of the following outcome categories: access to healthcare, quality of care, and health outcomes. We assessed studies' methodologic risk of bias (ROB) using the Risk Of Bias in Non-randomized Studies of Interventions (ROBINS-I) tool and piloted an appraisal supplement with signaling questions (e.g., was a transparent rationale provided for including or removing race and ethnicity?) to assess racial and ethnic equity-related ROB. Using this tool, ROB ratings on each

of seven domains are combined to generate an overall rating of Low, Moderate, or High ROB for each study. CQs 1-3 were addressed through supplemental literature reviews and conversations with Subject Matter Experts, Key Informants, and Technical Expert Panel members. For CQ 4, we evaluated, in-depth, the development approach, performance, and implementation of six algorithms not previously widely evaluated in the published literature for potential racial or ethnic bias.

Results

Fifty-eight studies met eligibility criteria. Fourteen studies addressed KQ 1, 41 studies addressed KQ 2, and 3 studies addressed both KQs and are presented in the results section for each KQ. For KQ 1, 17 studies examined 18 algorithms that inform decisions about ED care (3 studies), predict mortality to inform resource allocation in a crisis setting (e.g. crisis standards of care) (4 studies), predict future healthcare needs (1 study), allocate organs for transplant (2 studies), assess risk of lung cancer (3 studies), predict opioid misuse (1 study), predict risk of prostate cancer (2 studies), and predict risk of stroke (1 study). Four of the 18 algorithms included race or ethnicity as an input variable. All studies addressing KQ 1 were non-RCTs and were rated as moderate or high ROB due to concerns such as confounding, deviations from intended interventions, and missing data. Equity-based signaling questions changed domain-specific ROB in one instance (from Low to Moderate for the domain of bias due to selection of participants in one study) but did not change overall ROB for any KQ 1 studies. Most study designs employed a modeling approach to determine outcomes using either real-world or synthetic datasets for inputting into the algorithm. The studies found that algorithms may reduce racial and ethnic disparities (5 studies), perpetuate or exacerbate disparities (11 studies; 3 of these included an examination of methods to mitigate these disparities and thus addressed both KQ 1 and 2), or have no effect on disparities (1 study). Three studies examining algorithms that included race and ethnicity as an input variable found that these algorithms actually or potentially reduced racial and ethnic disparities. It is important to note that, in one of the three algorithms (the revised Kidney Allocation System), race and ethnicity input variables were included specifically to address existing racial and ethnic disparities.

For KQ 2, 21 of the 44 included studies focused on kidney function and evaluated efforts to mitigate potential harms associated with using the race correction in the estimation of glomerular filtration rate (eGFR). Seven studies examined algorithms that predict cardiovascular risk, four studies addressed kidney or liver donation or transplant, and three studies assessed algorithms that guide dosing of the anticoagulant warfarin. The remaining studies addressed the need for intensive or high-risk care management, assessment of lung function, and risk of stroke, lung cancer, postpartum depression, or opioid misuse. One study was a randomized controlled trial, 17 studies used cohort or pre-post designs, and 26 studies used modeling approaches. ROB was rated as Low for 8 studies, Moderate for 31 studies, and High for 5. In six studies, ROB ratings for individual domains of bias changed because of the equity-based signaling questions we added, but the overall ROB rating was changed in only one of these studies. The most common domain to receive a rating change (mostly from Low to Moderate) was bias in selection of study participants due to inconsistent reporting of racial and ethnic groups and inconsistent definitions and categories for race and ethnicity.

We identified six broad categories that describe the mitigation strategies used to address potential harms resulting from algorithms: removing an input variable (usually race and ethnicity) was used in 24 studies; replacing a variable with one or more different variables (5

studies); adding one or more input variables (9 studies); diversifying the racial and ethnic composition of the patient population used to train or validate an algorithm (4 studies); creating separate algorithms or thresholds for different populations (2 studies); and modifying the statistical or analytic techniques used for algorithm development (3 studies). Some studies compared more than one mitigation strategy. Evidence suggests that removing a race coefficient from eGFR may result in significantly more diagnoses of chronic and severe kidney disease among Black patients, which can then lead to increased eligibility for kidney transplant; however, this may also result in underuse or underdosing of other treatments. Further research is needed to better understand these implications across the wide range of outcomes and medical decisions that eGFR influences.

Although studies reported that mitigation approaches can improve algorithm calibration and may reduce disparities, they often relied on simulation and inference to estimate the effects of such strategies on patient outcomes. This may not adequately model potential biases occurring in algorithm translation, dissemination, and implementation, and further research is needed to quantify the real-world effects of using and modifying algorithms. Finally, we found the effectiveness of mitigation strategies is context-specific and may largely depend on the unique combination of algorithm, clinical condition, population, setting, and outcomes.

Findings from the CQs suggest the scope of algorithmic bias is difficult to quantify, but it clearly extends across the entire spectrum of medicine. Public awareness of healthcare algorithms and their potential effects on health and healthcare is very limited. We identified numerous efforts by regulatory, professional, and corporate stakeholders to develop standards for algorithms, often emphasizing the need for transparency, accountability, and representativeness.

Limitations

Our multipronged and multidisciplinary approach to conduct a comprehensive review of the use of algorithms and efforts to mitigate their potential contribution to racial and ethnic disparities enabled us to synthesize a broad array of evidence. Due to heterogeneity of included studies, conclusions about the effect of algorithms on exacerbating racial and ethnic disparities in health and healthcare outcomes varied across different clinical assessment areas. We included algorithms for many different clinical settings, and results in one setting do not necessarily apply to those in other settings. Furthermore, attempts to mitigate race disparities caused by algorithms are also highly context-specific. Included studies frequently used national datasets that typically provide a more representative distribution of races, yet some studies used an overly broad race categorization, such as White/Non-White, when presenting findings. While this may allow investigators the ability to study systemic racism, broadly speaking, there is often inadequate representation of specific racial and ethnic groups to identify subgroup-specific issues such as differences in effects across different populations. This is because virtually all “Non-White” people self-identify using more specific race designation(s); furthermore, most electronic health records and scoring systems use more specific designations. Other studies focused only on two races (e.g., Black/White); their results may be less relevant to those of other races, and often ethnicity was not specified by study authors (i.e., whether these categories include patients identifying as Hispanic/Latino).

The lack of studies evaluating the real-world effects of an algorithm or mitigation strategy is a limitation of the current evidence base. Only 7 of 58 studies (3 for KQ 1 and 4 for KQ 2) actually managed patients with an algorithm or reported real outcomes experienced by patients. The rest of the studies used outcome simulation, whereby the authors estimated an algorithm’s

(or mitigation strategy's) hypothetical influence. The applicability of such studies depends heavily on assumptions made, representativeness of the data sources analyzed, and whether the algorithm would actually be used in the manner hypothesized. Simulation may, however, provide the basis for future hypothesis-driven clinical research into the effects of algorithms on racial and ethnic differences.

Implications and Conclusions

Algorithms have been shown to potentially exacerbate, perpetuate, or reduce racial and ethnic disparities in health and healthcare outcomes. When race or ethnicity were incorporated into an algorithm to intentionally tackle known disparities in resource allocation (e.g., kidney transplant allocation) or healthcare delivery (e.g., prostate cancer screening historically led to Black men receiving more low-yield biopsies), disparities were reduced. However, when race or ethnicity were included without a clear and appropriate rationale, incorrect notions of race as biological may be reinforced; moreover, algorithms that inappropriately used race and ethnicity as a proxy for biological mechanisms have been shown to potentially perpetuate and exacerbate disparities (e.g., eGFR for kidney function measurement). Furthermore, some algorithms do not contain race or ethnicity as an input but could also affect disparities. Several modeling studies showed that applying algorithms out of context of original development (e.g., illness severity scores used for crisis standards of care) would exacerbate disparities. Conversely, algorithms may also reduce disparities by standardizing care (e.g., Lung Allocation Score for lung transplantation). In terms of strategies to mitigate racial and ethnic disparities associated with algorithms, many studies presented proximal outcomes, such as improvements of algorithmic accuracy within a single racial group, resulting in the need to infer or extrapolate effects on differences between racial and ethnic groups. No single strategy led to the greatest success, but several have been shown to successfully mitigate disparities. Results may be highly context-specific, relating to unique combinations of algorithm, clinical condition, population, setting, and outcomes.

Finally, we emphasize the challenge of determining cause and effect in this literature. Disparities in health and healthcare are well documented for BIPOC (Black, Indigenous, or People of Color) people, but assessing how, and how much, a particular algorithm may contribute to or redress a disparity needs to be assessed. Distal health outcomes are also influenced by multiple contributing clinical, health system and social factors. Important future steps include increasing transparency in algorithm development and implementation, increasing diversity of research and leadership teams, engaging diverse patient and community groups in the development to implementation lifecycle, promoting awareness by stakeholders (including patients) of potential algorithmic risk, and investing in real-world experiments to assess the effect of algorithms on racial and ethnic disparities before widespread implementation.

References

1. 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(9):874-82. doi: 10.1056/NEJMms2004740. PMID: 32853499.
2. Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. Lancet. 2020;396(10257):1125-8. doi: 10.1016/s0140-6736(20)32076-6. PMID: 33038972.
3. Schmidt IM, Waikar SS. Separate and unequal: race-based algorithms and implications for nephrology. J Am Soc Nephrol. 2021;32(3):529-33. doi: 10.1681/asn.2020081175. PMID: 33510038.
4. Eneanya ND, Yang W, Reese PP. Reconsidering the consequences of using race to estimate kidney function. JAMA. 2019;322(2):113-4. doi: 10.1001/jama.2019.5774. PMID: 31169890.
5. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi: 10.1126/science.aax2342. PMID: 31649194.
6. Amutah C, Greenidge K, Mante A, et al. Misrepresenting race - the role of medical schools in propagating physician bias. N Engl J Med. 2021;384(9):872-8. doi: 10.1056/NEJMms2025768. PMID: 33406326.

1. Introduction

1.1 Background

Healthcare algorithms are frequently used to guide clinical decision making both at the point of care and as part of resource allocation and healthcare management. For this review, algorithms are defined as mathematical formulas or models that combine different input variables or factors to inform a calculation or an estimate, such as an estimate of disease or risk of a particular health outcome. Algorithms are often incorporated into healthcare decision tools, such as clinical guidelines, pathways, clinical decision support programs in electronic health records (ERs), and operational systems used by health systems and payers; our use of "algorithm" includes algorithm-informed tools. End-users, such as clinicians, integrated delivery networks, payers, and consumers, use algorithms for at least six broad purposes: screening, risk prediction, diagnosis, prognosis, treatment planning, and resource allocation. While algorithms have long been derived from traditional statistical techniques, such as regression analysis, their use in predictive analyses is increasingly fueled by artificial intelligence techniques, including machine learning.

Algorithms commonly include clinical and sociodemographic input variables and measures of healthcare utilization. Race and ethnicity are often used as input variables;[1-3] however, because race and ethnicity are not biological concepts but socially constructed and represent a variety of other factors, their use in algorithms that influence clinical decision making can have a wide range of effects on patient outcomes. Some effects could be desirable (e.g., improved allocation of resources or access to care), and some could be harmful (e.g., exacerbation or perpetuation of health and healthcare disparities).[4-6] Many effects are unknown.

In a seminal review published in 2020, Vyas et al.[1] examined race-based algorithms commonly used in eight clinical specialties. The review observed that while use of race and ethnicity as an input variable was often driven by primary studies that noted a difference in health between racial and ethnic groups, little, if any, actual evaluation has measured potential race-based harms of using such algorithms. The authors noted that including race and ethnicity might direct more resources toward White patients and thus exacerbate health and healthcare inequities.

Algorithm developers often include race and ethnicity as input variables, intending to increase diagnostic or predictive accuracy by capturing genetic predispositions due to racial and ethnic differences that affect clinical outcomes. However, race and ethnicity are poor proxies for genetic predisposition. Greater genetic variation typically exists within groups classified as the same race or ethnicity than between them.[7-9] Numerous purported genetic predisposing differences between races and ethnicities regarding muscle mass, pain sensitivity, lung function, and similar biomarkers have been debunked.[10] Mounting research details nonbiological root causes of biological phenomena (e.g., epigenetics) and observed differences in health between racial and ethnic groups. Specifically, chronic exposure to interpersonal discrimination, coupled with structural racism or biases intrinsic to societal systems, create unearned disadvantage or advantage depending on one's identity. This leads to unequal opportunities for health and wellbeing through disparities in social determinants of health (SDOH).[11-14]

While individual self-identification of race and ethnicity is considered the preferred method for defining and collecting these data, the sensitivity and specificity of this method depends on the response categories presented, in particular for individuals who may identify with more than one racial and ethnic group (e.g., Hispanic and Asian populations).[15,16] This further highlights

challenges with operationalizing U.S.-centered, socially constructed racial and ethnic categories, such as those used by the federal Office of Management and Budget (OMB). In response to the challenges with OMB categories, the National Academy of Medicine in 2009 issued standards for optimal collection of Race, Ethnicity, and Language (REaL). REaL standards improved on existing OMB race categories by incorporating Hispanic ethnicity as a racial category and recommended capture of granular ethnicity to better approximate ancestry or country of origin.[16-18] However, implementation of REaL standards by health systems, researchers, and the broader medical community is variable. For this report, we use race and ethnicity to represent the socially constructed OMB categories of race and Hispanic ethnicity, with a multiple-choice option as distinct from granular ethnicity in accord with the 2009 standards. Unless otherwise specified, we follow the convention of capitalizing all racial and ethnic categories.

Algorithm developers often justify including race and ethnicity input variables by citing observational studies or *post hoc* analyses of randomized controlled trial data that demonstrated differences in outcomes among different race and ethnicity subgroups. These studies may be small and unrepresentative, serve to reinforce misconceptions, or assume that race and ethnicity are fundamental causes, even though other factors may be causative, confounding, or modifying the effects of race and ethnicity.[19,20] A robust example examines a "race-correction" coefficient that raises the threshold of concern or action for a given estimated glomerular filtration rate (eGFR), a key biomarker in examining kidney function and diagnosing and treating chronic kidney disease, only for Black patients. Recent studies have modeled the effect of removing the race-based coefficient from eGFR and concluded that Black patients would be more likely to receive more timely referrals for kidney transplants compared with race-based eGFR calculations.[21-23] However, controversy around this issue remained, as the evidence base lacks prospective trials comparing differing approaches to assessing kidney function and subsequent need for treatments, including transplant.[24-26] Accordingly, the National Kidney Foundation and the American Society of Nephrology convened a task force to address this topic. In September 2021, the task force released its final report recommending against use of the race coefficient, and supporting the use of a race-independent biomarker, cystatin C, to confirm eGFR.[27]

Algorithm input variables other than race and ethnicity may also perpetuate, contribute to, and/or exacerbate health disparities and inequities. For example, an algorithm used to allocate access to disease management support programs included prior healthcare costs as an input variable to serve as a proxy for clinical needs and subsequent healthcare utilization.[5] This algorithm led to a disproportionate enrollment of White patients with less severe disease into a chronic disease management program compared with Black patients with greater disease severity. These stark racial and ethnic disparities were a consequence of selecting a proxy for disease severity and healthcare needs, given that healthcare expenditure is higher, on average, for White patients than for Black patients with the same conditions reflecting barriers to accessing care. Replacing healthcare costs with a better indicator for disease severity corrected this race and ethnicity disparity for access to an indicated chronic disease management program. Therefore, due to structural racial and ethnic biases and other forms of racism in healthcare, algorithms that do not include race and ethnicity input variables may nonetheless contribute to or perpetuate healthcare and health disparities.

Evidence gaps regarding the impact of many algorithms that include race and ethnicity input variables on potential racial and ethnic disparities in healthcare delivery remain, with few studies comparing the effects of alternative algorithm strategies on important health outcomes. Nevertheless, several academic societies have issued position statements or guidelines supporting removal of race and ethnicity input variables in their algorithms. Notable examples

include recommendations to remove race from pediatric urinary tract infection treatment guidelines[10,19] and to remove race considerations in hypertension management guidelines.[28-31] Moreover, little is currently known about how algorithms that do not explicitly include race and ethnicity input variables may affect racial and ethnic health and healthcare disparities.

1.2 Purpose and Scope of the Review

In September 2020, the Agency for Healthcare Research and Quality (AHRQ) received a request from Congress to review the evidence on the use of race and ethnicity in clinical algorithms and the potential of algorithms to contribute to disparities in healthcare, and commissioned this evidence review. AHRQ also issued a public Request for Information[32] that generated responses related to algorithms from 42 organizations, agencies, and individuals.[33] This evidence review is intended to:

- Examine how algorithms, with or without race and ethnicity as input variables, affect racial and ethnic differences in access to care, quality of care, and health outcomes.
- Evaluate strategies to mitigate any racial and ethnic bias in the development and use of algorithms.
- Explore contextual factors, including the role of algorithm developers and end-users; identify available or emerging guidance on preventing racial and ethnic bias during algorithm development; clarify stakeholder awareness of and perspectives on potentially racially and ethnically biased algorithms; and determine incentives and barriers affecting use and evaluation of algorithms.

With guidance from AHRQ and in collaboration with Subject Matter Experts (SMEs), Key Informants (KIs), and Technical Expert Panel (TEP) members, we developed two Key Questions (KQs) and four Contextual Questions (CQs). KQ 1 assesses the effects of algorithms on racial and ethnic disparities in health and healthcare. An algorithm might create, perpetuate, exacerbate, reduce, or have no effect on health and healthcare disparities. We excluded studies of algorithms that did not examine their effect on disparities. KQ 2 focuses on strategies to mitigate algorithmic bias. Focal points for mitigation could be on datasets used to develop or train algorithms, input variables included in algorithms, or processes for validating, implementing, disseminating, or adapting algorithms. KQ 2 includes studies of algorithms that were redesigned to mitigate algorithmic bias in response to prior associations between these algorithms and racial and ethnic health or healthcare disparities. We aimed to identify and describe strategies to address potential racial and ethnic algorithmic bias and evaluate the effect on racial and ethnic health and healthcare disparities. To address KQs 1 and 2, we conducted a systematic literature search.

The CQs were designed to explore the context and capture insights on practical aspects of these issues. CQ 1 examines the problem's scope within healthcare. CQ 2 addresses recently emerging standards and guidance, from within and outside healthcare, on how racial and ethnic bias can be prevented or mitigated when healthcare and non-healthcare algorithms are developed and deployed. CQ 3 explores key stakeholders' knowledge and perspectives about the interaction of algorithms and racial and ethnic disparities in health and healthcare. CQ 4 conducted an in-depth analysis of a sample of six healthcare algorithms not previously widely evaluated in published literature to understand how their design and implementation might contribute to racial and ethnic health and healthcare disparities.

To address CQs 1, 2, and 3, we searched for studies, standards, frameworks, white papers, and other relevant resources and sought input from our SMEs, KIs, and TEP members. As a

supplement to KQs 1 and 2, a separate conceptual model was developed for CQ 4, and an objective four-step, *a priori* approach was used to identify the sample of algorithms currently in use whose effects on racial and ethnic health and healthcare disparities were not previously studied.

The conceptual model for CQ 4 was motivated by the fact that although the selected algorithms are being used to make clinical decisions, little is known about their development, stakeholders involved in development, validation, performance testing, translation and implementation into clinical practice, and process for updating. We therefore conducted a deeper analysis to understand these considerations, using a representative sample of algorithms that were likely to affect large populations. We focused on algorithms not identified in KQs 1 and 2 to contextualize the development, validation, and impact of algorithms. This evaluation is complementary to findings from the KQs and may provide a fuller understanding of the opportunities and challenges faced by policymakers and healthcare providers for more just and equitable care with the use of algorithms.

Findings from this review are intended to inform: (1) policymakers, providers, payers, health systems, and patients seeking to understand or address the role of algorithms in racial and ethnic health and healthcare disparities; (2) future research opportunities exploring the effects of algorithms on racial and ethnic health and healthcare disparities; and (3) current and emerging guidance and best practices for developing, validating, implementing, and evaluating algorithms to mitigate potential racial and ethnic bias.

For this report, we define commonly used terms in Table 1.

Table 1. Definition of terms and report-specific considerations

Category	Terms and Considerations	Definition
Disparities & inequities	Racial and ethnic health disparities	Differences in measures of health, such as burden of disease and health outcomes, between various racial and ethnic populations.[34,35]
	Racial and ethnic healthcare disparities	Differences in healthcare, such as provision and quality of care and/or treatment, between various racial and ethnic populations after accounting for equal access to care, clinical need, and patient preferences.[34,35]
	Inequities	Disparities that result from structural, institutional, and/or interpersonal biases that contribute to broad imbalances in power, justice, social structures, or resources.[34,35]
	Report-specific considerations	Throughout this report, we describe racial and ethnic differences as disparities rather than inequities because, when synthesizing a wide variety of studies that differ in specific details and contexts, "disparities" serves as a more inclusive and therefore accurate summary-level description. Nevertheless, algorithms can and do contribute to inequities as well.
Algorithms & mitigation strategies	Algorithm	Mathematical formulas or models that combine different input variables or factors to inform a calculation or an estimate, such as an estimate of disease or risk of a particular health outcome.
	Mitigation strategy	An approach to reduce racial and ethnic bias in algorithm development, validation, dissemination, and implementation.
	Report-specific considerations	Algorithms are often incorporated into healthcare decision tools, such as clinical guidelines, pathways, clinical decision support programs in electronic health records, and operational systems used by health systems and payers; our use of "algorithm" includes such tools. In this report, "algorithm" refers to healthcare algorithms unless otherwise specified. Mitigation strategies refer to approaches that focus on the components of an algorithm or the processes of development, validation, dissemination, or implementation.

Category	Terms and Considerations	Definition
Bias	Racial and ethnic bias	Racial and ethnic bias is an umbrella term encompassing 1) structural and institutional bias, or bias intrinsic to the design of societal and institutional, structures, and systems; and/or 2) interpersonal bias and or discrimination that act in concert to create unearned advantages or disadvantages in populations depending on their race and ethnicity.[12,36,37] In this context of algorithms, these biases can influence an algorithm's development, validation, translation, and dissemination (Table 3, Figure 2) to differentially impact populations by race and ethnicity.[35] This includes implicit and explicit bias intrinsic to algorithm developers, clinicians, and policymakers and bias intrinsic to the process of validating and operationalizing the algorithm.
	Algorithmic bias	Refers narrowly to attributes intrinsic to an algorithm that may result in differential model performance in different groups.[38]
	Risk of bias	This term refers to a methodologic assessment of the risk of bias in a study, also known as quality assessment or critical appraisal. This determines the potential for the reported results to be misleading due to methodologic issues in study design.
	Report-specific considerations	Racial and ethnic bias can contribute to algorithmic bias at the development stage and can also occur at the implementation phase due to provider implicit or explicit bias and therefore can impact the effect of algorithms discussed in both Key Question 1 and 2. The studies included in this report provide estimates of algorithmic bias; our risk-of-bias assessment is an indication of the likelihood that those estimates may be wrong due to limitations in study design. Please see the Methods section below for further details.

1.3 Key Questions

Key Question 1. What is the effect of healthcare algorithms on racial and ethnic differences in access to care, quality of care, and health outcomes?

Key Question 2. What is the effect of interventions, models of interventions, or other approaches to mitigate racial and ethnic bias in the development, validation, dissemination, and implementation of healthcare algorithms?

a. Datasets: What is the effect of interventions, models of interventions, or approaches to mitigate racial and ethnic bias in datasets used for development and validation of algorithms?

b. Algorithms: What is the effect of interventions, models of interventions, or approaches to mitigate racial and ethnic bias produced by algorithms or their dissemination and implementation?

1.4 Contextual Questions

Contextual Question 1. How widespread is the inclusion of input variables based on race and ethnicity in healthcare algorithms?

a. What types of algorithms used in healthcare include input variables based on race and ethnicity? How widely are they used?

b. Who develops algorithms used in healthcare that might include input variables based on race and ethnicity?
c. Who are the end-users of these algorithms used in healthcare? What incentives and barriers are there to implementation or de-implementation?
d. What patient populations are included?
e. What clinical conditions, processes of care, and healthcare settings are included?

Contextual Question 2. What are existing and emerging national or international standards or guidance for how algorithms should be developed, validated, implemented, and updated to avoid introducing bias that could lead to health and healthcare disparities?

a. Within these standards or guidance, what are the recommendations about the use of input variables or datasets that include race and ethnicity to develop or validate algorithms?
b. What are the recommendations about input variables used or sought in place of race and ethnicity (e.g., genetic markers and biomarkers, social determinants of health, the experience of individual and structural racism), including standards or guidance for how to define and collect data on these variables, and their impact on exacerbating or mitigating bias?
c. What are the recommendations for identifying and addressing other types of input variables that could introduce bias leading to disparities, such as measures of healthcare use or SDOH?
d. What are the recommendations regarding transparency or disclosure of information related to algorithm development, validation, use, and outcomes?

Contextual Question 3. To what extent are patients, providers (e.g., clinicians, hospitals, health systems), payers (e.g., insurers, employers), and policymakers (e.g., healthcare and insurance regulators, State Medicaid directors) aware of the inclusion of input variables based on race and ethnicity in healthcare algorithms?

a. Is there evidence of how these types of algorithms might contribute to biases in provider and payer perceptions of affected populations and their clinical care?

Contextual Question 4. Select a sample of approximately 5-10 healthcare algorithms that have the potential to impact racial and ethnic disparities in access to care, quality of care, or health outcomes and are not included in KQs 1 or 2. For each algorithm, describe the type of algorithm, its purpose (e.g., screening, risk prediction, diagnosis), its developer and intended end-users, affected patient population, clinical condition or process of care, healthcare setting, and information on outcomes, if available. This question's intent is to consider the use of healthcare algorithms that may be perpetuating racial and ethnic bias but have not been previously linked to disparities in health or healthcare.

a. If race and ethnicity is included as an input variable, how is it defined? Are definitions consistent with available standards, guidance, or important considerations identified in CQ 2?

b. For healthcare algorithms that include other input variables in place of or associated with race and ethnicity, how were these other variables defined? Are these definitions consistent with available standards, guidance, or important considerations as identified in CQ 2? Were racial and ethnic variables considered during initial development or validation?

c. For each healthcare algorithm, what methods were used for development and validation? What evidence, evidence quality, data sources, and study populations were used for development and validation?

d. Are development and validation methods consistent with available standards, guidance, and strategies to mitigate bias and reduce the potential of healthcare algorithms to contribute to health disparities?

e. What approaches and practices are there to implement, adapt, or update each healthcare algorithm?

1.5 Organization of This Report

In the following Methods section, we describe in detail the methods used to address the KQs and CQs. In the Results section, we first provide results of the literature searches for KQ 1 and KQ 2. This includes descriptions of eligible research studies, key points, and syntheses of findings. Results for CQs 1 through 4 follow KQ results. The Discussion section reviews key findings, examines general applicability of the findings, identifies evidence gaps, and describes strengths and limitations of the evidence review and evidence base. The report's main body is followed by six appendixes: Appendix A. Methods for Search Strategy; Appendix B. List of Excluded Studies; Appendix C. Characteristics of Key Question 1 and 2 Included Studies; Appendix D. Key Question 1 and 2 Evidence Tables; Appendix E. Contextual Question 4 Detailed Supplement; Appendix F. Appendix References.

2. Methods

This evidence review followed methods outlined in the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) Methods Guide for Effectiveness and Comparative Effectiveness Reviews (hereafter the "AHRQ Methods Guide") for Key Questions (KQs) 1 and 2.[39] We determined most methods *a priori* but needed additional input from Subject Matter Experts (SMEs) to finalize methods. A protocol was developed through a process that included collaborating with a Technical Expert Panel (TEP), Key Informants (KIs), federal partners, and public input on KQs and study eligibility criteria. For additional details, see the review protocol posted on the AHRQ Effective Health Care Program website (https://effectivehealthcare.ahrq.gov/products/racial-disparities-health-healthcare/protocol). The protocol was registered with PROSPERO (CRD42022335090).

2.1 Methods To Address Key Questions

2.1.1 Classification of Studies by Key Question

Studies were included in KQ 1 if they evaluated an algorithm's effect on health or healthcare outcomes stratified by racial and ethnic groups (i.e., studies reporting only model fit and accuracy were excluded). Studies were included in KQ 2 if they examined a strategy's ability to mitigate 1) racial and ethnic algorithmic bias or 2) a known racial and ethnic disparity associated with an algorithm. Studies that described both a racial and ethnic disparity associated with an algorithm, and an intervention on the algorithm to mitigate the disparity, were included in both KQ 1 and KQ 2.

2.1.2 Literature Search Strategies for Key Questions

EPC information specialists conducted a comprehensive literature search following established systematic review protocols. We searched the following databases using controlled vocabulary and text words from January 1, 2011, to February 7, 2023: Embase and Medline (via embase.com), PubMed (in-process citations to capture items not yet indexed in Medline), and the Cochrane Library. Based on guidance from SMEs, KIs, and TEP members, articles published before 2011 were considered unlikely to be contemporaneous to current algorithms. The search strategy included controlled vocabulary terms (e.g., MeSH, EMTREE), along with free-text words, related to race, ethnicity, algorithms, disparities, and inequities. Searches used a hedge to remove conference abstracts, editorials, letters, and news items; however, we retained some of these items in the final search to help inform the Contextual Questions. Information specialists independently peer reviewed searches using the Peer Review of Electronic Search Strategies Checklist. The search strategy for Embase and Medline is included in Appendix A. We also reviewed submissions to AHRQ's Supplemental Evidence and Data portal to identify other studies meeting protocol eligibility criteria.

Information specialists also conducted grey literature searches of the following resources: Association for Computing Machinery Digital Archives, medRxiv and bioRxiv preprint servers, ClinicalTrials.gov, and websites of relevant organizations (e.g., AHRQ, American Actuarial Association, American Hospital Association Institute for Diversity and Health Equity, American Medical Informatics Association, Centers for Disease Control and Prevention, Consumer Financial Protection Bureau, Healthcare Information and Management Systems Society, U.S. Food and Drug Administration, Health Resources and Services Administration, National Institute

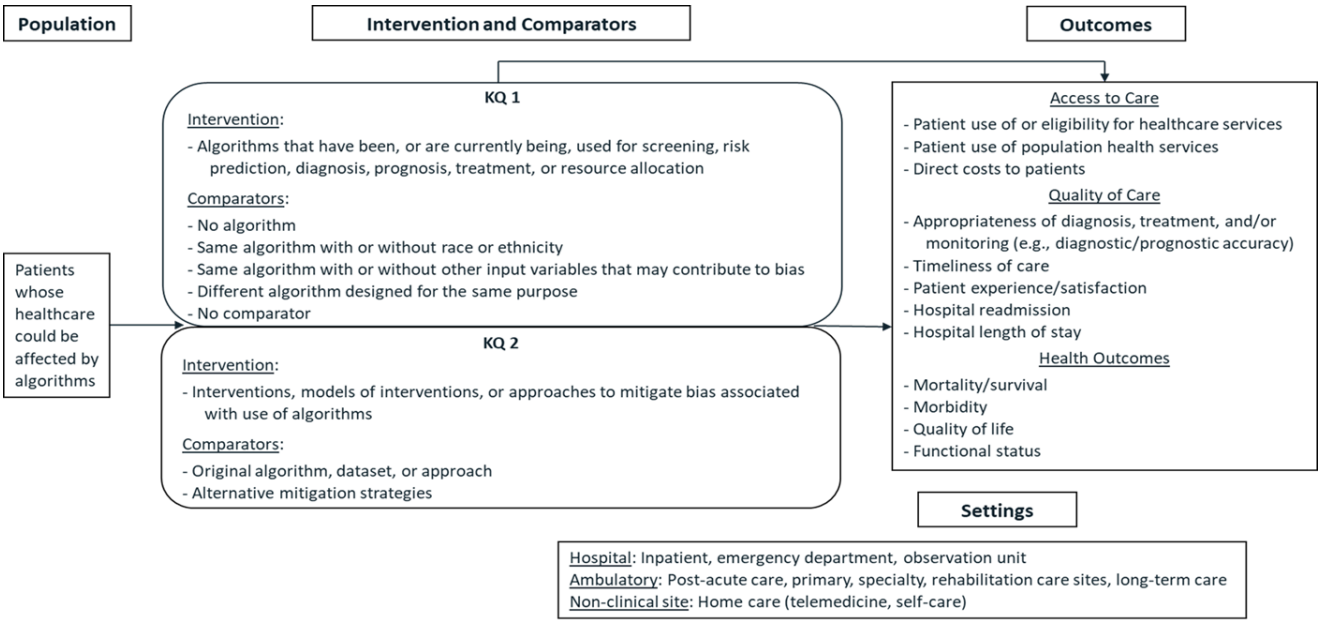
of Standards and Technology, Office of the National Coordinator for Health Information Technology, Observational Health Data Sciences and Informatics, and others as recommended by SMEs and TEP). We hand-searched published systematic reviews to identify any studies missed by our searches. Scopus was also used to identify related publications through citation tracking.

We screened eligible records using DistillerSR (Evidence Partners, Ottawa, Ontario, Canada). After screening titles, two analysts independently screened each abstract in duplicate for eligibility. We then retrieved eligible full-text articles screened for final eligibility, again in duplicate. All disagreements were resolved by consensus discussion between the two duplicate screeners.

2.1.3 Analytic Framework for Key Questions

KQs were addressed by a systematic review of published studies and grey literature. Figure 1 presents the analytic framework that displays the interaction between major components of the evidence base, organized according to the PICOTS (Population, Intervention, Comparator, Outcome, Timing, Setting) model.

Figure 1. Analytic framework for Key Questions



Abbreviations: KQ = Key Question

2.1.4 Inclusion and Exclusion Criteria for Key Questions

As suggested in the AHRQ Methods Guide,[39] we list eligibility criteria in several categories: publication type, study design, intervention characteristics, setting, and outcome.

2.1.4.1 Publication Criteria

1. We did not include abstracts or meeting presentations, which do not provide sufficient details about experimental methods to permit an evaluation of study design and conduct; they may also contain only a subset of measured outcomes.[40,41] Also, abstracts that are published as part of conference proceedings can have inconsistencies compared with the final study publication or may describe studies that are never published as full articles.[42-45]
2. We included studies published from 2011 to the present. Based on guidance from subject matter and technical experts, earlier articles were considered unlikely to be contemporaneous to current algorithms.
3. To avoid double-counting patients, when several reports of overlapping patients are available, we included outcome data only from the report with the most patients. We included data from a smaller, overlapping publication when it reported data on different racial and ethnic group(s), included an outcome not provided by the larger report, or reported longer follow-up data for an outcome.
4. This review's timeline did not permit translation of non-English-language articles.

2.1.4.2 Study Design Criteria

1. We included only full-length research studies; thus, we excluded narrative reviews, letters, guidelines, position statements, and commentaries. We used systematic reviews only to identify individual studies as a supplement to the full literature search (described above in the Literature Search Strategy).
2. We considered any study design with a relevant comparison or no comparator, as described in Table 2.
3. We included studies with prospective or retrospective patient identification or studies that modeled potential outcomes. Modeling studies used real-world or synthetic source data for calculation of algorithmic scores and outcomes that would have resulted from using the algorithm were simulated.
4. For KQ 1, the study must have measured an algorithm's effect on racial and ethnic disparities. For KQ 2, the study must have measured a mitigation strategy's effect.

2.1.4.3 Intervention Criteria

1. To be considered an "algorithm," a mathematical formula or model must combine different input variables or factors to produce a numerical score or scaled ranking or populate a classification scheme that may be used to guide healthcare decisions. We also included studies of algorithm-informed decision support tools, defined as any clinical guideline, pathway, clinical decision support intervention in an electronic health record (EHR), or operational system used by health systems and payers that is informed by an algorithm as defined above. We did not require that an algorithm explicitly use race or ethnicity as an input.
2. For KQ 1, the algorithm must have been applied to a patient/participant population other than the derivation population. We excluded newly developed algorithms evaluated only in a derivation population.
3. Three studies directly evaluated both the effect of an algorithm on racial and ethnic disparities and strategies to mitigate racial and ethnic bias; therefore, relevant data

from these studies were summarized within both KQs. Additionally, a few studies were applicable primarily to one of the KQs while indirectly addressing the other KQ. These studies were analyzed with the most appropriate KQ following consensus discussion among reviewers.

2.1.4.4 Setting Criteria

1. For representativeness, we included only studies of patients in the United States for KQ 1. For KQ 2, we did not restrict by country as strategies to mitigate bias outside the United States may be generalizable to settings in the United States.
2. We included any study conducted in a clinical or nonclinical site, as described in Table 2.

2.1.4.5 Outcome Criteria

1. For KQ 1, a study must have evaluated whether an algorithm affects a racial or ethnic disparity in outcomes. Studies must have reported outcomes separately for two or more races or ethnicities. For KQ 2, we allowed studies that reported outcomes for only one race or ethnicity. We did not require that reported effect sizes be statistically significant or that a study control or adjust for possible confounders (confounding is addressed in our narrative appraisal of the evidence).
2. For KQ 1, the study must have reported health or healthcare outcomes. Studies that reported only diagnostic or prognostic accuracy without specifying clinical implications were excluded.
3. For both KQs, a study must have reported race and ethnicity-based outcomes in at least one of three outcome categories (access to care, quality of care, and health outcomes).

Table 2 presents criteria that guided study eligibility and categorization of outcomes, organized according to the PICOTS framework.

Table 2. PICOTS (Population, Intervention, Comparator, Outcome, Timing, Setting) for Key Questions 1 and 2

Category	Definition
Population	Patients whose healthcare could be affected by algorithms (including algorithm-informed decision tools, e.g., clinical guidelines, pathways, clinical decision support programs in EHRs, operational systems used by health systems and payers).
Interventions/ Exposures	KQ 1: Algorithms that have been, or are currently being used for screening, risk prediction, diagnosis, prognosis, treatment, or resource allocation. They do not have to use race and ethnicity variables as inputs. KQ 2a: Interventions, models of interventions, or approaches to mitigate bias in the datasets used to develop or validate algorithms. KQ 2b: Interventions, models of interventions, or approaches to mitigate bias associated with use of algorithms. These strategies could focus on the components of an algorithm or the processes of development, validation, dissemination, or implementation.

Category	Definition
Comparators	KQ 1: Appropriate comparators include: • No algorithm • Same algorithm with or without race and ethnicity variables • Same algorithm with or without other input variable(s) that may contribute to bias (e.g., prior utilization, socioeconomic status, SDOH) • Different algorithm designed for the same clinical purpose, with or without input variable(s) based on race and ethnicity • No comparator (e.g., studies comparing outcomes of a single algorithm across individuals in different racial and ethnic categories) KQ 2: • Original algorithm, dataset, approach • Alternative mitigation strategies
Outcomes	Outcomes must be reported by race and ethnicity (2+ racial and ethnic groups for KQ 1, but could be just 1 group for KQ 2) <u>Access to care</u> • Patient use of or eligibility for healthcare services (e.g., primary care visits, specialty referrals and visits, emergency department visits, hospitalizations, post-acute care, medication use) • Patient use of population health services (e.g., screening, preventive care, chronic disease management) • Direct costs to patients <u>Quality of care</u> • Appropriateness of diagnosis, treatment, and/or monitoring • Timeliness of care • Patient experience/satisfaction • Hospital readmission • Hospital length of stay <u>Health outcomes</u> • Mortality / Survival • Morbidity • Quality of life • Functional status
Timing	No minimum follow-up
Setting	<u>Hospital care</u> • Inpatient • Emergency department • Observation unit <u>Non-hospital care</u> • Post-acute care, primary, specialty, rehabilitation care sites • Long-term care (e.g., assisted living facilities, nursing homes) <u>Non-clinical sites</u> • Home care (e.g., telemedicine, self-care) For KQ 1, studies conducted in populations outside the United States will be excluded.

Abbreviations: EHRs = electronic health records; KQ = Key Question; SDOH = social determinants of health

2.1.5 Data Abstraction and Data Management

Data were extracted into Microsoft Word and/or Excel. Elements abstracted included general study characteristics (e.g., study design, setting, number of patients enrolled), patient characteristics (e.g., age, sex, race and ethnicity, clinical condition), intervention details (e.g., study objective, type of algorithm, intent of algorithm, input variables used, data sources), and outcome data.

2.1.6 Assessment of Methodologic Risk of Bias and Data Synthesis

Some included studies were prediction modeling studies, so we first considered using PROBAST (Prediction model study Risk Of Bias Assessment Tool) to assess risk of bias (ROB).[46] After piloting PROBAST in our evidence base, we determined it was not applicable because our KQs addressed the *effect* of algorithms (KQ 1) or the *effect* of mitigation strategies (KQ 2) on clinical outcomes, which are not considered in PROBAST, which focuses on algorithm development. Therefore, based on EPC guidance,[39] we focused ROB assessment on how well a study measured the true effect of algorithms or mitigation strategies (neither overestimates nor underestimates). While a randomized controlled trial (RCT) would be the ideal design to measure this, only one of the included studies was an RCT. No ROB tools existed for studies of the effect of algorithms, so we used an existing tool, ROBINS-I (Risk Of Bias in Non-randomized Studies of Interventions) to assess ROB.[47] Using this tool involves rating a study's ROB on each of seven domains, listed below, and then combining the domain-specific ratings to categorize the study as being at Low, Moderate, or High ROB. The domains are as follows:

- Bias due to confounding
- Bias in selection of participants into the study
- Bias in classification of interventions
- Bias due to deviations from intended interventions
- Bias due to missing data
- Bias in measurement of outcomes
- Bias in selection of the reported result

Based on feedback from our TEP, there was consensus that studies evaluating algorithms' effects on racial and ethnic disparities should undergo ROB assessment in the context of several racial and ethnic-specific factors. Therefore, there was a need to incorporate racial and ethnic equity-related considerations as part of ROB assessment. For four of the seven ROBINS-I domains, we used additional ROB signaling questions related to racial and ethnic health equity, adapted from a prior AHRQ project by another EPC:[38]

- Bias due to confounding domain: "Was a transparent rationale provided for including or removing race and ethnicity as an input variable?"
- Bias in selection of participants into the study domain: "Were data on racial and ethnic groups gathered using consistent definitions or categories with adequate response options?"
- Bias due to missing data domain: "Were there sufficient outcomes occurring in specific racial and ethnic groups to assess model performance separately in these groups?"
- Bias in measurement of outcomes domain: "Were relevant model performance measures evaluated appropriately in racial and ethnic groups?"

A study was deemed at overall High ROB if any single domain (also considering the health equity signaling question in that domain) was judged to be at High ROB. A study was deemed at overall Low ROB if all domains were Low ROB. All others were moderate ROB.

Given variation in study designs of included studies, an acceptable response to the racial and ethnic equity-based signaling questions was "Not Applicable (N/A)." N/A ratings would not affect ROB assessment. Further, the racial and ethnic health equity question in the "bias due to missing data" domain was applied only to studies of algorithm derivation and internal validation. Therefore, this question was usually N/A for studies addressing KQ 1, all of which examined established, previously validated algorithms. Similarly, the health equity question added to the

"bias due to measurement of outcomes" domain was restricted to model performance measures addressing discrimination and calibration outcomes. For KQ 1, only studies evaluating the clinical effects of algorithm use rather than measures of discrimination and calibration were eligible. Therefore, this health equity signaling question was N/A to studies addressing KQ 1. If a study reported both model performance outcomes (e.g., model calibration) and clinical outcomes, we reported only the latter.

For KQ 1, we organized the evidence into various clinical assessment categories, described the purpose of the identified algorithms, and narratively summarized the evidence with a focus on three potential results of algorithms: exacerbation or introduction of race and ethnicity disparities, reduction of existing racial and ethnic disparities, or a report of no discernible effect related to race and ethnicity. For KQ 2, synthesis focused on the various types of mitigation strategies identified. We analyzed the extent of different mitigation approaches, examined and classified their key features, reviewed evidence of their effectiveness when available, and summarized interventions and approaches identified for mitigation of racial and ethnic bias.

2.2 Methods for Contextual Questions

2.2.1 Contextual Questions 1-3

In addition to the literature searches conducted to address the KQs, we conducted supplemental searches to identify studies, standards, frameworks, white papers, and other relevant resources that addressed Contextual Questions (CQs) 1, 2, and 3. We also reviewed responses to AHRQ's Request for Information (RFI)[32] and discussions with SMEs, TEP, and KIs to inform our analysis of the CQs.

2.2.2 Contextual Question 4

The algorithm development-to-clinical implementation lifecycle involves multiple steps, each of which has the potential to introduce racial and ethnic bias. The conceptual model in Figure 2 guided our analysis and helped describe and summarize mechanisms through which racial and ethnic bias can be introduced and result in disparities in access, quality, and health outcomes. This conceptual model is informed by the Sociotechnical Model for Studying Health Information Technology in Complex Adaptive Healthcare Systems[48] and the conceptual model for biases in healthcare proposed by Rajkomar et al.[49]

Race and ethnicity biases can be introduced at any step in the algorithm development-to-implementation process. Figure 2 organizes this process into two major steps: algorithm development (Figure 2a) and algorithm translation, dissemination, and implementation (2b). Table 3 details potential racial and ethnic biases that can be introduced during the algorithm development phase.

2. Methods

Figure 2. Conceptual model for understanding racial and ethnic biases introduced during algorithm/clinical decision-making tool development, translation, dissemination, and implementation

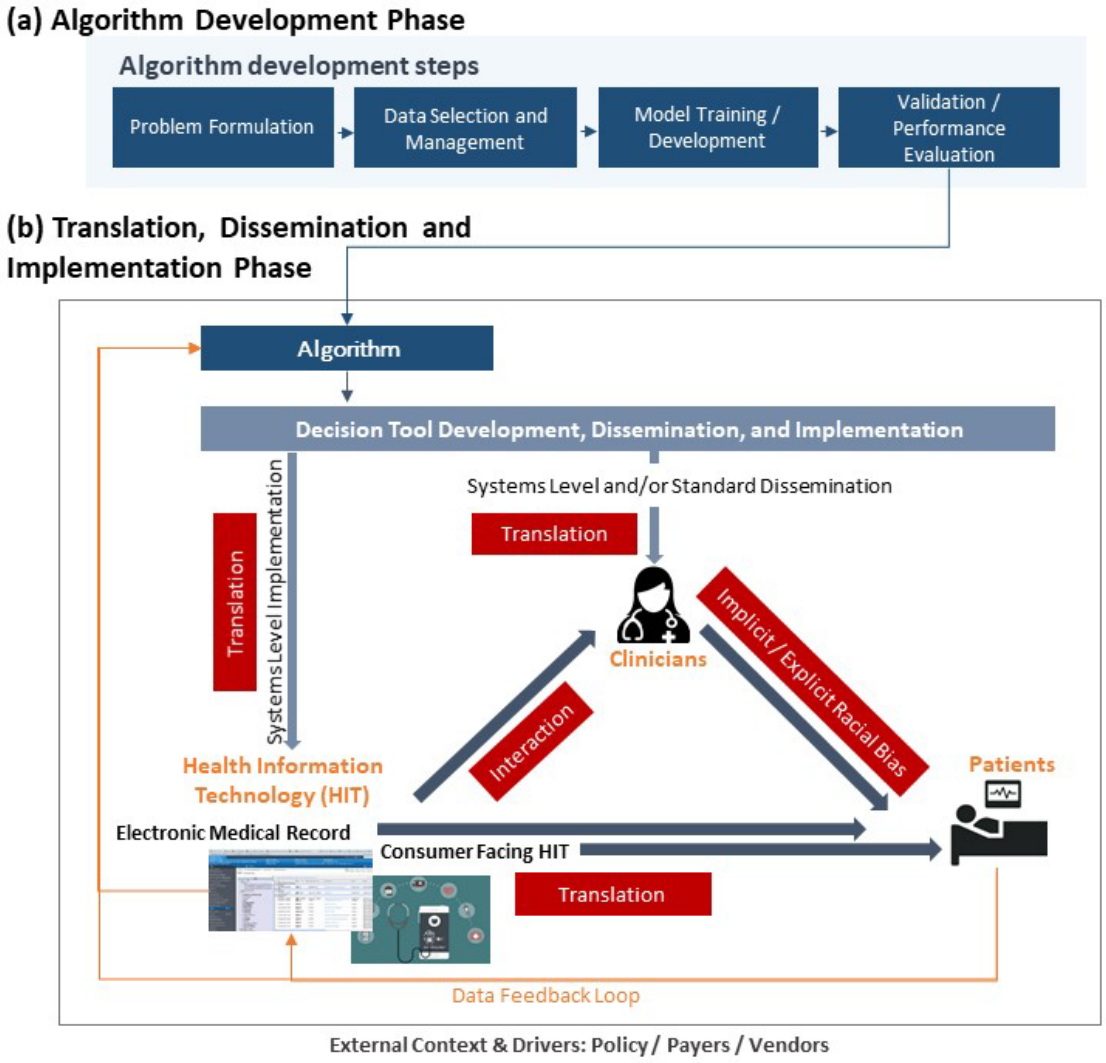


Table 3. Examples of racial and ethnic biases that can be introduced during algorithm development

Algorithm Development Phase	Examples of Biases
Data Selection and Management	• Biases in study inclusion criteria (e.g., using estimated glomerular filtration rate to select study participants) • Study data collection biases (including misclassification of race and ethnicity) • Lack of representation / selection in the dataset • Missing data • Biases in imputed data • Biases in learning and training data • Collapsing race variables • Lack of reporting for methodologic approach • Insufficient sample size • Labeling bias
Model Training/ Development	• Overfitting • Lack of reporting for methodologic approach • Interpretation bias • Correlation bias
Validation/ Evaluation	• Training-validation data skew • Lack of external validation • Lack of performance assessments, such as calibration and discrimination • Lack of reporting for methodologic approach

Racial and ethnic biases can be introduced de-novo during dissemination and implementation or carried over from the development phase. Dissemination focuses on the spreading of knowledge and evidence by passively *informing* audiences. Implementation is a more active initiative that focuses on *integrating* and incorporating guidance into clinical workflow, often with technological support. We outlined three vulnerabilities in which racial and ethnic bias can be newly introduced during implementation (Figure 2[b]). Racial and ethnic bias can be introduced first during **translation,** which is the process of operationalizing algorithms into decision tools or clinical processes. **Interaction** with an algorithm can result in racial and ethnic bias when a clinician is presented with guidance during care but chooses not to act. **Implicit/explicit bias** might occur, for example, when a clinician determines, on behalf of a mixed-race patient, which race-category to document in an EHR. Use of consumer-facing health information technology (HIT) may contribute to additional racial and ethnic biases. Examples are HIT design and language choices that do not account for differences in healthcare literacy, numeracy, and language. Furthermore, racial and ethnic bias can result when an algorithm is not updated as the evidence base evolves or changes.

The method by which algorithms are disseminated and implemented provides additional vulnerabilities for introduction of racial and ethnic bias. We organized dissemination and implementation methods into hierarchical tiers, each based on the increased impact on outcomes. **Standard dissemination** is defined as non-HIT-supported methods for providing guidance to clinicians. Standard dissemination requires a clinician to be aware of the existence of guidance,

understand the guidance and patient applicability, and understand how to integrate guidance into care. **Systems-level dissemination** is defined by the use of HIT to reach clinicians, such as through a cloud-based clinical pathways library.[50] This has a potentially larger impact on outcomes than standard dissemination, as use of HIT may increase the number of clinicians who use the algorithm.[51] **Systems-level implementation** is defined as the translation and integration of algorithms into clinical workflow, to display guidance at the right time, through the right system, to the right person, and in the right format to have the greatest likelihood to affect patient care and outcomes.

Racial and ethnic biases introduced during algorithm development can also be amplified, such as when an algorithm is incorporated in an EHR and clinicians interpret algorithm-based guidance with implicit or explicit biases. The magnitude and impact of racial and ethnic biases depend on the dissemination and implementation method (Figure 2[b]) as well as the interaction between the clinician user, dissemination and implementation method, and patient.

To inform CQ 4, we identified six algorithms, not evaluated in the studies included in KQs 1 or 2, to examine their potential impact on health and healthcare disparities. In selecting the algorithms, we considered a variety of patient populations, clinical conditions, types of algorithms, settings, and end-users. We prioritized algorithms by considering disease prevalence and burden in addition to conditions for which racial and ethnic disparities in healthcare and/or health outcomes are well-documented.

To assess the potential effects of the algorithms identified for CQ 4, we examined the health and intermediate outcomes delineated in Table 2. We also described development and validation methods and reported algorithm accuracy measures. We also documented whether algorithm developers explicitly considered potential racial and ethnic bias (e.g., by examining algorithm performance by race and ethnicity) or used any strategies that might mitigate racial and ethnic bias. Finally, we described key components of dissemination and implementation strategies used by algorithm developers and end-users and estimated the effects of these dynamics on racial and ethnic disparities. Findings for CQ 4 are available in the Results section of this report.

2.2.2.1 CQ 4 Sample Algorithm Identification and Selection

We employed five distinct approaches for identifying sample algorithms for CQ 4. Figure 3, step 1, depicts the flow and organization for these activities. First, we identified conditions with the highest disease burden and/or extreme racial and ethnic disparities in outcomes by examining available sources, such as the Centers for Disease Control and Prevention (CDC) mortality and morbidity reports and AHRQ's *National Healthcare Quality and Disparities Reports.*[52,53] Second, we reviewed findings of the searches for the KQs and performed supplemental searches as needed to identify algorithms and studies relevant to these conditions. Third, we reviewed our discussions with KIs, SMEs, and the TEP related to specific algorithms recommended for inclusion. We contacted select experts for follow-up when needed. Fourth, we reviewed responses to the RFI[32] and public posting of the KQs. Fifth, we queried select vendors to identify critical or high-use algorithms.

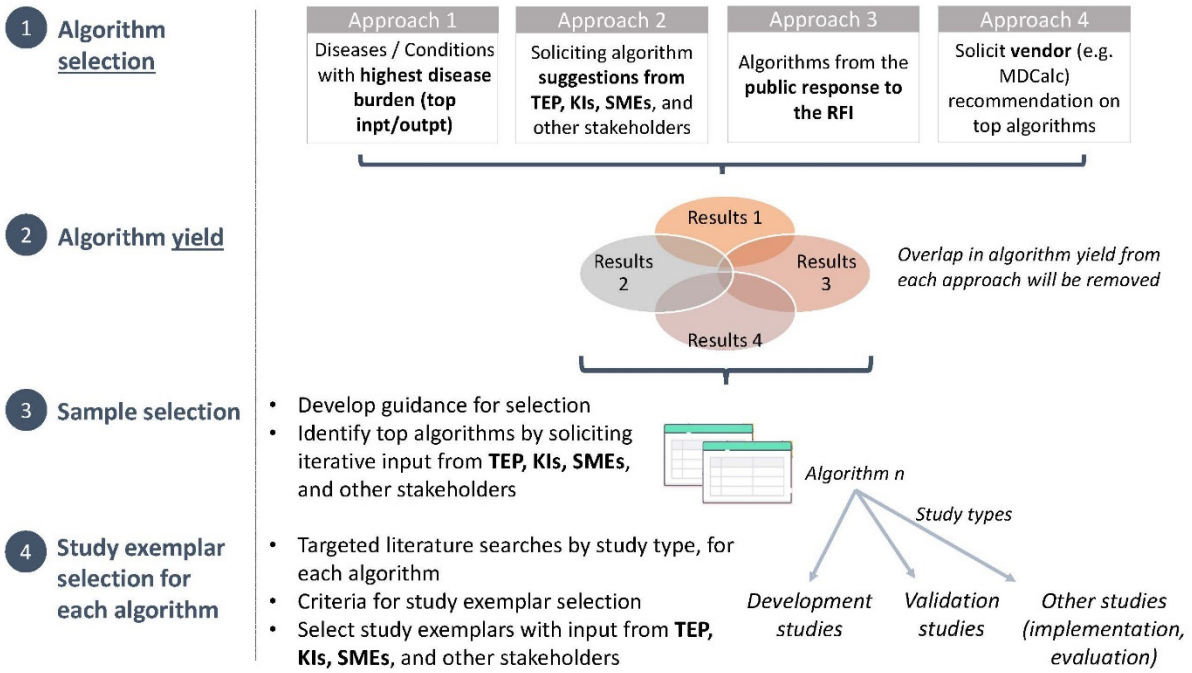
Results from each of the algorithm selection approaches were collated and duplicates removed. We constructed a database of algorithms from this pool and added key data, such as type of algorithm, intent of algorithm, developer/vendor, intended user, patient population, clinical condition, setting, and anticipated evidence base (e.g., citations). We used an iterative, consensus-driven approach to select the final six samples. Finally, we identified relevant and

representative exemplars by study type (e.g., development, validation, implementation, comparative effectiveness) for each algorithm in the sample.

2.2.2.2 Data Abstraction

For each algorithm in the CQ 4 sample, we abstracted technical specifications, such as input variables used, datasets used for development and validation, and types of outcomes produced. We also included, when available, details about processes used for development and validation along with outcome data. Finally, we documented, when possible, information about the extent of use in clinical practice, dissemination, and implementation activities (e.g., incorporated in a guideline or EHR), and years in use or since publication. Additional variables were abstracted depending on findings.

Figure 3. Framework for sample algorithm selection and data abstraction



Abbreviations: KI = Key Informant; RFI = Request for Information; SME = Subject Matter Expert; TEP = Technical Expert Panel

2.2.2.3 Algorithm Evaluation

Each sample algorithm was evaluated qualitatively and quantitatively, as feasible, to determine the likelihood of contributing to racial and ethnic disparities. We used existing evaluation tools, identified emerging standards, and identified gaps and deficiencies with our SMEs, TEP, KIs, and other stakeholders related to assessing racial and ethnic bias in algorithms. Descriptive data for each algorithm were summarized.

2.3 Peer Review and Public Commentary

Experts in clinical care, health equity, and bioinformatics along with individuals representing stakeholder and user communities provided external peer review of this evidence review; AHRQ

and an EPC program associate editor also reviewed draft reports. The draft report was posted on the AHRQ website for 4 weeks to elicit public comment. A disposition of comments table will be posted on the EHC website 3 months after AHRQ posts the final systematic review.

3. Results

3.1 Overview

To address the Key Questions (KQs), electronic searches for published scientific studies identified 11,500 citations. After we screened titles and abstracts, 336 articles were deemed eligible for full-text review and evaluated for KQ 1 and/or KQ 2 eligibility. After full-text review, 58 articles met inclusion criteria:

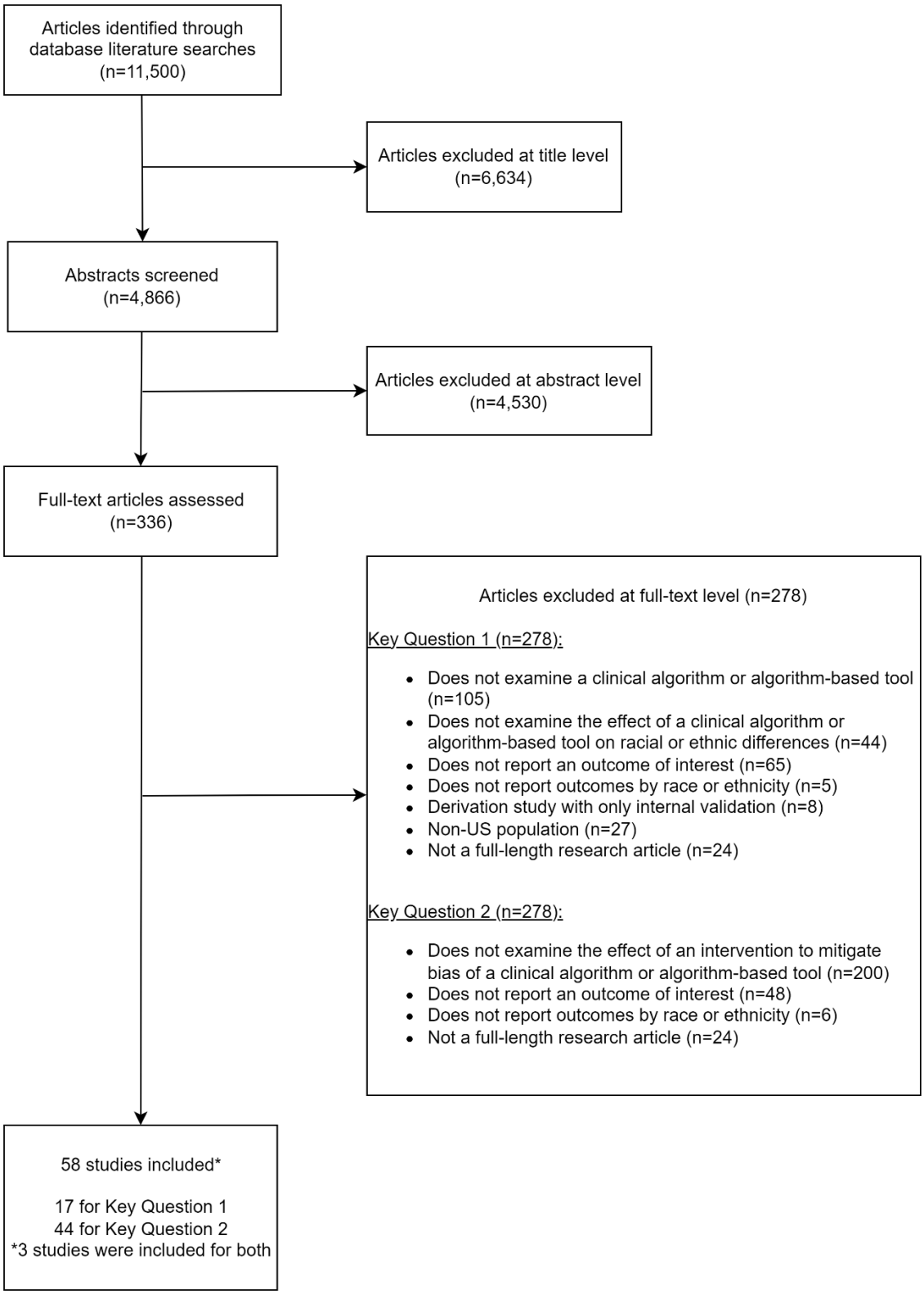
- Fourteen studies[54-67] examined the effect of algorithms on racial and ethnic disparities and did not evaluate potential mitigation strategies and therefore addressed *only* KQ 1.
- Forty-one studies[21,23,68-106] focused on strategies to mitigate racial and ethnic bias or known racial and ethnic disparities associated with healthcare algorithms and thus addressed *only* KQ 2.
- Three studies[5,107,108] presented evidence on both disparities and mitigation strategies and thus addressed both KQ 1 and KQ 2.

Figure 4 presents a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of study eligibility. The primary reasons for exclusion included the following: did not examine an algorithm (KQ 1), did not examine an intervention's ability to mitigate racial and ethnic bias of an algorithm (KQ 2), and did not report an outcome of interest (both KQs). Detailed results of the literature searches and excluded studies are in Appendixes A and B. Two submissions were received through the Agency for Healthcare Research and Quality (AHRQ) Supplemental Evidence and Data submission process but did not meet eligibility criteria. Detailed characteristics and outcomes of included studies are described in Appendixes C and D.

To address the Contextual Questions (CQs), we synthesized insights gathered during semi-structured interviews with 14 Key Informants (KIs), 10 members of our Technical Expert Panel (TEP), and 5 Subject Matter Experts (SMEs). We also summarized key points from 10 white papers, commentaries, and technical documents that our searches identified; these resources describe guidelines, standards, and best practices to reduce potential racial and ethnic bias related to algorithm development. For CQ 4, we examined six algorithms in depth.

3.1 Results, Overview

Figure 4. Study flow diagram



3.1 Results, Overview

Figure 5 and Figure 6 present an evidence map that summarizes the algorithms identified for KQ 1 and KQ 2. We include information about the type of clinical assessment and each study's design and categorized outcomes as defined *a priori* in our protocol. For each algorithm, we display the primary outcome(s) and directionality of impact on disparities as identified by the study. The direction of effect on the outcome of interest is represented by an arrow pointing up (an increase), down (a decrease), or a horizontal arrow (no effect). Further details on findings presented in the evidence map can be found in the KQ 1 and KQ 2 Summary of Findings.

Figure 5. Effect of algorithms on racial and ethnic disparities: Synthesis of KQ 1

Clinical Assessment	Algorithm	Study	Study Design[a]	Disparities in Health Outcome[b]	Disparities in Access[b]	Disparities in Quality[b]
Emergency department assessment	Rapid triage fast-track model based on ESI	Boley et al. 2022[63]	Retrospective cohort		↑	
	ESI	Metzger et al. 2022[64]	Retrospective cohort		↑	
	HEART Pathway	Snavely et al. 2021[58]	Pre-post	↔		↑[c]
High-risk care management	Novel algorithm for high-risk care management	Obermeyer et al. 2019[5d]	Modeling[e]		↑	
Kidney transplant allocation	Revised KAS[f]	Zhang et al. 2018[61]	Pre-post		↓	
Lung cancer risk	USPSTF-2013	Han et al. 2020[59]	Modeling[g]			↑
	USPSTF-2013	Pasquinelli et al. 2021[109]	Modeling[e]		↑	
	USPSTF-2013, USPSTF-2021, PLCOm2012(Race3L)	Williams et al. 2022[67]	Modeling[e]		↑	
Lung transplant allocation	LAS	Wille et al. 2013[62]	Pre-post		↓	
Opioid misuse risk	Natural language processing algorithm	Thompson et al. 2021[108d]	Modeling[e]			↑
Prostate cancer risk	PCPT[f]	Carbunaru et al. 2019[60]	Modeling[e]			↓
	KPCC RC[f]	Presti et al. 2021[56]	Modeling[e]			↓
Severity of illness measurement for crisis standards of Care	SOFA, LAPS2	Ashana et al. 2021[107d]	Modeling[e]	↑	↑	
	SOFA tiering systems	Miller et al. 2021[54]	Modeling[e]		↑	
	CSC algorithm based on SOFA and either comorbidities or physician assessment	Riviello et al. 2022[65]	Modeling[e]	↑		
	APACHE IVa, OASIS, SOFA	Sarkar et al. 2021[57]	Modeling[e]	↑		
Stroke risk	ACC/AHA atrial fibrillation guideline based on CHA2DS2-VASc (2020)	Yoo et al. 2023[66]	Modeling[e]	↓		

Abbreviations: AHA = American Heart Association; ACC = American College of Cardiology; APACHE IVa = Acute physiology and chronic health evaluation version IVa; CHA_2DS_2-VASc = congestive heart failure, hypertension, age ≥75 , diabetes, stroke , vascular disease, age 65 to 74 years, and sex category ; CSC = crisis standards of care; ESI = Emergency Severity Index; HEART = History, Electrocardiogram, Age, Risk factor for coronary artery disease, Troponin; KAS = Kidney Allocation System; KPPC RC = Kaiser Permanente prostate cancer risk calculator; LAPS2 = Laboratory-based Acute Physiology Score version 2; OASIS = Oxford Acute Severity of Illness Score; LAS = Lung Allocation System; PCPT RC = Prostate Cancer Prevention Trial risk calculator; PLCOm2012 = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Model 2012; PLCOm2012(Race3L) = PLCOm2012 with 3-level race; SOFA = Sequential Organ Failure Assessment; USPSTF = U.S. Preventive Services Task Force.

3.1 Results, Overview

[a]Study design types were categorized as: 1) modeling, 2) pre-post study, 3) retrospective cohort.
[b]Multiple outcomes were reported in studies. If there was an effect of increased disparities for any outcome in each category (e.g., access, health), it is shown as increased, even if other outcomes within that category were not. For more information on the specific outcome, please see Appendix D Table D-1. Please see Table 2 in the Methods for criteria that guided study eligibility and categorization of outcomes, organized according to the PICOTS (Population, Intervention, Comparator, Outcome, Timing, and Setting) framework.
[C]In Snavely 2021, which studied the impact of the HEART Pathway, non-White patients were less likely to receive objective cardiac testing or get hospitalized compared to White patients. However, this was felt to be appropriate care because cardiac health outcomes were not different between racial groups.
[d]Three studies addressed both Key Question 1 and 2: Ashana et al. 2021;[107] Thompson et al. 2021;[108] and Obermeyer et al. 2019.[5]
[e]Modeling study using real world datasets (i.e., source data for calculation of algorithmic scores are real world data and outcomes that would have resulted from using the algorithm are simulated).
[f]Indicates that race or ethnicity is included as an input variable.
[g]Modeling study using synthetic datasets (i.e., source data for calculation of algorithmic scores are synthetic data and outcomes that would have resulted from using the algorithm are simulated).
Note on direction of effect: arrows pointing up indicate increased outcome; arrows pointing down indicate decreased outcome; sideways arrows indicate no effect.
Note on shading: color is for emphasis only. Green indicates a desirable result; red indicates an undesirable result; yellow indicates no change.

3.1 Results, Overview

Figure 6. Effect of algorithms on racial and ethnic disparities: Synthesis of KQ 2

Mitigation Strategy	Algorithm	Study	Study Design[a,d]	Health Outcome[b]	Access to Care[b]	Quality of Care[b]
Removed race	eGFR for kidney function	Ahmed et al. 2021[21]	Modeling			↑ Reclassification to higher disease severity
	eGFR for kidney function	Bundy et al. 2022[103]	Prospective cohort			↑ Accuracy of predicting end-stage renal disease
	eGFR for kidney function	Casal et al. 2021[69]	Modeling		↓ Access to cancer treatment	↑ Reclassification to higher disease severity
	eGFR for kidney function	Diao et al. 2023[105]	Modeling		↑ Eligibility for kidney transplant and other treatments	↑ CKD diagnosis and Reclassification to higher severity
	eGFR for kidney function	Duggal et al. 2021[70]	Modeling			↑ CKD diagnosis
	eGFR for kidney function	Gutierrez et al. 2022[102]	Retrospective cohort			↑ Accuracy of predicting kidney graft failure and mortality
	eGFR for kidney function	Hoenig et al. 2022[72]	Modeling		↑ Eligibility for kidney transplant	
	eGFR for kidney function	Huang et al. 2022[100]	Modeling			↑ Accuracy of predicting acute kidney injury
	eGFR for kidney function	Inker et al. 2021[73]	Modeling			↑ Underestimation of GFR
	eGFR for kidney function	Mahmud et al. 2022[75]	Modeling			↔ Prediction of acute kidney injury
	eGFR for kidney function	Meeusen et al. 2022[99]	Retrospective cohort			↑ CKD diagnosis and accuracy of predicting severity
	eGFR for kidney function	Miller et al. 2021a[76]	Modeling			↑ Appropriate antibiotic dosing
	eGFR for kidney function	Muiru et al. 2023[101]	Modeling			↑ Accuracy of predicting progression of CKD
	eGFR for kidney function	Panchal et al. 2022[77]	Modeling	↓ Likelihood of SLKT ↑ Likelihood of only kidney or liver transplant	↑ Eligibility for transplant	↑ Reclassification to higher disease severity
	eGFR for kidney function	Schmeusser et al. 2022[104]	Modeling		↓ Eligibility for clinical trials	
	eGFR for kidney function	Shi et al. 2021[79]	Modeling			↑ Reclassification to higher disease severity
	eGFR for kidney function	Tsai et al. 2021[80]	Modeling		↑ Eligibility for care	↑ CKD diagnosis
	eGFR for kidney function	Yap et al. 2021[82]	Modeling			↑ CKD diagnosis
	eGFR for kidney function	Zelnick et al. 2021[83]	Modeling			↑ Diagnosis of severe CKD
	KDPI/KDRI for kidney donor suitability	Doshi et al. 2022[97]	Modeling		↓ Disparity in donor kidney availability	
	KDPI/KDRI for kidney donor suitability	Miller et al. 2022[98]	Modeling		↔ Non-use of donor kidneys	
	GLI spirometry equation for lung function	Baugh et al. 2022[68]	Modeling			↑ Accuracy of lung function evaluation
	GLI spirometry equation for lung function	Elmaleh-Sachs et al. 2021 [71]	Modeling	↔ Evaluation of lung function		
	Novel risk prediction algorithm for postpartum depression	Park et al. 2021[78]	Modeling			↓ Disparities in prediction of postpartum depression and use of mental health services
Replaced race with biological indicators	eGFR for kidney function	Coresh et al. 2019[86]	Modeling			↑ Accuracy of GFR estimation
	eGFR for kidney function	Inker et al. 2021[23]	Modeling			↑ Accuracy of GFR estimation
	eGFR for kidney function	Inker et al. 2021[73]	Modeling			↓ Disparity in GFR measurement
	KDRI for kidney donor suitability	Julian et al. 2017[89]	Modeling			↑ Accuracy of predicting kidney graft failure
Replaced healthcare outcomes	Novel risk prediction algorithm for complex healthcare needs	Obermeyer et al. 2019[9c]	Modeling		↓ Disparity in eligibility for care management	
Added race	FRS for CVD risk	Drawz et al. 2018[95]	Modeling			↔ Risk classification for heart disease
	FRS for CVD risk	Topel et al. 2012[87]	Modeling	↓ Disparities in subclinical vascular disease		
	CHA2DS2-VASc for stroke risk	Kabra et al. 2016[91]	Modeling			↓ Disparity in prediction of stroke risk
Added variables other than race	ASCVD for CVD risk	Weale et al. 2021[81]	Modeling			↑ Accuracy of predicting CVD
	Novel risk prediction algorithm for CVD	Fox et al. 2016[90]	Modeling		↔ Reclassification of severity of CV morbidity	
	Warfarin dosing	Kimmel et al. 2013[93]	RCT	↑ Disparity in time in therapeutic range		
	Warfarin dosing	Lindley et al. 2022[106]	Retrospective cohort	↑ Accuracy of warfarin dosing		
	USPSTF-2020 for lung cancer risk	Landy et al. 2021[74]	Modeling	↓ Disparity in lung cancer deaths	↓ Disparity in eligibility for screening	
	Novel risk prediction algorithm for complex healthcare needs	Hammond et al. 2020[85]	Modeling	↑ Accuracy of predicting hospitalization and mortality	↑ Accuracy of predicting health care costs	
Recalibrated after improving population representation	ASCVD for CVD risk	Fairman et al. 2020[84]	Modeling	↓ Disparity in statin prescribing	↓ Disparity in prediction of CVD morbidity	
	ASCVD for CVD risk	Yadlowsky et al. 2018[88]	Pre-post			↓ Disparity in prediction of CVD
	Novel risk prediction algorithm for postpartum depression	Park et al. 2021[78]	Modeling			↓ Disparities in prediction of postpartum depression and use of mental health services
	Donor Risk Index for liver transplant suitability	Shores et al. 2013[94]	Modeling			↑ Accuracy of predicting liver graft failure
Stratified algorithms for Black and White patients	CGAG for warfarin dosing	Limdi et al. 2015[92]	Prospective cohort			↑ Accuracy of warfarin dosing
	Novel risk prediction algorithm for opioid misuse	Thompson et al. 2021[108c]	Modeling		↓ Disparity in referral for education and treatment	
Statistical techniques	SOFA for severity of illness	Ashana et al. 2021[107c]	Modeling		↓ Disparity in eligibility for high-priority services	
	ASCVD for CVD	Foryciarz et al. 2022[96]	Modeling			Improved overall accuracy, reduced subgroup accuracy
	ASCVD for CVD	Yadlowsky et al. 2018[88]	Pre-post			↓ Disparity in prediction of CVD
	Novel risk prediction algorithm for postpartum depression	Park et al. 2021[78]	Modeling			↓ Disparities in prediction of postpartum depression and use of mental health services

3.1 Results, Overview

Abbreviations: ASCVD = atherosclerotic cardiovascular disease; CKD = chronic kidney disease; COAG = Clarification of Oral Anticoagulation through Genetics study; CVD = cardiovascular disease; eGFR = estimated glomerular filtration rate; FRS = Framingham Risk Score; GLI = Global Lung Function Initiative; KDPI = Kidney Donor Profile Index; KDRI = Kidney Donor Risk Index; RCT = randomized controlled trial; SLKT = simultaneous liver-kidney transplantation; SDOH = social determinants of health; SOFA = Sequential Organ Failure Assessment; USPSTF = United States Preventive Services Task Force

[a]Study design types were categorized as: 1) randomized controlled trial (RCT), 2) modeling, 3) prospective cohort, 4) retrospective cohort, 5) pre-post.

[b]For information on specific outcomes, please see Appendix D Table D-2. Please see Table 2 in the Methods for criteria that guided study eligibility and categorization of outcomes, organized according to the PICOTS (Population, Intervention, Comparator, Outcome, Timing, and Setting) framework.

[c]Three studies addressed both Key Question 1 and 2: Ashana et al. 2021;[107] Thompson et al. 2021;[108] and Obermeyer et al. 2019.[5]

[d]All modeling studies for KQ 2 used real world datasets (i.e., source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated).

Note on direction of effect: arrows pointing up indicate increased outcome; arrows pointing down indicate decreased outcome; sideways arrows indicate no effect.

Note on shading: color is for emphasis only. Green indicates a desirable result; red indicates an undesirable result; yellow indicates no change.

3.2 Key Question 1. What is the effect of healthcare algorithms on racial and ethnic differences in access to care, quality of care, and health outcomes?

3.2.1 Description of Included Evidence

For KQ 1, we included 17 studies: 11 modeling studies using real-world datasets,[5,54-57,60,65-67,107,108] 3 pre-post studies,[58,61,62] 2 retrospective cohort studies,[63,64] and 1 modeling study using synthetic datasets.[59] Most research was recent: 12 of 17 studies (71 percent) were published in 2021 or later.[54,56-58,63-67,107-109]

Studies in our evidence base examined algorithms that inform decisions about emergency department (ED) care,[58,63,64] measure severity of illness for crisis standard-of-care scenarios,[54,57,65,107] predict future healthcare needs,[5] allocate organs for transplant,[61,62] assess risk of lung cancer,[59,67,109] predict opioid misuse,[108] predict risk of prostate cancer,[56,60] and predict risk of stroke[66]. Some studies analyzed data from patients who were managed using an algorithm and examined differences in outcomes across racial and ethnic groups in real-world settings. Most studies employed a modeling approach, using patients who were not managed using the algorithm but who have data on all the input variables included in the algorithm, making it possible to determine what the algorithm's predictions and clinical recommendations would have been if applied to those patients. Eleven studies compared two or more algorithms,[5,54,56,57,59,60,66,67,107-109] three studies compared an algorithm with no algorithm (e.g., pre-implementation of algorithm),[58,61,62] and three studies examined algorithms in isolation, with no comparators.[63-65] Detailed information about the included studies is provided in Table C-1 in Appendix C and Table D-1 in Appendix D.

For KQ 1, studies had to report outcome data separately for more than one racial and ethnic group. Studies usually identified and selected patients from electronic health records (EHRs) or national databases (e.g., transplant registries); therefore, the reliability of race and ethnicity classifications depended on respective database collection methods. In 11 studies, participants self-reported race and ethnicity,[5,57,58,60,62-67,108] and the remaining studies did not specify how race or ethnicity was determined.[54,56,59,61,107,109] In five studies, analyses were restricted to patients categorized as African-American/Black or White,[5,54,62,63,107] and nine reported data for these two groups in addition to patients categorized as Asian or Hispanic;[56,57,59,61,65-67,108,109] One study reported only analyses comparing non-Hispanic White patients with non-White patients.[64] Eight studies included a heterogeneous non-White or Other patient group.[56,58,60,64,65,67,108,109] In some studies, either the database itself used an "Other" category or the authors chose to create this category for patients whose race and ethnicity was unknown (e.g., patients declined to respond) or who belonged to racial and ethnic groups with small sample sizes.

Overall risk of bias (ROB) ratings were Moderate in 12 studies[5,54,56,59,60,62-65,67,107,108] and High in 5 studies.[57,58,61,66,109] Equity-based signaling questions changed domain-specific ROB in only one instance (from Low to Moderate for bias due to selection of participants in one study)[64] and did not change overall ROB for any KQ 1 studies (Appendix Table D-3). In most studies, the algorithms examined were applied retrospectively to a single cohort of patients to model the effect of using the algorithm. In studies of this design, estimates of *overall* differences in the effects of one algorithm versus another are, by definition, not subject to ROB arising from

confounding or selection of participants (because the same patients simultaneously "receive" both algorithms) or to ROB due to deviations from intended interventions (because both algorithms are "applied" without deviation). However, other estimates – notably, estimates of differences in outcome between racial and ethnic groups – may be subject to ROB arising from confounding or selection of participants; thus, conclusions about the effect of different algorithms on such racial and ethnic differences may be subject to increased ROB. Furthermore, estimates of effects in modeling studies may have limited generalizability to actual clinical settings due to the inability to estimate the impact of real-world deviations from intended interventions (i.e., use of algorithm at the point of care). Studies varied in procedures used to gather race and ethnicity of enrolled patients, such as being self-reported or captured by an administrator. For most studies, it was unclear whether a consistent definition of race and ethnicity was used or if adequate response options were available. In addition, some studies reported outcomes for aggregate groups, such as those that identify as BIPOC (Black, Indigenous, or People of Color), likely resulting in missing data for some racial and ethnic groups. Other ROB concerns centered on problems related to missing data for algorithm score generation or outcomes (e.g., safety events) and variation in methods to measure outcomes (e.g., across different time periods).

Eighteen different clinical algorithms were examined across the studies included for KQ 1. Four algorithms included race and ethnicity as an input variable,[56,60,61,67] and 14 did not.[5,54,57-59,62-66,107-109] Five studies described 4 algorithms that were associated with reduced racial and ethnic disparities,[56,60-62,66] eleven studies found that 13 algorithms were associated with perpetuating or exacerbating racial and ethnic disparities (3 of 11 studies included an examination of methods to mitigate these disparities and thus addressed both KQ 1 and 2),[5,54,57,59,63-65,67,107-109] and 1 study reported no racial and ethnic disparities with or without the algorithm.[58] Three studies of algorithms that included race and ethnicity as an input variable found that disparities were reduced.[56,60,61] It is important to note that, in one of the three algorithms (the revised Kidney Allocation System), race and ethnicity input variables were included specifically to address existing racial and ethnic disparities. Lastly, other algorithms that included race and ethnicity input variables and perpetuated and exacerbated disparities (i.e., eGFR [estimated glomerular filtration rate], GLI [Global Lung Function Initiative] calculator for spirometry) are described below in the KQ 2 results rather than here, because the studies examining them focused primarily on reporting mitigation strategies.

3.2 Results, Key Question 1

Table 4. Effect of clinical algorithms on racial and ethnic disparities

Clinical Assessment	Study	Study Objective	Algorithm(s) (i.e., Intervention)	Comparator	Algorithm(s) Includes Race or Ethnicity? (Y/N)	Primary Outcome	Effect of Algorithm on Racial and Ethnic Disparities?
Emergency department assessment	Boley et al. 2022[63]	Examine impact of a rapid triage fast-track model on outcomes in Black non-Hispanic and White non-Hispanic patients presenting to the ED.	Rapid triage fast-track model based on the ESI	None	N	Assigned to fast-track or regular ED	Perpetuate / exacerbate
	Metzger et al. 2022[64]	Assess the association of race and language with ED triage scores.	ESI	None	N	Assigned ESI score	Perpetuate / exacerbate
	Snavely et al. 2021[58]	Compare the safety and effectiveness of the HEART pathway among white vs non-white patients, presenting to the ED with acute chest pain.	HEART Pathway	Pre-implementation of HEART Pathway	N	30-day death or myocardial infarction	No effect
High-risk care management	Obermeyer et al. 2019[5]	Quantify racial disparities in health care resource allocation produced by a widely used commercial risk prediction algorithm.	Novel risk prediction algorithm for complex healthcare needs	None	N	Eligibility for a care management program	Perpetuate
Kidney transplant allocation	Zhang et al. 2018[61]	Assess impact of the 2014 KAS policy change on waitlisting overall and evaluate whether racial/ethnic disparities in waitlisting in the United States changed following implementation.	KAS	Pre-implementation of KAS	Y	Waitlisting rate	Reduce

3.2 Results, Key Question 1

Clinical Assessment	Study	Study Objective	Algorithm(s) (i.e., Intervention)	Comparator	Algorithm(s) Includes Race or Ethnicity? (Y/N)	Primary Outcome	Effect of Algorithm on Racial and Ethnic Disparities?
Lung cancer risk	Han et al. 2020[59]	Characterize individuals who would be selected for lung cancer screening based on risk factors but would not be recommended for screening based on the current USPSTF guidelines.	USPSTF-2013	PLCOm2012	N (Y for comparator)	Lung cancer screening eligibility	Perpetuate / exacerbate
	Pasquinelli et al. 2021[109]	Compare 2 different lung cancer screening criteria, USPSTF 2013 and PLCOm2012.	USPSTF-2013	PLCOm2012	N (Y for comparator)	Lung cancer screening eligibility	Perpetuate / exacerbate
	Williams et al. 2022[67]	Compare number eligible for lung cancer screening between USPSTF criteria in 2013 with revised criteria in 2021, and with more detailed criteria from the PLCOm2012 model	USPSTF-2013, USPSTF-2021, PLCOm2012(Race3L)	Compared prediction models	N (USPSTF-2013, USPSTF-2021) Y (PLCOm2012 (Race3L))	Lung cancer screening eligibility	Perpetuate / exacerbate
Lung transplant allocation	Wille et al. 2013[62]	Compare ethnic disparities in lung transplantation rates and time to death on the wait list, before vs after introduction of the LAS.	LAS	Pre-implementation of LAS	N	Death while on waitlist or ineligibility due to morbidity while on waitlist	Reduce
Opioid misuse risk	Thompson et al. 2021[108]	Assess fairness and bias of a previously validated machine-learning opioid misuse classifier.	Novel risk prediction algorithm for opioid misuse	None	NR	Referral for education, treatment options, and care pathways	Perpetuate / exacerbate

3.2 Results, Key Question 1

Clinical Assessment	Study	Study Objective	Algorithm(s) (i.e., Intervention)	Comparator	Algorithm(s) Includes Race or Ethnicity? (Y/N)	Primary Outcome	Effect of Algorithm on Racial and Ethnic Disparities?
Prostate cancer risk	Carbunaru et al. 2019[60]	Compare the frequency of avoided biopsies and missed clinically significant prostate cancer resulting from use of two risk prediction algorithms across racial groups.	PCPT	PBCG	Y	Biopsies avoided and clinically significant prostate cancers missed	Reduce
	Presti et al. 2021[56]	Externally validate a newly developed prostate cancer risk prediction algorithm, and compare with two other calculators.	KPPC RC	Compared two versions of KPPC RC	Y	Biopsies avoided and clinically significant prostate cancers missed	Reduce
Severity of illness measurement for Crisis standards of Care	Ashana et al. 2021[107]	Assess the performance of the SOFA score and LAPS2 among Black and White patients admitted through the ED with sepsis or acute respiratory failure.	SOFA and LAPS2	Compared prediction models	N	In-hospital mortality	Perpetuate / exacerbate
	Miller et al. 2021[54]	Investigate whether using the SOFA is associated with deprioritization of Black patients in currently adopted CSC	SOFA tiering systems	Compared SOFA tiering systems	N	In-hospital mortality	Perpetuate / exacerbate
	Riviello et al. 2022[65]	Analyze the association of CSC scoring system with resource prioritization and estimated excess mortality by race, ethnicity, and residence in a socially vulnerable area.	CSC algorithm based on SOFA and either comorbidities or physician assessment	Compared prediction models	N	In-hospital mortality	Perpetuate / exacerbate

3.2 Results, Key Question 1

Clinical Assessment	Study	Study Objective	Algorithm(s) (i.e., Intervention)	Comparator	Algorithm(s) Includes Race or Ethnicity? (Y/N)	Primary Outcome	Effect of Algorithm on Racial and Ethnic Disparities?
	Sarkar et al. 2021[57]	Examine the performance of three severity scoring models.	SOFA, OASIS, APACHE IVa	Compared prediction models	N	In-hospital mortality	Perpetuate / exacerbate
Stroke risk	Yoo et al. 2023[66]	Evaluate how predictive performance of a clinical calculator affects downstream health outcomes.	ACC/AHA atrial fibrillation guideline based on CHA2DS2-VASc (2020)	ACC/AHA atrial fibrillation guideline based on CHA2DS2-VASc (2014)	N	Stroke	Reduce

Abbreviations: ACC = American College of Cardiology; AHA = American Heart Association; APACHE IVa = Acute physiology and chronic health evaluation version IVa; CHA_2DS_2-VASc = congestive heart failure, hypertension, age ≥75 (doubled), diabetes, stroke (doubled), vascular disease, age 65 to 74 and sex category (female); CSC = crisis standards of care; ED = emergency department; ESI = Emergency Severity Index; HEART = History, Electrocardiogram, Age, Risk factor for coronary artery disease, Troponin; KAS = Kidney Allocation System; KPPC RC = Kaiser Permanente prostate cancer risk calculator; LAS = Lung Allocation System; LAPS2 = Laboratory-based Acute Physiology Score version 2; N = no; NR = not reported; OASIS = Oxford Acute Severity of Illness Score; PBCG=Prostate Biopsy Collaborative Group; PCPT RC = Prostate Cancer Prevention Trial risk calculator; PLCOm2012 = Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Model 2012; PLCOm2012(Race3L) = PLCOm2012 with 3-level race; SOFA = Sequential Organ Failure Assessment; USPSTF = United States Preventive Services Task Force; Y = yes
Note: Three studies addressed both Key Question 1 and 2 and are described in Table C-1: Ashana et al. 2021;[107] Thompson et al. 2021;[108] and Obermeyer et al. 2019.[5]

3.2.2 Key Points

3.2.2.1 Emergency Department Assessment

- One study[58] assessed the impact of implementing the HEART (History, Electrocardiogram, Age, Risk Factors, Troponin) Pathway risk assessment. In this algorithm, which does not include race or ethnicity, high scores indicate the patient is at high risk for adverse outcomes resulting from acute coronary syndrome and should receive further testing. After implementation, early discharge rates increased, and hospitalization and objective cardiac testing rates decreased for BIPOC (i.e., Black or African American, Asian, American Indian, and Hawaiian or Pacific Islander, and other/unknown patients) and White patients. BIPOC patients were more likely to be classified as low risk than White patients. The difference in 30-day death and myocardial infarction rates between low-risk BIPOC and White patients was nonsignificant; therefore, authors concluded that the pathway did not impact health disparities.
- Two studies[63,64] examined the Emergency Severity Index (ESI), an algorithm triage providers use to assess a patient's level of acuity and to prioritize care. An assigned ESI score is based on an assessment of a patient's vital signs, primary reason for ED visit, and immediate needs. Race and ethnicity are not included as input variables. In a cohort of BIPOC and non-Hispanic (NH) White pediatric patients, one study examined the association between assigned ESI scores and the patient's race.[64] The second study assessed the effect of a rapid triage fast track (FT) model, an algorithm-informed care pathway based on ESI scores, on outcomes in NH Black and NH White adult patients.[63] Results of both studies indicated that BIPOC pediatric and adult patients were more likely than NH White patients to receive a lower acuity score indicating a less urgent need for care. Using the ESI alone or to inform a care pathway, may lead to placing BIPOC patients in a lower acuity care category compared with NH White patients at the same level of need, which may potentially exacerbate racial and ethnic disparities in access to care. Authors noted that BIPOC patients might present to the ED with less acute conditions than NH White patients. Thus, lower ESI scores may reflect true differences in illness severity between patients upon presentation to the ED as opposed to the impact of provider assessment to determine acuity level and care needs.

3.2.2.2 High-Risk Care Management

- One study[5] analyzed a widely used EHR-based algorithm (which does not include race or ethnicity as an input variable) designed to help determine whether patients should be placed in high-risk care management programs. The algorithm predicts future healthcare costs based on prior utilization; patients with high predicted costs are prioritized for placement in the programs. At any given level of actual healthcare need, the algorithm predicted lower costs for Black patients than for White patients, which could result in a race disparity in access to care.

3.2.2.3 Kidney Transplant Allocation

- One study[61] examined how implementing the revised Kidney Allocation System (KAS), which includes ethnicity as an input variable, affected the rate of waitlisting (i.e., placement on the national deceased donor waiting list). After implementation, the overall rate of waitlisting declined for all racial and ethnic groups, due largely to a reduction in inactive waitlisting (i.e., placing patients on the waitlist who are not in fact eligible for transplant for various reasons). The difference in waitlisting rates between Black and White patients was reduced after implementation, due partly to declines in inactive waitlisting and increases in active waitlisting among BIPOC (Black, Hispanic, and Asian) patients, but a difference in rates remained. Therefore, implementing the revised KAS reduced racial and ethnic disparities.

3.2.2.4 Lung Transplant Allocation

- One study[62] examined how implementing the Lung Allocation Score, which does not include race or ethnicity as input variables, affected outcomes for patients on the waitlist for lung transplantation. Before implementation, Black patients were more likely than White patients to die while on the waitlist or become too sick for transplantation within 3 years of listing (43.8 percent vs. 30.8 percent); after implementation, the rate of this outcome was reduced for both groups, and the difference between them became negligible (14.0 percent vs. 13.3 percent).

3.2.2.5 Lung Cancer Screening

- Two studies[59,109] examined the U.S. Preventive Services Task Force (USPSTF) lung cancer screening guidelines from 2013 and compared them with an algorithm (PLCOm2012) for determining eligibility for lung cancer screening. The algorithm includes race as an input variable; USPSTF guidelines include only age and smoking history. Results indicated that race differences in eligibility based on the USPSTF guidelines would be greatly reduced by expanding eligibility criteria to include individuals categorized as eligible by PLCOm2012.
- One study[67] examined racial and ethnic differences in the percentages eligible for lung cancer screening based on: (a) 2013 USPSTF guidelines; (b) the expanded 2021 USPSTF guidelines; (c) lung cancer risk as calculated by the PLCOm2012(Race3L) model (similar to the original PLCOm2012 but using 3-level race), using a value of 1.5 percent 6-year risk as the threshold for eligibility; and (d) PLCOm2012(Race3L) risk using a 1 percent threshold value. All four sets of criteria resulted in differences across racial and ethnic groups in percentages identified as eligible, with lower percentages among Black and Hispanic individuals than among White individuals and those of other races. Authors suggest that closing the gap between the proportions eligible by race might require inclusion of additional risk factors in risk-based lung cancer screening tools.

3.2.2.6 Opioid Misuse Risk

- One study[108] of a natural-language-processing classifier designed to identify individuals needing services to help them overcome opioid abuse found a higher

false-negative rate (i.e., patients who were misusing opioids but were not identified as such by the classifier) for Black patients (32 percent) and "Other" race and ethnicities (33 percent) than White or Hispanic/LatinX patients (17 percent), which could result in a racial and ethnic disparity in access to care for opioid misuse.

3.2.2.7 Prostate Cancer Risk

- Two studies[56,60] modeled the use of prostate cancer algorithms (both including race as an input variable) to inform the decision about whether to perform a biopsy. Under some model parameters, the net benefit of algorithms (defined in terms of unnecessary biopsies that would have been avoided versus clinically significant cancers that would have been missed) was higher for White patients than for Black patients, but this depended on which algorithm was used as well as the numerical threshold for recommending biopsies; other parameters led to a slightly higher net benefit for Black patients than for White patients.

3.2.2.8 Severity of Illness Measurement for Crisis Standards of Care

- Three studies[54,57,107] evaluated racial and ethnic differences in the performance of four illness-severity prediction models: Acute Physiology and Chronic Health Evaluation (APACHE IVa), Laboratory-based Acute Physiology Score version 2 (LAPS2), Oxford Acute Severity of Illness Score (OASIS), and Sequential Organ Failure Assessment (SOFA). In resource-constrained settings, these algorithms (none of which contain race or ethnicity as input variables) were proposed to be used to inform Crisis Standards of Care, which allocate resources preferentially to patients with better estimated chances of survival. In all three studies, the prediction models overestimated mortality in Black patients compared with White patients (i.e., at any given level of algorithm-predicted risk, Black patients had lower actual mortality than White patients). Using these prediction models therefore has the potential to lead to inappropriate deprioritization of Black patients.
- One study[65] examined a Crisis Standards of Care algorithm based on short-term mortality risk estimated by SOFA and long-term mortality risk estimated either by comorbidities or physician assessment and estimated excess deaths that might occur through use of this algorithm to allocate ventilators under conditions of resource shortage, by race. At certain risk-threshold values, the estimated excess mortality among Black patients was significantly higher than that among non-Black patients.

3.2.2.9 Stroke Risk

- One study[66] examined potential racial and ethnic disparities in health outcomes that could occur as a result of using an algorithm-informed decision tool, the American College of Cardiology/American Heart Association atrial fibrillation treatment guideline. The guideline recommendation for anticoagulant therapy is based on the CHA_2DS_2-VASc score, which predicts stroke risk in patients with atrial fibrillation. Race and ethnicity are not included as input variables. The study compared two versions of the guideline that use different CHA_2DS_2-VASc score thresholds to guide decision making. Using the 2014 guideline (CHA_2DS_2-VASc score > 1), among patients who would not have been offered anticoagulant therapy, 3.3 percent of

Hispanic patients had a stroke. Using the 2020 guideline (CHA_2DS_2-VASc score > 2 for males and > 3 for females), 1.78 percent of Hispanic patients had a stroke. Authors suggest that when using the 2020 guideline, the disparity in negative event frequency (stroke) in Hispanic patients was reduced.

3.2.3 Summary of Findings

Below, we present the findings in the following clinical categories:

- ED assessment[58,63,64]
- High-risk care management[5]
- Kidney transplant allocation[61]
- Lung transplant allocation[62]
- Lung cancer screening[59,67,109]
- Opioid misuse risk[108]
- Prostate cancer risk[56,60]
- Severity of illness measurement for crisis standards of care[54,57,65,107]
- Stroke risk[66]

Tables 5-15 describe the algorithms within each clinical category.

3.2.3.1 Emergency Department Assessment

Table 5. Description of the HEART Pathway

Algorithm	Description	Background
The HEART (History, Electrocardiogram, Age, Risk Factors, Troponin) Pathway	HEAR score (History, Electrocardiogram, Age, and Risk factor for coronary artery disease) and 0- and 3-hour troponin levels.[110] Clinicians use the algorithm to evaluate patients presented to the emergency department with acute chest pain to determine risk and triage. Race and ethnicity are not included in the score or pathway. High scores on the algorithm indicate higher clinical risk of short-term adverse outcomes resulting from acute coronary syndrome warranting further testing and evaluation.[110]	Prior studies, including a randomized controlled trial funded by the American Heart Association, have shown improvement in outcomes, including shortened hospital length of stay, increased early discharges, and a reduction in objective cardiac testing during 30 days, without increases in major adverse cardiac events.[111]

One study assessed the impact of implementing the HEART Pathway risk assessment over 24 months in 3 North Carolina EDs among White patients and non-White (BIPOC) patients (Table C-1 in Appendix C).[58] The BIPOC group included Black or African American, Asian, American Indian, and Hawaiian or Pacific Islander patients and a group of patients categorized as other/unknown (e.g., refused to provide information). Using an EHR database, the study examined data from 8474 White and BIPOC patients (n=3713 pre-implementation and n=4761 post-implementation). For several risk factors (e.g., cardiovascular disease), BIPOC patients already had lower rates than White patients before implementation.

In interpreting this study's findings, it should first be noted that the HEART Pathway identified significantly more BIPOC patients as low risk than White patients (35.6 percent vs. 28.0 percent; $p<0.0001$). But clinical outcomes, including death, were not higher among low-risk BIPOC patients; therefore, the authors agreed that patients were accurately classified as low risk.

3.2 Results, Key Question 1

Before HEART Pathway implementation, 30-day death or myocardial infarction rates were higher for White patients than for BIPOC patients, as were 30-day rates of hospitalization and objective cardiac testing. (Reduction in objective cardiac testing was a goal of the HEART Pathway, due to a high rate of unnecessary testing.) Post-implementation, hospitalization and objective cardiac testing rates decreased for both racial and ethnic groups, but the decrease was greater among BIPOC patients, while 30-day death or myocardial infarction rates decreased for BIPOC patients and increased for White patients. Thus, for these three outcomes, race differences increased post-implementation, with the difference-of-differences ranging from 1 to 5 percentage points (Table D-1 in Appendix D). For several adverse outcomes, BIPOC patients already had lower rates than White patients before implementation, and the disparities increased post-implementation. The rate of early discharge (proportion of patients discharged from the ED without objective cardiac testing, an outcome that the authors considered clinically appropriate) before implementation was lower for White patients than for BIPOC patients (36 percent vs. 40 percent). After HEART Pathway implementation, these rates were 39 percent for White patients and 49 percent for BIPOC patients. Although race differences increased from pre- to post-implementation for all four outcomes, the increase was statistically significant only for early discharge. Therefore, the authors concluded that implementing the HEART pathway did not worsen disparities for clinical outcomes, specifically 30-day death and myocardial infarction rates, and can be safely implemented. Authors suggested that pathway users should be cognizant that BIPOC patients are more likely to be classified as low-risk and therefore discharged early. This study did not capture long-term outcomes.

Table 6. Description of the Emergency Severity Index

Algorithm	Description	Background
Emergency Severity Index (ESI)	The ESI is a 5-level triage algorithm used to determine acuity level of patients presenting to the emergency department and prioritize resources. Triage providers assess the chief complaint and immediate needs, obtain vital signs, and then assign an ESI score of 1 to 5 as follows: • ESI 1 (Immediate medical attention) • ESI 2 (Emergency) • ESI 3 (Urgent) • ESI 4 (Nonurgent) • ESI 5 (Minor) Race and ethnicity are not included as input variables.	Prior studies have demonstrated an association between race and ethnicity and assigned triage scores, suggesting that Black patients and patients from other racial and ethnic groups are less likely than White patients to receive an immediate or urgent ESI score and are more likely to be assigned scores indicating care needs that are less urgent.[112]

One study examined the association between assigned ESI scores and the patient's race.[64] The study used an EHR to identify 8928 pediatric patients (3086 NH White; 5842 Non-White) with 10,815 ED visits. Authors categorized patients as non-White if they reported any race other than White/Caucasian (American Indian/Alaska Native, Asian, Black, Native Hawaiian/Pacific Islander, other, and patients with more than one race).

In analyses adjusting for illness severity (i.e., abnormal vital signs), non-White pediatric patients were significantly less likely than NH White pediatric patients to receive an ESI score of 2 (emergency; adjusted odds ratio [aOR] 0.40, 95 percent confidence interval [CI]: 0.33 to 0.49, $p<0.001$) or 3 (urgent; aOR 0.50, 95 percent CI: 0.45 to 0.56, $p<0.001$), but significantly more

likely to receive an ESI score of 5 (minor; aOR 1.34, 95 percent CI 1.07 to 1.69, p=0.012). That is, non-White patients were more likely than NH White patients to be assigned a lower acuity score (e.g., ESI 5), indicating a less urgent need for care. Subgroup analyses of the symptoms that caused patients to seek care (e.g., fever, headache) demonstrated similar results. For the outcome of ED length of stay, non-White patients had a higher discharge rate than NH White patients (adjusted hazard ratio [aHR] 1.08, 95 percent CI 1.03 to 1.14, p=0.002); the differences between racial and ethnic groups in time to provider (p=0.352) and hospital admission rates (p=0.094) were not significant. The authors stated that the observed pattern of results is consistent with the possibility that illness severity was inadequately controlled for – that is, there may have been a tendency, not fully adjusted for in the analyses, for non-White patients to present to the ED with less acute conditions than NH White patients. This explanation, if true, would lead one to expect that non-White patients would have lower ESI scores on average than NH White patients (as a reflection of objective differences in illness severity rather than algorithmic bias), but would not necessarily be undertreated compared with NH White patients (thus accounting for the non-White patients' shorter length of stay and the lack of racial differences in time to provider and hospital admission rates).

Table 7. Description of a Rapid Triage Fast Track model

Algorithm	Description	Background
Rapid Triage Fast Track (FT) Model	The Rapid Triage FT model is an algorithm-informed care pathway based on emergency severity index (ESI). A nurse assesses patients to determine the chief complaint and resource needs, measures vital signs (respiratory rate, heart rate, blood pressure, temperature, and oxygen saturation), and then uses the ESI to assign patients a score of 1 (most acute, needs immediate care) to 5 (least acute, does not need immediate care). After an ESI score is assigned, the nurse evaluates additional criteria to determine if a patient should be assigned to the FT area (a lower acuity area): 1. Patient is able to sit in a recliner 2. Patient is ambulatory and able to speak 3. Patient's ESI score is 3, 4, or 5 (least acute) 4. Patient is determined to be not critical based on the triage determination. Patients selected for FT are placed in a separate emergency department (ED) section and quickly evaluated and treated by a designated clinician (e.g., physician assistant, nurse practitioner). Patients assigned to the main ED wait until a bed is available and then receive care from an ED physician. Race and ethnicity are not included as input variables.	The rapid triage FT model uses a tiered approach involving an additional assessment after using the ESI to determine if patients should be assigned to the FT area or main ED. Prior studies have demonstrated an association between race and ethnicity and assigned triage scores, suggesting that Black patients and patients from other racial and ethnic groups are less likely than White patients to receive an immediate or urgent ESI score and are more likely to be assigned scores indicating care needs that are less urgent.[112]

One study assessed the impact of a rapid triage FT model, which is an algorithm-informed care pathway based on ESI. The authors studied this model's effect on outcomes in Black NH and White NH patients presenting to the ED of a tertiary care hospital in Minnesota.[63] Using EHR data, the study examined 9704 patients with 12,330 unique encounters (5151 Black NH and

7179 White NH, exact-matched on potential confounders, including presence of abnormal vital signs) during a 1-year period after implementation of the FT model. Race and ethnicity were self-reported. (Table C-1 in Appendix C).

Compared with White NH patients, Black NH patients were significantly more likely to be assigned to FT, a lower acuity area (22.6 percent vs. 18.5 percent; odds ratio [OR] 1.28, 95 percent CI 1.12 to 1.46, p<0.001), and significantly less likely to be categorized as a high-acuity patient (59.8 percent vs. 67 percent; OR 0.73, 95 percent CI 0.66 to 0.81, p<0.001). Among patients designated as high acuity, Black NH patients were also significantly more likely than White NH patients to be assigned to the FT area (3.4 percent vs. 2.5 percent; OR 1.40, 95 percent CI 1.05 to 1.87, p=0.024). The difference between Black NH and White NH low-acuity patients was not significant (p=0.934). In a subgroup analysis, Black NH patients with abdominal pain, shortness of breath, chest pain, or headache had an increased likelihood of being assigned to the FT area than White NH patients. The difference between racial and ethnic groups was significant for the chief issue of headache (OR: 2.10; 95 percent CI 1.01 to 4.39, p=0.048).

Black NH patients also had a significantly shorter wait time between ED arrival and being placed in a room than White NH patients (MD -3.47 minutes, 95 percent CI -6.56 to -0.37, p=0.028). Subgroup analyses also demonstrated significantly shorter wait times for Black NH patients than White NH patients with a chief issue of abdominal pain (mean difference [MD] -9.52 minutes, 95 percent CI -20.02 to -0.03, p=0.028) and chest pain (MD -18.82 minutes, 95 percent CI -28.93 to -8.72, p<0.001). Authors suggest that the shorter average wait time for Black NH patients may be associated with Black NH patients being more likely to be triaged to the FT area.

The rapid triage FT model, an algorithm-informed care pathway, involves ESI score assignment by a triage provider (e.g., nurse) that assesses a patient's acuity level. Study authors suggest that triage provider assessment might introduce implicit bias and potentially affect triage decisions.[63] Using the FT model, Black NH patients presenting to the ED were less likely than White NH patients to be categorized as needing immediate or urgent care and were more likely to be triaged to the FT area, which was designed to evaluate and manage lower-acuity patients. Authors concluded the FT model led to Black NH patients receiving lower-acuity scores compared with White NH patients at the same level of need, which may potentially exacerbate racial and ethnic disparities in access to care.

3.2.3.2 High-Risk Chronic Disease Care Management

Table 8. Description of a high-risk care management prediction algorithm

Algorithm	Description
Unnamed, but "widely used" commercial risk prediction algorithm used to identify high-risk patients for the health system's care management program	The algorithm, which does not include race as an input, is used to predict complex health needs in primary care patients enrolled in risk-based contracts; the goal is to direct additional resources to such patients. The outcome predicted by the algorithm is costs over the following year, and the allocation of additional resources is intended to reduce costs, although a health benefit presumably accrues to the prioritized patients as well. The algorithm's input consists of features of raw insurance claims data from the previous year, including age, sex, insurance type, diagnosis and procedure codes, medications, and detailed costs.

One study examined racial differences in healthcare resource allocation produced by what the authors termed a "widely used" commercial risk prediction algorithm and examined strategies to

reduce those differences.[5] This study's data came from a health system in which patients with scores on the algorithm above the 97th percentile are automatically identified for enrollment into the system's care management program. The study sample consisted of all primary care patients enrolled in risk-based contracts at a large academic hospital from 2013 to 2015 and self-identifying as either Black or as White without another race (n=49,618). For each patient, algorithmic risk scores generated annually by the health system were obtained, as well as actual costs per year. Also, the total number of chronic conditions was calculated for each patient as a measure of overall illness burden to examine the extent to which the algorithm had allocated additional resources to the patients with the greatest need.

At every level of algorithm-predicted risk, Black and White patients had similar actual costs in the following year. However, at a given level of *health* (measured, as described above, by number of chronic conditions), Black patients generated lower costs than White patients – on average, $1801 less per year. Thus, although the algorithm predicted costs equally well for Black and White patients, costs cannot be assumed to be a valid proxy for healthcare needs because the association between costs and health differs across racial and ethnic groups. At the cutoff score for automatic identification for enrollment into the care management program (97th percentile), Black patients had 26.3 percent more chronic conditions than White patients ($p<0.001$). Thus, use of the algorithm to determine program eligibility hypothetically leads to acceptance of White patients who have a lower level of actual need than Black patients (i.e., greater access to healthcare for Whites than Blacks). Further discussion of this study's subsequent mitigation strategy is described in KQ 2.

3.2.3.3 Kidney Transplant Allocation

Table 9. Description of the Kidney Allocation System

Algorithm	Description	Background
The revised Kidney Allocation System (KAS)	The revised KAS includes the Kidney Donor Profile Index (KDPI), which includes ethnicity as an input variable, and the Estimated Post Transplant Survival score (EPTS). KDPI contains 10 input variables representing donor characteristics such as age, height, weight, ethnicity (options include: American Indian or Alaska Native, Asian, Black or African American, Hispanic/Latino, Native Hawaiian or Other Pacific Islander, White, or Multi Racial), and other factors related to the donor's health; the scores range from 0% to 100%. Lower scores are associated with higher expected post-transplant longevity of donor kidneys. Input variables in the EPTS score include age, time on dialysis, current diabetes status, and prior solid organ transplant; the score ranges from 0% to 100%, with higher scores predicting lower post-transplant survival. Donor kidneys with a KDPI $\leq$ 20% (i.e., donor kidneys expected to function the longest post-transplant) are prioritized for candidates with an EPTS score $\leq$ 20% (i.e., candidates expected to live the longest post-transplant) followed by candidates with EPTS scores >20%.[113,114]	In 2014, KAS was revised to improve the process of allocating deceased donor kidneys and equity related to dialysis time.[61] As a result, transplant wait time begins at the earliest of either the start of dialysis or the date placed on the national deceased donor waitlist. BIPOC (Black, Indigenous, or People of Color) patients, who tend to spend more time on dialysis before receiving a referral for transplantation, were anticipated likely to benefit from the policy change.[61] An explicit goal of the revised KAS was to improve equity in kidney transplant allocation.

One study examined how the 2014 revised Kidney Allocation System (KAS) affected racial and ethnic differences in the waitlisting rate (i.e., placement on the national deceased donor

waiting list).[61] More information on the revised KAS is in Table C-1 in Appendix C. Authors selected data from the U.S. Renal Data System of 1,253,100 new (n=1,120,655 pre-KAS and n=132,445 post-KAS) and 1,556,954 existing patients on dialysis between 2005 and 2015.

Of note, the analyses reported in this study do not examine how implementing KAS affected transplantation rates (which is the clinical outcome that the KAS directly determines), but rather how the policy's existence affected the waitlisting rate, which is an "upstream" clinical outcome for a patient before the KAS comes into play. The rationale for examining this outcome is that awareness of the policy change, and thus the change in the anticipated likelihood of individual patients receiving transplants once on the waitlist, could have affected clinicians' decisions about whether to initiate the requisite screening process.

After implementing KAS, small to moderate declines occurred in the waitlisting rate for Black (4 percent), Asian (8 percent), Hispanic (10 percent), and White (11 percent) patients. The interaction of race and ethnicity with KAS implementation was significant ($p<0.0001$). Authors further examined the difference in waitlisting rates among incident and prevalent end-stage renal disease (ESRD) patients. Compared with White incident ESRD patients, Black incident ESRD patients had a 19 percent lower waitlisting rate before implementation of KAS (adjusted hazard ratio [aHR]: 0.81; 95 percent CI 0.80 to 0.82). Post-KAS, the difference between White and Black patients declined significantly to 12 percent (aHR: 0.88; 95 percent CI 0.85 to 0.90; $p<0.001$), partially due to more Black patients with incident ESRD placed on the waitlist. However, a difference in waitlisting rates remained. Before and after KAS implementation, Asian and Hispanic incident ESRD patients had higher waitlisting rates than White incident ESRD patients. The differences from pre- to post-KAS were not significant (Asian vs. White $p=0.27$; Hispanic vs. White $p=0.62$). Monthly waitlisting rates for prevalent dialysis patients decreased from pre- to post-implementation for all racial and ethnic groups, with a statistically significant decrease for White ($p=0.017$), Black ($p=0.011$), and Hispanic ($p=0.03$) patients.

Another analysis in this study examined active and inactive waitlisting rates before and after KAS implementation.[61] Actively listed patients can be called to receive a kidney transplant at any time, while patients listed as inactive are not eligible to be called for a transplant due to reasons such as health concerns. Active waitlisting rates were similar before and after KAS implementation ($p=0.601$), while inactive waitlisting rates declined significantly ($p <0.001$). The proportion of new actively waitlisted candidates (i.e., eligible to be called for transplantation) increased from pre- to post-KAS for Black (71.3 percent vs. 76.3 percent), Hispanic (72.2 percent vs. 78 percent), and Asian (72.7 percent vs. 73.5 percent) patients, while declining slightly for White patients (72.3 percent vs. 71.4 percent). Results also demonstrated a greater decline in inactive waitlisting counts (i.e., patients on the list but not eligible to be called for transplantation) among Black and Hispanic patients following KAS implementation ($p<0.0001$). For more information, see Table D-1 in Appendix D.

Study findings indicate that, post-KAS implementation, the overall waitlisting rate declined for all racial and ethnic groups, and the difference in rates between Black and White incident ESRD patients declined significantly but was not eliminated. Results suggest that the overall decline in waitlisting rates was due to a decline in inactive waitlisting, while rates of active waitlisting (i.e., patients actually eligible for transplant being placed on the waitlist, remained relatively stable). Similarly, the reduction in the difference between Black and White incident ESRD patients was due to both a decrease in inactive waitlisting and an increase in active waitlisting among BIPOC (Black, Hispanic, and Asian) patients.[61] That is, post-KAS, fewer BIPOC patients were listed as ineligible for transplantation and a greater proportion listed as

eligible. Authors also speculate that the decline in waitlisting rates might reflect a reduction in transplant referrals, which could negatively affect patients in need of resources and treatment.

3.2.3.4 Lung Transplant Allocation

Table 10. Description of the Lung Allocation Score

Algorithm	Description	Background
Lung Allocation Score (LAS)	The LAS is a numerical score based on survival models that estimate likelihood of survival both while on the waitlist and post-transplant; thus, it reflects transplantation's net benefit. LAS does not include race and ethnicity. The input variables are diagnosis, age, height, weight, cardiac index at rest, bilirubin, functional status, pulmonary artery systolic pressure, oxygen required at rest, 6-minute walk distance, continuous mechanical ventilation, partial pressure carbon dioxide (PCO_2), increase in PCO_2, and creatinine.	In 2005, the LAS became the predominant method for determining allocation of deceased donor lungs for transplantation in the United States; before then, time on the waitlist was the sole basis for allocation.

One study analyzed data from all White and Black non-Hispanic adults who were listed for lung transplantation during two time periods: pre-LAS (2000–2005; n=8765) and LAS (2005–2010; n=8806).[62] In the pre-LAS period, Black patients were far more likely than White patients to die or become too sick for transplantation within 3 years of listing (43.8 percent vs. 30.8 percent, adjusted OR 1.84; $p < 0.001$); the difference became negligible in the LAS period (14.0 percent vs. 13.3 percent, adjusted OR 0.93; $p = 0.74$). Black patients were 18 percent *more* likely than White patients to die while on the waitlist in the pre-LAS period (adjusted hazard ratio [HR] 1.18; 95 percent CI 0.99 to 1.40; p=0.06); in the LAS period, Black patients were 17 percent *less* likely than White patients to die (adjusted HR 0.83; 95 percent CI 0.62 to 1.10; p=0.18).

3.2.3.5 Lung Cancer Screening

Table 11. Description of lung cancer screening prediction models

Algorithms	Description	Background
• U.S. Preventive Services Task Force 2013 guidelines (USPSTF-2013) • USPSTF 2021 guidelines • PLCOm2012 Model • PLCOm2012(Race3L) Model	The 2013 USPSTF recommends annual low-dose computed tomography screening of individuals aged 55–80 years with at least 30 pack-years of smoking and within 15 years since cessation. The screening criteria are based on findings from the National Lung Cancer Screening Trial.[109] In 2021, new guidelines were issued lowering the minimum age to 50 and the pack-years to 20. The PLCOm2012 Model is a validated algorithm that predicts 6-year risk of lung cancer based on age, race, education, body mass index, chronic obstructive pulmonary disease, personal history of cancer, family history of lung cancer, and smoking variables (status, intensity, duration, and quit time). The original PLCOm2012 model represents race and ethnicity using four categories – Black non-Hispanic, White non-Hispanic, Hispanic, and other non-Hispanic. The PLCOm2012(Race3L) model differs from this in that Hispanics are pooled with individuals categorized as "White" or "Other".	Prior studies have shown that the USPSTF 2013 lung cancer screening guidelines may miss individuals at high-risk for lung cancer who do not meet pack-year or age criteria, in particular Black/African American individuals. This potentially leads to underscreening of African American individuals, which might exacerbate racial and ethnic disparities in screening outcomes.[59,115] Research suggests that risk-based models such as PLCOm2012 that incorporate additional factors (e.g., sociodemographic, medical history) might improve the ability to identify individuals at high risk of lung cancer and potentially reduce racial and ethnic disparities.[59,115]

Two studies[59,109] examined racial and ethnic differences in lung cancer screening recommendations between the U.S. Preventive Services Task Force (USPSTF) guidelines (2013 version) and a risk prediction algorithm, the PLCOm2012 Model. (Both studies were conducted before the 2020 revision of the guidelines, limiting the applicability of results to current clinical practice.)

In one study, patients (n=883) enrolled in a lung cancer cohort between 2010 and 2019 were selected for analysis.[109] Findings demonstrated that the PLCOm2012 prediction model (threshold: >1.7 percent/6-year risk) reduced the difference between Black and White patients in the percentages ineligible for screening based on the USPSTF-2013 criteria. The percentage of patients who were ineligible by USPSTF-2013 criteria was 35.3 percent among White patients and 48.3 percent among Black patients, expanding the eligibility criteria to include patients classified as being at risk by PLCOm2012 reduced the percentages to 26.0 percent and 26.3 percent, respectively.

The second study used a simulated dataset (n=100,000) representing the 1950 U.S. birth cohort and containing smoking history data generated by the CISNET (Cancer Intervention and Surveillance Modeling Network) Smoking History Generator and risk factor data generated by the Lung Cancer Risk Factor Generator.[59] For the PLCOm2012, a risk of >1.51 percent was used as the threshold for eligibility for screening.

Among individuals aged 50-54, 4.8 percent of White individuals and 15.6 percent of Black individuals were eligible for screening by PLCOm2012 but ineligible by USPSTF criteria. Among individuals aged 55-70, the percentages were 3.3 percent and 7 percent, respectively; among those aged 71-80, the percentages were 10.8 percent and 14.2 percent, respectively. In the youngest and oldest of the three age groups, the difference in percentages was significant, at $p < 0.001$; the p-value for the middle group was not reported. Results at varying risk thresholds are presented graphically; the authors described the proportion as "consistently higher in Black individuals compared with White individuals independently of risk threshold."

A sensitivity analysis was performed on a similar dataset representing the 1960 U.S. birth cohort. Differences persisted but were generally smaller than those in the 1950 cohort. Across all age groups in the 1960 cohort, 2.3 percent of White individuals and 5.8 percent of Black individuals were eligible for screening by PLCOm2012 but ineligible by USPSTF ($p<0.001$).

A third study examined racial and ethnic differences in the percentages of individuals eligible for screening under four different sets of criteria: (a) the 2013 USPSTF eligibility criteria; (b) the 2021 expanded USPSTF criteria; (c) lung cancer risk as calculated by the PLCOm2012(Race3L) model, using a value of 1.5 percent 6-year risk as the threshold for eligibility; and (d) PLCOm2012(Race3L) risk using a 1 percent threshold value.[67] Data came from the 2019 Centers for Disease Control Behavioral Risk Factor Surveillance System; the analysis sample included respondents who were 50+ years old and were current or former smokers. The sample included 41,544 individuals (88.5 percent non-Hispanic White, 5 percent non-Hispanic Black, 2 percent Hispanic, 4.5 percent other). Overall, the 2013 USPSTF criteria identified the lowest percentage of individuals as eligible for screening (21 percent), and the PLCOm2012(Race 3L) model using the 1 percent threshold identified the highest (45 percent), with the 2021 USPSTF criteria and the PLCOm2012(Race 3L) model using the 1.5 percent threshold identifying similar, intermediate percentages (34.7 percent and 35.3 percent, respectively). All four sets of criteria, however, resulted in differences across racial and ethnic groups in percentages identified as eligible, with lower percentages among Black and Hispanic individuals than among White individuals and those of other races. Using the 2013 USPSTF criteria, the percentages identified as eligible for screening among White, Black, Hispanic, and other individuals were 21.9, 16.0, 9.8, and 22.1, respectively; using the 2021 USPSTF criteria, the percentages were 35.8, 28.5, 18.0, and 39.3, respectively; using the PLCOm2012(Race 3L) model with a 1.5 percent threshold, the percentages were 36.2, 31.1, 15.0, and 43.4, respectively; and using the PLCOm2012(Race 3L) model with a 1 percent threshold, the percentages were 46.3, 39.3, 20.3, and 51.4, respectively.

3.2.3.6 Opioid Misuse Risk

Table 12. Description of a natural-language processing classifier

Algorithm	Description
Natural-language processing classifier	The natural-language processing classifier employs a neural network, using clinical notes in the electronic health record as input. The algorithm's goal is "to provide point-of-care education, treatment options, and care pathways to patients who misuse opioids." The algorithm's complete set of input variables, or "features," is not given, but the classifier's 10 most highly weighted features as originally developed are listed; the list does not include race.

One study[108] examined a natural-language-processing classifier. The study analyzed an external validation dataset of adult inpatient encounters (n=53,974). Patients' actual opioid misuse was assessed by screening questions administered at admission.

The key outcome was the false-negative rate (FNR): the percentage of actual opioid misusers whom the classifier missed. The FNR was considerably higher among Black patients (32 percent) and "Other" racial and ethnic groups (33 percent) than among White patients (17 percent) and Hispanic/LatinX patients (17 percent). This suggests a race disparity in resource allocation: while 83 percent of White and Hispanic/LatinX patients would receive needed resources based on the classifier, only 67 percent of Blacks/Others would.

3.2.3.7 Prostate Cancer Risk

Table 13. Description of prostate cancer screening algorithms

Algorithms	Description	Background
• Kaiser Permanente Prostate Cancer Risk Calculator (KPPC RC) version A and B • Prostate Cancer Prevention Trial (PCPT) 2.0 and the Prostate Biopsy Collaborative Group (PBCG) algorithms	Both versions of the KPPC RC include age, race (patient-reported), body mass index, family history of prostate cancer, number of prior biopsies, prostate-specific antigen (PSA) level, and digital rectal exam result; version B also contains prostate volume. For each version, 2 risk threshold values for biopsy recommendation (≥7.5% and ≥10%) were tested. Input variables in PCPT and PBCG include age, PSA level, digital rectal exam result, previous biopsy history, and (for PBCG) family history of prostate cancer; both algorithms also include race as an input.	Overdiagnosis is a major concern in prostate cancer, so a primary goal of the pertinent algorithms is to decrease the overall biopsy rate by reducing the number of unnecessary biopsies, while still detecting as many clinically significant cancers as possible.

Two studies applied algorithms retrospectively to cohorts of men who had received biopsies based on abnormal digital rectal exams and/or elevated levels of prostate-specific antigen (PSA).[56,60] The studies calculated, if the biopsy decisions had been based solely on the algorithms, how many negative biopsies would have been avoided, how many total biopsies would have been avoided, and how many cancers would have been missed. A "net benefit" calculation illustrates the key tradeoff, as it provides the number of negative biopsies avoided for each missed high-grade cancer.

In one study, a newly developed algorithm, the Kaiser Permanente Prostate Cancer Risk Calculator, was externally validated.[56] Results were presented by racial and ethnic category for two versions of the model (versions A and B).

The net benefit for members of each racial and ethnic category differed substantially depending on the model version and threshold value used. For example, using version A and a risk threshold of ≥10 percent, 9 percent of negative biopsies would have been avoided among White patients while missing 1 percent of high-grade cancers (i.e., 9:1 ratio); among Black patients, 25 percent of negative biopsies would have been avoided, but 6 percent of high-grade cancers would have been missed, yielding a ratio of only about 4:1. By contrast, using version B and a ≥10 percent cutoff, the percentages of unnecessary biopsies avoided and high-grade cancers missed would be 39 percent and 4 percent, respectively, for White patients (i.e., about a

10:1 ratio), and 61 percent and 5 percent, respectively, for Black patients (i.e., about a 12:1 ratio).

For Hispanic patients, the effect of using the algorithm would have been relatively low; using version B with a ≥10 percent cutoff, 24 percent of negative biopsies would have been avoided and 1 percent of high-grade cancers would have been missed. For Asian patients, the net benefit of using the algorithm would have been, in general, less positive than for the other groups. Using version B with a ≥10 percent cutoff, 51 percent of negative biopsies would have been avoided, but 9 percent of high-grade cancers would have been missed (i.e., 5.7:1 ratio).

The second study compared the Prostate Cancer Prevention Trial (PCPT) 2.0 and the Prostate Biopsy Collaborative Group (PBCG) algorithms.[60] Decision curve analysis was used to calculate the net benefit that would have accrued to the men in the sample if each algorithm had been used to determine whether they should have a biopsy, as well as the net benefit of conducting biopsies on all men. While the article does not define "net benefit," the term has a formal definition in decision curve analysis; it is a function of the true-positive rate, the false-positive rate, and the threshold risk value. Each strategy's net benefit was calculated, using a range of threshold risk values from 5 percent to 40 percent (described by the authors as the range that "patients and providers usually have"), separately for White men, Black men, and others (which were 75 percent Hispanic and 25 percent Asian). The authors depicted the results graphically in their Figure 3 (see original article). In general, the net benefit of all three strategies declined, or at best remained constant, as the threshold probability used increased. For both Black and White men, neither algorithm had a net benefit superior to that of the strategy of conducting a biopsy of all men at any risk threshold below 30 percent. For Black men, there was little difference in net benefit for any of the three strategies, except at the 40 percent risk threshold, at which the net benefit of performing a biopsy on all men became slightly negative while that of the two algorithms remained positive; PCPT's net benefit was slightly higher than that of PBCG at values above 30 percent. For White men, by contrast, PCPT's net benefit was lower than that of PBCG at all threshold probabilities, while the net benefit of performing a biopsy on all men was similar to that of PBCG except at threshold values over 30 percent, where it dropped below the two algorithms. The net benefit of all three strategies was higher for Black men than for White men at all threshold values.

For the men in the sample belonging to other racial and ethnic groups, the net benefit of two of the strategies (PBCG and biopsying all men) was lower than that for Black or White men at all threshold values. The net benefit of performing biopsy on all men was lower than that of PBCG for threshold values of 20 percent or above but was negative for both. At threshold values above 10 percent, PCPT's net benefit was higher than that of the other two (remaining slightly positive) and was comparable to PCPT's net benefit among White men.

3.2.3.8 Severity of Illness Measurements for Crisis Standards of Care

Table 14. Description of severity of illness measurements

Algorithms	Description	Background
• Acute Physiology and Chronic Health Evaluation (APACHE IVa) • Laboratory-based Acute Physiology Score version 2 (LAPS2) • Oxford Acute Severity of Illness Score (OASIS) • Sequential Organ Failure Assessment (SOFA)	These models are used to describe acute severity of illness and predict in-hospital mortality in hospitalized patients. The SOFA score was not designed to predict mortality but has been widely used for that purpose. Race and ethnicity are not included in these algorithms.	Crisis standards of care (CSC) direct ethical decision making when demand for resources in an intensive care unit exceeds the available supply; the underlying principle is to direct resources to those patients most likely to survive with appropriate care. Professional societies have advocated the use of these existing severity of illness models to inform CSC, although they were not originally developed for this use.

Three studies[54,57,107] retrospectively evaluated racial and ethnic differences in the performance of four models used to predict risk of mortality:APACHE IVa, LAPS2, OASIS, and SOFA. None of the models use race or ethnicity as an input.

All three studies selected patients from an EHR[107] or intensive care unit (ICU) database such as the eICU Collaborative Research database (eICU-CRD)[54,57] or the Medical Information Mart for Intensive Care-III (MIMIC-III) database[57]. Two studies limited analyses to Black and White patients,[54,107] and the third study also included Hispanic and Asian patients.[57] For more information, see Table C-1 in Appendix C.

One study compared the performance of SOFA and LAPS2 in Black and White patients (n=113,158) admitted to 27 hospital EDs between 2013 and 2018 with acute respiratory failure or sepsis (Table C-1 in Appendix C).[107] Most patients were White (75.6 percent), and White patients were older than Black patients (mean age: 67.1 vs. 61.7, $p<0.001$). Black patients had higher overall mean SOFA scores (3.1 [standard deviation (SD) 2.1] vs. 2.9 [SD 1.8], $p<0.001$) than White patients, indicating a lower predicted likelihood of survival, but lower mean LAPS2 scores (102.2 [SD 38.4] vs. 103.1 [SD 36.7], $p<0.001$), indicating a higher predicted likelihood of survival. However, authors found that at a given SOFA or LAPS2 score, Black patients had lower in-hospital mortality than White patients in almost every category, suggesting that both prediction models overestimated in-hospital mortality for Black patients and underestimated this outcome for White patients (Table D-1 in Appendix D). Use of these models to prioritize resource allocation in ICUs (with priority given to patients most likely to survive) would thus tend to lead to inappropriate deprioritization of Black patients.

This study also examined the subset of patients in the highest-priority category (i.e., SOFA <6), indicating a higher predicted likelihood of survival, and again found that Black patients had lower in-hospital mortality than White patients (5.3 percent vs. 6.9 percent).[107] To illustrate the extent of inappropriate deprioritization associated with this discrepancy, the authors performed a simulation analysis in which Black patients with SOFA scores ≥6 were sequentially reclassified into the highest-priority category until rates of in-hospital mortality for Black and White patients in that category were similar (6.7 percent vs. 6.9 percent). The Black patients thus reclassified were those with SOFA scores of 6 to 8 (n=2611), representing 9.4 percent of all Black patients and 81.6 percent of Black patients with SOFA >5 (Table D-1 in Appendix D). Overall, authors found the use of illness severity models in crisis standards of care, in particular SOFA, “may

divert critical care resources away from Black patients and lead to racial disparities in resource allocation."[107]

One study examined the APACHE IVa, OASIS, and SOFA illness severity models in patients admitted to the ICU.[57] Participants were selected from two ICU databases: the eICU-CRD (n=122,919) between 2014 and 2015, which contains APACHE IVa scores, or the MIMIC-III (n=43,823) between 2001 and 2012, which includes OASIS scores. Authors calculated SOFA scores for participants in both databases. Race and ethnicity data captured in each database were self-reported (Table C-1 in Appendix C). Both APACHE IVa and OASIS overestimated mortality for all racial and ethnic groups, and overestimates were worse for Black and Hispanic patients. Standardized mortality ratios of observed to predicted death rates for both prediction models were lower (i.e., overestimated mortality to a greater extent) for Black (0.67 for both models) and Hispanic (0.73 and 0.64) patients than for White (0.76 and 0.81) and Asian (0.77 and 0.95) patients, respectively. Among patients with SOFA scores 0 to 7, the ratio of the observed ethnicity-specific mortality rate to the mortality rate in the overall population was lower for Black (0.86 and 0.74) and Hispanic (0.96 and 0.62) patients than for White (1.02 and 1.04) and Asian (1.12 and 1.06) patients when calculated for both databases, eICU-CRD and MIMIC-III, respectively. Thus, to be placed in the lowest category of predicted risk (and therefore be assigned the highest priority), Black and Hispanic patients had to have lower true risk than White and Asian patients. See Table D-1 in Appendix D.

A third study examined whether use of the SOFA score is associated with inappropriate deprioritization of Black patients in currently adopted crisis standards of care.[54] For use in allocating resources, SOFA scores are collapsed into tiers; depending on the severity of the shortage, resources may be allocated only to patients in the highest-priority tier (i.e., those with the lowest scores and lowest risk of mortality), the two highest-priority tiers, etc. Authors evaluated three widely used tier systems, termed A (4 tiers, with scores <6 forming the highest-priority tier and scores ≥12 forming the lowest), B (3 tiers, scores <8 highest priority, ≥12 lowest), and C (4 tiers, scores <9 highest priority, ≥15 lowest). SOFA scores were retrospectively calculated for 111,885 patient encounters involving 95,549 unique patients in the eICU-CRD occurring between 2014 and 2015. The sample included 16,688 encounters with Black patients (14.9 percent) and 95,197 encounters with White patients (85.1 percent) (Table C-1 in Appendix C).

One analysis modeled actual in-hospital mortality using the continuous version of the SOFA score, race, and the interaction of race by SOFA score.[54] The interaction was significant (OR for Black vs. White, 0.98; 95 percent CI, 0.97 to 0.99; p<0.001). This indicated a small but statistically significant tendency for the SOFA score to overestimate the true risk of death among Black patients compared with White patients, thus hypothetically lowering their eligibility for resources compared with White patients. See Table D-1 in Appendix D.

Another analysis in this study examined the tier systems.[54] For each system, this analysis focused on the subset of patients in the highest-priority tier (i.e., those who would be prioritized for resources under conditions of severe shortage) and compared the adjusted odds of in-hospital mortality among Black versus White patients (Table D-1 in Appendix D). Black patients had significantly lower odds of in-hospital mortality than White patients in the highest tier of system A (OR, 0.65; 95 percent CI, 0.58 to 0.74; p <0.001), system B (OR, 0.70; 95 percent CI, 0.64 to 0.78; p <0.001), and system C (OR, 0.73; 95 percent CI, 0.67 to 0.80; p <0.001), indicating that, under each system, Black patients had to have a lower true risk of death to qualify for resources than White patients. The percentage of Black patients who would have been inappropriately

deprioritized (i.e., assigned to a lower-priority tier even though their true risk of death was lower than some patients in the highest-priority tier) was 15.6 percent for system A, 9.0 percent for system B, and 6.5 percent for system C. Across all systems and all levels of shortage, increasing the SOFA threshold by 2 points for Black patients would be necessary to equalize the adjusted odds of death for Black and White individuals who qualify for the high-priority tier.

A fourth study examined differences across racial groups in estimated rates of excess deaths that would have been caused by using a crisis standards of care algorithm to ration mechanical ventilators.[65] The sample consisted of patients who were admitted to the ICUs of 6 hospitals in a Boston-area hospital system in April and May of 2020 and received mechanical ventilation (n=244). The distribution of self-reported race in this sample was 16.8 percent Black, 49.1 percent White, 2.8 percent Asian, 10.6 percent of any other reported race, and 20.4 percent of unknown race. Priority scores were preemptively calculated for these patients in anticipation of resource shortages due to COVID-19 (which did not materialize), using state-issued guidelines. Scores were based on estimated likelihood of acute and long-term survival. Acute survival was estimated using the SOFA score grouped into four categories (1 point for best prognosis, 4 points for worst). Long-term survival was estimated using a 3-level score (0 points for best prognosis, 2 points for intermediate, 4 points for worst) based either on comorbidity data from the electronic medical record (through April 27) or on a clinical assessment by the attending physician (after April 27). The total score was the sum of the acute and long-term scores and was grouped into three tiers: highest priority (scores of 1 or 2), intermediate (3 to 5), and lowest (6 to 8). If this system had been used to allocate ventilators to the patients in the sample, 140 would not have received ventilation if ventilators had been allocated only to patients in the highest-priority tier, and 30 would not have received ventilation if ventilators had been allocated to patients in the highest and intermediate tiers. The analysis assumed that all patients who lived, but would not have received ventilators under these scenarios, would have died (i.e., excess deaths). At the cutoff of ≤2, the estimated number of excess deaths among Black patients was 18 (i.e., there were 18 patients who lived, but would not have received ventilation because they had scores >2). This represented 43.9 percent of all Black patients in the sample compared with a rate of 28.6 percent among the other 203 patients ($p = 0.05$). At a cutoff of ≤3, the estimated excess mortality among Black patients was 26.8 percent and 14.3 percent among all other patients ($p = 0.05$). There were no statistically significant differences in excess mortality between Black patients and all other patients at any other cutoff (p's ≥0.08) or between Black and White patients at any cutoff (p's ≥0.22).

In summary, findings from these four studies[54,57,65,107] indicate that illness severity models consistently overestimated mortality in Black patients compared with White patients. That is, at any given level of algorithm-predicted risk, Black patients had lower actual mortality than White patients. Using prediction models that overestimate mortality in Black patients can lead to inappropriate deprioritization and divert resources away from Black patients.

3.2.3.9 Stroke Risk

Table 15. Description of the CHA_2DS_2-VASc

Algorithm	Description	Background
CHA_2DS_2-VASc	CHA_2DS_2-VASc (congestive heart failure, hypertension, age ≥75, diabetes, stroke, vascular disease, age 65 to 74 years, sex category) predicts stroke risk in patients with atrial fibrillation. The algorithm starts at 0 and adds points for the following input variables: • Age 65 to 74 (+1) or > 75 (+2) • Female (+1) • Congestive heart failure history (+1) • Hypertension history (+1) • Stroke / Transient Ischemic Attack / thromboembolism history (+2) • Vascular disease history (+1) • Diabetes history (+1) The algorithm informs the American College of Cardiology (ACC)/American Heart Association (AHA) atrial fibrillation treatment guideline. Race and ethnicity are not included as input variables. 2014 ACC/AHA guideline recommendation: *do not* recommend antithrombotic therapy for male patients with a score of 0 or female patients with a score of 1. 2020 ACC/AHA guideline recommendation: *recommend* antithrombotic therapy for male patients with a score ≥ 2 and female patients with a score ≥ 3. *Consider* antithrombotic therapy for male patients with a score of 1 and female patients with a score of 2.	The 2020 ACC/AHA clinical practice guideline (CPG) for atrial fibrillation treatment uses a higher CHA_2DS_2-VASc threshold when recommending antithrombotic therapy than the threshold used in the 2014 CPG. In the 2020 version, guideline developers sought to acknowledge that biological sex does not increase the risk of stroke and increased the treatment threshold by 1 for female sex.[66] Although a new CPG is available, the previous version (2014) may still be in use.

One study examined potential racial and ethnic disparities in health outcomes that could occur as a result of using the CHA_2DS_2-VASc, which predicts stroke risk in patients with atrial fibrillation and is used to guide recommendations for anticoagulation treatment.[66]

The study data came from the Stanford Medicine Research Data Repository, which is composed of records from Stanford Health Care and the Lucile Packard Children's Hospital. Race and ethnicity were self-reported; the racial and ethnic groups included were White, Black, Hispanic, and Asian. Potential disparities in health outcomes were quantified by identifying individuals who would have been denied treatment based on CHA_2DS_2-VASc and ascertaining the frequency of negative events (stroke) among these individuals.

The sample consisted of 233,129 patients (176,278 White, 33,927 Asian, 13,578 Hispanic, 7323 Black, and 2023 other). The 2014 American College of Cardiology/American Heart Association atrial fibrillation treatment guideline, an algorithm-informed decision tool, uses a threshold of a CHA_2DS_2-VASc score > 1 to recommend anticoagulant therapy. The negative event frequency was 3.30 for Hispanic patients (i.e., 3.3 percent of Hispanic patients who would not have received a recommendation for anticoagulant therapy had a stroke), 2.26 for Asian patients, 2.21 for White patients, and 2.19 for Black patients. The 2020 guideline uses a higher threshold (CHA_2DS_2-VASc score >2 for males and > 3 for females) for recommending anticoagulant therapy. The negative event frequency when using the 2020 guideline was 2.21 for Black patients (i.e., 2.21 percent of Black patients who would not have received a

recommendation had a stroke), 2.15 for Asian patients, 2.14 for White patients, and 1.78 for Hispanic patients. Authors suggest that when using the 2020 guideline, the disparity in negative event frequency (stroke) in Hispanic patients was reduced.

This study also examined potential racial and ethnic disparities in health outcomes that could occur as a result of using the Model for End-Stage Liver Disease (MELD) calculator and simplified Pulmonary Embolism Severity Index (sPESI). However, authors reported that the algorithms performed poorly for different racial and ethnic groups. Due to limitations in subgroup performance, there was insufficient information about the potential impact of the MELD calculator and sPESI on racial and ethnic health outcomes; therefore, we present study findings only for the CHA_2DS_2-VASc scor

3.3 Key Question 2. What is the effect of interventions, models of interventions, or other approaches to mitigate racial and ethnic bias in the development, validation, dissemination, and implementation of healthcare algorithms?

a. Datasets: What is the effect of interventions, models of interventions, or approaches to mitigate racial and ethnic bias in datasets used for development and validation of algorithms?

b. Algorithms: What is the effect of interventions, models of interventions, or approaches to mitigate racial and ethnic bias produced by algorithms or their dissemination and implementation?

3.3.1 Description of Included Evidence

Our searches identified 44 studies (Table 16) published between 2011 and 2023 that met our inclusion criteria and evaluated strategies to mitigate racial and ethnic disparities associated with healthcare algorithms.[5,21,23,68-108] The evidence base included 1 randomized controlled trial,[93] 17 studies that used cohort, pre-post, or cross-sectional designs,[5,21,23,71,72,74-77,83,86-88,92,95,107,108] and 26 studies involving comparison models or simulated effects.[68-70,73,78-82,84,85,89-91,94,96-106] Detailed information about the included studies is provided in Table C-2 in Appendix C and Table D-2 in Appendix D.

Twenty-one studies measuring kidney function[21,23,69,70,72,73,75-77,79,80,82,83,86,99-105] and seven studies predicting cardiovascular risk[81,84,87,88,90,95,96] composed the majority of the research on mitigation strategies, but numerous other clinical issues were addressed as well. Four studies addressed organ donation,[89,94,97,98] three studies examined algorithms for appropriate warfarin dosing,[92,93,106] and the remaining studies addressed lung function,[68,71] stroke risk,[91] intensive care needs,[107] lung cancer screening,[74] postpartum depression,[78] opioid misuse,[108] and healthcare costs and utilization.[5,85]

Our searches identified numerous strategies used to mitigate bias in healthcare algorithms. Broadly, these strategies fall into six categories: removing an input variable (usually race and ethnicity) without changing an algorithm's other features; replacing an input variable with one or more different variables; adding one or more input variables without removing any; changing the racial and ethnic composition of the patient population used to train or validate a model; stratifying algorithms by race and ethnicity; or modifying the statistical or analytic techniques used by an algorithm. Three studies[73,78,88] used more than one strategy. The most common approach, used in 24 of 44 studies, was to remove race. Not surprisingly, this strategy was used predominantly in studies of eGFR, but removal of race-based variables was also evaluated in studies of lung function,[68,71] kidney donor suitability,[97,98] and postpartum depression.[78] Five studies replaced an input variable with something different; three of these replaced race in eGFR with biological indicators, such as cystatin-C or metabolic markers,[23,73,86] while a study of the Kidney Donor Risk Index (KDRI) replaced race with a genetic marker.[89] Lastly, in Obermeyer's landmark study of a healthcare needs algorithm that did not include race, the utilization variable that was identified as causing disparities was replaced with three other measures of patient needs that were not associated with outcome disparities.[5]

3.3 Results, Key Question 2

Eight studies added an input variable to improve algorithm performance: three of these added race to address disparities associated with initially race-free algorithms for risk of cardiovascular disease[87,95] or stroke;[91] four studies added genetic or other biological variables to cardiovascular risk prediction algorithms[81,90] or warfarin dosing algorithms;[93,106] one study added a measure of life-years gained to a screening algorithm for lung cancer;[74] and one study incorporated social determinants of health (SDOH) measures into an algorithm that predicted healthcare use, costs, and mortality.[85]

Four studies recalibrated models using datasets derived from a different mix of patients than those used for initial model development. These studies focused on cardiovascular risk for Black patients,[84,88] postpartum depression for women who receive Medicaid,[78] and Black liver donors with hepatitis C.[94] In two studies, the authors sought to address concerns about algorithms for warfarin dosing[92] and opioid misuse[108] by developing separate algorithms or thresholds for Black and White patients. Finally, in three studies that focused on postpartum depression[78] and cardiovascular risk,[88,96] the statistical methods used for model calibration were modified with innovative techniques designed to mitigate potential algorithmic bias.

As with KQ 1, studies addressing KQ 2 usually included patients from EHRs, clinical trials, or national databases such as the National Health and Nutrition Examination Survey. In 14 studies, race and ethnicity were self-reported,[5,68-70,81,85,87,99,102-106,108] while 27 studies did not describe how race and ethnicity were collected. Two studies[82,101] included a combination of self-reported and administratively designated classifications for race and ethnicity, and one study[91] employed an algorithm developed by the Research Triangle Institute that assigns race and ethnicity based on first and last name. In 29 studies, analyses were restricted to only 2 race and ethnicity groups (African American/Black and White/Non-Black), while 5 other studies reported data for these 2 groups in addition to patients categorized as Asian or Hispanic.[71,74,79,85,105] Because KQ 2 focused on mitigation strategies, we also identified 10 studies that reported outcomes only for Black patients.[69,70,75,77,80,82,83,89,90,104]

Overall, ROB was rated as Low for 8 studies, Moderate for 31 studies, and High for 5 studies. The strengths and limitations affecting ROB for the KQ 2 studies were similar to those for the KQ 1 studies, as described above. The equity-based signaling questions changed domain ROB in six studies, but only changed overall ROB in one study.[91] The most common domain to receive a change in ROB rating was for bias in selection of study participants due to inconsistent reporting of racial and ethnic groups (i.e., self-reported as ideal) and with inconsistent definitions and categories. Complete ROB ratings are in Appendix Table D-3.

3.3 Results, Key Question 2

Table 16. Mitigation strategies

Mitigation Strategy	Study	Initial Algorithm	Revised Algorithm	Algorithm Includes Race or Ethnicity	Effect of Mitigation Strategy
Removed race	19 studies[21,69,70,72,73,75-77,79,80,82,83,99-105]	eGFR for kidney function	eGFR without race coefficient	Initial algorithm	• Increased diagnosis of CKD or severe CKD: 8 studies[21,69,70,77,79,80,82,83,105] • Improved accuracy: 5 studies[99-103] • Increased access to care: 4 studies[72,77,80,105] • Reduced access: 2 studies[69,104] • Improved antibiotic dosing: 1 study[76] • Underestimated GFR: 1 study[73] • Mixed effects on organ transplant: 1 study[77] • No effect on prediction of acute kidney injury: 1 study[75]
	Doshi et al. 2022[97]	KDPI/KDRI for kidney donor suitability	KDPI/KDRI without race variable	Initial algorithm	Reduced disparity in availability of donor kidneys
	Miller et al. 2022[98]	KDPI/KDRI for kidney donor suitability	KDPI/KDRI without race variable	Initial algorithm	No significant effect on non-use of donor kidneys
	Baugh et al. 2022[68]	GLI spirometry equation for lung function	Same equation without race variable	Initial algorithm	Improved accuracy of evaluation of lung function
	Elmaleh-Sachs et al. 2021[71]	GLI spirometry equation for lung function	Same equation without race variable	Initial algorithm	No difference in evaluation of lung function
	Park et al. 2021[78]	Novel risk prediction algorithm for postpartum depression	Same algorithm without race variable	Initial algorithm	Reduced disparities in prediction of postpartum depression and likelihood to use mental health services
Replaced variable	Inker et al. 2021[23]	eGFR for kidney function	Replaced race with creatinine, cystatin-C, beta-trace protein, beta2-microglobulin	Initial algorithm	Improved accuracy of GFR estimation
	Inker et al. 2021[73]	eGFR for kidney function	Replaced race with cystatin-C	Initial algorithm	Reduced disparity in GFR
	Coresh et al. 2019[86]	eGFR for kidney function	Replaced race with metabolic panel	Initial algorithm	Improved accuracy of GFR estimation
	Julian et al. 2017[89]	KDRI for kidney donor suitability	Replaced race with apo lipoprotein L1 genotype	Initial algorithm	Improved accuracy of prediction kidney graft failure

3.3 Results, Key Question 2

Mitigation Strategy	Study	Initial Algorithm	Revised Algorithm	Algorithm Includes Race or Ethnicity	Effect of Mitigation Strategy
	Obermeyer et al. 2019[5]	Novel risk prediction algorithm for complex healthcare needs	Replaced outcome variable based on national costs with: local costs, avoidable costs, and severity of chronic conditions	None of the algorithms	Reduced disparity in eligibility for care management
Added race	Topel et al. 2012[87]	FRS for CVD risk	Added race variable to same algorithm	Revised algorithm	Reduced disparity in measures of subclinical vascular disease
	Drawz et al. 2018[95]	FRS for CVD risk	Added race variable to same algorithm	Revised algorithm	No effect on risk classification for heart disease
	Kabra et al. 2016[91]	CHA_2DS_2-VASc for stroke risk	Added race variable to same algorithm	Revised algorithm	Reduced disparity in prediction of stroke risk
Added non-race variables	Weale et al. 2021[81]	ASCVD for CVD risk	Added polygenic risk scores to same algorithm	Both algorithms	Improved accuracy of prediction of CVD
	Fox et al. 2016[90]	Novel risk prediction algorithm for CVD	Added 10 biomarkers	Both algorithms	No effect on reclassification of severity of CVD
	Kimmel et al. 2013[93]	Warfarin dosing	Added genotype data	Both algorithms	Exacerbated disparity in time in therapeutic warfarin range
	Lindley et al. 2022[106]	Warfarin dosing	Added genotype data	Both algorithms	Improved accuracy of prediction of therapeutic warfarin dose
	Landy et al. 2021[74]	USPSTF-2020 for lung cancer risk	Added life years from screening with computed tomography	Neither algorithm	Reduced disparity in prediction of lung cancer death
	Hammond et al. 2020[85]	Novel risk prediction algorithm for complex healthcare needs	Added 7 measures of SDOH: rural vs urban; alcohol abuse; access to care; economic status; financial strain; marital status; education	None of the algorithms	Improved accuracy of prediction of hospitalization, death, and healthcare costs
Algorithmic recalibration	Fairman et al. 2020[84]	ASCVD for CVD	Refined the algorithm using newer and more racially diverse patient cohorts	Both algorithms	Reduced disparities in statin prescribing and prediction of CVD events
	Yadlowsky et al. 2018[88]	ASCVD for CVD	Refined the algorithm using newer and more racially diverse patient cohorts	Both algorithms	Reduced disparity in prediction of CVD
	Park et al. 2021[78]	Novel risk prediction algorithm for postpartum depression	Refined the algorithm by reweighing key population groups during model training	Both algorithms	Reduced disparities in prediction of postpartum depression and likelihood to use mental health services
	Shores et al. 2013[94]	Donor Risk Index for liver transplant suitability	Refined the index using population of Black liver recipients with Hepatitis C	Both algorithms	Improved accuracy of prediction of liver graft failure
Population stratification	Limdi et al. 2015[92]	COAG for warfarin dosing	Used race-stratified analysis rather than race-combined algorithms	Both algorithms	Improved accuracy of warfarin dosing
	Thompson et al. 2021[108]	Novel risk prediction algorithm for opioid misuse	Used race-specific thresholds and recalibration by racial subgroup	All algorithms	Reduced disparity in referral to education and treatment

3.3 Results, Key Question 2

Mitigation Strategy	Study	Initial Algorithm	Revised Algorithm	Algorithm Includes Race or Ethnicity	Effect of Mitigation Strategy
Statistical adjustment	Foryciarz et al. 2022[96]	ASCVD for CVD	Used a group recalibrated model and an equalized odds model	Both algorithms	Increased overall accuracy of prediction of CVD, but reduced accuracy for racial groups
	Yadlowsky et al. 2018[88]	ASCVD for CVD	Used elastic net regularization to reduce model overfitting	Both algorithms	Reduced disparity in prediction of CVD
	Ashana et al. 2021[107]	SOFA for CVD risk	Reclassified threshold for intervention	Neither algorithm	Reduced disparity in eligibility for high-priority care
	Park et al. 2021[78]	Novel risk prediction algorithm for postpartum depression	Added a regularization term that adjusts the algorithm to limit the effect of race-based variables	Both algorithms	Reduced disparities in prediction of postpartum depression and likelihood to use mental health services

Abbreviations: ASCVD = atherosclerotic cardiovascular disease; CKD = chronic kidney disease; COAG = Clarification of Oral Anticoagulation through Genetics study; CHA2DS2-VASc = congestive heart failure, hypertension, age ≥75, diabetes, stroke, vascular disease, age 65 to 74 years, sex category; CVD = cardiovascular disease; eGFR = estimated glomerular filtration rate; FRS = Framingham Risk Score; GLI = Global Lung Function Initiative; KDPI = Kidney Donor Profile Index; KDRI = Kidney Donor Risk Index; SLKT = simultaneous liver-kidney transplantation; SDOH = social determinants of health; SOFA = Sequential Organ Failure Assessment; USPSTF = United States Preventive Services Task Force

3.3.2 Key Points

- We included 44 studies addressing a broad range of clinical assessment. The most frequently examined algorithms evaluated kidney function and cardiovascular risk.
- Six types of mitigation strategies were examined, with some studies testing multiple strategies. The most common approach was removal of race, which occurred in 24 studies. Five studies replaced race or another input variable with a different measure, while nine studies added an input variable to an algorithm. In four studies, an algorithm was recalibrated with a more representative patient population. Two studies developed stratified algorithms that assessed Black and White patients separately, and three studies evaluated the effect of different statistical techniques within algorithms.
- The evidence base featured considerable heterogeneity across patient populations, clinical conditions, healthcare settings, and primary outcomes.
- Removing race from eGFR may increase the likelihood of diagnosis of chronic kidney disease and severe kidney disease in Black patients. This could result in broader and earlier eligibility for kidney transplant. Conversely, removing race from eGFR might reduce access to other types of treatment, affect medication dosing for a broad range of conditions, and reduce enrollment of Black patients in clinical trials.
- Most published studies found that mitigation strategies may reduce racial and ethnic disparities and could improve outcomes for BIPOC patients. However, strategies that improve one outcome (e.g., eligibility for kidney transplant) may have undesired effects on other outcomes (e.g., medication dosing or eligibility for enrollment in clinical trials).
- A mitigation strategy's effectiveness may depend critically on a unique combination of algorithm, clinical condition, population, setting, and outcomes. It is unclear from the current evidence base if certain types of strategies are generally more effective than others, or what the implications are for both existing and future algorithms.

3.3.3 Summary of Findings

The sections below discuss six categories of mitigation strategies:

- Removing Input Variables
- Replacing Input or Outcome Variables
- Adding Input Variables
- Changing the Patient Mix Used for Development and Validation
- Developing Separate Algorithms by Race
- Refining Statistical and Analytic Techniques

3.3.3.1 Removing Input Variables

Our review identified 24 studies that examined the effect of removing race from an algorithm. Nineteen of these studies focused on kidney function as estimated by eGFR, two evaluated kidney donor suitability in the KDRI and Kidney Donor Profile Index (KDPI), two examined lung function, and one study addressed postpartum depression.

The 19 studies of eGFR were heterogeneous in their overall objectives and in the type and number of outcomes assessed. Ten studies examined the effect of removing race from eGFR on diagnosing kidney disease or classifying disease severity.[21,69,70,73,77,79,80,82,99,105] Seven studies

evaluated prediction of mortality,[102] kidney failure,[70,102] end-stage kidney disease,[103] progression of kidney disease in patients with human immunodeficiency virus,[101] and acute kidney injury in patients with cirrhosis[75] or following percutaneous coronary intervention.[100] Finally, several studies explored possible downstream effects of removing race, including changes to medication dosing,[76,105] appropriateness of drug therapy or other treatments,[69,105] potential enrollment in clinical trials,[104] and eligibility for kidney[21,72,80,83,105] or joint liver-kidney transplant.[77]

The effects of removing race from eGFR were consistent across most studies, although some variation in outcomes emerged. One analysis[70] of a national healthcare database found that the proportion of Black patients qualifying for a diagnosis of chronic kidney disease more than doubled when the race coefficient was removed. In a separate analysis of Veterans Administration patients in that same study, diagnosis in Black patients rose by 74 percent. A study by Shi et al.[79] demonstrated that after removing the race coefficient, between 16 percent and 38 percent of Black patients in every severity class (stages 1 through 4) were reclassified to a higher stage, and no Black patients were reclassified to a lower stage. Conversely, when patients of all races were combined, between 1 percent and 30 percent of patients moved to a lower stage while just 1 percent to 5 percent moved to a higher stage.

A study[77] examining data from a national transplant registry found that removing the race coefficient from eGFR led to a 26 percent increase in eligibility of Black patients for kidney transplant waitlists. Hoenig et al.[72] reported a more modest but still meaningful increase in transplant eligibility; the authors found that 15 percent of patients who were added to a transplant list after the race coefficient was removed would not have been eligible when the race coefficient was in use. Diao et al.[105] found that Black patients would have expanded access to nephrology referral and preemptive arteriovenous fistula, while Medicare coverage of kidney disease education and medical nutrition therapy would increase by 45 and 48 percent, respectively.

Two studies examined the effect of removing race in eGFR in patients with cirrhosis. As background, patients with cirrhosis are assessed using the Model for End-Stage Liver Disease score, which incorporates the race-based eGFR equation and can drive clinical decision making, including liver transplant eligibility. Mahmud et al.[75] found that inclusion of race in eGFR in a Veterans Administration dataset did not improve prediction of acute kidney injury events. Panchal et al.[77] used data from a national transplant registry and found that removing the race coefficient from eGFR could lead to a 26 percent increase in eligibility of Black patients for simultaneous liver-kidney transplantation.

Not all consequences are necessarily positive. A multivariate model[21] found that removing race from eGFR would result in no Black patients referred to transplant waitlists; the reason for that is unclear. Casal et al.[69] reported that, although 26 percent of Black patients who were undergoing cancer treatment were reclassified as having more severe kidney disease, 5 percent were newly deemed ineligible to receive cisplatin because their revised renal function estimate exceeded standard medication safety thresholds. Diao et al.,[105] using 18 years of National Health and Nutrition Examination Survey (NHANES) data, determined that 38 percent of Black patients would see a reduction in their dose of common medications (e.g., beta blockers, statins, opioids), with unknown implications. Finally, Schmeusser et al.[104] reported that eligibility of Black patients for participation in cancer clinical trials could decrease significantly when eGFR is estimated without a race coefficient.

Five studies removed race from algorithms other than eGFR. Baugh et al.[68] compared spirometry measures with and without race-based equations in patients with chronic obstructive pulmonary disease. The authors found that race-specific algorithms slightly overestimated

healthy lung function in Black patients compared with that of non-Hispanic White patients; predicted mean forced expiratory volume (FEV_1) was 5 percent higher and predicted mean forced vital capacity (FVC) was 2.3 percent higher in Black patients. Removing race led to more accurate assessment of lung function, with Black patients demonstrating a mean FEV_1 that was 7.9 percent lower and FVC 16.3 percent lower, than non-Hispanic White patients. Another study examined the impact of spirometry equations with or without race and ethnicity on predicting chronic lower respiratory disease (CLRD) events and all-cause mortality in Asian, Black, Hispanic, and White patients.[71] Findings suggest that percentage predicted FEV_1 and FVC with race-specific spirometry equations did not appear to improve the prediction of CLRD events or all-cause mortality compared with race-neutral equations.[71] The C-statistic for the standard race-specific spirometry equation predicting CLRD-events was 0.71 for FEV_1 and 0.61 for FVC. Authors found very similar C-statistics (0.72 and 0.62) for the race-neutral equation. Findings were similar for all-cause mortality.

Kidney donation was the subject of two studies that examined the KDRI and KDPI.[97,98] These are interrelated indices that predict graft failure following transplantation, using donor characteristics including race. Published in 2022 by separate research teams analyzing data from the Scientific Registry for Transplant Recipients, both studies had similar findings. They reported that removing race as a variable did not affect the predictive accuracy of the algorithms but did result in a small increase of approximately 50[98] to 70[97] kidneys from Black donors that might become available annually. Finally, a 2021 study[78] evaluated an algorithm designed to predict diagnosis and treatment needs associated with postpartum depression. Using data on Medicaid beneficiaries, the authors compared three mitigation strategies to improve the model and yield more accurate prediction. The study modeled three alternative versions of the algorithm: without race; with addition of a statistical adjustment designed to limit race-based effects, and following recalibration based on a reweighing of key population subgroups. The latter two strategies are described below in the respective sections addressing those approaches. The authors found that removing race improved the algorithm's accuracy, and this approach was more effective than adding a statistical adjustment but less effective than recalibrating the model with more diverse patient data.

3.3.3.2 Replacing Input or Outcome Variables

Five studies evaluated the impact of replacing initial algorithmic variables (either inputs or outcomes) with alternative variables. Three studies replaced race in eGFR with biological measures, and one replaced race with a biologically relevant genotype in the KDRI. The fifth study, by Obermeyer et al., demonstrated how variables other than race could unintentionally affect healthcare disparities.

Substantial interest in alternatives to race-based GFR estimation has led to much recent research. We identified three studies meeting our review criteria that also represent the current research addressing eGFR. A 2019 study[86] identified a panel of metabolites, excluding creatinine (and thereby the race-based coefficient), that estimated GFR as effectively as using either creatinine or cystatin C alone (although it was less effective than a combination of creatinine and cystatin C). This algorithm was developed using data only from Black patients and validated using a diverse population data set. In 2021, the Chronic Kidney Disease Epidemiology Collaboration examined the effect of removing race from eGFR calculations while adding cystatin C and creatinine.[73] They found that, when both cystatin C and creatinine were used, the new algorithm was more accurate than the previous race-based eGFR equation using only

creatinine. The new version with creatinine (but not cystatin C) increased the estimates of population-level chronic kidney disease for Black people, with similar or lower estimates among other racial and ethnic populations. Also in 2021, the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) published a study[23] examining the effect of replacing race in eGFR with measures of cystatin C, beta-trace protein, and beta2-microglobulin, with or without creatinine. They discovered that replacing race with this combination of input variables resulted in GFR estimation that is equivalent to that derived from the race-based eGFR algorithm.

On a related topic but addressing a different algorithm, a 2017 study sought to replace race in the KDRI.[89] The authors removed race and added a measure for the apolipoprotein L1 genotype, which is associated with kidney disease in Black patients. They found that replacing race with the genotype marker improved the index's ability to predict graft failure (area under the curve improved from 0.59 to 0.60 at 1-5 years after transplantation). Because the study enrolled only Black patients, it did not provide data on the effect on race differences.

Finally, the landmark study by Obermeyer et al.[5] is addressed above in KQ 1 because it found racial differences in access to a care management program, resulting from unintentional bias in model design that used cost as a proxy for clinical need. After identifying the problem, the model developers sought to mitigate the issue by replacing the previous model with three new algorithms that better predict clinical need: degree of chronic conditions, avoidable costs, and total costs. None of the algorithms used input variables based on race and ethnicity. The new algorithms significantly increased access to disease management resources for Black patients.

3.3.3.3 Adding Input Variables

Eight studies added input variables to mitigate or avoid potential bias resulting from clinical algorithms. Two of these studies added race and ethnicity to the Framingham Risk Score (FRS) equations to address concerns about underestimating cardiovascular risk in Black patients. In one study,[87] the authors found that the atherosclerotic cardiovascular disease equations (ASCVD) algorithm, which added race and other factors to FRS, resulted in a general reduction of differences between Black and White patients on two measures of subclinical vascular disease. The reductions in race differences were generally greater for low-risk patients than high-risk patients. In contrast, a study of patients with hypertension[95] found that adding race did not lead to improvements in cardiovascular risk classification for either Black or White patients. Another study[91] added an input variable labeled "African American ethnicity" to a stroke risk prediction tool that included age, gender, and morbidity. The authors found that the new model was slightly better (1.2 percent closer) at predicting true stroke risk in Black patients, while the algorithm's predictive ability for White patients was unchanged (<0.1 percent closer).

Race is not the only input variable that researchers have added to algorithms to address disparities. A 2021 study added polygenic risk scores to the ASCVD algorithm and evaluated patients in multiple racial and ethnic populations.[81] The authors found integrating polygenic risk resulted in significant net classification improvement for patients who self-reported as White, Black/African American/Black Caribbean/Black African, South Asian, or Hispanic. They suggest that incorporating genetic risk markers into ASCVD could lead to more accurate risk prediction for patients of all racial, ethnic, and ancestral backgrounds. Another study simulated the effect of adding up to 10 different biomarkers (measuring factors such as adiposity, inflammation, glycemic control, and more) to the ASCVD and FRS algorithms.[90] Using data based only on Black patients, the study found that incorporating the biomarkers provided no substantial benefit over the original algorithms for classifying cardiovascular risk.

Our review identified only one randomized controlled trial addressing the KQs. This 2013 multicenter study assessed incorporating genotype variables into warfarin dosing algorithms during the first 4 weeks of anticoagulation therapy.[93] Overall, the study found that algorithms informed by genotype information performed no better than traditional clinical algorithms. However, the genotype-informed algorithms led to poorer management in Black patients than in White patients. Thus, adding genotype variables would actually exacerbate disparities. Conversely, Lindley et al.[106] reported that standard warfarin dosing algorithms may overestimate dosing by 30 percent in patients of African ancestry with a specific allele variant. Incorporating this single-nucleotide variation could improve the safety of warfarin dosing and potentially reduce health disparities.

One study modified a lung cancer screening tool by adding a measure of life-years gained from screening.[74] The authors reported that the revised algorithm reclassified 3.5 million people as eligible for lung cancer screening. Importantly, 22 percent of the newly eligible were Black, and differences between Black and White patients would be greatly reduced by implementing this modified algorithm (from 13 percent difference to 0 percent in preventable lung cancer deaths, and from 16 percent difference to 1 percent in life-years gainable). Slight reductions occurred in the White-Hispanic difference (by 3-4 percent), but no change occurred in the White-Asian difference.

Finally, a study based on Medicare beneficiary data examined the effect of using SDOH to predict healthcare use, costs, and mortality.[85] The authors compared models that used four different sets of input variables: 1) sex and age only; 2) sex, age, and morbidity; 3) sex, age, morbidity, and seven SDOH measures; and 4) only the SDOH measures (which included education, economic status, financial strain, marital status, access to healthcare, rural or urban location, and alcohol abuse). The model that included SDOH in addition to the other input variables performed best at predicting risk of hospitalization and death for both Black and White patients. Moreover, the models without SDOH tended to underestimate risk of hospitalization and overestimate risk of death for Black patients, while overestimating the likelihood of hospitalization and underestimating risk of death for White patients.

3.3.3.4 Changing the Patient Mix Used for Development and Validation

An algorithm's components and construct are substantially affected by the characteristics of the patients used for derivation and validation. When relevant populations are not adequately represented during development, an algorithm may reflect and contribute to racial differences. Our review identified four studies that attempted to mitigate bias by recalibrating algorithms based on a different patient mix than initially used. Two of these studies focused on ASCVD equations for cardiovascular risk. In 2018, a seminal study by Yadlowsky et al.[88] revised the pooled cohort equations using more heterogeneous patient data. The authors reported that the original algorithm overestimated risk for most patients, leading to unnecessary treatment, while the new version was significantly better at predicting risk, especially for Black people. These findings were supported two years later by a study that found the revised algorithm eliminated significant differences in risk assessment and recommendations for statin use between Black and White patients.[84]

We briefly described Park et al.'s algorithm for predicting postpartum depression[78] in the section above on removing race-based input variables. The authors tested two additional mitigation strategies in their study: they incorporated a statistical technique to adjust the algorithm (described below), and they reweighed key population groups to better calibrate their

model. Reweighting the algorithm with diverse patient data proved the most effective of their three strategies, leading to more accurate predictions that were less likely to produce disparities.

Finally, we identified a study that modified a Donor Risk Index for liver transplant.[94] In this case, the initial algorithm had been developed using a diverse population and included Black race among seven input variables predicting risk of graft failure. The authors sought to create an algorithm that would better predict risk specifically in Black patients with a diagnosis of hepatitis C. They revised the algorithm using data drawn solely from that subpopulation and reported that the new strategy resulted in more accurate risk assessment, including reclassification of more than a quarter of patients.

3.3.3.5 Developing Separate Algorithms by Race

Two studies went a step beyond recalibration with representative data and developed different algorithms for Black and White patients. Limdi et al. compared two models for developing warfarin dosing algorithms based on multiple clinical and genetic factors.[92] The authors found that separate algorithms for Black and White patients were better at predicting correct dosing levels than traditional, combined algorithms that adjust for but are not stratified by race. A study of opioid misuse that was also described in KQ 1[108] took a related approach. The authors tested two mitigation strategies: develop separate thresholds for Black and White patients and recalibrate the model for each racial subgroup. They found that both approaches eliminated differences in false-negative predictions between Black and White patients. The first mitigation technique involved creating separate thresholds for each racial and ethnic group. Reducing the threshold from 0.3 to 0.2 for Black patients reduced the FNR to 0.25 (95 percent CI: 0.20 to 0.30) and "in closer approximation to the White subgroup with overlapping confidence intervals."[108] (Data for other racial and ethnic groups were presented only in graphical form, so exact values cannot be determined.) The second technique, which involved recalibration by racial and ethnic group, produced results virtually identical to those of the first: the FNR was 0.24 (95 percent CI: 0.19 to 0.29) among Black patients and 0.21 (95 percent CI: 0.15 to 0.27) among White patients.

3.3.3.6 Refining Statistical and Analytic Techniques

Modifying the technical aspects of algorithms, including statistical methodologies and analytic approaches, composes the final mitigation strategy described in two studies we reviewed. As described above, an algorithm to predict postpartum depression was tested against three mitigation approaches.[78] Removing race improved the algorithm's accuracy and may have reduced the likelihood of contributing to disparities, while recalibrating the model with diverse patient data was even more effective. The third strategy involved adding an adjustment term to the model intended to limit the impact of including a race-based input variable (what the authors termed "Prejudice Remover"). This modification had no significant effect on outcomes.

We also discussed above the work by Yadlowsky et al.[88] to update the ASCVD algorithm with more diverse patient data. The authors also adjusted the statistical methodology used in the equations, employing elastic net regularization to avoid overfitting the data and to address concerns about proportional hazards assumptions. They found that these adjustments improved accuracy but to a lesser degree than recalibration with diverse patient data. Foryciarz et al.[96] also addressed the ASCVD algorithm, adjusting estimation of risk through group calibration and equalized odds. They found that recalibrating by subgroups could increase accuracy for a given

group while increasing disparities between groups, and use of an equalized odds constraint led to poorer calibration for the overall model.

3.4 Contextual Question 1. How widespread is the inclusion of input variables based on race and ethnicity in healthcare algorithms?

The evidence base presented throughout this report offers an insightful but limited view of the landscape of race and ethnicity in healthcare algorithms. The 31 distinct algorithms (and their various iterations) examined in KQ 1 and KQ 2 and the 6 described in CQ4 affect cardiology, nephrology, oncology, hematology, neurology, hepatology, endocrinology, infectious disease, obstetrics, pulmonary medicine, transplant medicine, urology, addiction medicine, surgery, and mental health. They are used in primary care settings, hospital medicine, emergency medicine, and intensive care and address screening, diagnosis, treatment, prognosis, and the use and allocation of healthcare resources. Seventeen of the algorithms include race and ethnicity as an input variable, and five include measures such as SDOH, healthcare costs, or healthcare utilization that may correlate with or serve as proxies for race and ethnicity.

However, this is just the tip of the iceberg, because our review was limited to studies that met specific inclusion criteria (especially related to study design and reported health outcomes). To gain a wider perspective, we briefly examined excluded studies. Of the 278 studies excluded during full-text review, 156 were not included due to study design (these were usually derivation studies without external validation, indicating clinical algorithm development but not necessarily use) or because they did not report outcomes related to access to care, quality of care, or health. Similarly, the 6 algorithms examined in CQ 4 were selected from a final pool of 55 algorithms that were initially identified after reviewing hundreds of potential resources. Although a comprehensive analysis of the excluded studies and examples was beyond the scope of this report, a cursory review revealed that hundreds of them included clinical algorithms that were similar or identical to those that were included in the KQ 1 and 2 results. Also, several studies were conducted in specialties that were not included in the evidence for KQs 1 and 2, such as orthopedics, gastroenterology, and pain medicine. While we did not explore whether any of the excluded studies explicitly included race-based input variables within algorithms, algorithms clearly could affect health and healthcare disparities can be found in every medical specialty, healthcare setting, and patient population.

Our findings are reinforced by websites such as MDCalc, a widely used repository for healthcare algorithms, formulas, and calculators. MDCalc, which does not develop algorithms but aims to make them readily available to clinicians, has more than 700 entries. Despite this scope, as of the writing of this report, only 14 included race and ethnicity as an input variable. It is unknown how many algorithms include input variables that might be proxies for race and ethnicity.

We were able to ascertain the original source behind the development of many but not all the algorithms in our review and found that clinical research teams accounted for at least 12 of the algorithms. These were typically investigators managing clinical trials or large observational studies who then promulgated an algorithm derived from the data they collected. At least nine were developed by medical specialty societies or other organizations tasked with setting healthcare policy, such as the United Network for Organ Sharing and USPSTF. At least five of the algorithms were published by academic researchers using machine learning, artificial intelligence (AI), or other data-mining strategies to develop and validate risk prediction algorithms. Three algorithms were created by health plans using large member datasets, and two algorithms examined in CQ 4 were built by EHR vendors.

3.4 Results, Contextual Question 1

Our review was limited in scope; therefore, the algorithms we examined do not fully represent the larger environment. For example, we evaluated only two EHR algorithms, but our KIs, TEP, and SMEs indicated that there are probably hundreds of clinical algorithms embedded in the systems used by many academic medical centers. Obermeyer et al.[5] demonstrated how there may be unforeseen effects of algorithms that can influence patient care on a broad scale, but little has been published about these algorithms. Meanwhile, larger health systems are increasingly devoting resources to develop homegrown algorithms for managing healthcare delivery, often aimed at reducing readmissions or predicting which patients are at highest risk for sepsis or death. Smaller hospitals, although unlikely to have internal capacity for such efforts, may be likely to use algorithms already embedded by vendors in their EHRs.[116]

We also found that at least 18 of the algorithms we reviewed are or were previously endorsed by medical specialty societies, included in clinical practice guidelines, and/or used by quasi-regulatory agencies such as United Network for Organ Sharing. Such designations are important mechanisms to disseminate clinical algorithms, although it is difficult to evaluate the extent of their use.

Finally, AI might dramatically alter how algorithms are developed and used and how healthcare is delivered in countless ways. A recent journalistic investigation[117] revealed that AI-informed algorithms used by Medicare Advantage plans resulted in widespread denial of care to seniors. While AI and machine learning tools are recognized as a source of significant concern,[118-121] rigorous, real-world research is lacking. We identified only five studies meeting our inclusion criteria that evaluated algorithms that were derived or tested with the use of AI or machine learning tools, and these studies focused on AI tools used during development and validation of algorithms, rather than implementation. Nevertheless, research on AI is growing rapidly, and 36 of the 278 studies we excluded during full-text review (13 percent) presented research related to AI or machine learning in the context of algorithms. CQ2 highlights several efforts to address the challenges of new AI tools.

3.5 Contextual Question 2. What are existing and emerging national or international standards or guidance for how algorithms should be developed, validated, implemented, and updated to avoid introducing bias that could lead to health and healthcare disparities?

The recent evolution of AI as a major source of clinical algorithms has led to the emergence of nascent standards, principles, and frameworks to address growing concerns about AI's ethical, legal, and social impacts.[122-124] Discussions with our KIs and TEP revealed that every relevant sector of healthcare and health technology – from EHR vendors and medical device manufacturers to medical specialty societies and clinical guideline panels, from academic medical centers and community health centers to health plans and employers, from researchers and patient advocacy groups to federal and state agencies – all recognize both the value and inevitability of new standards for healthcare algorithms. In recent years, multiple federal agencies have grappled with the broad challenges of healthcare AI and the specific difficulties posed by algorithms.[125,126] In August 2022, California's Attorney General launched an investigation into potential racial and ethnic biases in healthcare algorithms used across the state.[127] In November 2021, the New York City Department of Health and Mental Hygiene launched a Coalition to End Racism in Clinical Algorithms (CERCA), designed in part "to end race adjustment, monitor the impact on racial health inequities, and engage patients whose care was negatively impacted by it".[128] Simultaneously, health technology companies, and the technology sector more generally, have sought to design self-regulatory strategies to reassure consumers and policymakers that they are acting responsibly.[129,130] Concurrent with these efforts has been the rise of a specialized field of research focused on strategies to identify, mitigate, and prevent harms associated with AI and healthcare algorithms.[49,118,120,121,131,132] Moreover, these trends are not unique to the United States but are highly visible in other countries as well.[133,134]

In Table 17, we summarize 10 policy briefs, white papers, and research articles that provide principles or frameworks that could help guide future development and evaluation of clinical algorithms. All have been published in the past 3 years and primarily reflect expert and consensus opinion. Several of these resources are not specific to healthcare or medicine but provide guidance that readily translates across disciplines. Two documents represent U.S. federal regulatory agencies: the Food and Drug Administration (FDA)[135] and the National Institute of Standards and Technology.[136] Our KIs and TEP repeatedly discussed FDA as a potential home for formal regulation of algorithms, similar to the agency's role with medical devices. We also identified two resources developed by Google[129] and Microsoft.[130] While not focused exclusively on healthcare, both documents include widely applicable recommendations.

Another publication reflected the work of the Algorithmic Justice League,[137] a research and advocacy organization that promotes awareness of AI-fueled bias and designs algorithms to mitigate harms. Finally, we include manuscripts and material prepared by five academic or nonprofit research institutions, most prominently the Algorithmic Bias Playbook[123] published by the Chicago Booth Center for Applied Artificial Intelligence. The Playbook lays out a step-by-step process for organizations to mitigate harmful consequences of biased algorithms.

Several themes, including fairness, transparency, representativeness, and accountability, are major principles cited throughout the guidance we reviewed. Multidisciplinary and diverse teams that include representatives of populations that may be most at risk for the harms caused by algorithmic bias are a key element as well. It is noteworthy that all these resources focus on developing algorithms using AI capabilities, although most of the algorithms we examined in

3.5 Results, Contextual Question 2

KQ 1 and 2 were derived and validated outside that context. Nevertheless, the principles and guidance presented here may be applicable to many types of clinical algorithms.

3.5 Results, Contextual Question 2

Table 17. Guidance, standards, and recommendations

Resources	Stakeholder	Summary of Content
Preventing bias and inequities in AI-enabled health tools.[122] 2022	Academic researchers at Duke University Margolis Center for Health Policy	Authors identified 4 areas of algorithmic bias: 1) Inequitable framing of the healthcare challenge 2) Unrepresentative training data 3) Biased training data 4) Insufficient care with choices in data selection, curation, preparation, and model development They also offer recommendations for key stakeholders: Developers should recognize the potential for harm, follow good machine learning practices, work with diverse teams, and develop an understanding of the problem being solved, the data being used, potential differences across subgroups, and how the algorithm is likely to be used. Purchasers and users should test algorithms in their populations immediately and over time, focusing on patient outcomes. Health systems/payers/other owners of large health datasets should prioritize standardization reduce bias in subjective descriptions, and note where their data may differ across groups. FDA and other agencies should ensure that devices that use AI perform well for all subgroups, require clear, accessible labeling, and build systems to monitor for biased outcomes.
The medical algorithmic audit.[133] 2022	Academic researchers based primarily in United Kingdom	Presents rationale (based on fairness and justice) and describes components of an algorithmic audit tailored to medicine. Expands on work of Raji by emphasizing intended use, intended impact, exploratory error analysis, subgroup testing, and adversarial testing in the context of healthcare.
Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem.[137] 2022	Algorithmic Justice League	Not specific to healthcare, focuses on AI. Presents 6 recommendations for policymakers: 1) Require the owners and operators of AI systems to engage in independent algorithmic audits against clearly defined standards 2) Notify individuals when they are subject to algorithmic decision-making systems 3) Mandate disclosure of key components of audit findings for peer review 4) Consider real-world harm in the audit process, including through standardized harm incident reporting and response mechanisms 5) Directly involve the stakeholders most likely to be harmed by AI systems in the algorithmic audit process 6) Formalize evaluation and, potentially, accreditation of algorithmic auditors.

3.5 Results, Contextual Question 2

Resources	Stakeholder	Summary of Content
Microsoft responsible AI standard, v2: general requirements.[130] 2022	Microsoft	Microsoft's detailed standards for AI algorithms. Shaped around 6 core goals: accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness. Numerous principles relevant to healthcare disparities, including: F2.1) Identify and prioritize demographic groups, including marginalized groups, that may be at risk of being differentially affected by the system based on intended uses and geographic areas where the system will be deployed. Include: 1) groups defined by a single factor and 2) groups defined by a combination of factors. F2.2) Evaluate all data sets to assess inclusiveness of identified demographic groups and collect data to close any gaps. F2.1) Reassess the system design, including the choice of training data, features, objective function, and training algorithm, to pursue the goals of minimizing differences between the rates at which resources and opportunities are allocated to identified demographic groups, paying particular attention to those that exceed the target maximum difference, while recognizing that doing so may appear to affect system performance and it is seldom clear how to make such tradeoffs. F2.1.1) For North America, use Best Practices for Age, Gender Identity, and Ancestry to help identify demographic groups and methods for collecting demographic information. F2.1.2) Work with user researchers to understand variations in demographic groups across intended uses and geographic areas. F2.1.3) Work with domain-specific subject matter experts to understand the facts that impact performance of your system and how they vary across identified demographic groups in this domain. F2.1.4) Work with members of identified demographic groups to understand risks of and impacts associated with differences between the rates at which resources and opportunities are allocated. F3.1) Identify and prioritize demographic groups, including marginalized groups, that may be at risk of being subject to stereotyping, demeaning, or erasing outputs of the system. Include: 1) groups defined by a single factor, and 2) groups defined by a combination of factors. F3.5) Reassess the system design, including the choice of training data, features, objective function, and training algorithm, to pursue the goal of minimizing the potential for stereotyping, demeaning, and erasing the identified demographic groups.
Towards a standard for identifying and managing bias in artificial intelligence.[136] 2022	National Institute of Standards and Technology (NIST)	NIST has developed the groundwork for consensus standards on bias in AI. This report: • describes the stakes and challenge of bias in artificial intelligence and provides examples of how and why it can chip away at public trust • identifies three categories of bias in AI - systemic, statistical, and human - and describes how and where they contribute to harms • describes three broad challenges for mitigating bias - datasets, testing and evaluation, and human factors - and introduces preliminary guidance for addressing them

3.5 Results, Contextual Question 2

Resources	Stakeholder	Summary of Content
Algorithmic Bias Playbook[123] 2021	Academic researchers at the University of Chicago Booth School of Medicine and the University of California Berkley School of Public Health	Describes a 4-step process (with multiple sub-steps) for researchers and institutions investigating any type of algorithm. Focuses heavily on harms associated with label bias. Step 1: Inventory algorithms 1A) Talk to relevant stakeholders about how and when algorithms are used. 1B) Designate a "steward" to maintain and update the inventory. Step 2: Screen for bias 2A) Articulate the ideal target (what the algorithm should be predicting) vs. the actual target (what it is actually predicting). 2B) Analyze and interrogate bias. Step 3: Retrain biased algorithms (or throw them out) 3A) Try retraining the model on a label closer to the ideal target. 3B) Consider alternative options. 3C) Consider suspending or discontinuing use of the algorithm. Step 4: Set up structures to prevent future bias 4A) Implement best practices for organizations working with algorithms.
Good machine learning practice for medical device development: guiding principles.[135] 2021	US Food and Drug Administration, Health Canada, Medicines and Healthcare Products Regulatory Agency	Brief overview of 10 principles for medical device development driven by machine learning but broadly applicable to algorithms. 1) Multidisciplinary expertise is leveraged throughout the total product life cycle. 2) Good software engineering and security practices are implemented. 3) Clinical study participants and data sets are representative of the intended patient population. 4) Training data sets are independent of test sets. 5) Selected reference datasets are based on best available methods. 6) Model design is tailored to the available data and reflects the device's intended use. 7) Focus is placed on the performance of the human-AI team. 8) Testing demonstrates device performance during clinically relevant conditions. 9) Users are provided clear, essential information. 10) Deployed models are monitored for performance, and retraining risks are managed.
Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing.[129] 2020	Authors affiliated with Google	Seminal paper introduces framework for auditing algorithms that were developed with AI. 5 key stages: Scoping, Mapping, Artifact Collection, Testing, Reflection (SMACTR). Examples from healthcare, aerospace, finance.
A governance model for the application of AI in health care.[134] 2020	Academic researchers based in Australia	Presents governance model for healthcare AI. 4 main components: Fairness, Transparency, Trustworthiness, and Accountability. Recommendations include: "… a data governance panel constituted by AI developers that includes patient and target group representatives, clinical experts, and people with relevant AI, ethical, and legal expertise. The panel would review datasets used for training AI to ensure the data is representative and sufficient to inform requisite model outcomes… The panel's remit would also be to review algorithms – noting that data and algorithms go together in developing AI models… Normative standards for the application of AI in healthcare should be developed by governmental bodies and healthcare institutions as part of governance."

3.5 Results, Contextual Question 2

Resources	Stakeholder	Summary of Content
Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms.[124] **2019**	The Brookings Institution	Policy brief looks broadly at algorithms across industries and disciplines. Emphasizes need to focus on context, highlights tradeoffs between fairness and accuracy. Authors propose mitigation strategies and general principles to consider: 1) Nondiscrimination and other civil rights laws should be updated to interpret and redress online disparate impacts. 2) Operators of algorithms must develop a bias impact statement. This should include questions such as: What will the automated decision be? How will potential bias be detected? What are the operator incentives? How are other stakeholders being engaged? Has diversity been considered in the design and execution? 3) Operators of algorithms should regularly audit for bias. 4) Operators of algorithms must rely upon cross-functional work teams and expertise. 5) Increase human involvement in the design and monitoring of algorithms. 6) Congress should implement regulatory sandboxes and safe harbors to curb online biases. 7) Consumers need better algorithmic literacy.

Abbreviations: AI=artificial intelligence; FDA=Food and Drug Administration

3.6 Contextual Question 3. To what extent are patients, providers (e.g., clinicians, hospitals, health systems), payers (e.g., insurers, employers), and policymakers (e.g., healthcare and insurance regulators, State Medicaid directors) aware of the inclusion of input variables based on race and ethnicity in healthcare algorithms?

Patients, providers, payers, and policymakers all have vital roles in addressing the challenges inherent at the intersection of race, healthcare, technology, and society. Recently published research[33] and discussions with our KIs, TEP, and SMEs explored the perspectives of these key stakeholder groups and highlighted several important considerations.

The KIs and TEP were in consensus that patients are generally unaware of healthcare algorithms and how they might lead to racial and ethnic disparities in health and healthcare. People typically view healthcare through the lens of their own experiences as patients (or as family members and friends of patients); their perspectives may be shaped strongly by their interactions with doctors and nurses, hospital and health clinics, pharmacies, and insurers. Low health literacy is also a barrier to understanding health and healthcare. Moreover, our KIs and TEP suggested that many Americans (including BIPOC communities) do not conceptualize race as a social construct or understand the mechanisms and effects of structural racism. Not surprisingly, algorithms, often complex and embedded in EHRs and clinical guidelines, are not on patients' minds. Recent controversies about eGFR and other algorithms may have attracted broad attention, but our KIs and TEP did not believe that this has significantly affected public opinion or patient awareness. These perspectives were reinforced in a recent qualitative study that interviewed patients about race and healthcare algorithms.[138] The authors reported that few participants were aware that race may be included in common algorithms, and patients were almost universally opposed to the concept of using race to modify clinical equations.

Patient perspectives on AI-informed tools in healthcare, construed broadly, have also been studied recently. A survey of 926 people, conducted in 2019 and published in 2022 by researchers at the Yale School of Medicine and Weill Cornell Medical College, found that 55 percent believed that AI will eventually make healthcare somewhat or much better.[139] However, 91 percent of respondents were somewhat or very concerned about AI's potential to result in misdiagnosis, and 71 percent expressed privacy concerns. Additionally, while White and non-White patients did not differ in their overarching opinions regarding AI in healthcare, non-White participants were more likely to be very concerned about potentially negative consequences. Another survey, conducted in December 2022 by the Pew Research Center,[140] attracted attention for reporting that 60 percent of patients were uncomfortable with their healthcare provider relying on AI tools to support care. Thirty-eight percent of respondents believed that AI would lead to better care overall, while 33 percent thought outcomes will worsen, and 27 percent expected no change. Seventy-five percent worried that healthcare will adopt AI too quickly, without fully understanding the risks that patients may face. However, 70 percent of people thought that racially biased treatment is a major or minor problem in healthcare, and 51 percent of these respondents were optimistic that AI can reduce bias (33 percent thought things were unlikely to change, and 15 percent expected AI to lead to more biased care).

Two current healthcare trends may have the potential to further shape patient perspectives on healthcare algorithms, in the view of our KIs and TEP. First, increased emphasis on patient-centered care and shared decision making has begun to expand the types of conversations that patients have with their providers, ideally leading to greater trust and transparency. Second,

scientific advances continue to pave the way for personalized medicine and, along with broad public interest in personal genetic profiles, may lead to patients (and providers) asking more questions about interactions between genetics, ancestry, race, ethnicity, and health. Indeed, some of our KIs reported that BIPOC patients are increasingly seeking information about treatments that might have unique benefits, or harms, for specific populations. Taken together, these conditions may enable patients to better understand the role of algorithms in guiding their care and to initiate important conversations about how algorithms are developed and used.

Compared with patients, providers (e.g., clinicians, hospitals, health systems) have greater familiarity with healthcare algorithms, but their understanding is also limited in significant ways. Our discussions revealed that front-line clinicians routinely use algorithms in much the same way as imaging devices or pharmaceuticals: they learn how and when to incorporate these algorithms or treatments into medical practice (typically during their medical training or, later in their careers, from colleagues or vendors) without needing to know every component or ingredient or understanding in-depth how such items are developed, tested, or manufactured. Clinicians generally defer to the trusted institutions of their field, such as regulatory agencies, specialty societies, and academic medical centers, to vet the safety and assess the utility of the drugs they prescribe and devices they use. Algorithms are largely viewed in a similar manner—as tools that can be used effectively without knowing how their input variables are selected, defined, or adjusted, or what patient populations were used to develop and test their efficacy. Hospitals and health systems increasingly deploy algorithms that are embedded in EHR systems but rarely seek to review algorithmic formulas or review underlying evidence of effectiveness or possible algorithmic bias.

At the same time, many hospitals and health systems adapt existing algorithms or develop and test homegrown algorithms, but this does not necessarily lead to greater scrutiny of potential bias. Healthcare institutions have only recently begun to recognize the potentially harmful role of algorithms. Regulatory or professional guidance on these concerns is only beginning to emerge (as demonstrated in CQ 2). Unfortunately, as with American society broadly, many healthcare professionals view race as a biological concept rather than a social construct, potentially reinforcing common biases.

Our KIs and TEP identified medical education as a vital locus for changing these dynamics. Medical school curricula and graduate medical education can begin to emphasize critical thinking about algorithms (and clinical practice guidelines and EHRs that often embed them in practice). More attention can be given to teaching human genetics. And medical schools and academic medical centers could endeavor to debunk historical stereotypes about race and biology. Efforts already underway by the Association of American Medical Colleges and American Medical Association to address entrenched institutional racism in medicine and medical education are major steps in this direction.[141-143]

Healthcare payers, especially insurers and government-funded healthcare programs, represent a key sector responsible for developing algorithms that often focus on cost reduction and resource allocation. However, they may also tend to lack a sophisticated understanding of how algorithms may contribute to bias and disparities. They often rely on the data they collect to lead them in the right direction, perhaps assuming that patterns linking patient characteristics, healthcare use, and outcomes sufficiently reflect real healthcare needs, without considering the complex social systems that influence access and barriers to care. The recent revelations about potential harms associated with AI-informed algorithms used by Medicare Advantage plans underscores such concerns. Additionally, payers may not be ideal settings for driving change.

3.6 Results, Contextual Question 3

Health insurers must grapple with the decentralized nature of their operations and the challenges of conducting business across each state's varied regulatory system. When we sought input on this report from commercial insurers, we found deeply held concerns about the proprietary nature of their operations and data. We also discovered that state-level entities of large, national insurers are substantially autonomous and may not always coordinate or align innovations.

Finally, our discussions addressed the roles of policymakers in confronting the challenges of healthcare algorithms. As described above in CQ 2, FDA and the National Institute of Standards and Technology have taken on critical roles in leading federal activity. Our KIs and TEP agreed that both agencies, especially FDA given its specific role in the healthcare sector, are well-positioned to address these issues. The Center for Medicare & Medicaid Services and the Office of the National Coordinator for Healthcare Information Technology were also suggested as sources of leadership and innovation. Our experts agreed strongly that the federal government will inevitably need to play a key role in setting standards and guidance to ensure that healthcare algorithms do good without exacerbating disparities. Moreover, we heard that all sectors of healthcare – including algorithm developers, commercial vendors, and end-users – anticipate federal guidance, and would generally prefer a stable regulatory environment to the current state of uncertainty.

3.7 Contextual Question 4. Select a sample of approximately 5-10 healthcare algorithms that have the potential to impact racial and ethnic disparities in access to care, quality of care, or health outcomes and are not included in KQs 1 or 2. For each algorithm, describe the type of algorithm, its purpose (e.g., screening, risk prediction, diagnosis), its developer and intended end-users, affected patient population, clinical condition or process of care, healthcare setting, and information on outcomes, if available.

We identified six algorithms to explore for CQ 4, which are described in detail in Tables 18-24. Selected algorithms encompassed four conditions: heart failure, end-stage renal disease, cardiac surgery, and HIV. Two focused on EHR vendor-developed algorithms (Cerner Corp. and Epic Systems Corp.) and were selected based on topics of critical relevance to healthcare inpatient settings and patients: 30-day hospital readmissions (Cerner)[144] and Pediatric Hospital Admissions and ED visits (Epic Systems).[145] We reviewed each algorithm to understand the extent of racial and ethnic biases that may have been introduced during the problem formulation and variable inclusion justification, algorithm development, and translation, dissemination, and implementation phases. In general, while algorithm developers did provide information on the rationale for the algorithm and the intended use for clinical practice, most developers did not provide an adequate justification for included variables (e.g., literature support, expert panel). Related to the development phase (Table 3, stage 3), all identified studies consistently reported on internal and external validation. However, several elements related to data selection and management and model training and development (Table 3, stage 1 and 2) were inconsistently addressed. For instance, for missing data, some developers simply stated that missing data were omitted and did not describe trends observed in missing data (less desirable for reproducibility and transparency), whereas other developers provided information on the distribution of missing data and how missing data were imputed (more desirable for reproducibility and transparency). In one case (Cerner, 30-day hospital readmission risk prediction model), the developers provided no information for missing data. The impact on race and ethnic biases for this algorithm may be compounded depending on the number of hospitals implementing this algorithm and whether these hospitals perform their own external validation. Furthermore, algorithm developers did not list input variables used and instead referred to categories of variables (e.g., “demographic,” “lab”). This precluded extensive analysis, which was completed for the other described algorithms. We performed limited searches of the literature to assess race and ethnic biases introduced during the algorithm translation, dissemination, and implementation phase. While several calculators were developed to support widespread use of the algorithms (Appendix Table E-1 through E-3), these calculators lack detail to understand how the algorithm may be used in a clinical setting. We found only one instance (unpublished) that described how the algorithm was implemented in a clinical setting (including in an EHR decision support design) (Appendix Figure E-4); the authors of this work did not describe outcomes by race or ethnicity. Lack of published data prevented us from assessing race and ethnic biases that may result from provider interpretation or lack of action (i.e., interaction bias). Overall, vendor-developed algorithms (Cerner or Epic) had the least information available in the published literature, likely due to the work’s proprietary nature, which hampered our assessment of race and ethnic biases at any phase.

3.7 Results, Contextual Question 4

We did not identify any prospective clinical validations for any models, let alone a subgroup analysis by race, which limits our understanding of how these models actually affect care. This leaves a gap related to establishing effectiveness for any group, especially marginalized subgroups. None of these models are FDA-approved and likely do not qualify for regulation under current federal standards. Results are organized into seven sections, including tables:

- Table 18. Overview of Algorithms Included to Address CQ 4
- Table 19. Potential Scale and Reach of Algorithm Impact on Populations
- Table 20. Race and Ethnicity Definitions and Standards
- Table 21. Algorithm Model Performance
- Table 22. Evidence, Evidence Quality, Data Sources, and Study Populations Used for Algorithm Development and Validation
- Table 23. Bias Mitigation Strategies Completed by Algorithm Developers
- Table 24. Approaches and Practices for Implementing, Adapting, or Updating Algorithms as Specified by Algorithm Developers

Table 18 summarizes key information about the algorithms: developer, year, clinical setting, intended user, key outcome, race variables included, whether algorithm developers included definitions for race and ethnicity, and the algorithm's potential impact on racial and ethnic disparities when translated for use in clinical practice. See Appendix E for calculator input variables and results by race.

3.7 Results, Contextual Question 4

Table 18. Overview of algorithms included to address Contextual Question 4

Algorithm (Common Name) Year Developed/ Published	Clinical Specialty Clinical Condition	Setting Intended End-User	Intended Use	Algorithm Output	Included Race Variables Race Defined	Impact or Implementation Studies	Effect of Race on Algorithm Results and Clinical Interpretation
Get with the Guidelines Heart Failure Risk Score[146] (GWTG-HF) 2010	Cardiology Heart failure	In-hospital Clinicians; Hospital administrators	Point-of-care to facilitate patient triage or care approach	Mortality risk	B (y/n); Race undefined	None found	Black patients will receive a lower mortality risk score. When this algorithm is used in clinical practice for triage, Black patients may be less likely to be prioritized for care or to receive additional care or resources than White patients with similar risk factors.
Development and Validation of Prediction Scores for Early Mortality at Transition to Dialysis[147] (Dialysis Mortality Risk) 2018	Nephrology End-stage renal disease	Ambulatory Clinicians; Patients and families, hospital administrators	2 separate prediction scores stratified by low and high eGFR (eGFR <15 and eGFR ≥15 mL/min/1.73 m^2). Goal is to individualize treatment and support shared decision making about whether to select maintenance dialysis therapy or conservative care.	Mortality risk	B, W, A, AI, O (y/n) and H (y/n); Race undefined	None found	The algorithm will yield the highest risk score for patients and providers who select “other race” (e.g., Native Hawaiian, Other Pacific Islander, Alaska Native), meaning that these patients may be deemed the poorest candidates for dialysis.

3.7 Results, Contextual Question 4

Algorithm (Common Name) Year Developed/ Published	Clinical Specialty Clinical Condition	Setting Intended End-User	Intended Use	Algorithm Output	Included Race Variables Race Defined	Impact or Implementation Studies	Effect of Race on Algorithm Results and Clinical Interpretation
Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models[148,149] (STS CABG) 2018	Cardiac surgery Coronary artery bypass graft	In-hospital, Ambulatory Clinicians; Hospital administrators and quality personnel	Benchmark participant outcomes compared with national aggregate data; individual surgeon composite performance measures, clinical evaluation	Risk of (1) operative mortality; (2) stroke; (3) renal failure; (4) prolonged ventilation or reintubation; (5) mediastinitis/ deep sternal wound infection; (6) reoperation for any cardiac reason; (7) major morbidity or Mortality; (8) prolonged postoperative length of stay (PLOS); (9) short PLOS	B, A, H; NA; PI; other, including non-Hispanic White Race defined (online calculator only)*	Implementation study available[150]	When assuming the same risk factors[a]: White patients have the lowest calculated risks across all outcomes. Black patients have the highest calculated risks for mortality, stroke, mortality and morbidity composite score, and longer length of stay. They are calculated to have the lowest probability of a short length of stay. Native Hawaiian /Pacific Islander patients have the highest calculated risks for prolonged ventilation and reoperation. Hispanic / Latino / Spanish ethnicity patients have the highest calculated risks for mediastinitis/ deep sternal wound infection. Absolute risk results are relatively close across all races (e.g., risk of mortality is 1.264% and 1.503% for White and Black patients, respectively). The potential impact on disparities across races is difficult to assess as information on clinical applicability or consequences of benchmarking is not readily available.

3.7 Results, Contextual Question 4

Algorithm (Common Name) Year Developed/ Published	Clinical Specialty Clinical Condition	Setting Intended End-User	Intended Use	Algorithm Output	Included Race Variables Race Defined	Impact or Implementation Studies	Effect of Race on Algorithm Results and Clinical Interpretation
Denver Human Immunodeficiency Virus (HIV) Risk Score for Targeted HIV Screening[151,152] (Denver HIV Risk) 2012	Infectious disease HIV	ED, Outpatient setting Clinicians (nurses, advanced practice providers, physicians)	Instrument to identify patients at risk for HIV infection. Can be used to consider additional interventions or services for higher-risk patients.	Risk for HIV infection	B, W, H, O; Race undefined in study	Presentation on implementation at a national conference[153]	White or Other races will receive a lower HIV risk score than Black patients with similar clinical risk factors. In clinical practice, when this score is used to determine eligibility for resources and/or testing, White or Other race patients may be less likely to be prioritized for care or receive additional care or resources compared with Black patients with similar risk factors.
EHR Sample Cerner: 30-day hospital readmission risk prediction model[144] (Cerner Hospital Readmission Risk) 2013	Hospital medicine Readmission risk	In-hospital Clinicians (nurses, advanced practice providers, physicians)	Instrument to identify all-cause 30-day hospital readmission risk. Intended to be used at admissions and before discharge.	All cause 30-day hospital readmission risk	Unknown	None found	2 predictive models were developed: 1 at admission and 1 before discharge. Authors describe categories of input variables, including "demographics and social characteristics," including race. Authors do not specify if race and ethnicity were included in the final model. Final model input variables are listed at the category level (e.g., demographics, lab). Data exclusions for derivation dataset included inpatient admissions for psychiatry.

3.7 Results, Contextual Question 4

Algorithm (Common Name) Year Developed/ Published	Clinical Specialty Clinical Condition	Setting Intended End-User	Intended Use	Algorithm Output	Included Race Variables Race Defined	Impact or Implementation Studies	Effect of Race on Algorithm Results and Clinical Interpretation
Epic Systems Corporation Pediatric Hospital Admissions and ED Visits[145] (Epic Peds Admission and ED Visit Risk) 2017	Pediatrics Hospital admission and ED visit risk	Ambulatory, ED Clinicians (nurses, advanced practice providers, physicians), care managers	Instrument to identify patients at risk for hospital admission and/or ED visit. Can be used to triage interventions to proactively treat highest-risk patients.	At least 1 hospital admission or ED visit within 6 months of prediction	AA or H	None found	Unable to assess race's effect on algorithm results as algorithm information, such as input variables and weights, was not available. Dataset inclusion criteria limited data to patients with previous healthcare utilization (1 in-person ambulatory encounter and at least 1 additional ambulatory encounter, ED visit, or hospitalization in the 2 years before the prediction).

Abbreviations: AA = African American; AI = American Indian; A = Asian; B = Black; ED = emergency department; eGFR = estimated glomerular filtration rate; EHR = electronic health record H=Hispanic; GWTG-HF = Get with the Guidelines® Heart Failure; HIV = human immunodeficiency virus; NA = Native American; O = Other; PI = Pacific Islander; PLOS = postoperative length of stay; STS CABG = Society of Thoracic Surgeons Coronary Artery Bypass Graft; W = White;
[a]See in-depth discussion on rationale for the inclusion of race completed by Shahian et al. in 2022.[154,155]

3.7 Results, Contextual Question 4

Results in Table 19 assessed the scale of potential impact of the algorithms on patient populations. A rubric was developed that accounted for condition prevalence, whether the algorithm was recommended by a professional clinical society or government organization, whether the algorithm was implemented within an EHR system, publication metrics (e.g., citations, downloads), and whether the algorithm had been implemented in a widely used online point-of-care clinical resource. All algorithms, except for the Dialysis Mortality Risk calculator,[147] were rated as having large potential patient impact. The impact scale for the Dialysis Mortality Risk calculator and the two EHR vendor algorithms was unable to be assessed due to several unknown key elements.

3.7 Results, Contextual Question 4

Table 19. Potential scale and reach of impact on populations

Algorithm & Potential Scale and Reach on Clinical Populations	Condition Prevalence	Professional Society or Government Guidelines	Available in EHR System; Online Point of Care Resource	Publication level Dissemination Measures	Comments
GWTG-HF[146] Large Impact	High[156]	American Heart Association (AHA)	EHR Unknown MDCalc	Altmetric score: 14 - top 25% of all research outputs; High Attention Score (87th percentile)	Guidelines: Association with AHA may result in clinicians ascribing more weight to this algorithm's utility compared with other elements of decision-making inputs (i.e., clinician judgment), especially in settings of high clinical volume and clinician caseload, where cognitive loading and fatigue are more likely to be high and where cognitive biases are more likely to manifest.[157] Clinical online resource: MDCalc warns about including race (described in an eye-catching orange box) in this algorithm but does not list information on external validations. Of note, the version of the algorithm in MDCalc differs from the published version (e.g., ethnicity is omitted in MDCalc).
Dialysis Mortality Risk[147] Unknown impact	High[158,159]	Not addressed	EHR Unknown Online risk calculator developed by authors	11 citations	Clinical online resource: Implemented at DialysisScore.com by authors to facilitate "practical implementation."
STS CABG[148,149] Large impact	Moderate[160]	Society of Thoracic Surgeons (STS); 2021 ACC/AHA/SCAI Coronary Revascul-arization Guideline	EHR Unknown Online risk calculator available from STS; users can select multiple races	Web of Science: Core Collection; Highly Cited Publication	Guidelines: 2021 ACC/AHA/SCAI Coronary Revascularization Guideline, recommendation to use the STS calculator to stratify patient risk. (Class of recommendation: strong; 1 Level of evidence: B-non-randomized). Publication Indicators: Top 1% of the academic field of clinical medicine based on a highly cited threshold for the field and publication year, as of March 2022. Clinical online resource: Implemented at riskcalc.sts.org

3.7 Results, Contextual Question 4

Algorithm & Potential Scale and Reach on Clinical Populations	Condition Prevalence	Professional Society or Government Guidelines	Available in EHR System; Online Point of Care Resource	Publication level Dissemination Measures	Comments
Denver HIV Risk[151,152] Large Impact	High, in select patient populations[161]	Health & Human Services, State of Rhode Island; Denver Prevention Training Center	Epic, University of Colorado Denver, Aurora, CO. Nursing driven algorithm in the ED used to identify people at risk for HIV infection. MDCalc	Altmetric score: 5 - Good Attention Score compared to outputs of the same age (76th percentile); Above-average Attention Score compared with outputs of the same age and source (61st percentile)[151]	(1) Available from Health & Human Services, State of Rhode Island to identify patients who are candidates for early intervention services. Implemented as a pdf/form. Guidance displayed at the top of the form indicates "While this tool was developed in Denver, it is validated for use in any jurisdiction." (2) Available from Denver Prevention Training Center, a program within the Public Health Institute at Denver Health, which resides within the Denver Health and Hospital Authority. This program is supported by a cooperative agreement with CDC. (3) Randomized controlled trial of Denver HIV Tool in 2021.[162] This study was funded by an investigator-initiated grant from the National Institute of Allergy and Infectious Diseases (No. R01AI106057).
Cerner Hospital Readmission Risk[144] Unknown Impact	Varies based on local prevalence		Cerner available; unknown how many Cerner customers have implemented the model; Other online algorithms NA.	73 citations	The potential scale and reach likely large due to market share of this EHR vendor; however, we are unable to determine how many vendor clients use this tool.
Epic Peds Admission and ED Visit Risk[145] Unknown Impact	Varies based on local prevalence		Epic available; unknown how many Epic customers have implemented the model; Other online algorithms NA.	NA	The potential scale and reach likely large due to market share of this EHR vendor; however, we are unable to determine how many vendor clients use this tool.

ThAbbreviations: ACC = American College of Cardiology; AHA = American Heart Association; CABG = coronary artery bypass graft; CDC= Center for Disease Control and Prevention; ED=emergency department; EHR = electronic health record; GWTG-HF = Get With The Guidelines®-Heart Failure; HIV=human immunodeficiency virus; NA = not available; SCAI = Society for Cardiovascular Angiography and Interventions; STS = Society of Thoracic Surgeons

3.7.1 Contextual Question 4a

a. If race and ethnicity is included as an input variable, how is it defined? Are definitions consistent with available standards, guidance, or important considerations identified in CQ 2?

Table 20 below contains information on how each algorithm developer defined race and ethnicity and whether those definitions or categories were consistent with available standards. The last column provides context as to how data were captured at the time of algorithm development and relevant data quality. Except for the Society of Thoracic Surgeons' (STS) algorithm for CABG, algorithm developers did not specify race and ethnicity definitions, nor were race and ethnicity consistent with available standards for race categories. Four algorithms were developed or validated using data from datasets containing data abstracted from multiple EHR systems or from multiple clinical settings.[146-148,152] In three cases, developers did not specify how data were collected or whether participants were directly asked to provide race and ethnicity responses.[146,147,152]

Table 20. Race and ethnicity definitions and standards

Algorithm	Race and Ethnicity Definitions and Consistency With Available Standards	Race and Ethnicity Data Collection and Definitions
GWTG-HF[146]	**Race and ethnicity definitions** not provided	Black (yes/no) was the race variable included in the algorithm. Ethnicity was not included. The authors do not describe how patients were asked about self-identification during data collection. Race category is not consistent with available standards. EHR data were abstracted by trained personnel using a proprietary Patient Management Tool and entered in the GWTG-HF registry by trained abstractors. Standardized data elements and definitions were used. Authors did not specify whether participating hospitals were required to use the same definition for race and ethnicity at the source hospital or how the data abstractor may have transformed race and ethnicity data to comply with any GWTG-HF standard definitions. Authors do not provide information on which version of the instrument was used to collect data from 2005-2007. Other GWTG programs (e.g., stroke) data collection instruments have changed over the years. For instance, the 2010 version of the data collection instrument found the GWTG stroke program included the following race and ethnicity options: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, or unable to be determined; ethnicity: Hispanic, yes, no, or unable to be determined. These definitions were consistent with the 1997 Office of Management and Budget (OMB) Standards for the Classification of Federal Data on Race and Ethnicity definitions (most recent update).

3.7 Results, Contextual Question 4

Algorithm	Race and Ethnicity Definitions and Consistency With Available Standards	Race and Ethnicity Data Collection and Definitions
Dialysis Mortality Risk[147]	**Race and ethnicity definitions** not provided	Black, White, Asian, Native American, and others were included in the algorithm. Hispanic (yes/no) was presented as a separate variable. The authors do not describe how patients were asked about self-identification during data collection. Race categories are not consistent with available standards. Authors described the lack of racial and ethnic diversity in both datasets. The validation dataset (Kaiser Permanente Southern California [KPSC]) included 11% Asian. Asian is a broad category, and the 2009 Subcommittee on Standardized Collection of Race and Ethnicity Data for Healthcare Quality Improvement recommended using more granular definitions when available.
STS CABG[148,149]	Race and ethnicity definitions are provided in the online calculator. Definitions are consistent with available standards.	Race definitions provided by the authors are displayed in the online risk calculator. The definitions use the 2010 Census Redistricting Data (Public Law 94-171) Summary File and are detailed. For example, American Indian / Alaskan Native "refers to a person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment." This category includes people who indicated their race(s) as "American Indian or Alaska Native" or reported their enrolled or principal tribe, such as Navajo, Blackfeet, Inupiat, Yup'ik, or Central American Indian groups or South American Indian groups. Hispanic, Latino, or Spanish Ethnicity "refers to a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race." Race and ethnicities had multiple categories and potential combinations, parameterized into 6 major categories: (1) Black, (2) Asian, (3) Hispanic, (4) Native American, (5) Pacific islander, (6) other, including non-Hispanic White. In developing the new STS risk algorithms, authors included input variables based on "empiric findings" and algorithm intended purpose (i.e., case mix adjustment). Authors stated that "race has an empiric association with outcomes and has the potential to confound the interpretation of a hospital's outcomes, although we do not know the underlying mechanism (e.g., genetic factors, differential effectiveness of certain medications, rates of certain associated diseases such as diabetes and hypertension, and potentially SES for some outcomes such as readmission)." The original publication did not provide supporting references for input variables.
Denver HIV Risk[151,152]	**Race and ethnicity definitions** not provided	Black, White, Other, and Hispanic variables were included in the model and subsequent validation studies. Ethnicity was combined with race, such that a participant cannot select Black and Hispanic. Race categories are not consistent with available standards. Algorithm developers obtained data from all CDC-funded HIV testing sites throughout the United States for external validation.[152] A standardized data-collection instrument was completed by individuals providing testing services. The authors did not provide information on which instrument version was used to collect data, nor did they specify whether participating sites were required to use the same definition for race and ethnicity as the source site. We noted 2 implementations of the algorithm. In the 2018 study by Dunlevey et al., we noted the authors modified the race categories (Other and White race were combined).[153] In a later study by the original algorithm authors, White and Other race were also combined.[162] In both cases, the authors did justify collapsing race variables. Race and ethnicity variables used were not consistent with all available standards for the classification of race and ethnicity.

3.7 Results, Contextual Question 4

Algorithm	Race and Ethnicity Definitions and Consistency With Available Standards	Race and Ethnicity Data Collection and Definitions
Cerner Hospital Readmission Risk[144]	**Race and ethnicity definitions** not provided	Authors describe categories of input variables, including demographics and social characteristics, including race.
Epic Peds Admission and ED Visit Risk[145]	**Race and ethnicity definitions** not provided	African American or Hispanic are included in the model. No additional information is provided. Race categories are not consistent with available standards. Algorithm derivation data were obtained from 3 Epic sites.

Abbreviations: CABG = coronary artery bypass graft; CDC = Centers for Disease Control and Prevention; ED = emergency department; EHR = electronic health record; GWTG-HF = Get With The Guidelines®-Heart Failure; HIV = human immunodeficiency virus; KPSC = Kaiser Permanente Southern California; OMB = Office of Management and Budget; SES = socioeconomic status; STS = Society of Thoracic Surgeons

b. For healthcare algorithms that include other input variables in place of or associated with race and ethnicity, how were these other variables defined? Are these definitions consistent with available standards, guidance, or important considerations as identified in CQ 2? Were racial and ethnic variables considered during initial development or validation?

All non-EHR vendor sample algorithms included race or ethnicity. Inclusion criteria for the Pediatric Hospital Admission and ED Visit risk algorithm[145] limit data to patients with previous healthcare utilization (one in-person ambulatory encounter and at least one additional ambulatory encounter, ED visit, or hospitalization in the 2 years before the prediction). The algorithm includes Medicaid status as an input variable. The Cerner Hospital Readmission Risk algorithm[144] developers describe analysis of prior hospital utilization data but do not elaborate on how these data were used in the derivation process.

c. For each healthcare algorithm, what methods were used for development and validation? What evidence, evidence quality, data sources, and study populations were used for development and validation?

Tables 21 and 22 contain information on algorithm performance metrics and the underlying data used to develop the algorithm. Performance data were mostly consistently reported. Most studied reported on goodness of fit, C-statistics, and predicted versus observed probability plots. Sensitivity and positive predictive value were provided only for the Epic Pediatric Admissions and ED Visit Risk algorithm.[145] All studies performed validations.

3.7 Results, Contextual Question 4

Table 21. Algorithm model performance

Algorithm & Stage	Model Type	C-Statistic or AUC Curve (Discrimination)	Calibration Goodness of Fit; P/O Plots	P/O Plots (Calibration)[a]	Overfitting	Commentary
GWTG-HF[146] DR & VL-I	GEE with exchangeable working correlation matrix	Derivation: c-statistic=0.75 (CI not provided). Internal validation: c-statistic=0.75 (CI not provided).	Derivation: H/L Test: p=0.189, internal validation: H/L Test: p=0.604. Graphically consistent results.	Not documented	Not documented	Dataset: The same dataset was used for both derivation and internal validation. Authors did not complete an external validation. Analysis: Subgroup analysis by stratifying by LV function (preserved (EF >40%) and impaired (EF <40%)) was completed, results were similar: c-index: 0.75 and 0.74, respectively; Hosmer-Lemeshow test, P: 0.888 and P: 0.852, respectively.
Dialysis Mortality Risk[147] DR & VL-I	Cox proportional hazards	c-statistic= 0.71 (95% CI, 0.70–0.72), 0.66 (95% CI, 0.65–0.67) for low and high eGFR, respectively. Graphically consistent results.	Goodness of Fit: p>0.1 and p>0.5, for low and high eGFR.	Graphically consistent results	Bootstrapping; shrinkage factor used to adjust model coefficients	Dataset: The same dataset was used for both derivation and internal validation (Veterans Affairs Medical Centers). Model development: 2 models, stratified by eGFR: low eGFR <15 mL/min/1.73 m^2 and high eGFR >15 mL/min/1.73 m^2. Analysis: Results for subgroup analysis were consistent except for patients 65 years or older: 0.66 (95% CI, 0.65-0.68) and 0.63 (95% CI, 0.61-0.64) for low and high eGFR respectively. Overfitting: Authors estimated a linear shrinkage factor using 100 bootstrap samples and adjusted the risk score by the shrinkage factor.

3.7 Results, Contextual Question 4

Algorithm & Stage	Model Type	C-Statistic or AUC Curve (Discrimination)	Calibration Goodness of Fit; P/O Plots	P/O Plots (Calibration)[a]	Overfitting	Commentary
Dialysis Mortality Risk[147] VL-E	N/A	c-index (low eGFR) = 0.77 (95% CI, 0.74–0.79) and 0.74 (95% CI, 0.71–0.76) among men and women, respectively. c-index (higher eGFR): 0.71 (95% CI, 0.67–0.74) and 0.67 (95% CI, 0.62–0.72) among men and women, respectively.	Not documented. Predicted survival is higher than observed in the high-risk groups.	Predicted survival is higher than observed in the high-risk groups.	N/A	Dataset: A separate dataset was used for external validation (Kaiser Permanente Southern California).
STS CABG[148,149] DR & DV-I	Logistic regression	c-indices for derivation and validation models, respectively: Operative mortality=0.806, 0.804; Stroke=0.721, 0.697; Renal failure= 0.810, 0.826; Prolonged ventilation= 0.773, 0.772; Reoperation= 0.627, 0.621; Composite mortality/morbidity=0.627, 0.738; Prolonged LOS (PLOS) 0.779, 0.777; Short LOS: 0.721, 0.716; DSWI: 0.721, 0.681.	H/L Test not provided. Calibration assessed by graphically comparing predicted vs. Observed. Graphically consistent results, except for DSWI (risk was under-estimated).	Not documented	Bootstrapping	Dataset: Single dataset used for derivation and validation

3.7 Results, Contextual Question 4

Algorithm & Stage	Model Type	C-Statistic or AUC Curve (Discrimination)	Calibration Goodness of Fit; P/O Plots	P/O Plots (Calibration)[a]	Overfitting	Commentary
Denver HIV Risk[151] DR & VL-E	Multi-variable logistic Regression	Derivation: AUC = 0.85, (95% CI: 0.83–0.88). Validation: AUC = 0.75 (95% CI: 0.70–0.78).	H/L Test not provided. Authors provided regression slope. Derivation slope: = 0.95, R2 = 0.94; Validation slope = 1.07, R^2 = 0.98. P/O in the performance is better in validation than in derivation.	Not documented	Unconditional bootstrapping	Overfitting: unconditional bootstrapping approach was used to estimate 95% CI for the regression coefficients of the final model. Subgroup analysis was performed stratifying by geographic region.
Denver HIV Risk[152] VL-E	N/A	AUC= 0.77 (95% CI: 0.77–0.77)[b]	H/L Test not provided. Authors provide regression slope results Slope = 1.09. Graphically consistent results.	Graphically consistent results	N/A	None

3.7 Results, Contextual Question 4

Algorithm & Stage	Model Type	C-Statistic or AUC Curve (Discrimination)	Calibration Goodness of Fit; P/O Plots	P/O Plots (Calibration)[a]	Overfitting	Commentary
Cerner Hospital Readmission Risk[144] DR/ VL-I & VL-E	Logistic regression	Derivation and Internal Validation: c-indices: 0.76 and 0.75. External Validation (after recalibration): c-indices: 0.76 and 0.78.	Derivation and Internal Validation H/L: 36.0 (p<0.001) and 23.5 (p=0.0027), External Validation: 6.1 (p=0.641) and 14.3 (p=0.074).	Not documented	Bootstrapping random sample of 2/3 of derivation dataset	Bootstrapping repeated 500 times; averaged coefficients used for validation Dataset. External validation completed using Cerner HealthFacts data, a de-identified patient database that includes over 480 providers across the U.S. with a majority from the Northeast (44%), having more than 500 beds (27%), and are teaching facilities (63%). Brier Scores, derivation and internal validation: 0.062 (7.6% improvement), 0.063 (6.6% improvement). Brier Scores, external internal validation: 0.061 (8.9% improvement) and 0.060 (9.1% improvement).

3.7 Results, Contextual Question 4

Algorithm & Stage	Model Type	C-Statistic or AUC Curve (Discrimination)	Calibration Goodness of Fit; P/O Plots	P/O Plots (Calibration)[a]	Overfitting	Commentary
Epic Peds Admission and ED Visit Risk[145] DR & VL-E	Logistic regression	C-indices for derivation test sites: 0737 (95% CI: 0.726–0.748), 0.725 (95% CI: 0.710–0.740), 0.783 (95% CI: 0.773–0.793). C-index for validation site: 0.731 (95% CI: 0.728–0.734).	Not documented	Not documented	Not documented	Input variables were selected if they were significant in at least 2 of 3 datasets. 50 cross-validation iterations were performed. The coefficients for the final model were an average of the coefficients from the individual models. Authors indicated that the final model's performance was superior to that of any individual model. Sensitivity and positive predictive value (PPV). There are 3 recommended thresholds: (1) Low risk, scores < 10%; (2) medium risk, scores 10%-20%; high risk, scores ≥20%. For high-risk patients, the PPV ranged from 0.34 to 0.45.

, Abbreviations: AUC = area under the curve; CI = confidence interval; DSWI = deep sternal wound infection; DR = derivation; EF = ejection fraction; eGFR = estimated glomerular filtration rate; GEE = generalized estimating equations; GWTF-HF = Get With The Guidelines® Heart Failure; H/L = Hosmer Lemeshow; LV = left ventricle; N/A = not available; PLOS = prolonged length of stay; P/O = plot predicted versus observed probability plots; PPV = positive predictive value; ROC = Receiver Operating Characteristic curve; STS CABG = Society of Thoracic Surgeons Coronary Artery Bypass Graft;
[a]Graphically consistent results: results are in line with the 45-degree line, indicating that observed is close to expected.
[b] Reproduced from the original source. VL-I = internal validation; VL-E = external validation.

3.7 Results, Contextual Question 4

3.7 Results, Contextual Question 4

Table 22. Evidence, evidence quality, data sources, and study populations used for algorithm development and validation

Algorithm	Candidate Input Variables	Evidence Quality	Population Data Sources	Missing Data	Dataset Split for Derivation and Validation	Commentary
GWTG-HF[146]	Selected *a priori* based on available evidence and clinical relevance	References not provided	Registry, retrospective	Patients with missing data were omitted	Randomly divided into derivation and validation sets, 70% and 30%, respectively	Data source: database registry of electronic health record data from 287 hospitals voluntarily participating in the GWTG-HF. Final study cohort included 198 hospitals (n=39,783) after removing data for missingness and exclusion criteria. Missing data: data missing from 85 hospitals (29.8% of all hospitals in GWTG-HF registry). Authors dropped these from the sample. Differences in characteristics were not assessed. Derivation / Validation Dataset: Randomization increases the risk for overoptimistic results as data are very similar.
Dialysis Mortality Risk[147]	Selected *a priori* based on available evidence	References not provided	Registry, retrospective	Imputed select variables in derivation cohort (mean values). External validation cohort restricted to those without missing data.	Randomly divided derivation and validation sets, 66.3% and 33.3%, respectively. External validation performed with separate dataset.	Input Variables & Quality: Authors selected input variables less likely to be intentionally modified over time. Medications and easily modifiable variables (i.e., hemoglobin, potassium, calcium, phosphorus, intact parathyroid hormone, bicarbonate, ferritin, lipids) were excluded. Data source: Transitions of Care in Chronic Kidney Disease Special Study, part of the United States Renal Data System . Authors used 2 historical cohorts from the TC-CKD dataset: VA cohort, used for development and validation (n=85,505, 10/1/2007 – 3/31/2014) and a Kaiser Permanente Southern California (KPSC) cohort, used for external validation (n=9,700, 1/1/2007 – 9/30/2015). Differences assessed between the derivation / internal validation cohorts and the external validation cohort KPSC. Analysis showed differences between the datasets on several key input variables. Missing data: minimal percentage of missing data were reported. Authors assessed for differences in characteristics between included vs. excluded patients for each cohort.

3.7 Results, Contextual Question 4

Algorithm	Candidate Input Variables	Evidence Quality	Population Data Sources	Missing Data	Dataset Split for Derivation and Validation	Commentary
STS CABG[148,149]	Initial input variable list based on data analysis. Selection for model based, in part, on surgeon feedback.	References not provided	STS Database Registry	Imputed to the most common category of binary or categorical variables and to the median or subgroup-specific median of continuous variables.	Dataset split 53.7%, derivation; 46.3% validation.	Input Variables & Quality: 10-member surgeon working group independently reviewed a list of 187 relevant preprocedure factors. Each person rated their *a priori* assessment of each variable's prognostic potential. Variables selected by at least 4 surgeons were retained. The authors published 2 subsequent articles in 2022 describing a nonsystematic literature review on the association between race and ethnicity on outcomes;[154 155] however, we did not find reporting of this in the initial development. Data source: Registry derived from participating hospitals, from July 1, 2011, to December 31, 2016. Missing data: Covariate data are missing in fewer than 5% of cases in each procedure population.
Denver HIV Risk[151,152]	Epidemiologically known or hypothesized associations between patient characteristics and HIV infection.	References not provided	Derivation: Prospectively collected from the Denver Metro Health Clinic Internal validation: Retrospective, Academic medical center ED External validation: all US CDC-funded HIV testing sites	Derivation: Small amount of missing data. Authors used Markov chain Monte Carlo approach to multiple imputation External validation sample: missing data: 29% of sample	No splitting for derivation and validation External validation (#1) performed with separate dataset (2012)[151] External validation (#2) performed with separate dataset (2015)[152]	Input Variables & Quality: Clinical researcher experience and gestalt. Authors further state that the final algorithm input variables reflect national demographic and risk behavior estimates from CDC. Data source: Derivation sample was from a large, prospectively collected dataset from the Denver Metro Health Clinic, a sexually transmitted disease clinic administered by Denver Public Health. This clinic is 1 of the largest in the region, serving over 10,000 patients annually, with an undiagnosed HIV prevalence of approximately 0.5%. Consecutive patients aged 13 years or older, between January 1, 1996, and December 31, 2008 (n=92,635), were used. HIV infection prevalence was 0.54%. External Validation (#1) sample included observations between January 1, 1998, and June 30, 2010, from the ED at the University of Cincinnati Medical Center (n=22,983). HIV infection prevalence was 0.73%. External validation (#2) sample was collected from all CDC-funded HIV testing sites

3.7 Results, Contextual Question 4

Algorithm	Candidate Input Variables	Evidence Quality	Population Data Sources	Missing Data	Dataset Split for Derivation and Validation	Commentary
						throughout the United States, January 1, 2008, to December 31, 2010. All patients aged 13 years or older were included. Data included a wider range of settings than the derivation and internal validation samples, including EDs, hospitals, outpatient clinics, sexually transmitted diseases and HIV counseling and testing sites, community-based organizations, blood banks, plasma centers, and correctional facilities. Missing data: External validation: authors performed a sensitivity analysis to estimate the effect of missing data on complete-case results. Multiple imputation was not feasible given the limited number of variables in the dataset.
Cerner Hospital Readmission Risk[144]	Selected *a priori* based on available evidence and qualitative interviews with clinicians and care managers	References not provided	Data from Cerner customer site and Cerner HealthFactsdata	Variables with missing data not included	Dataset split: derivation (75%), validation (25%)	Input Variables & Quality: Input variables considered were based on literature reviews and a mixed-method qualitative data. Clinicians and care managers identified readmission risk factors. Data source: derivation and internal validation sample from an existing 8 hospital sites. Retrospective data. External validation completed with large, de-identified large multihospital dataset. Missing data: authors do not specify the extent of missingness. They describe a large amount of missing data for social determinants.
Epic Peds Admission and ED Visit Risk[145]	Primary input variable selection used a LASSO penalized logistic regression. LASSO tuning parameter and 10-fold cross-validation to produce the model.	References not provided	Data from existing Epic customer sites. No additional information provided.	Missing categorical data imputed as "unknown" (training/ derivation). Categorical variables with missing data were excluded from the final model.	Unclear. It appears the dataset was not split between derivation and validation.	Data on patient characteristics for each derivation/ training and validation cohort were not provided.

3.7 Results, Contextual Question 4

Abbreviations: CABG = coronary artery bypass graft; CDC = Centers for Disease Control and Prevention; ED = emergency department; GWTG-HF = Get With The Guidelines®-Heart Failure; HIV = human immunodeficiency virus; KPSC = Kaiser Permanente Southern California; LASSO = least absolute shrinkage and selection operator; STS = Society of Thoracic Surgeons;

d. Are development and validation methods consistent with available standards, guidance, and strategies to mitigate bias and reduce the potential of healthcare algorithms to contribute to health disparities?

Table 23 summarizes the bias mitigation activities completed by algorithm developers. We defined mitigation strategies as evidence of external validation in a population that differed from the derivation and internal validation sample datasets, evidence of subgroup analysis by race and ethnicity.

Table 23. Bias mitigation strategies completed by algorithm developers

Algorithm	External Validation Performed by Authors	Data Sources for Derivation and Validation	Race-Specific Subgroup Analysis	Commentary
GWTG-HF[146]	External validation not performed	Single dataset used for derivation and internal validation	No information provided	The publication did not explicitly describe bias mitigation strategies.
Dialysis Mortality Risk[147]	External validation performed	Single dataset used for derivation and internal validation Separate dataset used for external validation	White vs. Other	Authors examined the operating characteristics by White and other race. Though not ideal, consideration for model fairness and bias was evaluated.
STS CABG[148,149]	External validation not performed	Single dataset used for derivation and validation	No information provided	The publication did not explicitly describe bias mitigation strategies.
Denver HIV Risk[151,152]	External validation performed	Separate datasets used for derivation and 2 rounds of external validation	No information provided	The publication did not explicitly describe bias mitigation strategies.
Cerner Hospital Readmission Risk[144]	External validation performed	Separate datasets used for derivation/internal validation and external validation	No information provided	Demographics not provided for external validation cohort.
Epic Peds Admission and ED Visit Risk[145]	External validation performed	Separate dataset used for derivation and external validation	No information provided	Little information provided on differences between the datasets used for derivation/training and validation.

Abbreviations: CABG = coronary artery bypass graft; ED = emergency department; GWTG-HF = Get With The Guidelines®-Heart Failure; HIV = human immunodeficiency virus; STS = Society of Thoracic Surgeons

e. What approaches and practices are there to implement, adapt, or update each healthcare algorithm?

Table 24 below summarizes the approaches and practices developed by algorithm developers for implementing, adapting, or updating algorithms. We performed a limited search for each

algorithm to identify subsequent publications by algorithm developers that contained guidance for general use, adaptation, or updating. We found examples of developer-suggested input variables to include in subsequent implementation or adaptations.[147] Only EHR algorithm authors provided recommended clinical thresholds for use in clinical care.

Table 24. Approaches and practices for implementing, adapting, or updating algorithms as specified by algorithm developers

Algorithm	Implement/Adapt/ Update	Comments
GWTG-HF[146]	Recommendations for clinical thresholds not provided	Limited searches did not identify implementation or impact studies.
Dialysis Mortality Risk[147]	Recommendations for clinical thresholds not provided	Limited searches did not identify implementation or impact studies. Authors developed an online risk score calculator (http://www.dialysisscore.com/) to facilitate "practical implementation" that provides the predicted risk of mortality at months 3, 6, 9, and 12 after dialysis initiation. Authors noted that implementation in an EHR is facilitated, given that ICD-9-CM diagnostic codes were used for model development. They cautioned that the resultant risk score and use in clinical practice may depend on timing; selected input variables may change over time, which may change the risk score. For example, the algorithm uses the last eGFR measurement before dialysis initiation. Timing for this measurement varied within and across cohorts, and the algorithm had better performance results in patients who had a shorter lag period. Authors suggested the addition of other input variables to improve algorithm performance, including socioeconomic status, medication adherence, and timing of referral to nephrologists. In particular, the authors recommended adding physical and cognitive variables to improve performance for older adults, in whom the risk score showed C-statistics of less than 0.7. Authors provided literature citations to support these additions. Given the lack of racial and ethnic diversity in the population (the KPSC cohort included 11% Asian), the authors recommended further validation in other populations. Authors caution that study results should be interpreted with the understanding that algorithm performance may depend on the data's accuracy on comorbid conditions. ICD-9-CM codes were used in this study; however, authors were not able to confirm their accuracy. External validation was performed, which may mitigate this risk.
STS CABG[148,149]	Recommendations for clinical thresholds not provided	Our limited review identified publications that provided guidance for use of the STS algorithms in practice; however, none was published by the original algorithm authors. Huckaby et al. describe the development of a surgeon committee for evaluating high-risk patients.[150] Risk thresholds were provided; postoperative mortality outcome was assessed. The Jin et al. publication was intended for statisticians and less so for clinical users for results interpretation.[163] Maiga et al. was intended for thoracic surgeon users and was a general primer to critically appraise risk prediction algorithms and interpret results for clinical practice.[164]

3.7 Results, Contextual Question 4

Algorithm	Implement/Adapt/ Update	Comments
Denver HIV Risk[151,152]	Recommendations for clinical thresholds not provided	Limited searches did not identify implementation or impact studies. Rosenberg et al. presented a commentary of this model and conclude that the model's predictive utility, as evidenced by the key performance metrics, "strongly endorse [CDC] guidelines for general HIV screening with highly sensitive tests".[165] However, the authors also have concerns about the population selection (i.e., patients who presented to a dedicated sexually transmitted infection clinic) and potential bias. They argue that this model may not be applicable to the general population and conclude that the data source presents a significant limitation on the generalizability of this risk score. These risks would be mitigated with additional external validation studies in a population in which patients have a broader risk distribution, but our review did not find any such studies.
Cerner Hospital Readmission Risk[144]	Recommendations for clinical thresholds provided	Probability threshold for high-risk patients: 11% (based on statistical analysis)
Epic Peds Admission and ED Visit Risk[145]	Recommendations for clinical thresholds provided	Limited searches did not identify implementation or impact studies. Authors recommend 3 thresholds: (1) Low-risk, scores <10%; (2) medium-risk, scores 10%-20%; high-risk, scores ≥20%. Authors estimate that 5% of patients were considered high-risk, 10% were medium-risk, and the remaining patients were low-risk. Authors provided no information on patient characteristics for each derivation/training and validation cohort.

Abbreviations: CABG = coronary artery bypass graft; CDC = Centers for Disease Control and Prevention; ED = emergency department; eGFR = estimated glomerular filtration rate; EHR = electronic health record; GWTG-HF = Get With The Guidelines®-Heart Failure; HIV = human immunodeficiency virus; ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification; KPSC = Kaiser Permanente Southern California; STS = Society of Thoracic Surgeons

4. Discussion

We have examined the evidence on (1) whether and how algorithms (including algorithm-informed decision tools) exacerbate, perpetuate, reduce, or have no effect on racial and ethnic disparities in health outcomes and healthcare access and quality, and (2) strategies that mitigate racial and ethnic bias in the development and use of algorithms. We also explored contextual concerns, including the roles of algorithm developers and end-users; available or emerging guidance on preventing bias during development; stakeholder awareness of and perspectives on potentially biased algorithms; incentives and barriers affecting use and evaluation of algorithms; and algorithms currently in development or in use but not yet sufficiently studied to assess their effects on racial and ethnic health and healthcare disparities. This project was originally envisioned to include only "race-based" algorithms (i.e., those that explicitly use race as an input) and whether they affect race differences. During topic refinement, it became clear that non-race-based algorithms could also contribute to race-based differences, and we expanded the scope accordingly. This expansion proved important as we identified both race-based and non-race-based algorithms that may affect racial and ethnic disparities. Fifty-eight studies (all but 1 nonrandomized) evaluating 31 algorithms addressed the Key Questions (KQs). Also, a sample of six algorithms not assessed in KQ 1 or 2 were examined to enable deeper understanding of how algorithms are designed and implemented and how they can become potential mechanisms affecting race-based differences.

4.1 Summary of Findings

Most of the evidence was published since 2018, and new research continues to emerge rapidly. The algorithms we examined for KQ 1 and 2 address a wide range of clinical assessments, including measurement of kidney and lung function; risk prediction for cardiovascular disease, heart failure, lung cancer, prostate cancer, opioid misuse, postpartum depression, and stroke; suitability for kidney, liver, and lung transplant; evaluation of critical illness severity and severity of metabolic syndrome; assessment of need for high-risk care management; and guidance on warfarin dosing. Studies most often employed a modeling approach, using data from patients whose care was not actually managed with the algorithm, but for whom data were available on the relevant algorithmic variables. These studies usually hypothetically assessed what the predictions, clinical recommendations, and outcomes would have been if the algorithm had been used in their care. Studies typically identified and selected patients from electronic health records (EHRs), clinical trials, or national databases (e.g., transplant registries). Eight studies were rated as Low risk of overall bias, 41 were rated as Moderate, and 9 were High risk of bias.

Twenty-five studies included patients who self-reported race and ethnicity, and the remaining studies did not specify how race and ethnicity were defined. While self-identification of race and ethnicity are generally preferred to other approaches, research suggests reliability of self-identification varies by race and ethnicity. Additionally, the accuracy and completeness of patient characteristics in EHR data may vary based on local data collection methods. Therefore, the validity of outcome data assessed by race and ethnicity depended on the quality of patient self-identification and data collection processes.

The evidence for both KQs was complex. For KQ 1, we reviewed 17 studies examining the effect of 18 algorithms on racial and ethnic disparities. Results varied between studies and across outcomes within studies. In some cases, algorithms were shown to have the potential to

exacerbate disparities, while other times they may have reduced disparities, and frequently they yielded no identified effect on disparities. Four of the 18 algorithms included race or ethnicity as an input variable, and these studies reported reductions in racial and ethnic disparities. In some algorithms (e.g., revised Kidney Allocation System), race and ethnicity input variables were included specifically to address existing racial and ethnic disparities. Therefore, we suggest that when race is included as the best available proxy variable for systemic racism, an algorithm might potentially reduce a disparity, particularly when reallocating resources to previously under-resourced groups (e.g., organ allocation). However, insofar as race and ethnicity is a poor proxy for genetic predisposition, including racial and ethnic variables may introduce or exacerbate disparities in care quality and outcomes, as well as perpetuating the false notion of race and ethnicity as biological concepts. Of the 14 algorithms that did not include race or ethnicity, studies found that 13 may have potentially contributed to racial and ethnic disparities. Therefore, all types of algorithms might contribute to racial and ethnic disparities, especially when they reflect effects of systemic racism.

For KQ 2, we identified six broad categories of mitigation strategies:

- Removing input variables
- Replacing input or outcome variables
- Adding input variables
- Changing the patient mix used for development and validation
- Developing separate algorithms by race
- Refining statistical and analytic techniques

Most of the 44 studies we reviewed demonstrated that mitigation approaches resulted in better algorithmic calibration and might reduce racial and ethnic disparities and improve patient outcomes. No single strategy consistently performed better or worse than another, and we were unable to discern any specific aspects of these approaches that might be associated with greater success in reducing potential bias. It is likely that the effect of any mitigation approach depends on the combination of algorithm, clinical condition, population, setting, and outcomes evaluated.

Changes to an algorithm can also affect a wide array of clinical decisions and may lead to unforeseen or less desirable results. For example, modifying estimated glomerular filtration rate (eGFR) could reduce disparities in critical aspects of renal care (e.g., eligibility for kidney transplant or earlier referral to nephrology care), while also reducing access to various medications or eligibility for clinical trials. Interpreting these effects is fraught with difficulty. Less access to a drug or exclusion from a trial may represent harmful consequences for patients who might benefit; alternatively, these restrictions may be safer for patients whose kidney function might be insufficient to support such interventions. Ultimately, anticipating and interpreting the downstream effects of algorithm modifications can be enormously complex with social as well as medical implications. Research that both models and observes the real-world effects of such decisions are necessary to assess changes to existing algorithms.

Contextual Question (CQ) 1 enabled us to delve deeper into the landscape of healthcare algorithms. Although only 31 algorithms were evaluated in studies that met our inclusion criteria, a broader scan reveals that thousands of algorithms exist and can affect every medical specialty, care setting, and population. Algorithms that include race or ethnicity as an input variable appear to represent a relatively small set of the entire environment; however, numerous other variables may serve as proxies for race and ethnicity, with unknown effects on disparities.

We also observed that algorithms are developed and implemented by a wide range of stakeholders. Significantly, the types of entities that potentially affect patient care on the largest scale – EHR vendors, payers, and integrated delivery systems – often provide the least transparency into how they develop, validate, assess, and deploy their algorithms. Finally, we noted that most of the current research literature examined algorithms developed through traditional statistical or analytic methods, while artificial intelligence (AI) capabilities promise to rapidly and dramatically change the algorithmic landscape. Evidence on the effects of such changes is still emerging. Meanwhile, patient care remains heavily influenced by longstanding algorithms that are not dependent on AI tools. Therefore, efforts to address disparities should aim to balance considerations of past as well as future concerns.

CQ 2 further emphasized this challenge. Warnings about potential harms associated with AI have become ubiquitous, and Congress recently initiated a series of hearings to explore regulatory action. Numerous federal agencies have been examining AI tools and algorithmic bias for several years, and state and local governments have also become active participants seeking to address these issues. Academic researchers have invested significant effort over the past decade to develop tools for evaluating algorithms, and leading advocacy organizations have worked to ensure that equity and fairness are cornerstones of this landscape. Global technology companies have sought to allay concerns about AI by promulgating ethical principles to govern their activities. In this report, we highlighted ten frameworks that represent recent efforts to promote an algorithmic environment that is beneficial and minimizes harm. Core concepts that are widely shared across these frameworks include fairness, transparency, accountability, privacy, representative data, and diverse teams that explicitly incorporate affected populations throughout algorithm development, evaluation, and implementation.

CQ 3 further explored the perspectives of patients, providers, payers, and policymakers. Not surprisingly, most patients are not well informed about how algorithms already affect their healthcare, and they generally have limited understanding of genetics and race. The public also has expressed concerns about how AI tools might influence healthcare in the near future, specifically identifying misdiagnosis and loss of privacy as major considerations. Providers typically use a wide array of algorithms without knowing how such tools were developed, tested, or evaluated. Similarly, hospitals and health systems routinely deploy algorithms through EHRs but rarely examine potential sources of bias that might result in negative consequences for their patients. Payers frequently focus on developing algorithms to manage patients more efficiently or reduce healthcare costs, and may have few incentives to consider harms associated with racial and ethnic disparities. Payers and EHR vendors may also tend to assume that their own data sources are inherently neutral, without recognizing the complex factors that embed biases in their algorithms. However, they are usually resistant to share their models with external experts, citing proprietary concerns. Policymakers have begun to confront many of these issues in recent years, but it is unclear whether the myriad efforts now underway can be aligned.

Finally, the conceptual model depicted in Figure 2 describes sources of bias that may be introduced during algorithm development and translation for dissemination and implementation into clinical practice. Findings from CQ 4 elucidated the dimensions identified in Figure 2 for a select sample of algorithms: how algorithm developers selected data, conducted model training and validation, and the types of performance evaluation tests conducted. Details critical for understanding an algorithm's validity were notably missing from many in the sample (e.g., a detailed accounting for how missing data were treated, justification for variable selection and rationale for race inclusion, rationale for not conducting external validation). Our analysis of CQ

4 found little published literature on implementing algorithms in the sample, the least being for EHR-based algorithms. Customers, users, and those affected by these algorithms may benefit from greater transparency, although this may be difficult to achieve without regulation. Overall, the findings from our in-depth analysis of these algorithms show that standards for reporting and subgroup assessments on factors that are key to uncovering potential impacts on health equity and disparities, before algorithms are used on patients, are severely needed.

4.2 Applicability

We included algorithms for many different clinical settings, and results in one setting do not necessarily apply to those in other settings. The lack of generalizability of algorithms' effects on racial and ethnic disparities necessitates critical assessment of any real-world implementation across various demographic groups and settings. Furthermore, attempts to mitigate disparities caused by algorithms are also highly context-specific.

Only 7 studies[58,61,62,92,93,103,106] actually managed patients with an algorithm or reported real outcomes experienced by patients. The other 51 studies used modeling or simulation, whereby the authors estimated the hypothetical influence of an algorithm or mitigation strategy. Applicability of such studies depends heavily on assumptions made, representativeness of the data sources analyzed, and whether the algorithm would actually be used in the manner hypothesized. Modeling studies may provide, however, the basis for future hypothesis-driven clinical research on the effects of algorithms on racial and ethnic differences.

Included studies frequently used national data (e.g., National Health Interview Survey [NHIS], Cancer Intervention and Surveillance Modeling Network [CISNET], U.S. Census, United Network for Organ Sharing waitlist, electronic intensive care unit [eICU] Collaborative Research Database, Medical Information Mart for Intensive Care III [MIMIC-III]) rather than data from a few local hospitals, potentially increasing applicability of findings to broad populations. National datasets provide a more representative distribution of races, algorithms typically used widely available input variables (and thresholds), and the levels of access to care and health outcomes more accurately reflect the United States as a whole. Although national datasets may provide a more representative distribution of populations, racial and ethnic biases may be introduced in the data selection and management process (e.g., misclassification of race and ethnicity, collapsing race variables into one category) and perpetuated or exacerbated in subsequent steps of the algorithm development and implementation process.

Some studies used an overly broad race categorization, such as White and Non-White. While their results may be useful for academic purposes, they may only motivate further investigation. Virtually all "Non-White" people self-identify using more specific race designation(s), and virtually all EHRs and scoring systems use more specific designations. Other studies focused only on two races (e.g., Black and White); their results may be less relevant to other races and are vague regarding applicability related to ethnicity.

Algorithms are often modified or adapted by end-users in ways that the developer may not have intended or anticipated. For example, an algorithm could be applied to a different patient population (e.g., symptomatic vs. asymptomatic or high-risk vs. low-risk), used to support a different kind of decision (e.g., initiating vs. terminating therapy, screening vs. diagnosis), or the components could be altered, which occurred in recent years when some clinicians began manually calculating glomerular filtration rate (GFR) without the race coefficient before such a change was officially endorsed or implemented in EHRs. Such departures from intended practice are not uncommon, and academic medical centers frequently adapt EHR algorithms to fit their

own needs. These types of "off-label" uses may further complicate interpretation of the potential risks and benefits of using algorithms.

4.3 Evidence Gaps

The evidence gaps are many and varied. We found only one randomized controlled trial, but perhaps more importantly, 51 of the 58 studies we reviewed did not examine the actual use of algorithms to manage patients; rather, algorithms were applied theoretically to patient datasets to evaluate how they would have performed if implemented (e.g., modeling simulations). Such studies are useful and instructive, but research evaluating algorithms as they are actually used during patient management are needed.

Many patient populations are poorly represented in the evidence we reviewed. Of the racial and ethnic groups that were reported, there was not often transparent information to how data were collected (e.g., self-reported or not). Most studies compared Black patients with White patients without examining or reporting on the effects of algorithms on other groups. Moreover, patients are often combined inappropriately into "Other" categories that mask variations between groups and can lead to skewed analyses. We also note that only one of the studies examined by KQ 1 or 2 included children. Some of our excluded studies addressed algorithms in pediatric care, but the paucity of literature highlights the need for additional research.

Critically, most of the algorithms we assessed did not include race and ethnicity as an input variable. For the studies that did include race and ethnicity, there was not often a transparent rationale regarding their inclusion, which limits deeper understanding of the algorithm and its effects—for example, was the inclusion of race or ethnicity in the algorithm perpetuating an incorrect notion that race is a biologic construct, or was it used purposefully as a proxy for the effects of systemic racism to improve disparities? Studies that focus on existing race-based algorithms are needed to further our understanding of how they might affect differences in health outcomes. At the same time, potential proxy variables for race and ethnicity, such as social determinants of health, have been studied even less frequently in the context of healthcare algorithms.

Another major gap is the lack of published studies assessing algorithms developed by EHR vendors, payers, and large health systems. Evaluation of these algorithms is critical, and the varying degree of rigor observed for algorithm development highlights the urgent need for development and reporting standards. As shown in CQ 4, many algorithms in use are not necessarily externally validated and there is no transparent warning for use in clinical care. Similarly, we found few studies of algorithms developed or implemented through AI processes. The next few years promises an explosion of AI-based algorithms, with unknown consequences, and we anticipate the research literature will soon reflect this growth.

The literature on mitigation strategies is also deeply lacking. Many of the studies we reviewed did not directly evaluate a mitigation approach's effect on disparities in health or healthcare but rather presented proximal outcomes, such as within-group improvements to algorithmic accuracy, leaving us to infer or extrapolate effects on differences between racial and ethnic groups. Modification of existing algorithms can also have numerous downstream effects for patients that may extend beyond the initial intent of a mitigation strategy; research on such consequences is also lacking.

4.4 Strengths and Limitations

This evidence review attempted to cover an enormous amount of ground, which is likely a strength and a weakness. We have presented the results from a comprehensive review of the use of algorithms and efforts to mitigate their potential contribution to racial and ethnic disparities. We also discussed a wide range of relevant issues with a broad pool of experts represented by our Key Informants and Technical Expert Panel. Finally, we conducted a deep dive into six algorithms that we determined have not been well studied previously to gain a better understanding of their dynamics. Our multipronged and multidisciplinary approach has hopefully enabled us to synthesize a broad array of evidence and perspectives. Our literature searches yielded over 11,000 potential studies published in just the past 12 years, emphasizing the scope of material we reviewed as well as the challenges of addressing such a complex topic.

To narrow the scope, we developed study eligibility criteria that were intentionally highly restrictive. By excluding studies that did not report health outcomes or focus on disparities, we excluded evidence on many algorithms targeting a wide variety of clinical conditions and patient populations. The list of algorithms we assessed should be viewed as representative rather than comprehensive. We discovered that taking a complete census of algorithms that might contribute to, or help redress, disparities is not feasible given the published literature's sheer volume (and exponential growth), in addition to the unknown number of algorithms developed by homegrown teams in academic medical centers or deployed in proprietary programs operated by EHRs, health insurers, chronic disease management companies, benefit managers, and more.

A major limitation we encountered is a lack of well-developed and tested tools to assess the quality of the studies we reviewed. The ROBINS-I (Risk Of Bias In Non-randomized Studies of Interventions) tool we used is a high-quality instrument, but its fit for this evidence base was often awkward. While seven studies examined the actual use of an algorithm or mitigation strategy, most studies employed a modeling approach (i.e., examined the estimated result of an algorithm or strategy). The type of data used for analysis, the methods used to control for missing data, the potential use of ancillary management strategies, and other factors may differ between modeling studies and other types of study designs, making it more difficult to assess and compare study quality. PROBAST (Prediction model Risk Of Bias Assessment Tool) is a tool gaining increasing use for evaluating artificial intelligence and machine learning models; a pilot application of this tool indicated that the study designs were not fully compatible with this tool either because studies were examining the algorithm's effect, not the original algorithm development. In collaboration with others outside this project team, we piloted some equity-focused items as part of our risk-of-bias assessment and hope this is an early step toward developing useful tools in the future. Nevertheless, our limited evaluation of risk of bias within this review should be recognized as an important caveat.

Another important limitation is the focus on racial and ethnic bias and disparities. While constructs of race have always played a central (and usually negative) role in healthcare and medicine, many people are vulnerable to personal and systemic bias for a wide range of other factors, including sex, gender identity, sexual preference, disability, age, and religion, and the intersectionality of these characteristics along with an individual's race can further magnify such biases. Although many of the principles we examined in this review may apply to biases based on other factors, we did not address those considerations.

Finally, we emphasize the challenge of evaluating causality in the context of differences in group-level outcomes. Disparities in health and healthcare are well documented in many clinical contexts for BIPOC (Black, Indigenous, or People of Color) people. However, assessing the

specific role of algorithms in exacerbating, perpetuating, or reducing disparities is very difficult and is not well supported by the current evidence base. This is especially true for health outcomes that are influenced by a wide variety of factors. More research on these issues will be an important and necessary step toward building effective, fair, and just healthcare algorithms.

4.5 Future Directions

In March 2023, the Agency for Healthcare Research and Quality (AHRQ) and the National Institute on Minority Health and Health Disparities convened a two-day meeting to explore the current use of algorithms in healthcare, their impact on racial and ethnic disparities in care, and approaches to identify and mitigate existing biases.[166] The meeting was designed to inform an expert panel of stakeholders developing a set of guiding principles to help recognize the potential for algorithms to contribute to racial and ethnic bias, how to identify and/or prevent biases before implementation, and how to mitigate biases discovered after implementation.[167] The meeting materials can be accessed via AHRQ's Effective Health Care Program website (https://effectivehealthcare.ahrq.gov/news/meetings).

To address the gaps and limitations of the evidence base, key stakeholders should consider the following strategies.

Researchers:

- Develop diverse research teams and engage patients and community members representing historically marginalized racial and ethnic groups throughout the research process.
- Provide sufficient detail in primary studies or evidence syntheses on input variables within an algorithm.
- When addressing race and ethnicity, use inclusive language168 and evidence-based recommended categories.
- Provide a clear rationale when race and ethnicity are included in an algorithm. Failure to explain how and why race and ethnicity are used can reinforce harmful stereotypes and perpetuate false notions about race.
- Research implementing healthcare algorithms to examine real-world impacts on racial and ethnic disparities. This includes evaluating the potential tradeoffs between beneficial and harmful effects of using or modifying algorithms (e.g., expanded eligibility for kidney transplant versus reduced access to drug therapy).
- Report context-specific details (e.g., unique combinations of algorithm, clinical condition, population, setting, and outcomes) about how healthcare algorithms were implemented in primary studies or evidence syntheses.

Health Systems and Healthcare Providers:

- Develop diverse leadership teams (e.g., clinicians, administrators) to evaluate effects of algorithms currently in use to identify possible biases and effects on racial and ethnic populations.
- Establish meaningful and collaborative relationships with local community members and patients representing historically marginalized racial and ethnic groups.
- Assess future algorithms for biases before implementation.

- Educate providers about common forms of algorithmic bias and potential effects on patient outcomes.
- Give providers resources that they can use to engage and educate patients about algorithmic bias in healthcare.

Algorithm Developers:

- Establish diverse development and implementation teams and engage patients and community members representing historically marginalized racial and ethnic groups in the algorithm development to implementation lifecycle.
- Increase transparency of algorithm development and implementation, addressing points, such as:
  - Clear justification for variable inclusion, especially race and ethnicity
  - Representativeness of data sets used for training and validation
  - Approaches for addressing missing data
  - External validation, or justification for using the same dataset for development and validation
- Routinely assess algorithms for effects by race and ethnicity
- Develop and use checklists and reporting standards to improve consistency, transparency, and accountability.
- Warn against use in clinical practice for algorithms that have not been externally validated and have not reported on differential outcomes by race and ethnicity.

Policymakers:

- Support research to assess the effect of healthcare algorithms on racial and ethnic disparities before widespread implementation.
- Provide guidance on best practices to examine algorithms for racial and ethnic biases.
- Develop or promote incentives for external algorithm validation.
- Fund implementation science (e.g., contextual) studies to understand adoption, cost, penetration, appropriateness, and sustainability of algorithms through a health equity lens.

Medical Associations and Specialty Societies:

- Promote stakeholder awareness (including patients) of potential algorithmic risk.
- Work with policymakers to review clinical algorithms, and address those that result in racial and ethnic inequities.
- Ensure that algorithms included in clinical guidelines and recommendations statements are assessed from a health equity lens and that methods are adequately reported.

4.6 Conclusions

Healthcare algorithms have been shown to potentially exacerbate, perpetuate, or reduce racial and ethnic disparities in health outcomes and healthcare access and quality. When race or ethnicity are incorporated into an algorithm to intentionally tackle known racial and ethnic disparities in resource allocation (e.g., kidney transplant allocation) or disparities in unequal care (e.g., prostate cancer screening historically led to Black men receiving more low-yield biopsies),

disparities were reduced. However, when race or ethnicity was included in an algorithm without clear rationale, it was often perpetuating the incorrect notion that race is a biologic construct, and algorithms were shown to have the potential to perpetuate and exacerbate disparities (e.g., estimated glomerular filtration rate for kidney function measurement). Furthermore, some algorithms do not contain race or ethnicity as an input variable but can also perpetuate or reduce disparities. Several modeling studies showed that applying algorithms out of context of original development (e.g., illness severity scores used for crisis standards of care) would perpetuate or exacerbate racial and ethnic disparities. On the other hand, algorithms that standardize care and reduce opportunities for implicit bias (e.g., Lung Allocation Score for lung transplantation) may also reduce disparities. In terms of strategies to mitigate racial and ethnic disparities associated with healthcare algorithms, no clear single strategy led to greatest success, but several have been shown to successfully mitigate disparities.

We emphasize the challenge of inferring causality and determining attribution of a particular healthcare algorithm on racial and ethnic disparities. Results may be highly context-specific, relating to unique combinations of algorithm, clinical condition, population, setting, and outcomes. Important future steps include increasing transparency in algorithm development and implementation, increasing diversity of research and leadership teams, engaging diverse patient and community groups in the development to implementation lifecycle, promoting awareness by stakeholders (including patients) of potential algorithmic risk, and investing in real-world experiments to assess the effect of healthcare algorithms on racial and ethnic disparities before widespread implementation.

References

1. 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(9):874-82. doi: 10.1056/NEJMms2004740. PMID: 32853499.
2. Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. Lancet. 2020;396(10257):1125-8. doi: 10.1016/s0140-6736(20)32076-6. PMID: 33038972.
3. Schmidt IM, Waikar SS. Separate and unequal: race-based algorithms and implications for nephrology. J Am Soc Nephrol. 2021;32(3):529-33. doi: 10.1681/asn.2020081175. PMID: 33510038.
4. Eneanya ND, Yang W, Reese PP. Reconsidering the consequences of using race to estimate kidney function. JAMA. 2019;322(2):113-4. doi: 10.1001/jama.2019.5774. PMID: 31169890.
5. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi: 10.1126/science.aax2342. PMID: 31649194.
6. Amutah C, Greenidge K, Mante A, et al. Misrepresenting race - the role of medical schools in propagating physician bias. N Engl J Med. 2021;384(9):872-8. doi: 10.1056/NEJMms2025768. PMID: 33406326.
7. Fujimura JH, Bolnick DA, Rajagopalan R, et al. Clines without classes: how to make sense of human variation. Sociological Theory. 2014;32(3):208-27. doi: 10.1177%2f0735275114551611.
8. Roberts D. Fatal invention: how science, politics, and big business re-create race in the twenty-first century. New York (NY): The New Press; 2011. https://thenewpress.com/books/fatal-invention
9. Yudell M, Roberts D, DeSalle R, et al. Science and society. Taking race out of human genetics. Science. 2016;351(6273):564-5. doi: 10.1126/science.aac4951. PMID: 26912690.
10. Wright JL, Davis WS, Joseph MM, et al. Eliminating race-based medicine. Pediatrics. 2009;150(1):e2022057998. doi: 10.1542/peds.2022-057998. PMID: 35491483.
11. Smedley BD. The lived experience of race and its health consequences. Am J Public Health. 2012;102(5):933-5. doi: 10.2105/ajph.2011.300643. PMID: 22420805.
12. Williams DR, Lawrence JA, Davis BA. Racism and health: evidence and needed research. Annu Rev Public Health. 2019;40:105-25. doi: 10.1146/annurev-publhealth-040218-043750. PMID: 30601726.
13. Bailey ZD, Krieger N, Agénor M, et al. Structural racism and health inequities in the USA: evidence and interventions. Lancet. 2017;389(10077):1453-63. doi: 10.1016/s0140-6736(17)30569-x. PMID: 28402827.
14. Paradies Y, Ben J, Denson N, et al. Racism as a determinant of health: a systematic review and meta-analysis. PLoS ONE. 2015;10(9):e0138511. doi: 10.1371/journal.pone.0138511. PMID: 26398658.
15. Waldo DR. Accuracy and bias of race/ethnicity codes in the Medicare enrollment database. Health Care Financ Rev. 2004;26(2):61-72. https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/HealthCareFinancingReview/Downloads/HCF_Review_2004_Winter_pg61.pdf. PMID: 25371985.
16. Klinger EV, Carlini SV, Gonzalez I, et al. Accuracy of race, ethnicity, and language preference in an electronic health record. J Gen Intern Med. 2015;30(6):719-23. doi: 10.1007/s11606-014-3102-8. PMID: 25527336.
17. Race, ethnicity, and language data: standardization for health care quality improvement. Washington (DC): Institute of Medicine, The National Academies Press; 2009. doi: 10.17226/12696.

18. Race, ethnicity, and language data: standardization for health care quality improvement. AHRQ publication no. 10-0058-EF. Rockville (MD): Agency for Healthcare Research and Quality; 2010. https://www.ahrq.gov/research/findings/final-reports/iomracereport/index.html. Accessed on August 14, 2022.

19. Kowalsky RH, Rondini AC, Platt SL. The case for removing race from the American Academy of Pediatrics clinical practice guideline for urinary tract infection in infants and young children with fever. JAMA Pediatr. 2020;174(3):229-30. doi: 10.1001/jamapediatrics.2019.5242. PMID: 31930353.

20. Vyas DA, Jones DS, Meadows AR, et al. Challenging the use of race in the vaginal birth after cesarean section calculator. Womens Health Issues. 2019;29(3):201-4. doi: 10.1016/j.whi.2019.04.007. PMID: 31072754.

21. Ahmed S, Nutt CT, Eneanya ND, et al. Examining the potential impact of race multiplier utilization in estimated glomerular filtration rate calculation on African-American care outcomes. J Gen Intern Med. 2021;36(2):464-71. doi: 10.1007/s11606-020-06280-5. PMID: 33063202.

22. Diao JA, Wu GJ, Taylor HA, et al. Clinical implications of removing race from estimates of kidney function. JAMA. 2021;325(2):184-6. doi: 10.1001/jama.2020.22124. PMID: 33263721.

23. Inker LA, Couture SJ, Tighiouart H, et al. A new panel-estimated GFR, including ß(2)-microglobulin and ß-trace protein and not including race, developed in a diverse population. Am J Kidney Dis. 2021;77(5):673-83.e1. doi: 10.1053/j.ajkd.2020.11.005. PMID: 33301877.

24. Norris KC, Eneanya ND, Boulware LE. Removal of race from estimates of kidney function: first, do no harm. JAMA. 2021;325(2):135-7. doi: 10.1001/jama.2020.23373. PMID: 33263722.

25. Levey AS, Titan SM, Powe NR, et al. Kidney disease, race, and GFR estimation. Clin J Am Soc Nephrol. 2020;15(8):1203-12. doi: 10.2215/cjn.12791019. PMID: 32393465.

26. Powe NR. Black kidney function matters: use or misuse of race? JAMA. 2020;324(8):737-8. doi: 10.1001/jama.2020.13378. PMID: 32761164.

27. Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. Am J Kidney Dis. 2022;79(2):268-88.e1. doi: 10.1053/j.ajkd.2021.08.003. PMID: 34563581.

28. Westby A, Okah E, Ricco J. Race-based treatment decisions perpetuate structural racism. Am Fam Physician. 2020;102(3):136-7. https://www.aafp.org/afp/2020/0315/p341.html. PMID: 32735444.

29. Gopal DP, Okoli GN, Rao M. Re-thinking the inclusion of race in British hypertension guidance. J Hum Hypertens. 2022;36(3):333-5. doi: 10.1038/s41371-021-00601-9. PMID: 34508156.

30. Holt HK, Gildengorin G, Karliner L, et al. Differences in hypertension medication prescribing for Black Americans and their association with hypertension outcomes. J Am Board Fam Med. 2022;35(1):26-34. doi: 10.3122/jabfm.2022.01.210276. PMID: 35039409.

31. Williams SK, Ravenell J, Seyedali S, et al. Hypertension treatment in blacks: discussion of the U.S. clinical practice guidelines. Prog Cardiovasc Dis. 2016;59(3):282-8. doi: 10.1016/j.pcad.2016.09.004. PMID: 27693861.

32. AHRQ seeking input to inform new systematic review, use of clinical algorithms that have the potential to introduce racial/ethnic bias into healthcare delivery. Rockville (MD): Effective Health Care Program, Agency for Healthcare Research and Quality; 2021. https://effectivehealthcare.ahrq.gov/news/algorithm-bias-rfi. Accessed on August 14, 2022.

33. Jain A, Brooks JR, Alford CC, et al. Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms. JAMA Health Forum. 2023 Jun;4(6):e231197. doi: 10.1001/jamahealthforum.2023.1197. PMID: 37266959.

34. Braveman P. Health disparities and health equity: concepts and measurement. Annu Rev Public Health. 2006;27:167-94. doi: 10.1146/annurev.publhealth.27.021405.102103. PMID: 16533114.

35. Smedley BD, Stith AY, Nelson AR, et al. Unequal treatment: confronting racial and ethnic disparities in health care. Washington (DC): National Academy of Sciences; 2003. http://www.ncbi.nlm.nih.gov/books/NBK220358/pdf/TOC.pdf

36. Jones CP. Levels of racism: a theoretic framework and a gardener's tale. Am J Public Health. 2000;90(8):1212-5. doi: 10.2105/ajph.90.8.1212. PMID: 10936998.

37. Braveman PA, Arkin E, Proctor D, et al. Systemic and structural racism: definitions, examples, health damages, and approaches to dismantling. Health Aff (Millwood). 2022;41(2):171-8. doi: 10.1377/hlthaff.2021.01394. PMID: 35130057.

38. Evans C, Johnson E, Lin J. Assessing algorithmic bias and fairness in clinical prediction models for preventive services: a health equity methods project for the U.S. Preventive Services Task Force. 2022. Unpublished work.

39. Methods guide for effectiveness and comparative effectiveness reviews. Effective Health Care Program, Agency for Healthcare Research and Quality; 2018. https://effectivehealthcare.ahrq.gov/products/collections/cer-methods-guide. Accessed on August 14, 2022.

40. Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. JAMA. 1990;263(10):1401-5. doi: 10.1001/jama.1990.03440100117017. PMID: 2304219.

41. Neinstein LS. A review of Society for Adolescent Medicine abstracts and Journal of Adolescent Health Care articles. J Adolesc Health Care. 1987;8(2):198-203. doi: 10.1016/0197-0070(87)90265-8. PMID: 3818406.

42. Dundar Y, Dodd S, Williamson P, et al. Case study of the comparison of data from conference abstracts and full-text articles in health technology assessment of rapidly evolving technologies: does it make a difference? Int J Technol Assess Health Care. 2006;22(3):288-94. doi: 10.1017/s0266462306051166. PMID: 16984055.

43. De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. Ann Oncol. 1992;3(3):187-91. doi: 10.1093/oxfordjournals.annonc.a058147. PMID: 1586615.

44. Scherer RW, Meerpohl JJ, Pfeifer N, et al. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev. 2018. doi: 10.1002/14651858.mr000005.pub4. PMID: 30480762.

45. Yentis SM, Campbell FA, Lerman J. Publication of abstracts presented at anaesthesia meetings. Can J Anaesth. 1993;40(7):632-4. doi: 10.1007/bf03009700. PMID: 8403137.

46. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51-8. doi: 10.7326/m18-1376. PMID: 30596875.

47. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919. doi: 10.1136/bmj.i4919. PMID: 27733354.

48. Sittig DF, Singh H. A new socio-technical model for studying health information technology in complex adaptive healthcare systems. In: Patel V, Kannampallil T, Kaufman D, eds. Cognitive Informatics for Biomedicine Health Informatics. Springer International Publishing; 2015:59-80. doi:10.1007/978-3-319-17272-9_4.

49. Rajkomar A, Hardt M, Howell MD, et al. Ensuring fairness in machine learning to advance health equity. Ann Intern Med. 2018;169(12):866-72. doi: 10.7326/m18-1990. PMID: 30508424.

50. Flores EJ, Mull NK, Lavenberg JG, et al. Using a 10-step framework to support the implementation of an evidence-based clinical pathways programme. BMJ Qual Saf. 2019;28(6):476-85. doi: 10.1136/bmjqs-2018-008454. PMID: 30463885.

51. Bartlett AH, Makhni S, Ruokis S, et al. Use of clinical pathways integrated into the electronic health record to address the COVID-19 pandemic. Infect Control Hosp Epidemiol. 2022:Online ahead of print. doi: 10.1017/ice.2022.64. PMID: 35314010.

52. National Healthcare Quality and Disparities Reports. Rockville (MD): Agency for Healthcare Research and Quality; 2013. https://www.ahrq.gov/research/findings/nhqrdr/index.html. Accessed on August 14, 2022.

53. 2021 National Healthcare Quality and Disparities Report [AHRQ Publication 21(22)-0054-EF]. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2021. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr21/index.html

54. Miller WD, Han X, Peek ME, et al. Accuracy of the sequential organ failure assessment score for in-hospital mortality by race and relevance to crisis standards of care. JAMA Netw Open. 2021;4(6):e2113891. doi: 10.1001/jamanetworkopen.2021.13891. PMID: 34143190.

55. Pasquinelli MM, Tammemägi MC, Kovitz KL, et al. Brief report: risk prediction model versus United States Preventive Services Task Force 2020 draft lung cancer screening eligibility criteria-reducing race disparities. JTO Clin Res Rep. 2021;2(3):100137. doi: 10.1016/j.jtocrr.2020.100137. PMID: 34590000.

56. Presti JC, Alexeeff S, Horton B, et al. Prospective validation of the Kaiser Permanente prostate cancer risk calculator in a contemporary, racially diverse, referral population. Urol Oncol. 2021;39(11):783.e11-.e19. doi: 10.1016/j.urolonc.2021.03.023. PMID: 33962850.

57. Sarkar R, Martin C, Mattie H, et al. Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. Lancet Digit Health. 2021;3(4):e241-e9. doi: 10.1016/s2589-7500(21)00022-4. PMID: 33766288.

58. Snavely AC, Hendley N, Stopyra JP, et al. Sex and race differences in safety and effectiveness of the HEART pathway accelerated diagnostic protocol for acute chest pain. Am Heart J. 2021;232:125-36. doi: 10.1016/j.ahj.2020.11.005. PMID: 33160945.

59. Han SS, Chow E, Ten Haaf K, et al. Disparities of National Lung Cancer Screening Guidelines in the US population. J Natl Cancer Inst. 2020;112(11):1136-42. doi: 10.1093/jnci/djaa013. PMID: 32040195.

60. Carbunaru S, Nettey OS, Gogana P, et al. A comparative effectiveness analysis of the PBCG vs. PCPT risks calculators in a multi-ethnic cohort. BMC Urol. 2019;19(1):121. doi: 10.1186/s12894-019-0553-6. PMID: 31771578.

61. Zhang X, Melanson TA, Plantinga LC, et al. Racial/ethnic disparities in waitlisting for deceased donor kidney transplantation 1 year after implementation of the new national kidney allocation system. Am J Transplant. 2018;18(8):1936-46. doi: 10.1111/ajt.14748. PMID: 29603644.

62. Wille KM, Harrington KF, Deandrade JA, et al. Disparities in lung transplantation before and after introduction of the lung allocation score. J Heart Lung Transplant. 2013;32(7):684-92. doi: 10.1016/j.healun.2013.03.005. PMID: 23582477.

63. Boley S, Sidebottom A, Vacquier M, et al. Investigating racial disparities within an emergency department rapid-triage system. Am J Emerg Med. 2022 Jul;60:65-72. doi: 10.1016/j.ajem.2022.07.030. PMID: 35907271.

64. Metzger P, Allum L, Sullivan E, et al. Racial and language disparities in pediatric emergency department triage. Pediatr Emerg Care. 2022 Feb;38(2):e556-e62. doi: 10.1097/PEC.0000000000002439. PMID: 34009885.

65. Riviello ED, Dechen T, O'Donoghue AL, et al. Assessment of a crisis standards of care scoring system for resource prioritization and estimated excess mortality by race, ethnicity, and socially vulnerable area during a regional surge in COVID-19. JAMA Netw Open. 2022;5(3):e221744. doi: 10.1001/jamanetworkopen.2022.1744. PMID: 35289860.

66. Yoo RM, Dash D, Lu JH, et al. Investigating real-world consequences of biases in commonly used clinical calculators. Am J Manag Care. 2023 Jan;29(1):e1-e7. doi: 10.37765/ajmc.2023.89306. PMID: 36716157.

67. Williams RM, Li T, Luta G, et al. Lung cancer screening use and implications of varying eligibility criteria by race and ethnicity: 2019 behavioral risk factor surveillance system data. Cancer. 2022 May;128(9):1812-9. doi: 10.1002/cncr.34098. PMID: 35201610.

68. Baugh AD, Shiboski S, Hansel NN, et al. Reconsidering the utility of race-specific lung function prediction equations. Am J Respir Crit Care Med. 2022;205(7):819-29. doi: 10.1164/rccm.202105-1246OC. PMID: 34913855.

69. Casal MA, Ivy SP, Beumer JH, et al. Effect of removing race from glomerular filtration rate-estimating equations on anticancer drug dosing and eligibility: a retrospective analysis of National Cancer Institute phase 1 clinical trial participants. Lancet Oncol. 2021;22(9):1333-40. doi: 10.1016/s1470-2045(21)00377-6. PMID: 34399096.

70. Duggal V, Thomas IC, Montez-Rath ME, et al. National estimates of CKD prevalence and potential impact of estimating glomerular filtration rate without race. J Am Soc Nephrol. 2021;32(6):1454-63. doi: 10.1681/asn.2020121780. PMID: 33958490.

71. Elmaleh-Sachs A, Balte P, Oelsner EC, et al. Race/ethnicity, spirometry reference equations and prediction of incident clinical events: the multi-ethnic study of atherosclerosis (MESA) lung study. Am J Respir Crit Care Med. 2022;205(6):700-10. doi: 10.1164/rccm.202107-1612OC. PMID: 34913853.

72. Hoenig MP, Mann A, Pavlakis M. Removal of the Black race coefficient from the estimated glomerular filtration equation improves transplant eligibility for Black patients at a single center. Clin Transplant. 2022;36(2):e14467. doi: 10.1111/ctr.14467. PMID: 34605076.

73. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. N Engl J Med. 2021;385(19):1737-49. doi: 10.1056/NEJMoa2102953. PMID: 34554658.

74. Landy R, Young CD, Skarzynski M, et al. Using prediction-models to reduce persistent racial/ethnic disparities in draft 2020 USPSTF lung-cancer screening guidelines. J Natl Cancer Inst. 2021;113(11):1590-4. doi: 10.1093/jnci/djaa211. PMID: 33399825.

75. Mahmud N, Asrani SK, Reese PP, et al. Race adjustment in eGFR equations does not improve estimation of acute kidney injury events in patients with cirrhosis. Dig Dis Sc. 2022;67(4):1399-408. doi: 10.1007/s10620-021-06943-1. PMID: 33761091.

76. Miller J, Knorr JP. Impact of removing the race coefficient in renal function estimate equations on drug dosage recommendations. Ann Pharmacother. 2022;56(1):44-51. doi: 10.1177/10600280211010228. PMID: 33866823.

77. Panchal S, Serper M, Bittermann T, et al. Impact of race-adjusted glomerular filtration rate estimation on eligibility for simultaneous liver-kidney transplantation. Liver Transpl. 2022;28(6):959-68. doi: 10.1002/lt.26310. PMID: 34558791.

78. Park Y, Hu J, Singh M, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. JAMA Netw Open. 2021;4(4):e213909. doi: 10.1001/jamanetworkopen.2021.3909. PMID: 33856478.

79. Shi J, Lindo EG, Baird GS, et al. Calculating estimated glomerular filtration rate without the race correction factor: observations at a large academic medical system. Clin Chim Acta. 2021;520:16-22. doi: 10.1016/j.cca.2021.05.022. PMID: 34052206.

80. Tsai JW, Cerdeña JP, Goedel WC, et al. Evaluating the impact and rationale of race-specific estimations of kidney function: estimations from U.S. NHANES, 2015-2018. EClinicalMedicine. 2021;42:101197. doi: 10.1016/j.eclinm.2021.101197. PMID: 34849475.

81. Weale ME, Riveros-Mckay F, Selzam S, et al. Validation of an integrated risk tool, including polygenic risk score, for atherosclerotic cardiovascular disease in multiple ethnicities and ancestries. Am J Cardiol. 2021;148:157-64. doi: 10.1016/j.amjcard.2021.02.032. PMID: 33675770.

82. Yap E, Prysyazhnyuk Y, Ouyang J, et al. The implication of dropping race from the MDRD equation to estimate GFR in an African American-only cohort. Int J Nephrol. 2021;2021:1880499. doi: 10.1155/2021/1880499. PMID: 34824870.

83. Zelnick LR, Leca N, Young B, et al. Association of the estimated glomerular filtration rate with vs without a coefficient for race with time to eligibility for kidney transplant. JAMA Netw Open. 2021;4(1):e2034004. doi: 10.1001/jamanetworkopen.2020.34004. PMID: 33443583.

84. Fairman KA, Romanet D, Early NK, et al. Estimated cardiovascular risk and guideline-concordant primary prevention with statins: retrospective cross-sectional analyses of US ambulatory visits using competing algorithms. J Cardiovasc Pharmacol Ther. 2020;25(1):27-36. doi: 10.1177/1074248419866153. PMID: 31353942.

85. Hammond G, Johnston K, Huang K, et al. Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death. Circ Cardiovasc Qual Outcomes. 2020;13(6):e006752. doi: 10.1161/circoutcomes.120.006752. PMID: 32412300.

86. Coresh J, Inker LA, Sang Y, et al. Metabolomic profiling to improve glomerular filtration rate estimation: a proof-of-concept study. Nephrol Dial Transplant. 2019;34(5):825-33. doi: 10.1093/ndt/gfy094. PMID: 29718360.

87. Topel ML, Shen J, Morris AA, et al. Comparisons of the Framingham and pooled cohort equation risk scores for detecting subclinical vascular disease in Blacks versus whites. Am J Cardiol. 2018;121(5):564-9. doi: 10.1016/j.amjcard.2017.11.031. PMID: 29361288.

88. Yadlowsky S, Hayward RA, Sussman JB, et al. Clinical implications of revised pooled cohort equations for estimating atherosclerotic cardiovascular disease risk. Ann Intern Med. 2018;169(1):20-9. doi: 10.7326/m17-3011. PMID: 29868850.

89. Julian BA, Gaston RS, Brown WM, et al. Effect of replacing race with apolipoprotein L1 genotype in calculation of kidney donor risk index. Am J Transplant. 2017;17(6):1540-8. doi: 10.1111/ajt.14113. PMID: 27862962.

90. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk prediction models for cardiovascular events in black adults: the Jackson heart study cohort. JAMA Cardiol. 2016;1(1):15-25. doi: 10.1001/jamacardio.2015.0300. PMID: 27437649.

91. Kabra R, Girotra S, Vaughan Sarrazin M. Refining stroke prediction in atrial fibrillation patients by addition of African-American ethnicity to CHA(2)DS(2)-VASc score. J Am Coll Cardiol. 2016;68(5):461-70. doi: 10.1016/j.jacc.2016.05.044. PMID: 27470453.

92. Limdi NA, Brown TM, Yan Q, et al. Race influences warfarin dose changes associated with genetic factors. Blood. 2015;126(4):539-45. doi: 10.1182/blood-2015-02-627042. PMID: 26024874.

93. Kimmel SE, French B, Kasner SE, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. New England Journal of Medicine. 2013;369(24):2283-93. doi: 10.1056/NEJMoa1310669. PMID: 24251361.

94. Shores NJ, Dodge JL, Feng S, et al. Donor risk index for African American liver transplant recipients with hepatitis C virus. Hepatology. 2013;58(4):1263-9. doi: 10.1002/hep.26478. PMID: 23696235.

95. Drawz PE, Baraniuk S, Davis BR, et al. Cardiovascular risk assessment: addition of CKD and race to the Framingham equation. Am Heart J. 2012;164(6):925-31.e2. doi: 10.1016/j.ahj.2012.09.003. PMID: 2012701650, 23194494.

96. Foryciarz A, Pfohl SR, Patel B, et al. Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation. BMJ Health Care Inform. 2022 Apr;29(1):e100460. doi: 10.1136/bmjhci-2021-100460. PMID: 35396247.

97. Doshi MD, Schaubel DE, Xu Y, et al. Clinical utility in adopting race-free kidney donor risk index. Transplant Direct. 2022 Jul;8(7):e1343. doi: 10.1097/TXD.0000000000001343. PMID: 35747522.

98. Miller J, Lyden GR, McKinney WT, et al. Impacts of removing race from the calculation of the kidney donor profile index. Am J Transplant. 2023 May;23(5):636-41. doi: 10.1016/j.ajt.2022.12.016. PMID: 36695678.

99. Meeusen JW, Kasozi RN, Larson TS, et al. Clinical impact of the refit CKD-EPI 2021 creatinine-based eGFR equation. Clinical Chemistry. 2022;68(4):534-9. doi: 10.1093/clinchem/hvab282. PMID: 35038721.

100. Huang C, Murugiah K, Li X, et al. Effect of removing race correction factor in glomerular filtration rate estimation on predicting acute kidney injury after percutaneous coronary intervention. medRxiv. 2022 Jan. doi: 10.1101/2022.01.18.22269155.

101. Muiru AN, Madden E, Scherzer R, et al. Effect of adopting the new race-free 2021 chronic kidney disease epidemiology collaboration estimated glomerular filtration rate creatinine equation on racial differences in kidney disease progression among people with human immunodeficiency virus: an observational study. Clin Infect Dis. 2023 Feb;76(3):461-8. doi: 10.1093/cid/ciac731. PMID: 36069064.

102. Gutiérrez OM, Sang Y, Grams ME, et al. Association of estimated GFR calculated using race-free equations with kidney failure and mortality by black vs non-black race. JAMA. 2022 Jun;327(23):2306-16. doi: 10.1001/jama.2022.8801. PMID: 35667006.

103. Bundy JD, Mills KT, Anderson AH, et al. Prediction of end-stage kidney disease using estimated glomerular filtration rate with and without race: a prospective cohort study. Ann Intern Med. 2022 Mar;175(3):305-13. doi: 10.7326/M21-2928. PMID: 35007146.

104. Schmeusser BN, Palacios AR, Midenberg ER, et al. Race-free renal function estimation equations and potential impact on Black patients: implications for cancer clinical trial enrollment. Cancer. 2023 Mar;129(6):920-4. doi: 10.1002/cncr.34637. PMID: 36606692.

105. Diao JA, Wu GJ, Wang JK, et al. National projections for clinical implications of race-free creatinine-based GFR estimating equations. J Am Soc Nephrol. 2023 Feb;34(2):309-21. doi: 10.1681/ASN.2022070818. PMID: 36368777.

106. Lindley KJ, Limdi NA, Cavallari LH, et al. Warfarin dosing in patients with CYP2C9*5 variant alleles. Clin Pharmacol Ther. 2022;111(4):950-5. doi: 10.1002/cpt.2549. PMID: 35108398.

107. Ashana DC, Anesi GL, Liu VX, et al. Equitably allocating resources during crises: racial differences in mortality prediction models. Am J Respir Crit Care Med. 2021;204(2):178-86. doi: 10.1164/rccm.202012-4383OC. PMID: 33751910.

108. Thompson HM, Sharma B, Bhalla S, et al. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. J Am Med Inform Assoc. 2021;28(11):2393-403. doi: 10.1093/jamia/ocab148. PMID: 34383925.

109. Pasquinelli MM, Tammemagi MC, Kovitz KL, et al. Risk prediction model versus United States Preventive Services Task Force lung cancer screening eligibility criteria: reducing race disparities. J Thorac Oncol. 2020;15(11):1738-47. doi: 10.1016/j.jtho.2020.08.006. PMID: 32822843.

110. Brady W, de Souza K. The HEART score: A guide to its application in the emergency department. Turk J Emerg Med. 2018;18(2):47-51. doi: 10.1016/j.tjem.2018.04.004. PMID: 29922729.

111. Mahler SA, Riley RF, Hiestand BC, et al. The HEART pathway randomized trial: identifying emergency department patients with acute chest pain for early discharge. Circ Cardiovasc Qual Outcomes. 2015;8(2):195-203. doi: 10.1161/circoutcomes.114.001384. PMID: 25737484.

112. Zhang X, Carabello M, Hill T, et al. Trends of racial/ethnic differences in emergency department care outcomes among adults in the United States from 2005 to 2016. Front Med (Lausanne). 2020;7:300. doi: 10.3389/fmed.2020.00300. PMID: 32671081.

113. The new Kidney Allocation System (KAS). Frequently asked questions. OPTN: Organ Procurement and Transplantation Network. https://optn.transplant.hrsa.gov/media/1235/kas_faqs.pdf. Accessed on December 12, 2022.

114. A guide to calculating and interpreting the Estimated Post-Transplant Survival (EPTS) score used in the Kidney Allocation System (KAS). OPTN: Organ Procurement and Transplantation Network; 2020. https://optn.transplant.hrsa.gov/media/1511/guide_to_calculating_interpreting_epts.pdf. Accessed on December 12, 2022.

115. Haddad DN, Sandler KL, Henderson LM, et al. Disparities in lung cancer screening: a review. Ann Am Thorac Soc. 2020;17(4):399-405. doi: 10.1513/AnnalsATS.201907-556CME. PMID: 32017612.

116. Rojas JC, Rohweder G, Guptill J, et al. Predictive analytics programs at large healthcare systems in the USA: a national survey. J Gen Intern Med. 2022;37(15):4015-7. doi: 10.1007/s11606-022-07517-1. PMID: 35396657.

117. Ross C, Herman B. Denied by AI: how Medicare Advantage plans use algorithms to cut off care for seniors in need. STAT; 2023. https://www.statnews.com/2023/03/13/medicare-advantage-plans-denial-artificial-intelligence/. Accessed on May 5, 2023.

118. Dorr DA, Adams L, Embí P. Harnessing the promise of artificial intelligence responsibly. JAMA. 2023;329(16):1347-8. doi: 10.1001/jama.2023.2771. PMID: 36972068.

119. Röösli E, Bozkurt S, Hernandez-Boussard T. Peeking into a black box, the fairness and generalizability of a MIMIC-III benchmarking model. Sci Data. 2022 Jan;9(1):24. doi: 10.1038/s41597-021-01110-7. PMID: 35075160.

120. Hane CA, Wasserman M. Designing equitable health care outreach programs from machine learning patient risk scores. Med Care Res Rev. 2023;80(2):216-27. doi: 10.1177/10775587221098831. PMID: 35685000.

121. Rojas JC, Fahrenbach J, Makhni S, et al. Framework for integrating equity into machine learning models: a case study. Chest. 2022;161(6):1621-7. doi: 10.1016/j.chest.2022.02.001. PMID: 35143823.

122. Locke T, Parker V, Thoumi A, et al. Preventing bias and inequities in AI-enabled health tools. Washington (DC): Duke-Margolis Center for Health Policy; 2022. https://healthpolicy.duke.edu/publications/preventing-bias-and-inequities-ai-enabled-health-tools. Accessed on August 14, 2022.

123. Obermeyer Z, Nissan R, Stern M, et al. Algorithmic bias playbook. Chicago (IL): The University of Chicago Booth School of Business, The Center for Applied Artificial Intelligence; 2021. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook

124. Turner Lee N, Resnick P, Resnick P, et al. Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms. Washington (DC): The Brookings Institution; 2019. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/. Accessed on August 14, 2022.

125. Shachar C, Gerke S. Prevention of bias and discrimination in clinical practice algorithms. JAMA. 2023;329(4):283-4. doi: 10.1001/jama.2022.23867. PMID: 36602791.

126. Goodman KE, Morgan DJ, Hoffmann DE. Clinical algorithms, antidiscrimination laws, and medical device regulation. JAMA. 2023;329(4):285-6. doi: 10.1001/jama.2022.23870. PMID: 36602795.

127. Attorney General Bonta launches inquiry into racial and ethnic bias in healthcare algorithms. Office of the Attorney General. State of California Department of Justice; 2022. https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-racial-and-ethnic-bias-healthcare. Accessed on May 16, 2023.

128. Khazanchi R, Morse M. New York City Coalition to End Racism in Clinical Algorithms (CERCA). Inaugural Report. New York City Department of Health and Mental Hygiene, Office of the Chief Medical Officer; 2022. https://www.nyc.gov/assets/doh/downloads/pdf/cmo/cerca-report.pdf. Accessed on May 5, 2023.

129. Raji ID, Smart A, White RN, et al. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. Conference on Fairness, Accountability, and Transparency (FAT* '20); January 27–30, 2020; Barcelona, Spain. Association for Computing Machinery. https://doi.org/10.1145/3351095.

130. Microsoft responsible AI standard, v2. General requirements. Redmond (WA): Microsoft Corporation; 2022. https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf. Accessed on August 14, 2022.

131. Landers RN, Behrend TS. Auditing the AI auditors: a framework for evaluating fairness and bias in high stakes AI predictive models. Am Psychol. 2023;78(1):36-49. doi: 10.1037/amp0000972. PMID: 35157476.

132. Makhni S, Chin MH, Fahrenbach J, et al. Equity challenges for artificial intelligence algorithms in health care. Chest. 2022;161(5):1343-6. doi: 10.1016/j.chest.2022.01.009. PMID: 35526892.

133. Liu X, Glocker B, McCradden MM, et al. The medical algorithmic audit. Lancet Digit Health. 2022;4(5):e384-e97. doi: 10.1016/s2589-7500(22)00003-6. PMID: 35396183.

134. Reddy S, Allan S, Coghlan S, et al. A governance model for the application of AI in health care. J Am Med Inform Assoc. 2020;27(3):491-7. doi: 10.1093/jamia/ocz192. PMID: 31682262.

135. Good machine learning practice for medical device development: guiding principles. Silver Spring (MD): U.S. Department of Health and Human Services, Food and Drug Administration; 2021. https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles. Accessed on August 14, 2022.

136. Schwartz R, Vassilev A, Greene K, et al. Towards a standard for identifying and managing bias in artificial intelligence. NIST special publication 1270: National Institute of Standards and Technology, U.S. Department of Commerce; 2022. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf

137. Costanza-Chock S, Raji ID, Buolamwini J. Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency; 2022 Seoul, Republic of Korea. Association for Computing Machinery; pp. 1571–83. doi: 10.145/3531146.3213.

138. Schmidt IM, Shohet M, Serrano M, et al. Patients' perspectives on race and the use of race-based algorithms in clinical decision-making: a qualitative study. J Gen Intern Med. 2023:Online ahead of print. doi: 10.1007/s11606-023-08035-4. PMID: 36811702.

139. Khullar D, Casalino LP, Qian Y, et al. Perspectives of patients about artificial intelligence in health care. JAMA Netw Open. 2022;5(5):e2210309. doi: 10.1001/jamanetworkopen.2022.10309. PMID: 35507346.

140. Tyson A, Pasquini G, Spencer A, et al. 60% of Americans would be uncomfortable with provider relying on AI in their own health care. Pew Research Center; 2023. https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/. Accessed on May 5, 2023.

141. Alberti P, Fair M, Skorton DJ. Now is our time to act: why academic medicine must embrace community collaboration as its fourth mission. Acad Med. 2021;96(11):1503-6. doi: 10.1097/ACM.0000000000004371. PMID: 34432717.

142. Addressing and eliminating racism at the AAMC and beyond. Association of American Medical Colleges (AAMC). https://www.aamc.org/addressing-and-eliminating-racism-aamc-and-beyond. Accessed on May 16, 2023.

143. Diversity, equity, and inclusion competencies across the learning continuum. Association of American Medical Colleges (AAMC); 2022. https://www.aamc.org/data-reports/report/diversity-equity-and-inclusion-competencies-across-learning-continuum. Accessed on May 16, 2023.

144. Choudhry SA, Li J, Davis D, et al. A public-private partnership develops and externally validates a 30-day hospital readmission risk prediction model. Online J Public Health Inform. 2013;5(2):219. doi: 10.5210/ojphi.v5i2.4726. PMID: 24224068.

145. Cognitive computing model brief. Pediatric Hospital admissions and ED visits. Verona (WI): Epic Systems Corporation; 2021. https://galaxy.epic.com/Redirect.aspx?DocumentID=3763630&Version=Epic%202018. Accessed on August 14, 2022.

146. Peterson PN, Rumsfeld JS, Liang L, et al. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. Circ Cardiovasc Qual Outcomes. 2010;3(1):25-32. doi: 10.1161/circoutcomes.109.854877. PMID: 20123668.

147. Obi Y, Nguyen DV, Zhou H, et al. Development and validation of prediction scores for early mortality at transition to dialysis. Mayo Clin Proc. 2018;93(9):1224-35. doi: 10.1016/j.mayocp.2018.04.017. PMID: 30104041.

148. O'Brien SM, Feng L, He X, et al. The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: part 2-statistical methods and results. Ann Thorac Surg. 2018;105(5):1419-28. doi: 10.1016/j.athoracsur.2018.03.003. PMID: 29577924.

149. Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: part 1-background, design considerations, and model development. Ann Thorac Surg. 2018;105(5):1411-8. doi: 10.1016/j.athoracsur.2018.03.002. PMID: 29577925.

150. Huckaby LV, Gleason TG, Ferdinand FD, et al. High-risk committee for cardiac surgery decision-making: results from 110 consecutive patients. Ann Thorac Surg. 2021;112(2):582-8. doi: 10.1016/j.athoracsur.2020.09.014. PMID: 33127404.

151. Haukoos JS, Lyons MS, Lindsell CJ, et al. Derivation and validation of the Denver human immunodeficiency virus (HIV) risk score for targeted HIV screening. Am J Epidemiol. 2012;175(8):838-46. doi: 10.1093/aje/kwr389. PMID: 22431561.

152. Haukoos JS, Hopkins E, Bucossi MM, et al. Brief report: validation of a quantitative HIV risk prediction tool using a national HIV testing cohort. J Acquir Immune Defic Syndr. 2015;68(5):599-603. doi: 10.1097/qai.0000000000000518. PMID: 25585300.

153. Dunlevy H, Robins M, Ashwood E, et al. Targeted HIV-testing in the emergency department with linkage to care using an HIV risk score. Aurora (CO): 2018 National Ryan White Conference on HIV Care & Treatment; 2018. https://targethiv.org/sites/default/files/supporting-files/11016_Dunlevy_508.pdf. Accessed on August 1, 2022.

154. Shahian DM, Badhwar V, O'Brien SM, et al. Social risk factors in Society of Thoracic Surgeons risk models. Part 1: concepts, indicator variables, and controversies. Ann Thorac Surg. 2022;113(5):1703-17. doi: 10.1016/j.athoracsur.2021.11.067. PMID: 34998732.

155. Shahian DM, Badhwar V, O'Brien SM, et al. Social risk factors in Society of Thoracic Surgeons risk models. Part 2: empirical studies in cardiac surgery; risk model recommendations. Ann Thorac Surg. 2022;113(5):1718-29. doi: 10.1016/j.athoracsur.2021.11.069. PMID: 34998735.

156. Roger VL. Epidemiology of heart failure: a contemporary perspective. Circ Res. 2021;128(10):1421-34. doi: 10.1161/circresaha.121.318172. PMID: 33983838.

157. Whelehan DF, Conlon KC, Ridgway PF. Medicine and heuristics: cognitive biases and medical decision-making. Ir J Med Sci. 2020;189(4):1477-84. doi: 10.1007/s11845-020-02235-1. PMID: 32409947.

158. United States Renal Data System. 2022 USRDS annual data report: epidemiology of kidney disease in the United States. Bethesda (MD): National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases; 2022. https://usrds-adr.niddk.nih.gov/2022. Accessed on May 16, 2023.

159. Diabetes Report Card 2021. Atlanta (GA): Centers for Disease Control and Prevention. US Department of Health and Human Services; 2022. https://www.cdc.gov/diabetes/library/reports/reportcard.html. Accessed on August 22, 2022.

160. Mullan CW, Mori M, Pichert MD, et al. United States national trends in comorbidity and outcomes of adult cardiac surgery patients. J Card Surg. 2020;35(9):2248-53. doi: 10.1111/jocs.14764. PMID: 33448476.

161. Hiv.gov. HIV basics. Overview: data and trends: U.S. statistics. Washington (DC): Office of Infectious Disease and HIV/AIDS Policy. https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics. Accessed on August 22, 2022.

162. Haukoos JS, Lyons MS, Rothman RE, et al. Comparison of HIV screening strategies in the emergency department: a randomized clinical trial. JAMA Netw Open. 2021;4(7):e2117763. doi: 10.1001/jamanetworkopen.2021.17763. PMID: 34309668.

163. Jin R, Furnary AP, Fine SC, et al. Using Society of Thoracic Surgeons risk models for risk-adjusting cardiac surgery results. Ann Thorac Surg. 2010;89(3):677-82. doi: 10.1016/j.athoracsur.2009.10.078.

164. Maiga A, Farjah F, Blume J, et al. Risk prediction in clinical practice: a practical guide for cardiothoracic surgeons. Ann Thorac Surg. 2019;108(5):1573-82. doi: 10.1016/j.athoracsur.2019.04.126. PMID: 31255609.

165. Rosenberg ES, Delaney KP, Branson BM, et al. Re: "Derivation and validation of the Denver Human Immunodeficiency Virus (HIV) risk score for targeted HIV screening". Am J Epidemiol. 2012;176(6):567-8; author reply 8. doi: 10.1093/aje/kws305. PMID: 22899828.

166. Meetings examine impact of healthcare algorithms on racial and ethnic disparities in health and healthcare. Rockville (MD): Effective Health Care Program, Agency for Healthcare Research and Quality; 2023. https://effectivehealthcare.ahrq.gov/news/meetings. Accessed on July 5, 2023.

167. Opportunity for feedback: principles to address the impact of healthcare algorithms on racial and ethnic disparities in health and healthcare. Rockville (MD): Effective Health Care Program, Agency for Healthcare Research and Quality; 2023. https://effectivehealthcare.ahrq.gov/news/opportunity-feedback. Accessed on July 5, 2023.

168. Inclusive language guidelines. American Psychological Association; 2021. https://www.apa.org/about/apa/equity-diversity-inclusion/language-guidelines.pdf. Accessed on May 16, 2023.

Abbreviations and Acronyms

ACC/AHA	American College of Cardiology/American Heart Association
APACHE Iva	Acute physiology and chronic health evaluation
ASCVD	Atherosclerotic cardiovascular disease
AUC	Area under the curve
BIPOC	Black, Indigenous, or People of Color
CDC	Centers for Disease Control and Prevention
CKD	Chronic kidney disease
CKD-EPI	Chronic Kidney Disease Epidemiology Collaboration
COAG	Clarification of Oral Anti-coagulation through Genetics study
CSC	Crisis standards of care
eGFR	Estimated glomerular filtration rate
FDA	Food and Drug Administration
FEV_1	Forced expiratory volume
FNR	False-negative rate
FRS	Framingham Risk Score
FVC	Forced vital capacity
GLI	Global Lung Function Initiative
HIT	Health information technology
HIV	Human immunodeficiency virus
iGFR	Iothalamate glomerular filtration rate
IQR	Interquartile rate
IRR	Incidence rate ratio
KAS	Kidney Allocation System
KPPC RC	Kaiser Permanente prostate cancer risk calculator
LAPS2	Laboratory-based Acute Physiology Score version 2
MDRD	Modification of Diet in Renal Disease study
NRI	Net Reclassification Index
OASIS	Oxford Acute Severity of Illness Score
OMB	Office of Management and Budget
OR	Odds ratio
PBCG	Prostate Biopsy Collaborative Group
PCE	Pooled cohort equations
PCPT RC	Prostate Cancer Prevention Trial risk calculator
PSA	Prostate-specific antigen
ROB	Risk of bias
SD	Standard deviation

SDOH	Social determinants of health
SOFA	Sequential Organ Failure Assessment
USPSTF	United States Preventive Services Task Force

Appendix A. Methods for Search Strategy

Search Details and Data Sources

The search strategy was designed and conducted by an experienced systematic review/medical reference Librarian with input from the investigators. Two other Librarians peer reviewed the search strategies using the PRESS Checklist. We consulted with SMEs, Key Informants, and Technical Experts to identify additional relevant keywords and concepts. We tested the final search against seven key articles identified during the project's Topic Refinement phase (Miller 2021, Park 2021, Zelnick 2021, Obermeyer 2019, Limdi 2015, Kimmel 2013, Hankinson 2010) to ensure that the strategy was sensitive enough to capture these articles. We also searched SCOPUS to identify articles that cited the key articles, plus one additional article (Vyas et al. 2020). We applied the following limits or filters to the database searches:

- *Date:* Our SMEs recommended a search parameter of at least 10 years. Earlier articles are unlikely to reflect current algorithms.
- *Language*: Publications were excluded if they were written in a language other than English due to resource constraints.
- *Publication status:* We searched for published, unpublished, and ongoing studies.
- *Human or organism:* The search was limited to human studies.
- *Study design:* The search was not restricted to any study type; however, a hedge was applied to remove animal studies, books, case reports, conference materials, editorials, letters, and news items.
- *Study location:* Retrieval was limited to studies published in the United States and/or studies using data from populations in the United States. Concepts of race and ethnicity as well as racial/ethnic make-up differ across countries and regions, and this decision was made to focus our efforts to the needs of the requestor (U.S. Congress).

We conducted a comprehensive literature search in January 2022 (updated in February 2023). We searched the following databases:

- Embase and MEDLINE (searched simultaneously in Embase.com) (2011 to February 7, 2023) Dates searched: January 6, 2022; updated February 7, 2023
- PubMed (publisher supplied/in process citations/PubMed not Medline) (2011 to February 7, 2023) Dates searched: January 6, 2022; updated February 7, 2023
- Cochrane Database of Systematic Reviews (2011 to February 7, 2023) Dates searched: January 12, 2022; updated February 7, 2023

We also searched the following resources to identify additional relevant materials published in open-access journals and the computing literature:

- Association for Computing Machinery (ACM) Digital Library. https://dl.acm.org/ (2011 to February 7, 2023) Dates searched: January 25, 2022; updated February 7, 2023
- PubMed Central (PMC). https://www.ncbi.nlm.nih.gov/pmc (2011 to February 7, 2023). Dates searched: February 25, 2022; updated February 7, 2023

- Nature Digital Medicine (searched via PMC search engine 2018* to February 7, 2023). **This journal was not included in PMC prior to 2018.* Dates searched: February 25, 2022; updated February 7, 2023

We searched the grey literature in January 2022 (selected resources updated in February 2023 as indicated below) to address the key questions in the Systematic Review and the Contextual Questions. Searches were executed using the search function of the website, browsing the menu items, and searching the website via Google. The following search terms were used: algorithm, machine learning, artificial intelligence, bias, race, racial, racism, ethnic, ethnicity, disparities, inequities. Retrieval was at the searcher's discretion and focused on white papers, monographs, reports, recommendations, policies, guidelines, regulatory information, ongoing clinical trials, and original research outside the scope of health that may not have been included in the bibliographic databases searched for this project. The following resources were included in the grey literature search:

Strategies: browse menu items, use websites search engine, search via Google

- *Trials/research registries*.
  - ClinicalTrials.gov www.clinicaltrials.gov Date searched: January 25, 2022. Methods: see strategy below
- *Preprint servers*. To locate unpublished studies, we searched preprint servers listed below for items posted since 2020. Methods: see strategies below
  - MedRxiv. www.medrxiv.org/ (searched using the MedRxivr tool: medrxivr (shinyapps.io)) Date searched: January 11, 2022; updated February 7, 2023
  - BioRxiv. www.biorxiv.org/ Date searched: January 12, 2022; updated February 7, 2023
- *Web search engines/specific websites*. We searched the following associations for relevant materials posted/published 2011 through January 2022 (selected resources updated in February 2023 as indicated below). Methods: Searches were conducted using the website search and browsing capabilities and the Google site search tool.
  - Agency for Healthcare Research and Quality (AHRQ) Website. www.ahrq.gov/ Date searched Jan 14, 2022; updated February 7, 2023
  - American Academy of Actuaries. www.actuary.org/ Date searched January 20, 2022
  - American Health Information Management Association www.ahima.org/ Date searched January 26, 2022
  - American Hospital Association Institute for Diversity and Health Equity (IFDHE) https://ifdhe.aha.org/ Date searched January 18, 2022
  - American Medical Informatics Association (AMIA) www.amia.org/ Date searched Jan 18, 2022
  - Association for Computing Machinery Website www.acm.org/ Date searched January 14, 2022
  - Centers for Disease Control and Prevention www.cdc.gov/ Date searched January 20, 2022
  - Consumer Financial Protection Bureau www.consumerfinance.gov/ Date searched January 19, 2022
  - Food and Drug Administration (FDA). www.fda.gov/ Date searched January 21, 2022; updated February 8, 2023

- Health Resources and Services Administration. www.hrsa.gov/ Date searched January 21, 2022
- Healthcare Information and Management System Society www.himss.org/. Date searched January 20, 2022
- Health Resources and Services Administration www.hrsa.gov/. Date searched January 21, 2022
- National Institute of Standards and Technology www.nist.gov/ Date searched January 21, 2022; updated February 8, 2023
- Observational Health Data Sciences and Informatics www.ohdsi.org/ Date searched January 21, 2022
- Office of the National Coordinator for Health Information Technology (ONC) www.healthit.gov/ Date searched January 24, 2022; updated February 8, 2023

- *International websites*. We searched the following international resources to identify possible legislation and research outside the United States.
  - AlgorithmWatch. https://algorithmwatch.org/en/ Date searched January 27, 2022
  - European Commission Website https://ec.europa.eu/info/index_en Date searched January 26, 2022
  - International Medical Informatics Association https://imia-medinfo.org/wp/ Date searched January 27, 2022
  - United Kingdom Government Website http://www.uk.gov/ Date searched January 26, 2022
- *Guideline repositories*
  - ECRI Guidelines Trust. https://guidelines.ecri.org/ Date searched January 27, 2022
  - TRIP Database (search limited to guidelines only). https://www.tripdatabase.com/ Date searched January 27, 2022

Database Search Strategies

Embase.com Strategy: (Combines Medline and EMBASE) January 1, 2011, through February 7, 2023

1 'ancestry group'/exp OR 'ethnic group'/exp OR 'ethnic or racial aspects'/de OR 'ethnicity'/mj OR 'race'/de OR race:ti OR racial*:ti OR 'ethnic group*':ti OR ethnicit*:ti

2 'multiracial person'/exp OR 'asian american'/exp OR 'black person'/exp OR 'african american'/exp OR 'hispanic'/exp OR 'alaska native'/exp OR 'american indian'/exp OR 'pacific islander'/exp OR (((arab OR asian OR african OR indian* OR indigenous) NEXT/3 american*):ti,ab,kw) OR ((native NEAR/2 (American* OR Alaskan*)):ti,ab,kw) OR (((black OR brown) NEXT/2 (person* OR people OR patient* OR American*)):ti,ab,kw) OR blacks:ti,ab,kw OR hispanic*:ti,ab,kw OR latino*:ti,ab,kw OR latina*:ti,ab,kw OR latinx:ti,ab,kw OR (pacific NEXT/2 islander*):ti,ab,kw OR 'non caucasian*':ti,ab,kw OR noncaucasian*:ti,ab,kw OR 'non white*':ti,ab,kw OR nonwhite*:ti,ab,kw OR ((mexican* NEAR/5 (america* OR us OR usa)):ti,ab,kw) OR (mixed NEAR/2 (ethnic* OR race*)):ti,ab,kw OR Multiracial:ti,ab,kw OR Multi-racial:ti,ab,kw OR biracial:ti,ab,kw OR multiethnic*:ti,ab,kw OR multi-ethnic*:ti,ab,kw OR (multiple NEXT/1 (ethnic* OR race*)):ti,ab,kw OR bipoc:ti,ab,kw OR ((ethnic* OR

	race* OR racial) NEXT/1 group*):ti,ab,kw OR ((ethnic* OR race* OR racial) NEAR/2 ('sub group*' OR subgroup*)):ti,ab,kw
3	1 OR 2
4	'algorithm'/exp OR 'algorithm bias'/exp OR algorithm*:ti,ab,kw
5	'artificial intelligence'/exp OR 'computer model'/exp OR 'machine learning'/exp OR 'computer prediction'/exp OR 'data mining'/exp OR 'artificial neural network'/exp OR 'computer assisted diagnosis'/de OR 'computer analysis'/exp OR 'statistical model'/exp OR 'information processing'/mj OR ((artificial NEXT/2 intelligence):ti,ab,kw) OR (((computer OR machine OR deep) NEXT/2 (learning OR predict*)):ti,ab,kw) OR ((neural NEXT/2 network*):ti,ab,kw) OR ((data NEXT/2 (mine OR mined OR mining)):ti,ab,kw) OR ((dataset* OR 'data set*' OR model OR models) NEAR/5 (train OR training OR mitigat* OR bias*)):ti,ab,kw OR 'training data':ti,ab,kw
6	'calculation'/exp/mj OR 'rating scale'/exp/mj OR 'model'/mj OR 'disease model'/exp/mj OR 'scoring system'/exp/mj OR 'prediction and forecasting'/exp/mj OR scale:ti,kw OR scales:ti,kw OR instrument*:ti,kw OR index*:ti,kw OR indices:ti,kw OR measure*:ti,kw OR metric*:ti,kw OR calculat*:ti,kw OR score*:ti,kw OR formula:ti,kw OR formulas:ti,kw OR variable*:ti,kw OR coefficient*:ti,kw OR 'co-efficient*':ti,kw OR equation*:ti,kw OR proxy:ti,ab,kw OR proxies:ti,ab,kw OR tool*:ti,kw OR ((correction NEXT/2 factor*):ti,ab,kw) OR ((data NEXT/2 driven):ti,ab,kw) OR ((big NEXT/2 data):ti,ab,kw) OR ((predict* NEXT/2 (model* OR analytic*)):ti,ab,kw)
7	4 OR 5 OR 6
8	'bias'/de OR 'prejudice'/exp OR 'health disparity'/exp OR 'health care disparity'/exp OR 'disparity'/exp OR 'health equity'/exp OR 'race difference'/exp OR 'racism'/exp OR 'ethnic difference'/exp OR equity:ti,ab,kw OR disparit*:ti,ab,kw OR discrimination:ti,kw OR bias*:ti,ab,kw OR unequal*:ti,ab,kw OR inequal*:ti,ab,kw OR inequit*:ti,ab,kw OR disproportionat*:ti,ab,kw OR prejudice*:ti,ab,kw OR imbalance*:ti,ab,kw OR fairness:ti,ab,kw OR underserved:ti,ab,kw OR ((under NEXT/2 served):ti,ab,kw) OR marginalized:ti,ab,kw OR (((race* OR racial* OR ethnic* OR ancestries OR ancestry) NEAR/5 (differen* OR discrimination*)):ti,ab,kw) OR racism:ti,ab,kw OR racist:ti,ab,kw OR reclassif*:ti,ab,kw OR misestimat*:ti,ab,kw OR misrepresent*:ti,ab,kw OR "less likely":ti,ab OR "more likely":ti,ab OR ((with OR without) NEXT/3 (race OR ethnic* OR racial)):ti,ab OR (compared NEAR/6 (white OR whites OR Caucasian*)):ti,ab OR (underrepresent* OR overrepresent*):ti,ab
9	3 AND 7 AND 8
10	(('algorithm'/exp OR 'algorithm bias'/exp OR algorithm*:ti,kw) AND ('race'/de OR 'race difference'/exp OR 'racism'/exp OR race:ti,kw OR racial*:ti,kw OR ethnicity:ti,kw)) OR (Algorithm* NEAR/10 (race OR racial* OR ethnic* OR racis*))
11	9 OR 10
12	11 NOT ('book'/de OR 'case report'/de OR 'conference paper'/exp OR 'editorial'/de OR 'letter'/de OR (book OR chapter OR conference OR editorial OR letter):it OR [conference abstract]/lim OR [conference paper]/lim OR [conference review]/lim OR [editorial]/lim OR [letter]/lim OR (abstract OR annual OR conference OR congress OR meeting OR proceedings OR sessions OR symposium):nc OR ((book NOT series) OR 'conference proceeding'):pt OR ('case report' OR comment* OR editorial OR letter OR news):ti OR ((protocol AND (study OR trial)) NOT ('therapy protocol*' OR 'treatment protocol*')):ti)

13 12 NOT (([animals]/lim NOT [humans]/lim) OR ((animal OR animals OR canine* OR dog OR dogs OR feline OR hamster* OR lamb OR lambs OR mice OR monkey OR monkeys OR mouse OR murine OR pig OR piglet* OR pigs OR porcine OR primate* OR rabbit* OR horse OR horses OR rat OR rats OR rodent* OR sheep* OR swine OR veterinar*) NOT (human* OR patient*)):ti)
14 13 AND [english]/lim AND [2011-2022]/py
15 14 AND ('united states'/exp OR 'united states' OR usa OR American*)

Embase.com Syntax
**** = truncation***
/exp = explode to include all terms in the tree
/mj = limit to terms indexed as major concepts
/de = search term without exploding
:ti = search in the title field
:kw = search in the author keywords field
:ab = search in the abstract field
NEAR/# - search the terms within # of each other in any order
NEXT/# - search terms within # of each other in the specified order.

PubMed (In Process Citations): January 1, 2011, through February 7, 2023
1 Race*[ti] OR racial*[ti] OR ethnic*[ti] OR "ethnic group*"[tiab] OR "asian american*"[tiab] OR "black person*"[tiab] OR "black people"[tiab] OR "black american*"[tiab] OR blacks OR "african american*"[tiab] OR "brown person*"[tiab] OR "brown people"[tiab] OR "American indian*"[tiab] OR "native american*"[tiab] OR "native Alaskan*"[tiab] OR "Alaskan native*"[tiab] OR "pacific islander*"[tiab] OR ((arab[tiab] OR Asian[tiab] OR African[tiab] OR indian*[tiab] OR indigenous[tiab] OR mexican[tiab]) AND american*[tiab]) OR hispanic*[tiab] OR latino*[tiab] OR Latina*[tiab] OR latinx[tiab] OR "mixed ethnic*"[tiab] OR "mixed race*"[tiab] OR "non caucasian*"[tiab] OR noncaucasian*[tiab] OR "non white*" [tiab] OR nonwhite*[tiab] OR multiracial[tiab] OR "multi racial"[tiab] OR "multiethnic*"[tiab] OR "multi ethnic*"[tiab] OR bipoc[tiab] OR "multiple ethnic*"[tiab] OR "multiple race*"[tiab] OR ((ethnic*[ti] OR race*[ti] OR racial[ti]) AND (group*[ti] OR subgroup*[ti]))
2 algorithm*[tiab] OR "artificial intelligence"[tiab] OR "computer learning"[tiab] OR "computer model*"[tiab] OR "machine learning"[tiab] OR "computer predict*"[tiab] OR "machine predict*"[tiab] OR "data mining"[tiab]OR " neural network"[tiab] OR "computer assisted"[tiab] OR "computer analysis"[tiab] OR "statistical model*"[tiab] OR "deep learning"[tiab] OR ((dataset*[tiab] OR "data set*"[tiab] OR model[tiab] OR models[tiab]) AND (train[tiab] OR training[tiab] OR mitigat*[tiab] OR bias*[tiab])) OR "training data"[tiab]
3 calculat*[ti] OR scale[ti] OR scales[ti] OR model*[ti] OR score*[ti] OR predict*[ti] OR instrument*[ti] OR index*[ti] OR indices[ti] OR measure*[ti] OR metric*[ti] OR formula[ti] OR formulas*[ti] OR variable*[ti] OR coefficient*[ti] OR "co-efficient"[ti] OR equation*[ti] OR proxy[ti] OR proxies[ti] OR tool[ti] OR tools[ti] OR "correction factor*"[tiab] OR "data driven"[tiab] OR "big data"[tiab] OR "predictive analytic*"[tiab] OR "prediction analytic*"[tiab] OR "predictive model*"[tiab] OR "prediction model*"[tiab]
4 2 OR 3

5	bias*[tiab] OR prejudice*[tiab] OR racist[tiab] OR racism[tiab] OR disparit*[tiab] OR equity[tiab] OR unequal*[tiab] OR equality[tiab] OR inequal*[tiab] OR inequit*[tiab] OR disproportionat*[tiab] OR imbalance*[tiab] OR fairness[tiab] OR underserved[tiab] OR marginalized[tiab] OR reclassif*[tiab] OR misestimate*[tiab] OR misrepresent*[tiab] OR "less likely"[tiab] OR "more likely"[tiab] OR underrepresent*[tiab] OR overrepresent*[tiab] OR ((race[tiab] OR racial*[tiab] OR ethnic*[tiab] OR ancestry*[tiab]) AND (difference*[tiab] OR discriminat*[tiab]))
6	1 AND 4 AND 5
7	Algorithm* AND (racist OR racism OR race OR racial OR ethnic*)
8	6 OR 7
9	8 NOT (booksdocs[Filter] OR "case reports"[pt] OR comment[pt] OR congress[pt] OR editorial[pt] OR letter[pt] OR "case report"[ti] OR comment*[ti] OR editorial[ti] OR letter[ti] OR news[ti] OR ((protocol[ti] AND (study[ti] OR trial[ti])) NOT ("therapy protocol*"[ti] OR "treatment protocol*"[ti])))
10	9 NOT ((animal[ti] OR animals[ti] OR canine*[ti] OR dog[ti] OR dogs[ti] OR feline[ti] OR hamster*[ti] OR lamb[ti] OR lambs[ti] OR mice[ti] OR monkey[ti] OR monkeys[ti] OR mouse[ti] OR murine[ti] OR pig[ti] OR piglet*[ti] OR pigs[ti] OR porcine[ti] OR primate*[ti] OR rabbit*[ti] OR rat[ti] OR rats[ti] OR rodent*[ti] OR sheep*[ti] OR swine[ti] OR veterinar*[ti]) NOT (human*[ti] OR patient*[ti]))
11	10 AND (inprocess[sb] OR publisher[sb] or pubmednotmedline[sb])
12	11 AND english[la] AND ("2011"[Date - Publication] : "3000"[Date - Publication])
13	12 AND ("united states" OR "USA" OR American*)

PubMed Syntax
**** = truncation***
[ti]= search in the title field
[tiab]= search in the title and abstract
[la] = search in the language field
[sb] = subset

Cochrane Library (Cochrane Database of Systematic Reviews, Clinical Answers, Special Collections): January 1, 2011, through February 7, 2023

1	(Race* OR racial* OR ethnic* OR multiracial OR "asian american*" OR "black person*" OR "black people" OR "black american*" OR "african american*" OR "brown person*" OR "brown people" OR "American indian*" OR "native american*" OR "native Alaskan*" OR "Alaskan native*" OR "pacific islander*" OR ((arab OR Asian OR African OR indian* OR indigenous OR mexican) AND american*) OR hispanic OR latino* OR Latina* OR latinx OR "non white*" OR nonwhite* OR multiracial OR "multi racial" OR "multiethnic*" OR "multi ethnic*" OR bipoc):ti,ab,kw
2	(algorithm* OR "artificial intelligence" OR "computer learning" OR "computer model*" OR "machine learning" "machine based learning" OR "computer predict*" OR "machine predict*" OR "data mining" OR "artificial neural network" OR "computer assisted" OR "computer analysis" OR "statistical model*" OR "deep learning" OR ((computer OR machine OR deep) NEAR/2 (learning OR predict*)) OR ((dataset* OR 'data set*' OR model OR models) NEAR/5 (train OR training OR mitigat* OR bias*)) OR "training data"):ti,ab,kw
3	(calculat* OR scale OR scales OR model* OR score* OR predict* OR instrument* OR index OR indices OR measure* OR metric* OR formula OR formulas* OR variable* OR coefficient* OR "co-efficient" OR equation* OR proxy OR proxies OR tool OR tools):ti
4	("correction factor*" OR "data driven" OR "big data" OR "predictive analytic*" OR "prediction analytic*" OR "predictive model*" OR "prediction model*" OR "correction factor*"):ti,ab,kw
5	2 OR 3 OR 4
6	(bias* OR prejudice* OR racist OR racism OR disparit* OR equity OR unequal* OR equality OR inequal* OR inequit* OR disproportionat* OR imbalance* OR fairness OR underserved OR "under-served" OR "under served" OR marginalized OR reclassif* OR misestimate* OR misrepresent* OR "less likely" OR "more likely" OR underrepresent* OR overrepresent* OR ((race OR racial* OR ethnic* OR ancestr*) AND (differen* OR discriminat*)) OR ((with OR without) NEAR/3 (race OR ethnic* OR racial)) OR (compared NEAR/6 (white OR whites OR Caucasian*))):ti,ab,kw
7	1 AND 5 AND 6
8	Algorithm* AND (racist OR racism OR race OR racial OR ethnic*):ti,ab,kw
9	7 OR 8
10	9 AND Cochrane Library publication date from Jan 2011 to Jan 2022, in Cochrane Reviews, Clinical Answers and Special Collections
11	Browsed *Highlighted Reviews, Editorials, and Special Collections* from the main website (https://www.cochranelibrary.com/), and reviews from the *Methodology* group

Cochrane Library Syntax

*** = truncation**

:ti = search in the title field

:ti,ab,kw= search in the title, abstract, and keyword field

Association for Computing Machinery (ACM) Digital Library: January 1, 2011, through February 7, 2023

Title search:
(algorithm* OR "machine learning" OR "artificial intelligence") AND (race OR racial* OR ethnic* OR racist OR racism)

(algorithm* OR "machine learning" OR "artificial intelligence") AND bias* AND mitigat*

Abstract searches
(algorithm* OR "machine learning" OR "artificial intelligence") AND (race OR racial* OR ethnic* OR racist OR racism) AND (disparit* OR inequit* OR unequal OR bias*)

(algorithm* OR "machine learning" OR "artificial intelligence") AND (race OR racial* OR ethnic* OR racist OR racism) AND mitigat*

ACM Syntax
**** = truncation***

PubMed Central (PMC): January 1, 2011, through February 7, 2023

1 (algorithm*[Title] OR "artificial intelligence"[Title] OR "machine learning"[Title]) AND (race[Title] OR racial*[Title] OR ethnic*[Title]) AND (bias*[Title] OR prejudice*[Title] OR racist[Title] OR racism[Title] OR disparit*[Title] OR equity[Title] OR unequal*[Title] OR equality[Title] OR inequal*[Title] OR inequit*[Title] OR disproportionat*[Title] OR imbalance*[Title] OR fairness[Title] OR underserved[Title] OR marginalized[Title] OR reclassif*[Title] OR misestimate*[Title] OR misrepresent*[Title] OR "less likely"[Title] OR "more likely"[Title] OR underrepresent*[Title] OR overrepresent*[Title] OR mitigat*[Title])

2 (algorithm*[Abstract] OR "artificial intelligence"[Abstract] OR "machine learning"[Abstract]) AND (race[Abstract] OR racial*[Abstract] OR ethnic*[Abstract]) AND (bias*[Abstract] OR prejudice*[Abstract] OR racist[Abstract] OR racism[Abstract] OR disparit*[Abstract] OR equity[Abstract] OR unequal*[Abstract] OR equality[Abstract] OR inequal*[Abstract] OR inequit*[Abstract] OR disproportionat*[Abstract] OR imbalance*[Abstract] OR fairness[Abstract] OR underserved[Abstract] OR marginalized[Abstract] OR reclassif*[Abstract] OR misestimate*[Abstract] OR misrepresent*[Abstract] OR "less likely"[Abstract] OR "more likely"[Abstract] OR underrepresent*[Abstract] OR overrepresent*[Abstract] OR mitigat*[Abstract]) (332)

3 (algorithm*[Title] OR "artificial intelligence"[Title] OR "machine learning"[Title]) AND (race[Title] OR raci*[Title] OR ethnic*[Title] OR bias*[Title] OR mitigat*[Title])

4 1 OR 2 OR 3

5 4 AND ("2011/01/01"[Publication Date] : "3000"[Publication Date])

6 4 AND ("2011/01/01"[Publication Date] : "3000"[Publication Date]) Filters: MEDLINE journals

7 5 NOT 6

PMC Syntax

**** = truncation***

Nature Digital Medicine (searched via PubMed Central): January 1, 2011, through February 7, 2023

1 (algorithm* OR "artificial intelligence" OR "machine learning") AND (race OR racial*OR racism OR racist* OR ethnic* OR bias* OR mitigat*) AND "2398 6352"[Journal]
2 1 AND ("2011/01/01"[Publication Date] : "3000"[Publication Date])
3 2 NOT (PMC8302667 OR PMC8169744 OR PMC7511400 OR PMC7441407 OR PMC6700078 OR PMC6555808) (265) Note: to remove records from this journal that were already identified in the PubMed In Process search

PMC Syntax
**** = truncation***

BioRxiv: January 1, 2020, through February 7, 2023

Advanced Search, In Title and Abstract Fields:
race algorithm (all words)
racial algorithm (all words)
ethnic algorithm (all words)
ethnicity algorithm (all words)
race bias (all words)
racial bias (all words)
ethnic bias (all words)
ethnicity bias (all words)
artificial intelligence race (all words)
artificial intelligence racial (all words)
artificial intelligence ethnic (all words)
artificial intelligence ethnicity (all words)
machine learning race (all words)
machine learning racial (all words)
machine learning ethnic (all words)
machine learning ethnicity (all words)

MedRxIv (searched using the MedRxIVr tool https://mcguinlu.shinyapps.io/medrxivr/): January 1, 2011, through February 7, 2023

1 ([Rr]ace OR [Rr]acial OR [Rr]acism OR [Rr]acist OR [Ee]thnic OR [Dd]isparit OR [Ii]nequit) AND [Aa]lgorithm
2 ([Rr]ace OR [Rr]acial OR [Rr]acism OR [Rr]acist OR [Ee]thnic OR [Dd]isparit OR [Ii]nequit) AND ([Mm]achine OR [Aa]rtificial) AND ([Ll]earning OR [Ii]ntelligen)
3 1 OR 2

MedRxIVr syntax:
[Aa] = search instances of the word where the first letter is either capitalized or lowercase

ClinicalTrials.gov: January 25, 2022

Search in "other terms"
("machine learning" OR "artificial intelligence" OR algorithm) AND (race OR racial OR racism OR ethnic OR ethnicity) AND (disparity OR disparities OR inequity OR inequities OR unequal OR bias)

Advanced Search:
algorithm [in intervention]; race OR racial OR racism OR ethnic OR ethnicity [in other terms]

Expert Search

(EXPAND[Concept] "machine learning" OR EXPAND[Concept] "machine based learning" OR EXPAND[Concept] "artificial intelligence" OR EXPAND[Concept] "computer aided" OR algorithm) AND (race OR racial OR racism OR ethnic OR ethnicity OR EXPAND[None] "black" OR EXPAND[None] "caucasian" OR EXPAND[None] "African American" OR latino OR Latina OR latinx OR Asian OR indian OR arab OR multiracial OR multi-racial OR minority OR minorities) AND (disparity OR disparities OR inequity OR inequities OR unequal OR bias)

Appendix B. List of Excluded Studies

Key Question 1 Exclusion Reasons

Does not examine a clinical algorithm or algorithm-based tool

Afrose S, Song W, Nemeroff CB, et al. Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction. Commun Med (Lond). 2022 Sep;2:111. doi: 10.1038/s43856-022-00165-w. PMID: 36059892.

Afrose S, Song W, Nemeroff CB, et al. Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction. medRxiv. 2021 Apr:1-48. doi: 10.1101/2021.03.26.21254401.

Akintoye E, Mahmoud K, Shokr M, et al. Racial/ethnic differences in the prognostic utility of left ventricular mass index for incident cardiovascular disease. Clin Cardiol. 2018 Apr;41(4):502-9. doi: 10.1002/clc.22914. PMID: 29663526.

Allen A, Mataraso S, Siefkas A, et al. A racially unbiased, machine learning approach to prediction of mortality: algorithm development study. JMIR Public Health Surveill. 2020 Oct;6(4):e22400. doi: 10.2196/22400. PMID: 33090117.

Al-Mallah MH, Qureshi WT, Keteyian SJ, et al. Racial differences in the prognostic value of cardiorespiratory fitness (results from the Henry Ford Exercise Testing Project). Am J Cardiol. 2016 May;117(9):1449-54. Doi: 10.1016/j.amjcard.2016.02.013. PMID: 26976790.

Awasthi S, Grass GD, Torres-Roca J, et al. Genomic testing in localized prostate cancer can identify subsets of African-Americans with aggressive disease. J Natl Cancer Inst. 2022 Dec;114(12):1656-64. doi: 10.1093/jnci/djac162. PMID: 36053178.

Awasthi S, Mahal BA, Park JY, et al. Substantial Gleason reclassification in Black men with national comprehensive cancer network low-risk prostate cancer – a propensity score analysis. Prostate Cancer Prostatic Dis. 2022 Sep;25(3):547-52. doi: 10.1038/s41391-022-00510-z. PMID: 35194179.

Blaga V, Seth K, Valentim C, et al. Opioid prescription in ophthalmology and the impact of a decision support tool in reducing excess dosing. Am J Ophthalmol. 2022 Nov;243:34-41. doi: 10.1016/j.ajo.2022.06.019. PMID: 35809659.

Burton MJ, Sunesara I, Penman A, et al. Comparing the aspartate aminotransferase (AST) to platelet ratio index (APRI) between African American and White veterans with chronic hepatitis C. South Med J. 2011 May;104(5):309-14. doi: 10.1097/SMJ.0b013e318213cf52. PMID: 21606706.

Castro VM, Apperson WK, Gainer VS, et al. Evaluation of matched control algorithms in EHR-based phenotyping studies: a case study of inflammatory bowel disease comorbidities. J Biomed Inform. 2014 Dec;52:105-11. doi: 10.1016/j.jbi.2014.08.012. PMID: 25196084.

Castro Y, Kendzor DE, Businelle MS, et al. Structural and predictive equivalency of the Wisconsin smoking withdrawal scale across three racial/ethnic groups. Nicotine Tob Res. 2011 Dec;13(7):548-55. doi: 10.1093/ntr/ntr039. PMID: 21454912.

Chapman CH, Schechter CB, Cadham CJ, et al. Identifying equitable screening mammography strategies for Black women in the United States using simulation modeling. Ann Intern Med. 2021 Dec;174(12):1637-46. doi: 10.7326/M20-6506. PMID: 34662151.

Chen Z, Cao B, Edwards A, et al. A deep imputation and inference framework for estimating personalized and race-specific causal effects of genomic alterations on PSA. J Bioinform Comput Biol. 2021 Aug;19(4):2150016. doi: 10.1142/S0219720021500165. PMID: 34225568.

Clark CR, Ommerborn MJ, Moran K, et al. Predicting self-rated health across the life course: health equity insights from machine learning models. J Gen Intern Med. 2021 May;36(5):1181-8. doi: 10.1007/s11606-020-06438-1. PMID: 33620624.

Cohn T, Miller A, Fogg L, et al. Impact of individual and neighborhood factors on cardiovascular risk in White Hispanic and non-Hispanic women and men. Res Nurs Health. 2017 Apr;40(2):120-31. doi: 10.1002/nur.21778. PMID: 27862050.

Coley RY, Johnson E, Simon GE, et al. Racial/ethnic disparities in the performance of prediction models for death by suicide after mental health visits. JAMA Psych. 2021 Jul;78(7):726-34. doi: 10.1001/jamapsychiatry.2021.0493. PMID: 33909019.

Congdon HB, Eldridge BH, Truong HA. Development and implementation of a navigator- facilitated care coordination algorithm to improve clinical outcomes of underserved Latino patients with uncontrolled diabetes. J Health Care Poor Underserved. 2013 Nov;24(4):1604-13. doi: 10.1353/hpu.2013.0181. PMID: 24185156.

Cooper DL, Manago J, Patel V, et al. Universal posttransplant cyclophosphamide after allogeneic transplant, a retrospective single institution study. Leuk Res. 2022 Nov;122:10693. doi: 10.1016/j.leukres.2022.106934. PMID: 36084368.

Coram MA, Candille SI, Duan Q, et al. Leveraging multi-ethnic evidence for mapping complex traits in minority populations: an empirical Bayes approach. Am J Hum Genet. 2015 May;96(5):740-52. doi: 10.1016/j.ajhg.2015.03.008. PMID: 25892113.

Cullen J, Lynch JA, Klein EA, et al. Multicenter comparison of 17-gene genomic prostate score as a predictor of outcomes in African American and Caucasian American men with clinically localized prostate cancer. J Urol. 2021 Apr;205(4):1047-54. doi: 10.1097/JU.0000000000001484. PMID: 33493001.

Devick KL, Valeri L, Chen J, et al. The role of body mass index at diagnosis of colorectal cancer on Black-White disparities in survival: a density regression mediation approach. Biostatistics. 2022 Apr;23(2):449-66. doi: 10.1093/biostatistics/kxaa034. PMID: 32968805.

Dixon SE, Haas SA, Klopp A, et al. A quality improvement project: using the STOP-BANG tool in a military population to improve equity in preoperative screening. J Perianesth Nurs. 2016 Oct;31(5):371-80. doi: 10.1016/j.jopan.2014.12.002. PMID: 27667343.

Dutta A, Ikwuezunma A, Castellanos MI, et al. An evidence-based, risk-adapted algorithm for antifungal prophylaxis reduces risk for invasive mold infections in children with hematologic malignancies. Pediatr Blood Cancer. 2021 Dec;68(12):e29228. doi: 10.1002/pbc.29228. PMID: 34268879.

Fremont A, Weissman JS, Hoch E, et al. When race/ethnicity data are lacking: using advanced indirect estimation methods to measure disparities. Rand Health Q. 2016 Jun;6(1):16. doi: 10.7249/RR1162. PMID: 28083444.

Gadrey SM, Mohanty P, Haughey SP, et al. Overt and occult hypoxemia in patients hospitalized with novel coronavirus disease 2019 [Preprint]. medRxiv. 2022 Jun. doi: 10.1101/2022.06.14.22276166. PMID: 35734082.

Gallifant J, Zhang J, Del Pilar Arias Lopez M, et al. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. Br J Anaesth. 2022 Feb;128(2):343-51. doi: 10.1016/j.bja.2021.09.025. PMID: 34772497.

Gao Y, Cui Y. Deep transfer learning provides a Pareto improvement for multi-ancestral clinico-genomic prediction of diseases [Preprint]. bioRxiv. 2022 Nov. doi: 10.1101/2022.09.22.509055.

Gao Y, Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nat Commun. 2020 Oct;11(1):5131. doi: 10.1038/s41467-020-18918-3. PMID: 33046699.

Gianattasio KZ, Ciarleglio A, Power MC. Development of algorithmic dementia ascertainment for racial/ethnic disparities research in the US Health and Retirement Study. Epidemiology. 2020 Jan;31(1):126-33. doi: 10.1097/EDE.0000000000001101. PMID: 31567393.

Gim J, An J, Sung J, et al. A between ethnicities comparison of chronic obstructive pulmonary disease genetic risk. Front Genet. 2020;11:329. doi: 10.3389/fgene.2020.00329. PMID: 32373161.

Goldie SJ, Daniels N. Model-based analyses to compare health and economic outcomes of cancer control: inclusion of disparities. J Natl Cancer Inst. 2011 Sep;103(18):1373-86. doi: 10.1093/jnci/djr303. PMID: 21900120.

Goldstein JR, Atherwood S. Improved measurement of racial/ethnic disparities in COVID-19 mortality in the United States. medRxiv. 2020 Jun:[Preprint]. doi: 10.1101/2020.05.21.20109116. PMID: 32511557.

Gu T, Han Y, Duan R. A transfer learning approach based on random forest with application to breast cancer prediction in underrepresented populations. Pac Symp Biocomput. 2023;28:186-97. doi: 10.1142/9789811270611_0018. PMID: 36540976.

Gulbahce HE, White S, Herget KA, et al. 21-gene recurrence score testing utilization among older women from different races: a population-based study. J Geriatr Oncol. 2021 Mar;12(2):206-11. doi: 10.1016/j.jgo.2020.06.004. PMID: 32646620.

Gurinovich A, Bae H, Farrell JJ, et al. PopCluster: an algorithm to identify genetic variants with ethnicity-dependent effects. Bioinformatics. 2019 Sep;35(17):3046-54. doi: 10.1093/bioinformatics/btz017. PMID: 30624692.

Gurka MJ, Lilly CL, Oliver MN, et al. An examination of sex and racial/ethnic differences in the metabolic syndrome among adults: a confirmatory factor analysis and a resulting continuous severity score. Metabolism. 2014 Feb;63(2):218-25. doi: 10.1016/j.metabol.2013.10.006. PMID: 24290837.

Hall EC, Massie AB, James NT, et al. Effect of eliminating priority points for HLA-B matching on racial disparities in kidney transplant rates. American Journal of Kidney Diseases. 2011 Nov;58(5):813-6. doi: 10.1053/j.ajkd.2011.05.023. PMID: 21802805.

Han JH, Bilker WB, Edelstein PH, et al. Derivation and validation of clinical prediction rules for reduced vancomycin susceptibility in staphylococcus aureus bacteraemia. Epidemiol Infect. 2013 Jan;141(1):165-73. doi: 10.1017/S0950268812000295. PMID: 22490228.

Han Y, Miao ZF, Lian M, et al. Racial and ethnic disparities in 21-gene recurrence scores, chemotherapy, and survival among women with hormone receptor-positive, node-negative breast cancer. Breast Cancer Res Treat. 2020 Dec;184(3):915-25. doi: 10.1007/s10549-020-05902-0. PMID: 32929567.

Hernandez SE, Sylling PW, Mor MK, et al. Developing an algorithm for combining race and ethnicity data sources in the Veterans Health Administration. Mil Med. 2020 Apr;185(3):e495-e500. doi: 10.1093/milmed/usz322. PMID: 31603222.

Hernandez W, Gamazon ER, Aquino-Michaels K, et al. Integrated analysis of genetic variation and gene expression reveals novel variant for increased warfarin dose requirement in African Americans. J Thromb Haemost. 2017 Apr;15(4):735-43. doi: 10.1111/jth.13639. PMID: 28135054.

Hivert MF, Christophi CA, Jablonski KA, et al. Genetic ancestry markers and difference in A1c between African American and White in the diabetes prevention program. J Clin Endocrinol Metab. 2018 Nov;104(2):328-36. doi: 10.1210/jc.2018-01416. PMID: 30358859.

Hsia DS, Rasouli N, Pittas AG, et al. Implications of the hemoglobin glycation index on the diagnosis of prediabetes and diabetes. J Clin Endocrinol Metab. 2020 Mar;105(3):e130-e8. doi: 10.1210/clinem/dgaa029. PMID: 31965161.

Homkrailas P, Bunnapradist S. Association between ethnicity and kidney transplant waitlist outcomes beyond estimated post-transplant survival score. Transpl Int. 2021 Oct;34(10):1837-44. doi: 10.1111/tri.13965. PMID: 34192375.

Hu H, Huff CD, Yamamura Y, et al. The relationship between Native American ancestry, body mass index and diabetes risk among Mexican-Americans. PLoS ONE. 2015 Oct;10(10):e0141260. doi: 10.1371/journal.pone.0141260. PMID: 26501420.

Hu T, Mortensen K, Chen J. Medicaid managed care in Florida and racial and ethnic disparities in preventable emergency department visits. Med Care. 2018;56(6):477-83. doi: 10.1097/MLR.0000000000000909. PMID: 29629922.

Hussain A, Virani SS, Zheng L, et al. Potential impact of 2017 American College of Cardiology/American Heart Association hypertension guideline on contemporary practice: a cross-sectional analysis from NCDR PINNACLE registry. J Am Heart Assoc. 2022 Jun;11(11):e024107. doi: 10.1161/JAHA.121.024107. PMID: 35656989.

Inker LA, Shafi T, Okparavero A, et al. Effects of race and sex on measured GFR: the multi-ethnic study of atherosclerosis. Am J Kidney Dis. 2016 Nov;68(5):743-51. doi: 10.1053/j.ajkd.2016.06.021. PMID: 20160831039, 27555103.

Jialal I, Remaley AT, Adams-Huet B. The triglyceride-waist circumference index is a valid biomarker of metabolic syndrome in African Americans. Am J Med Sci. 2023;365(2):184-8. doi: 10.1016/j.amjms.2022.11.003. PMID: 36435217.

Jiang X, Morgenstern LB, Cigolle CT, et al. Multiple chronic conditions explain ethnic differences in functional outcome among patients with ischemic stroke. Stroke. 2022 Jan;53(1):120-7. doi: 10.1161/STROKEAHA.120.032595. PMID: 34517767.

Johnston V, Bao Y. Race/ethnicity-related and payer-related disparities in the timeliness of emergency care in U.S. emergency departments. J Health Care Poor Underserved. 2011 May;22(2):606-20. doi: 10.1353/hpu.2011.0050. PMID: 21551937.

Kaspers M, Llamocca E, Quick A, et al. Black and Hispanic women are less likely than white women to receive guideline-concordant endometrial cancer treatment. Am J Obstet Gynecol. 2020 Sep;223(3):398.e1-.e18. doi: 10.1016/j.ajog.2020.02.041. PMID: 32142825.

Kasturi SN, Park J, Wild D, et al. Predicting COVID-19-related health care resource utilization across a statewide patient population: model development study. J Med Internet Res. 2021 Nov;23(11):e31337. doi: 10.2196/31337. PMID: 34581671.

Kay DM, Langfelder-Schwind E, Decelie-Germana J, et al. Utility of a very high IRT/No mutation referral category in cystic fibrosis newborn screening. Pediatr Pulmonol. 2015 Aug;50(8):771-80. doi: 10.1002/ppul.23222. PMID: 26098992.

Kim JS, Gao X, Rzhetsky A. RIDDLE: Race and ethnicity Imputation from Disease history with Deep LEarning. PLoS Comput Biol. 2018;14(4):e1006106. doi: 10.1371/journal.pcbi.1006106. PMID: 29698408.

Krissberg J, Kaufmann M, Gupta A, et al. Racial disparities in pediatric kidney transplantation under the new kidney allocation system in the United States. Clin J Am Soc Nephrol. 2021 Oct;16(12):1862-71. doi: 10.2215/CJN.06740521. PMID: 34670797.

Leonard SA, Main EK, Lyell DJ, et al. Obstetric comorbidity scores and disparities in severe maternal morbidity across marginalized groups. Am J Obstet Gynecol MFM. 2021 Nov;4(2):100530. doi: 10.1016/j.ajogmf.2021.100530. PMID: 34798329.

Li J, Bzdok D, Chen J, et al. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. Sci Adv. 2022 Mar;8(11):eabj1812. doi: 10.1126/sciadv.abj1812. PMID: 35294251.

Liu R, Li X, Zhang W, et al. Comparison of nine statistical model based warfarin pharmacogenetic dosing algorithms using the racially diverse international warfarin pharmacogenetic consortium cohort database. PLoS ONE. 2015 Aug;10(8):e0135784. doi: 10.1371/journal.pone.0135784. PMID: 26305568.

Liu S, Pan H, Xia J, et al. Bridging continual reassessment method for phase I clinical trials in different ethnic populations. Stat Med. 2015 May;34(10):1681-94. doi: 10.1002/sim.6442. PMID: 25626429.

Lynch JA, Berse B, Coomer N, et al. 21-gene recurrence score testing among Medicare beneficiaries with breast cancer in 2010-2013. Genet Med. 2017 Oct;19(10):1134-43. doi: 10.1038/gim.2017.19. PMID: 28333918.

Marini S, Lena UK, Crawford KM, et al. Comparison of genetic and self-identified ancestry in modeling intracerebral hemorrhage risk. Front Neurol. 2018;9:514. doi: doi.org/10.3389/fneur.2018.00514. PMID: 30034361.

Maziarz M, Black RA, Fong CT, et al. Evaluating risk of ESRD in the urban poor. J Am Soc Nephrol. 2015 Jun;26(6):1434-42. doi: 10.1681/ASN.2014060546. PMID: 25475746.

McCall CJ, DeCaprio D, Gartner J. The measurement and mitigation of algorithmic bias and unfairness in healthcare AI models developed for the CMS AI health outcomes challenge [Preprint]. medRxiv. 2022 Oct. doi: 10.1101/2022.09.29.22280537.

.McCoy D, Mgbara W, Horvitz N, et al. Ensemble machine learning of factors influencing COVID-19 across US counties. Sci Rep. 2021 Jun;11(1):11777. doi: 10.1038/s41598-021-90827-x. PMID: 34083563.

Miller V. COVID-19 surveillance and Black American substance use disorder: an examination of data and policy. J Subst Abuse Treat. 2021 Apr;123:108243. doi: 10.1016/j.jsat.2020.108243. PMID: 33612203.

Nayan M, Salari K, Bozzo A, et al. Predicting survival after radical prostatectomy: variation of machine learning performance by race. Prostate. 2021 Dec;81(16):1355-64. doi: 10.1002/pros.24233. PMID: 34529282.

Noseworthy PA, Attia ZI, Brewer LC, et al. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. Circ Arrhythm Electrophysiol. 2020 Mar;13(3):e007988. doi: 10.1161/CIRCEP.119.007988. PMID: 32064914.

Okwuosa TM, Greenland P, Ning H, et al. Distribution of coronary artery calcium scores by Framingham 10-year risk strata in the MESA (multi-ethnic study of atherosclerosis): potential implications for coronary risk assessment. J Am Coll Cardiol. 2011 May;57(18):1838-45. doi: 10.1016/j.jacc.2010.11.053. PMID: 21527159.

Patel A, García-Closas M, Olshan AF, et al. Gene-level germline contributions to clinical risk of recurrence scores in black and white patients with breast cancer. Cancer Res. 2022 Jan;82(1):25-35. doi: 10.1158/0008-5472.CAN-21-1207. PMID: 34711612.

Payne NR, Puumala SE. Racial disparities in ordering laboratory and radiology tests for pediatric patients in the emergency department. Pediatr Emerg Care. 2013 May;29(5):598-606. doi: 10.1097/PEC.0b013e31828e6489. PMID: 23603649.

Perez Alday EA, Rad AB, Reyna MA, et al. Age, sex and race bias in automated arrhythmia detectors. J Electrocardiol. 2022 Sep-Oct;74:5-9. doi: 10.1016/j.jelectrocard.2022.07.007. PMID: 35878534.

Pressman A, Jacobson A, Eguilos R, et al. Prevalence of migraine in a diverse community - electronic methods for migraine ascertainment in a large integrated health plan. Cephalalgia. 2016 Jul;36(4):325-34. doi: 10.1177/0333102415590242. PMID: 26069243.

Pressman AR, Lockhart SH, Shen Z, et al. Measuring and promoting SARS-CoV-2 vaccine equity: development of a COVID-19 vaccine equity index. Health Equity. 2021 Jul;5(1):476-83. doi: 10.1089/heq.2021.0047. PMID: 34316531.

Rappaport D, Chuu A, Hullett C, et al. Assessment of alcohol withdrawal in native American patients utilizing the clinical institute withdrawal assessment of alcohol revised scale. J Addict Med. 2013 May-Jun;7(3):196-9. doi: 10.1097/ADM.0b013e31828b3cc3. PMID: 23579238.

Rayford W, Greenberger M, Bradley RV. Improving risk stratification in a community-based African American population using cell cycle progression score. Transl Androl Urol. 2018 Sep;7(Suppl 4):S384-S91. doi: 10.21037/tau.2018.03.09. PMID: 30363476.

Reuland CP, Collins J, Chiang L, et al. Oregon's approach to leveraging system-level data to guide a social determinants of health-informed approach to children's healthcare. BMJ Innovations. 2021 Jan;7(1):18-25. doi: 10.1136/bmjinnov-2020-000452.

Röösli E, Bozkurt S, Hernandez-Boussard T. Peeking into a black box, the fairness and generalizability of a MIMIC-III benchmarking model. Sci Data. 2022 Jan;9(1):24. doi: 10.1038/s41597-021-01110-7. PMID: 35075160.

Rutkowski R, Salemi J, Tanner JP, et al. Assessing the impact of different race-bridging algorithms on the reported rate of birth defects. J Registry Manag. 2017 Dec;44(4):146-56. PMID: 30133431.

Salahuddin M, Mandell DJ, Lakey DL, et al. Maternal comorbidity index and severe maternal morbidity during delivery hospitalizations in Texas, 2011-2014. Birth. 2020 Mar;47(1):89-97. doi: 10.1111/birt.12465. PMID: 31659788.

Santamaria-Barria JA, Graff-Baker AN, Chang SC, et al. Disparities in the impact of the AJCC 8th edition staging system on differentiated thyroid cancer outcomes. Head Neck. 2022 Oct;44(10):2129-41. doi: 10.1002/hed.27122. PMID: 35766292.

Schindler SE, Karikari TK, Ashton NJ, et al. Effect of race on prediction of brain amyloidosis by plasma Aß42/Aß40, phosphorylated tau, and neurofilament light. Neurology. 2022 Jul;99(3):e245-e57. doi: 10.1212/WNL.0000000000200358. PMID: 35450967.

Schultz D, Lovejoy S, Peet E. Tackling persistent and large disparities in birth outcomes in Allegheny County, Pennsylvania. Matern Child Health J. 2022 May;26(5):978-84. doi: 10.1007/s10995-021-03289-y. PMID: 34982343.

Segar MW, Jaeger BC, Patel KV, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multi-cohort analysis. Circulation. 2021 Jun;143(24):2370-83. doi: 10.1161/CIRCULATIONAHA.120.053134. PMID: 33845593.

Shachar BZ, Mayo JA, Lee HC, et al. Effects of race/ethnicity and BMI on the association between height and risk for spontaneous preterm birth. Am J Obstet Gynecol. 2015 Nov;213(5):700e1-e9. doi: 10.1016/j.ajog.2015.07.005. PMID: 26187451.

Shah A, Polascik TJ, George DJ, et al. Implementation and impact of a risk-stratified prostate cancer screening algorithm as a clinical decision support tool in a primary care network. J Gen Intern Med. 2021 Jan;36(1):92-9. doi: 10.1007/s11606-020-06124-2. PMID: 32875501.

Shin J, Cao D. Comparison of warfarin pharmacogenetic dosing algorithms in a racially diverse large cohort. Pharmacogenomics. 2011 Jan;12(1):125-34. doi: 10.2217/pgs.10.168. PMID: 21174627.

Shulman E, Aagaard P, Kargoli F, et al. Validation of PR interval length as a criterion for development of atrial fibrillation in non-Hispanic whites, African Americans and Hispanics. J Electrocardiol. 2015 Jul;48(4):703-9. doi: 10.1016/j.jelectrocard.2015.04.015. PMID: 26025203.

Srivatsan S, Guduguntla V, Young KZ, et al. Clinical versus patient-reported measures of depression in bariatric surgery. Surg Endosc. 2018 Aug;32(8):3683-90. doi: 10.1007/s00464-018-6101-8. PMID: 29435747.

Stevens J, Erber-Oakkar E, Cui Z, et al. Cardiovascular disease risk by assigned treatment using the 2013 and 1998 obesity guidelines. Obesity. 2016 Jul;24(7):1554-60. doi: 10.1002/oby.21496. PMID: 27184463.

Sughayer M, Alaaraj R, Alsughayer A. Applying new Magee equations for predicting the Oncotype Dx recurrence score. Breast Cancer. 2018 Sep;25(5):597-604. doi: 10.1007/s12282-018-0860-x. PMID: 29691722.

Sullivan BA, Hochheimer CJ, Chernyavskiy P, et al. Impact of race on heart rate characteristics monitoring in very low birth weight infants. Pediatr Res. 2023 Jan:Online ahead of print. doi: 10.1038/s41390-023-02470-z. PMID: 36650306.

Taber DJ, Hamedi M, Rodrigue JR, et al. Quantifying the race stratified impact of socioeconomics on graft outcomes in kidney transplant recipients. Transplantation. 2016 Jun;100(7):1550-7. doi: 10.1097/TP.0000000000000931. PMID: 26425875.

Tallaj JA, Pamboukian SV, George JF, et al. Have risk factors for mortality after heart transplantation changed over time? Insights from 19 years of cardiac transplant research database study. J Heart Lung Transplant. 2014 Dec;33(12):1304-11. doi: 10.1016/j.healun.2014.08.014. PMID: 25443871.

Tan-McGrory A, Bennett-AbuAyyash C, Gee S, et al. A patient and family data domain collection framework for identifying disparities in pediatrics: results from the pediatric health equity collaborative. BMC Pediatr. 2018 Jan;18(1):18. doi: 10.1186/s12887-018-0993-2. PMID: 29385988.

Tiwari A, Dadhania AV, Ragunathrao VAB, et al. Using machine learning to develop a novel COVID-19 vulnerability index (C19VI). Sci Total Environ. 2021 Jun;773:145650. doi: 10.1016/j.scitotenv.2021.145650. PMID: 33940747.

Tolksdorf J, Kattan MW, Boorjian SA, et al. Multi-cohort modeling strategies for scalable globally accessible prostate cancer risk tools. BMC Med Res Methodol. 2019 Oct;19(1):191. doi: 10.1186/s12874-019-0839-0. PMID: 31615451.

Toseef M, Li X, Wong KC. Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning. Brief Bioinform. 2022 May;23(3):bbac078. doi: 10.1093/bib/bbac078. PMID: 35323862.

Valera P, McClernon FJ, Burkholder G, et al. A pilot trial examining African American and White responses to algorithm-guided smoking cessation medication selection in persons living with HIV. AIDS Behav. 2017 Jul;21(7):1975-84. doi: 10.1007/s10461-016-1634-0. PMID: 27942999.

Wang AY, Wong MS, Humbyrd CJ. Eligibility criteria for lower extremity joint replacement may worsen racial and socioeconomic disparities. Clin Orthop Relat Res. 2018;476(12):2301-8. doi: 10.1097/CORR.0000000000000511. PMID: 30303879.

Wang J, Qiao Y, Tina Shih YC, et al. Potential health implications of racial and ethnic disparities in meeting MTM eligibility criteria. Res Social Adm Pharm. 2014 Jan;10(1):106-25. doi: 10.1016/j.sapharm.2013.03.007. PMID: 23759673.

Wong RJ, Devaki P, Nguyen L, et al. Increased long-term survival among patients with hepatocellular carcinoma after implementation of model for end-stage liver disease score. Clin Gastroenterol Hepatol. 2014 Sep;12(9):1534-40.e1. doi: 10.1016/j.cgh.2013.12.008. PMID: 24361414.

Worrell FC, Mendoza-Denton R, Wang A. Introducing a new assessment tool for measuring ethnic-racial identity: the cross ethnic-racial identity scale-adult (CERIS-A). Assessment. 2019 Apr;26(3):404-18. doi: 10.1177/1073191117698756. PMID: 29214847.

Yoon S, Davis N, Odlum M, et al. Applying artificial intelligence to predict self-reported poor health among Black and Hispanic caregivers with mild cognitive impairment. Stud Health Technol Inform. 2020 Jun;272:433-6. doi: 10.3233/SHTI200588. PMID: 32604695.

Zavorsky GS, Almamary AS, Alqahtani MK, et al. The need for race-specific reference equations for pulmonary diffusing capacity for nitric oxide. BMC Pulm Med. 2021 Jul;21(1):232. doi: 10.1186/s12890-021-01591-7. PMID: 34256739.

Does not examine the effect of a clinical algorithm or algorithm-based tool on racial/ethnic differences

Arnold SV, Spertus JA, Jones PG, et al. Predicting adverse outcomes after myocardial infarction among patients with diabetes mellitus. Circ Cardiovasc Qual Outcomes. 2016;9(4):372-9. doi: 10.1161/CIRCOUTCOMES.115.002365. PMID: 27220369.

Buckley A, Sestito S, Ogundipe T, et al. Racial and ethnic disparities among women undergoing a trial of labor after cesarean delivery: performance of the VBAC calculator with and without patients' race/ethnicity. Reprod Sci. 2022 Jul;29(7):2030-8. doi: 10.1007/s43032-022-00959-2. PMID: 35534768.

Bundy JD, Rahman M, Matsushita K, et al. Risk prediction models for atherosclerotic cardiovascular disease in patients with chronic kidney disease: the CRIC study. J Am Soc Nephrol. 2022 Mar;33(3):601-11. doi: 10.1681/ASN.2021060747. PMID: 35145041.

Christine PJ, Young R, Adar SD, et al. Individual- and area-level SES in diabetes risk prediction: the multi-ethnic study of atherosclerosis. Am J Prev Med. 2017 Aug;53(2):201-9. doi: 10.1016/j.amepre.2017.04.019. PMID: 28625713.

Cohen JL, Thompson E, Sinvani L, et al. Assessment of warfarin algorithms for hospitalized adults: searching for a safe dosing strategy. J Thromb Thrombolysis. 2019 Nov;48(4):570-9. doi: 10.1007/s11239-019-01902-0. PMID: 31228039.

Collin LJ, Yan M, Jiang R, et al. Oncotype DX recurrence score implications for disparities in chemotherapy and breast cancer mortality in Georgia. npj Breast Cancer. 2019 Sep;5:32. doi: 10.1038/s41523-019-0129-3. PMID: 31583272.

Conran CA, Shi Z, Resurreccion WK, et al. Assessing the clinical utility of genetic risk scores for targeted cancer screening. J Transl Med. 2021 Jan;19(1):41. doi: 10.1186/s12967-020-02699-w. PMID: 33482857.

DeFilippis AP, Young R, Carrubba CJ, et al. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. Ann Intern Med. 2015 Feb;162(4):266-75. doi: 10.7326/M14-1281. PMID: 25686167.

Dipnall JF, Pasco JA, Berk M, et al. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. PLoS ONE. 2016 Feb;11(2):e0148195. doi: 10.1371/journal.pone.0148195. PMID: 26848571.

Eapen ZJ, McCoy LA, Fonarow GC, et al. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. 2015 May;8(3):473-80. doi: 10.1161/CIRCHEARTFAILURE.114.001879. PMID: 25747700.

Gershengorn HB, Patel S, Shukla B, et al. Predictive value of sequential organ failure assessment score across patients with and without COVID-19 infection. Ann Am Thorac Soc. 2022 May;19(5):790-8. doi: 10.1513/AnnalsATS.202106-680OC. PMID: 34784497.

Gobardhan SN, Dimitriu-Leen AC, van Rosendael AR, et al. Prevalence by computed tomographic angiography of coronary plaques in South Asian and White patients with type 2 diabetes mellitus at low and high risk using four cardiovascular risk scores (UKPDS, FRS, ASCVD, and JBS3). Am J Cardiol. 2017 Mar;119(5):705-11. doi: 10.1016/j.amjcard.2016.11.029. PMID: 28024655.

Goodson DA, Chalupsky MR, Wiegley N, et al. GFR estimation in potential living kidney donors: race- and nonrace-based equations and measured GFR. Kidney Medicine. 2022 Dec;4(12):100558. doi: 10.1016/j.xkme.2022.100558. PMID: 36471819.

Harari Y, Harari Y, O'Brien MK, et al. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. J Neuroengineering Rehabil. 2020 Jun;17(1):71. doi: 10.1186/s12984-020-00704-3. PMID: 32522242.

Hoppmann AL, Chen Y, Landier W, et al. Individual prediction of nonadherence to oral mercaptopurine in children with acute lymphoblastic leukemia: results from COG AALL03N1. Cancer. 2021 Oct;127(20):3832-9. doi: 10.1002/cncr.33760. PMID: 34161608.

Inoue Y, Howard AG, Stickley A, et al. Sex and racial/ethnic differences in the association between childhood attention-deficit/hyperactivity disorder symptom subtypes and body mass index in the transition from adolescence to adulthood in the United States. Pediatr Obes. 2019 May;14(5):e12498. doi: 10.1111/ijpo.12498. PMID: 30629806.

Jasem J, Amini A, Rabinovitch R, et al. 21-gene recurrence score assay as a predictor of adjuvant chemotherapy administration for early-stage breast cancer: an analysis of use, therapeutic implications, and disparity profile. J Clin Oncol. 2016 Jun;34(17):1995-2002. doi: 10.1200/JCO.2015.65.0887. PMID: 27001563.

Karmali KN, Goff DC, Ning H, et al. A systematic examination of the 2013 ACC/AHA pooled cohort risk assessment tool for atherosclerotic cardiovascular disease. J Am Coll Cardiol. 2014;64(10):959-68. doi: 10.1016/j.jacc.2014.06.1186. PMID: 25190228.

Keeney BJ, Koenig KM, Paddock NG, et al. Do aggregate socioeconomic status factors predict outcomes for total knee arthroplasty in a rural population? J Arthroplasty. 2017 Dec;32(12):3583-90. doi: 10.1016/j.arth.2017.07.002. PMID: 28781014.

Kostick-Quenet KM, Cohen IG, Gerke S, et al. Mitigating racial bias in machine learning. J Law Med Ethics. 2022;50(1):92-100. doi: 10.1017/jme.2022.13. PMID: 35243993.

Ku KC, Li J, Ha NB, et al. Chronic hepatitis B management based on standard guidelines in community primary care and specialty clinics. Dig Dis Sc. 2013 Dec;58(12):3626-33. doi: 10.1007/s10620-013-2889-1. PMID: 24122622.

Lubelski D, Alentado V, Nowacki AS, et al. Preoperative nomograms predict patient-specific cervical spine surgery clinical and quality of life outcomes. Clin Neurosurg. 2018 Jul;83(1):104-13. doi: 10.1093/neuros/nyx343.

Mahle WT, Simpson SA, Fye P, et al. Management of warfarin in children with heart disease. Pediatr Cardiol. 2011 Dec;32(8):1115-9. doi: 10.1007/s00246-011-9984-x. PMID: 21499856.

Matsushita K, Mahmoodi BK, Woodward M, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. JAMA. 2012 May;307(18):1941-51. doi: 10.1001/jama.2012.3954. PMID: 22570462.

Moss HE, An R, Nelson T, et al. Risk of atherosclerotic cardiovascular disease among US adults: use of 1999-2014 NHANES data. J Prim Prev. 2019 Oct;40(5):569-73. doi: 10.1007/s10935-019-00564-1. PMID: 31571032.

Panchal HJ, Durinka JB, Patterson J, et al. Survival outcomes in liver transplant recipients with model for end-stage liver disease scores of 40 or higher: a decade-long experience. HPB (Oxford). 2015 Dec;17(12):1074-84. doi: 10.1111/hpb.12485. PMID: 26373873.

Pathak P, Panday SB, Ahn J. Artificial neural network model effectively estimates muscle and fat mass using simple demographic and anthropometric measures. Clin Nutr. 2022 Jan;41(1):144-52. doi: 10.1016/j.clnu.2021.11.027. PMID: 34879301.

Pollock BD, Hu T, Chen W, et al. Utility of existing diabetes risk prediction tools for young Black and White adults: evidence from the Bogalusa heart study. J Diabetes Complications. 2017 Jan;31(1):86-93. doi: 10.1016/j.jdiacomp.2016.07.025. PMID: 27503406.

Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012 Dec;40(6):1324-43. doi: 10.1183/09031936.00080312. PMID: 22743675.

Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. J Arthroplasty. 2019 Apr;34(4):632-7. doi: 10.1016/j.arth.2018.12.030. PMID: 30665831.

Razavi AC, Potts KS, Kelly TN, et al. Pooled cohort equations heart failure risk score predicts cardiovascular disease and all-cause mortality in a nationally representative sample of US adults. BMC Cardiovasc Disord. 2020 Apr;20(1):202. doi: 10.1186/s12872-020-01485-2. PMID: 32334524.

Robinson A, Hirode G, Wong RJ. Ethnicity and insurance-specific disparities in the model for end-stage liver disease score at time of liver transplant waitlist registration and its impact on mortality. J Clin Exp Hepatol. 2021 Mar;11(2):188-94. doi: 10.1016/j.jceh.2020.07.011. PMID: 33746443.

Roche C, Kumar V, Overman S, et al. Validation of a machine learning-derived clinical metric to quantify outcomes after total shoulder arthroplasty. J Shoulder Elbow Surg. 2021 Oct;30(10):2211-24. doi: 10.1016/j.jse.2021.01.021. PMID: 33607333.

Schwartz JB, Lai J, Lizaola B, et al. A comparison of measured and calculated free 25(OH) vitamin D levels in clinical populations. Journal of Clinical Endocrinology and Metabolism. 2014 May;99(5):1631-7. doi: 10.1210/jc.2013-3874. PMID: 24483159.

Shin J, Kayser SR. Accuracy of the pharmacogenetic dosing table in the warfarin label in predicting initial therapeutic warfarin doses in a large, racially diverse cohort. Pharmacotherapy. 2011 Sep;31(9):863-70. doi: 10.1592/phco.31.9.863. PMID: 21923587.

Steiner HE, Giles JB, Patterson HK, et al. Machine learning for prediction of stable warfarin dose in US Latinos and Latin Americans. Front Pharmacol. 2021 Oct;12:749786. doi: 10.3389/fphar.2021.749786. PMID: 34776967.

Sung JH, Yeboah J, Lee JE, et al. Diagnostic value of coronary artery calcium score for cardiovascular disease in African Americans: the Jackson Heart Study. Br J Med Med Res. 2016;11(2):BJMMR/2016/21449. doi: 10.9734/BJMMR/2016/21449. PMID: 26949662.

Thurlapati A, Velez-Martinez CS, Hirani S, et al. Do the 2013 United States Preventive Services Task Force guidelines for lung cancer screening fail high-risk African American smokers? An institutional retrospective observational cohort study. Eur J Cancer Prev. 2021 Sep;30(5):375-81. doi: 10.1097/CEJ.0000000000000652. PMID: 34010237.

Vigil JM, Alcock J, Coulombe P, et al. Ethnic disparities in emergency severity index scores among U.S. Veteran's Affairs Emergency Department patients. PLoS ONE. 2015 May;10(5):e0126792. doi: 10.1371/journal.pone.0126792. PMID: 2015154600, 26024515.

Wang Y, Katzmarzyk PT, Horswell R, et al. Comparison of the heart failure risk stratification performance of the CKD-EPI equation and the MDRD equation for estimated glomerular filtration rate in patients with type 2 diabetes. Diabet Med. 2016 May;33(5):609-20. doi: 10.1111/dme.12859. PMID: 26202081.

Warsch JRL, Rundek T, Paik MC, et al. Association between northern Manhattan study global vascular risk score and successful aging. J Am Geriatr Soc. 2013 Apr;61(4):519-24. doi: 10.1111/jgs.12166. PMID: 23527874.

Wiley LK, Vanhouten JP, Samuels DC, et al. Strategies for equitable pharmacogenomic-guided warfarin dosing among European and African American individuals in a clinical population. Pac Symp Biocomput. 2017;22:545-56. doi: 10.1142/9789813207813_0050. PMID: 27897005.

Wuerz TH, Kent DM, Malchau H, et al. A nomogram to predict major complications after hip and knee arthroplasty. J Arthroplasty. 2014 Jul;29(7):1457-62. doi: 10.1016/j.arth.2013.09.007. PMID: 24793891.

Zhou X, Ning Q, Jin K, et al. Development and validation of a preoperative nomogram for predicting survival of patients with locally advanced prostate cancer after radical prostatectomy. BMC Cancer. 2020 Feb;20(1):97. doi: 10.1186/s12885-020-6565-5. PMID: 32019501.

Does not report an outcome of interest (i.e., access to care, quality of care, or health outcomes)

Abate M, Jandovitz N, Hirsch JS, et al. The effect of race coefficients on preemptive listing for kidney transplantation. Clin Kidney J. 2022 May;15(5):942-50. doi: 10.1093/ckj/sfab287.

Abate M, Jandovitz N, Hirsch JS, et al. The effect of race coefficients on preemptive listing for kidney transplantation. Clin Kidney J. 2022 May;15(5):942-50. doi: 10.1093/ckj/sfab287. PMID: 35498880.

Aminsharifi A, Schulman A, Anderson J, et al. Primary care perspective and implementation of a multidisciplinary, institutional prostate cancer screening algorithm embedded in the electronic health record. Urol Oncol. 2018 Nov;36(11):502.e1-.e6. doi: 10.1016/j.urolonc.2018.07.016. PMID: 30170982.

Arnall DA, Nelson AG, Hearon CM. Maximal respiratory pressure reference values for Hopi children ages 4 to 13. Cardiopulm Phys Ther J. 2022 Jul;33(3):123-9. doi: 10.1097/CPT.0000000000000195. PMID: 36090687.

Belouali A, Bai H, Raja K, et al. Impact of social determinants of health on improving the LACE index for 30-day unplanned readmission prediction. JAMIA Open. 2022 Jun;5(2):ooac046. doi: 10.1093/jamiaopen/ooac046. PMID: 35702627.

Bennett R, Mulla ZD, Parikh P, et al. An imbalance-aware deep neural network for early prediction of preeclampsia. PLoS ONE. 2022 Apr;17(4):e0266042. doi: 10.1371/journal.pone.0266042. PMID: 35385525.

Bhattacharya A, García-Closas M, Olshan AF, et al. A framework for transcriptome-wide association studies in breast cancer in diverse study populations. Genome Biol. 2020 Feb;21(1):42. doi: 10.1186/s13059-020-1942-6. PMID: 32079541.

Borgese M, Joyce C, Anderson EE, et al. Bias assessment and correction in machine learning algorithms: a use-case in a natural language processing algorithm to identify hospitalized patients with unhealthy alcohol use. AMIA Annu Symp Proc. 2022 Feb;2021:247-54. https://knowledge.amia.org/74229-amia-1.4622266/t003-1.4626466/t003-1.4626467/3575666-1.4626789/3575666-1.4626790?qr=1. PMID: 35308909.

Bose A, O'Neal WT, Bennett A, et al. Relation between estimated cardiorespiratory fitness and atrial fibrillation (from the reasons for geographic and racial differences in stroke study). Am J Cardiol. 2017 Jun;119(11):1776-80. doi: 10.1016/j.amjcard.2017.03.008. PMID: 28390681.

Bowerman C, Bhatka NR, Brazzale D, et al. A race-neutral approach to the interpretation of lung function measurements. Am J Respir Crit Care Med. 2023 Mar;207(6):768-74. doi: 10.1164/rccm.202205-0963OC. PMID: 36383197.

Butt JH, Adamson C, Docherty KF, et al. Eligibility for pharmacological therapies in heart failure with reduced ejection fraction: implications of the new chronic kidney disease Epidemiology Collaboration creatinine equation for estimating glomerular filtration rate. Eur J Heart Fail. 2022 May;24(5):861-6. doi: 10.1002/ejhf.2460. PMID: 35199418.

Chase EC, Bryant AK, Sun Y, et al. Development and validation of a life expectancy calculator for US patients with prostate cancer. BJU Int. 2022 Oct;130(4):496-506. doi: 10.1111/bju.15740. PMID: 35373440.

Chen LY, Norby FL, Chamberlain AM, et al. CHA_2DS_2-VASc score and stroke prediction in atrial fibrillation in whites, blacks, and hispanics. Stroke. 2019;50(1):28-33. doi: 10.1161/STROKEAHA.118.021453. PMID: 30580712.

Chen Q, Ayer T, Nastoupil LJ, et al. Population-specific prognostic models are needed to stratify outcomes for African-Americans with diffuse large B-cell lymphoma. Leuk Lymphoma. 2016 Apr;57(4):842-51. doi: 10.3109/10428194.2015.1083098. PMID: 26415108.

Deng J, Vozmediano V, Rodriguez M, et al. Genotype-guided dosing of warfarin through modeling and simulation. Eur J Pharm Sci. 2017 Nov;109S:S9-S14. doi: 10.1016/j.ejps.2017.05.017. PMID: 28502675.

Dennis JA. Racial/ethnic disparities in triage scores among pediatric emergency department fever patients. Pediatr Emerg Care. 2021 Dec;37(12):e1457-e61. doi: 10.1097/PEC.0000000000002072. PMID: 32150002.

Echouffo-Tcheugui JB, Greene SJ, Papadimitriou L, et al. Population risk prediction models for incident heart failure: a systematic review. Circ Heart Fail. 2015 May;8(3):438-47. doi: 10.1161/CIRCHEARTFAILURE.114.001896. PMID: 25737496.

Ekström M, Mannino D. Research race-specific reference values and lung function impairment, breathlessness and prognosis: analysis of NHANES 2007–2012. Respir Res. 2022 Oct;23(1):271. doi: 10.1186/s12931-022-02194-4. PMID: 36182912.

Fiscella K, Winters P, Farah S, et al. Do lung cancer eligibility criteria align with risk among blacks and hispanics? PLoS ONE. 2015 Nov;10(11):e0143789. doi: 10.1371/journal.pone.0143789. PMID: 26618478.

Fudim M, Zalawadiya S, Patel DK, et al. Data on coronary artery calcium score performance and cardiovascular risk reclassification across gender and ethnicities. Data Brief. 2016 Mar;6:578-81. doi: 10.1016/j.dib.2016.01.002. PMID: 26909370.

Guo D, Wang C, Wang B, et al. Learning fair representations via distance correlation minimization. IEEE Trans Neural Netw Learn Syst. 2022 Aug:Online ahead of print. doi: 10.1109/TNNLS.2022.3187165. PMID: 35969542.

Hong C, Pencina MJ, Wojdyla DM, et al. Predictive accuracy of stroke risk prediction models across Black and White race, sex, and age groups. JAMA. 2023 Jan;329(4):306-17. doi: 10.1001/jama.2022.24683. PMID: 36692561.

Howell CR, Zhang L, Yi N, et al. Race versus social determinants of health in COVID-19 hospitalization prediction. Am J Prev Med. 2022 Jul;63(1 Suppl 1):S103-S8. doi: 10.1016/j.amepre.2022.01.034. PMID: 35725136.

Inker LA, Levey AS, Tighiouart H, et al. Performance of glomerular filtration rate estimating equations in a community-based sample of Blacks and Whites: the multiethnic study of atherosclerosis. Nephrol Dial Transplant. 2018 Mar;33(3):417-25. doi: 10.1093/ndt/gfx042. PMID: 28505377.

Ku E, Amaral S, McCulloch CE, et al. Comparison of 2021 CKD-EPI equations for estimating racial differences in preemptive waitlisting for kidney transplantation. Clin J Am Soc Nephrol. 2022 Oct;17(10):1515-21. doi: 10.2215/CJN.04850422. PMID: 36122938.

Kurian AW, Hughes E, Simmons T, et al. Performance of the IBIS/Tyrer-Cuzick model of breast cancer risk by race and ethnicity in the women's health initiative. Cancer. 2021 Oct;127(20):3742-50. doi: 10.1002/cncr.33767. PMID: 34228814.

Kowlgi GN, Gunda S, Padala SK, et al. Comparison of frequency of atrial fibrillation in Blacks versus Whites and the utilization of race in a novel risk score. Am J Cardiol. 2020 Nov;135:68-76. doi: 10.1016/j.amjcard.2020.08.029. PMID: 32866451.

Li F, Wu P, Ong HH, et al. Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction. J Biomed Inform. 2023 Feb;138:104294. doi: 10.1016/j.jbi.2023.104294. PMID: 36706849.

Lin YC, Mallia D, Clark-Sevilla AO, et al. Preeclampsia predictor with machine learning: a comprehensive and bias-free machine learning pipeline [Preprint]. medRxiv. 2022 Jun. doi: 10.1101/2022.06.08.22276107.

3Loomba RS, Raskin A, Gudausky TM, et al. Role of the Egami score in predicting intravenous immunoglobulin resistance in Kawasaki disease among different ethnicities. Am J Ther. 2016 Nov;23(6):e1293-e9. doi: 10.1097/MJT.0000000000000045. PMID: 25611359.

Márquez-Luna C, Loh PR, Price AL, et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol. 2017 Dec;41(8):811-23. doi: 10.1002/gepi.22083. PMID: 29110330.

McCarthy AM, Liu Y, Ehsan S, et al. Validation of breast cancer risk models by race/ethnicity, family history and molecular subtypes. Cancers. 2021 Dec;14(1):45. doi: 10.3390/cancers14010045. PMID: 35008209.

Mehta A, Pandey A, Ayers CR, et al. Predictive value of coronary artery calcium score categories for coronary events versus strokes: impact of sex and race: MESA and DHS. Circ Cardiovasc Imaging. 2020 Aug;13(8):e010153. doi: 10.1161/CIRCIMAGING.119.010153. PMID: 32806939.

Mercado CI, Cogswell ME, Loria CM, et al. Validity of predictive equations for 24-h urinary potassium excretion based on timing of spot urine collection among adults: the MESA and CARDIA Urinary Sodium Study and NHANES Urinary Sodium Calibration Study. Am J Clin Nutr. 2018 Sep;108(3):532-47. doi: 10.1093/ajcn/nqy138. PMID: 30535091.

Mifsud F, Saint-Martin C, Dubois-Laforgue D, et al. Monogenic diabetes in adults: a multi-ancestry study reveals strong disparities in diagnosis rates and clinical presentation. Diabetes Res Clin Pract. 2022 May;188:109908. doi: 10.1016/j.diabres.2022.109908. PMID: 35533745.

Ng DK, Furth SL, Warady BA, et al. Self-reported race, serum creatinine, cystatin C, and GFR in children and young adults with pediatric kidney diseases: a report from the chronic kidney disease in children (CKiD) study. Am J Kidney Dis. 2022 Aug;80(2):174-85.e1. doi: 10.1053/j.ajkd.2021.10.013. PMID: 34974031.

Oliveira N, Doyle LE, Atlas RO, et al. External validity of first-trimester algorithms in the prediction of pre-eclampsia disease severity. Ultrasound Obstet Gynecol. 2014 Sep;44(3):286-92. doi: 10.1002/uog.13433. PMID: 24912952.

Orandi BJ, Kumar V, Reed RD, et al. Reclassification of CKD in living kidney donors with the refitted race-free eGFR formula. Am J Surg. 2023 Feb;225(2):425-8. doi: 10.1016/j.amjsurg.2022.09.024. PMID: 36167624.

Park J, Artin MG, Lee KE, et al. Deep learning on time series laboratory test results from electronic health records for early detection of pancreatic cancer. J Biomed Inform. 2022 Jul;131:104095. doi: 10.1016/j.jbi.2022.104095. PMID: 35598881.

Park JI, Bozkurt S, Park JW, et al. Evaluation of race/ethnicity-specific survival machine learning models for Hispanic and Black patients with breast cancer. BMJ Health Care Inform. 2023 Jan;30(1):e100666. doi: 10.1136/bmjhci-2022-100666. PMID: 36653067.

Pasquinelli MM, Tammemägi MC, Kovitz KL, et al. Risk prediction model versus United States Preventive Services Task Force lung cancer screening eligibility criteria: reducing race disparities. J Thorac Oncol. 2020 Nov;15(11):1738-47. doi: 10.1016/j.jtho.2020.08.006. PMID: 32822843.

Philibert R, Long JD, Mills JA, et al. A simple, rapid, interpretable, actionable and implementable digital PCR based mortality index. Epigenetics. 2021 Oct;16(10):1135-49. doi: 10.1080/15592294.2020.1841874. PMID: 33138668.

Porterhouse MD, Paul S, Lieberenz JL, et al. Black women are less likely to be classified as high-risk for breast cancer using the Tyrer-Cuzick 8 model. Ann Surg Oncol. 2022 Oct;29(10):6419-25. doi: 10.1245/s10434-022-12140-9. PMID: 35790586.

Poventud-Fuentes I, Garnett E, Akcan-Arikan A, et al. Comparison of cystatin C and creatinine-based equations with measured glomerular filtration rate in a diverse pediatric population. J Appl Lab Med. 2022 Sep;7(5):1016-24. doi: 10.1093/jalm/jfac043. PMID: 35671191.

Radjef R, Peterson EL, Michaels A, et al. Performance of the meta-analysis global group in chronic heart failure score in black patients compared with whites. Circ Cardiovasc Qual Outcomes. 2019 Jul;12(7):e004714. doi: 10.1161/CIRCOUTCOMES.118.004714. PMID: 31266369.

Reeves M, Bhat HS, Goldman-Mellor S. Resampling to address inequities in predictive modeling of suicide deaths. BMJ Health Care Inform. 2022 Apr;29(1):e100456. doi: 10.1136/bmjhci-2021-100456. PMID: 35396246.

Rocco MV, Chapman A, Chertow GM, et al. Chronic kidney disease classification in systolic blood pressure intervention trial: comparison using modification of diet in renal disease and CKD-epidemiology collaboration definitions. Am J Nephrol. 2016 Sep;44(2):130-40. doi: 10.1159/000448722. PMID: 27513312.

Sajal IH, Chowdhury M, Wang T, et al. CBCRisk-Black: a personalized contralateral breast cancer risk prediction model for black women. Breast Cancer Res Treat. 2022 Jul;194(1):179-86. doi: 10.1007/s10549-022-06612-5. PMID: 35562619.

Schaich CL, Yeboah J, Espeland MA, et al. Association of vascular risk scores and cognitive performance in a diverse cohort: the multi-ethnic study of atherosclerosis. J Gerontol A Biol Sci Med Sci. 2022 Jun;77(6):1208-15. doi: 10.1093/gerona/glab189. PMID: 34216214.

Segar MW, Hall JL, Jhund PS, et al. Machine learning-based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with heart failure. JAMA Cardiol. 2022 Aug;7(8):844-54. doi: 10.1001/jamacardio.2022.1900. PMID: 35793094.

Shaikh N, Lee M, 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 Jun;176(6):569-75. doi: 10.1001/jamapediatrics.2022.0700. PMID: 35435935.

Singh K, Valley TS, Tang S, et al. Evaluating a widely implemented proprietary deterioration index model among hospitalized patients with COVID-19. Ann Am Thorac Soc. 2021 Jul;18(7):1129-37. doi: 10.1513/AnnalsATS.202006-698OC. PMID: 33357088.

Starlard-Davenport A, Allman R, Dite GS, et al. Validation of a genetic risk score for Arkansas women of color. PLoS ONE. 2018 Oct;13(10):e0204834. doi: 10.1371/journal.pone.0204834. PMID: 30281645.

Stevens LA, Claybon MA, Schmid CH, et al. Evaluation of the chronic kidney disease epidemiology collaboration equation for estimating the glomerular filtration rate in multiple ethnicities. Kidney Int. 2011 Mar;79(5):555-62. doi: 10.1038/ki.2010.462. PMID: 21107446.

Syn NLX, Lee SC, Brunham LR, et al. Pharmacogenetic versus clinical dosing of warfarin in individuals of Chinese and African-American ancestry: assessment using data simulation. Pharmacogenet Genomics. 2015 Sep;25(10):491-500. doi: 10.1097/FPC.0000000000000165. PMID: 26230382.

Tammemagi CM, Pinsky PF, Caporaso NE, et al. Lung cancer risk prediction: prostate, lung, colorectal and ovarian cancer screening trial models and validation. J Natl Cancer Inst. 2011 Jul;103(13):1058-68. doi: 10.1093/jnci/djr173. PMID: 21606442.

Vasan RS, van den Heuvel E. Differences in estimates for 10-year risk of cardiovascular disease in Black versus White individuals with identical risk factor profiles using pooled cohort equations: an in silico cohort study. Lancet Digital Health. 2022 Jan;4(1):e55-e63. doi: doi.org/10.1016/S2589-7500(21)00236-3. PMID: 34952676.

Wang D, Willis DR, Yih Y. The pneumonia severity index: assessment and comparison to popular machine learning classifiers [Preprint]. medRxiv. 2021 Dec. doi: 10.1101/2021.12.06.21267390.

Weissman GE, Teeple S, Eneanya ND, et al. Effects of neighborhood-level data on performance and algorithmic equity of a model that predicts 30-day heart failure readmissions at an urban academic medical center. J Card Fail. 2021 Sep;27(9):965-73. doi: 10.1016/j.cardfail.2021.04.021. PMID: 34048918.

Williams RM, Kareff SA, Sackstein P, et al. Race & sex disparities related to low-dose computed tomography lung cancer screening eligibility criteria: a lung cancer cases review. Lung Cancer. 2022 Jul;169:55-60. doi: 10.1016/j.lungcan.2022.05.008. PMID: 35644087.

Witonsky J, Elhawary JR, Eng C, et al. Race- and ethnicity-based spirometry reference equations: are they accurate for genetically admixed children? Chest. 2022 Jul;162(1):184-95. doi: 10.1016/j.chest.2021.12.664. PMID: 35033507.

Wu Q, Xiao X, Xu Y. Performance of FRAX in predicting fractures in US postmenopausal women with varied race and genetic profiles. J Clin Med. 2020 Jan;9(1):285. doi: 10.3390/jcm9010285. PMID: 31968614.

Zhang L, Zhang Z, Zhang Y, et al. Evaluation of Finnish diabetes risk score in screening undiagnosed diabetes and prediabetes among U.S. adults by gender and race: NHANES 1999-2010. PLoS ONE. 2014 May;9(5):e97865. doi: 10.1371/journal.pone.0097865. PMID: 24852786.

Zilberberg MD, Chaudhari P, Nathanson BH, et al. Development and validation of a bedside risk score for MRSA among patients hospitalized with complicated skin and skin structure infections. BMC Infect Dis. 2012 Jul;12:154. doi: 10.1186/1471-2334-12-154. PMID: 22784260.

Zook HG, Kharbanda AB, Flood A, et al. Racial differences in pediatric emergency department triage scores. J Emerg Med. 2016 May;50(5):720-6. doi: 10.1016/j.jemermed.2015.02.056. PMID: 26899520.

Does not report outcomes by race/ethnicity

Fan L, Levey AS, Gudnason V, et al. Comparing GFR estimating equations using cystatin C and creatinine in elderly individuals. J Am Soc Nephrol. 2015 Aug;26(8):1982-9. doi: 10.1681/ASN.2014060607. PMID: 25527647.

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-.e7. doi: 10.1016/j.ajog.2021.05.021. PMID: 34043983.

Raghu VK, Walia AS, Zinzuwadia AN, et al. Validation of a deep learning-based model to predict lung cancer risk using chest radiographs and electronic medical record data. JAMA Netw Open. 2022 Dec;5(12):e2248793. doi: 10.1001/jamanetworkopen.2022.48793. PMID: 36576736.

Ricci BC, Sachs J, Dobbertin K, et al. Risk stratification in primary care: value-based contributions of provider adjudication. J Gen Intern Med. 2022 Feb;37(3):601-7. doi: 10.1007/s11606-021-06896-1. PMID: 34100237.

Yeboah J, McClelland RL, Polonsky TS, et al. Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. JAMA. 2012 Aug;308(8):788-95. doi: 10.1001/jama.2012.9624. PMID: 22910756.

Derivation study with only internal validation

Agniel D, Martino SC, Burkhart Q, et al. Incentivizing excellent care to at-risk groups with a health equity summary score. J Gen Intern Med. 2021 Jul;36(7):1847-57. doi: 10.1007/s11606-019-05473-x. PMID: 31713030.

Davoudi A, Sajdeya R, Ison R, et al. Fairness in the prediction of acute postoperative pain using machine learning models. Front Digit Health. 2023 Jan;4:970281. doi: 10.3389/fdgth.2022.970281. PMID: 36714611.

Im EO, Chee W. The DSCP-CA: a decision support computer program-cancer pain management. Comput Inform Nurs. 2011 May;29(5):289-96. doi: 10.1097/NCN.0b013e3181f9dd23. PMID: 20975538.

Lakoski SG, Mallick H, McClure LA, et al. A risk algorithm for assessing short-term mortality for obese black and white men and women. Obesity. 2014 Apr;22(4):1142-8. doi: 10.1002/oby.20622. PMID: 24115735.

Meredith ME, Steimle LN, Stanhope KK, et al. Racial/ethnic differences in pre-pregnancy conditions and adverse maternal outcomes in the nuMoM2b Cohort [Preprint]. medRxiv. 2022 Nov. doi: 10.1101/2022.11.02.22281812.

Oliveira N, Poon LC, Nicolaides KH, et al. First trimester prediction of HELLP syndrome. Prenat Diagn. 2016 Jan;36(1):29-33. doi: 10.1002/pd.4694. PMID: 26402854.

Pierson E, Cutler DM, Leskovec J, et al. An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nat Med. 2021 Jan;27(1):136-40. doi: 10.1038/s41591-020-01192-7. PMID: 33442014.

Stevens J, Ou FS, Cai J, et al. Prediction of percent body fat measurements in Americans 8 years and older. Int J Obes (Lond). 2016 Apr;40(4):587-94. doi: 10.1038/ijo.2015.231. PMID: 26538187.

Non-U.S. population

Bernardez-Pereira S, Gioli-Pereira L, Marcondes-Braga FG, et al. Genomic ancestry as a predictor of haemodynamic profile in heart failure. Open Heart. 2016;3(2):e000434. doi: 10.1136/openhrt-2016-000434. PMID: 27547430.

Brewster LM, Van Valkengoed I, van Montfrans GA. African ancestry vs. creatine kinase to predict hypertension control: time for a change? Am J Hypertens. 2021 Dec;34(12):1264-8. doi: 10.1093/ajh/hpab114. PMID: 34272843.

Bukabau JB, Sumaili EK, Cavalier E, et al. Performance of glomerular filtration rate estimation equations in Congolese healthy adults: the inopportunity of the ethnic correction. PLoS ONE. 2018 Mar;13(3):e0193384. doi: 10.1371/journal.pone.0193384. PMID: 29499039.

Delanaye P, Vidal-Petiot E, Björk J, et al. Performance of creatinine-based equations to estimate glomerular filtration rate in White and Black populations in Europe, Brazil, and Africa. Nephrol Dial Transplant. 2023 Jan;38(1):106-18. doi: 10.1093/ndt/gfac241. PMID: 36002032.

Ekladious SMM, Issac MSM, Sharaf S, et al. Validation of a proposed warfarin dosing algorithm based on the genetic make-up of Egyptian patients. Mol Diagn Ther. 2013 Dec;17(6):381-90. doi: 10.1007/s40291-013-0046-3. PMID: 23839801.

Feng Y, Min Y, Chen H, et al. Construction and validation of a nomogram for predicting cervical lymph node metastasis in classic papillary thyroid carcinoma. J Endocrinol Invest. 2021 Oct;44(10):2203-11. doi: 10.1007/s40618-021-01524-5. PMID: 33586026.

Galvez JM, Restrepo CM, Contreras NC, et al. Creating and validating a warfarin pharmacogenetic dosing algorithm for Colombian patients. Pharmacogenomics Person Med. 2018;11:169-78. doi: 10.2147/PGPM.S170515. PMID: 30410385.

Gama RM, Clery A, Griffiths K, et al. Estimated glomerular filtration rate equations in people of self-reported black ethnicity in the United Kingdom: inappropriate adjustment for ethnicity may lead to reduced access to care. PLoS ONE. 2021 Aug;16(8):e0255869. doi: 10.1371/journal.pone.0255869. PMID: 34383841.

Haas Pizarro M, Conte Santos D, Gomes Nunes Melo L, et al. Glomerular filtration rate estimated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation in type 1 diabetes based on genomic ancestry. Diabetol Metab Syndr. 2020 Aug;12:71. doi: 10.1186/s13098-020-00578-4. PMID: 32821292.

Hosein A, Stoute V, Chadee S, et al. Evaluating cardiovascular disease (CVD) risk scores for participants with known CVD and non-CVD in a multiracial/ethnic Caribbean sample. PeerJ. 2020 Mar;8:e8232. doi: 10.7717/peerj.8232. PMID: 32195041.

Jayakumar T, Savithri SR. Effect of geographical and ethnic variation on Dysphonia Severity Index: a study of Indian population. J Voice. 2012 Jan;26(1):e11-e6. doi: 10.1016/j.jvoice.2010.05.008. PMID: 20951546.

Mancini GBJ, Ryomoto A. Comparison of cardiovascular risk assessment algorithms to determine eligibility for statin therapy: implications for practice in Canada. Can J Cardiol. 2014 Jun;30(6):661-6. doi: 10.1016/j.cjca.2014.04.001. PMID: 24882538.

Masekela R, Hall GL, Stanojevic S, et al. An urgent need for African spirometry reference equations: the paediatric and adult African spirometry study. Int J Tuberc Lung Dis. 2019 Aug;23(8):952-8. doi: 10.5588/ijtld.18.0442. PMID: 31533886.

Nakash O, Gerber Y, Goldbourt U, et al. Ethnicity and long-term prognosis after myocardial infarction: a population-based cohort study. Med Care. 2013 Feb;51(2):137-43. doi: 10.1097/MLR.0b013e318270bab5. PMID: 23032353.

Omuse G, Maina D, Mwangi J, et al. Comparison of equations for estimating glomerular filtration rate in screening for chronic kidney disease in asymptomatic black Africans: a cross sectional study. BMC Nephrol. 2017 Dec;18:369. doi: 10.1186/s12882-017-0788-y. PMID: 29262800.

Pastor-Barriuso R, Ascunce N, Ederra M, et al. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. Breast Cancer Res Treat. 2013 Feb;138(1):249-59. doi: 10.1007/s10549-013-2428-y. PMID: 23378108.

Pei L, Tian X, Long Y, et al. Establishment of a Han Chinese-specific pharmacogenetic-guided warfarin dosing algorithm. Medicine (Baltimore). 2018 Sep;97(36):e12178. doi: 10.1097/MD.0000000000012178. PMID: 30200121.

Perini W, Snijder MB, Agyemang C, et al. Eligibility for cardiovascular risk screening among different ethnic groups: the HELIUS study. Eur J Prev Cardiol. 2020 Jul;27(11):1204-11. doi: 10.1177/2047487319866284. PMID: 31345055.

Praditpornsilpa K, Townamchai N, Chaiwatanarat T, et al. The need for robust validation for MDRD-based glomerular filtration rate estimation in various CKD populations. Nephrology Dialysis Transplantation. 2011 Sep;26(9):2780-5. doi: 10.1093/ndt/gfq815. PMID: 21357214.

Pylypchuk R, Wells S, Kerr A, et al. Cardiovascular disease risk prediction equations in 400?000 primary care patients in New Zealand: a derivation and validation study. Lancet. 2018 May;391(10133):1897-907. doi: 10.1016/S0140-6736(18)30664-0. PMID: 29735391.

Rocha AD, Garcia S, Santos AB, et al. No race-ethnicity adjustment in CKD-EPI equations is required for estimating glomerular filtration rate in the Brazilian population. Int J Nephrol. 2020 Jul;2020:2141038. doi: 10.1155/2020/2141038. PMID: 32733708.

Sasano M, Ohno M, Fukuda Y, et al. Verification of pharmacogenomics-based algorithms to predict warfarin maintenance dose using registered data of Japanese patients. Eur J Clin Pharmacol. 2019 Jul;75(7):901-11. doi: 10.1007/s00228-019-02656-7. PMID: 30852642.

Seape T, Gounden V, van Deventer HE, et al. Cystatin C- and creatinine-based equations in the assessment of renal function in HIV-positive patients prior to commencing highly active antiretroviral therapy. Ann Clin Biochem. 2016 Jan;53(1):58-66. doi: 10.1177/0004563215579695. PMID: 25766385.

Segarra A, de la Torre J, Ramos N, et al. Assessing glomerular filtration rate in hospitalized patients: a comparison between CKD-EPI and four cystatin c-based equations. Clin J Am Soc Nephrol. 2011 Oct;6(10):2411-20. doi: 10.2215/CJN.01150211. PMID: 21852668.

Teo BW, Koh YY, Toh QC, et al. Performance of the CKD-EPI creatinine-cystatin C glomerular filtration rate estimation equations in a multiethnic Asian population. Singapore Med J. 2014 Dec;55(12):656-9. doi: 10.11622/smedj.2014181. PMID: 25630321.

Teo BW, Xu H, Wang D, et al. Estimating glomerular filtration rates by use of both cystatin C and standardized serum creatinine avoids ethnicity coefficients in Asian patients with chronic kidney disease. Clin Chem. 2012 Feb;58(2):450-7. doi: 10.1373/clinchem.2011.172346. PMID: 22205693.

Teo BW, Xu H, Wang D, et al. GFR estimating equations in a multiethnic Asian population. Am J Kidney Dis. 2011 Jul;58(1):56-63. doi: 10.1053/j.ajkd.2011.02.393. PMID: 21601325.

Not a full-length primary study (e.g., narrative review, commentary)

Benjamin R. Assessing risk, automating racism: a health care algorithm reflects underlying racial bias in society. Science. 2019 Oct;366(6464):421-2. doi: 10.1126/science.aaz3873. PMID: 31649182.

Boulware LE, Purnell TS, Mohottige D. Systemic kidney transplant inequities for Black individuals: examining the contribution of racialized kidney function estimating equations. JAMA Netw Open. 2021 Jan;4(1):e2034630. doi: 10.1001/jamanetworkopen.2020.34630. PMID: 33443577.

Bozkurt S, Cahan EM, Seneviratne MG, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. J Am Med Inform Assoc. 2020 Dec;27(12):1878-84. doi: 10.1093/jamia/ocaa164. PMID: 32935131.

Bradley P, Ryan D, Craig C, et al. 62 - The use of PLCOm2012 vs PLCOm2012noRace risk prediction models in a UK lung cancer screening programme. Lung Cancer. 2021 Jun;156(Suppl 1):S25. doi: 10.1016/S0169-5002(21)00260-9.

Byrne L, Toland AE. Polygenic risk scores in prostate cancer risk assessment and screening. Urol Clin North Am. 2021 Aug;48(3):387-99. doi: 10.1016/j.ucl.2021.03.007. PMID: 34210493.

Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, et al. Putting the data before the algorithm in big data addressing personalized healthcare. NPJ Digit Med. 2019;2(1):78. doi: 10.1038/s41746-019-0157-2. PMID: 31453373.

Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. Am J Kidney Dis. 2022 Feb;79(2):268-88.e1. doi: 10.1053/j.ajkd.2021.08.003. PMID: 34563581.

Han Y, Zhang X, Bragg-Gresham J, et al. 113 The impact of removing the EGFR race coefficient on CKD care practices among Black Veterans receiving care in US. Am J Kidney Dis. 2021 Apr;77(4):602. doi: 10.1053/j.ajkd.2021.02.118.

Hanaway MJ, MacLennan PA, Locke JE. Exacerbating racial disparities in kidney transplant: the consequences of geographic redistribution. JAMA Surg. 2020 Aug;155(8):679-81. doi: 10.1001/jamasurg.2020.1455. PMID: 32492119.

Hauspurg A. Reducing maternal morbidity and racial disparities in hypertension care using an automated care pathway. Obstet Gynecol. 2021 Feb;137(2):209-10. doi: 10.1097/AOG.0000000000004265. PMID: 33416282.

Lennon JC. Machine learning algorithms for suicide risk: a premature arms race? Gen Psychiatr. 2020;33(6):e100269. doi: 10.1136/gpsych-2020-100269. PMID: 33089067.

Lucas A, Wyatt CM, Inker LA. Removing race from GFR estimates: balancing potential benefits and unintended consequences. Kidney Int. 2021 Jul;100(1):11-3. doi: 10.1016/j.kint.2021.02.017. PMID: 33647323.

Medlock S, Ravelli ACJ, Tamminga P, et al. Prediction of mortality in very premature infants: a systematic review of prediction models. PLoS ONE. 2011 Sep;6(9):e23441. doi: 10.1371/journal.pone.0023441. PMID: 21931598.

Obeng AO, Kaszemacher T, Abul-Husn NS, et al. Implementing algorithm-guided warfarin dosing in an ethnically diverse patient population using electronic health records and preemptive CYP2C9 and VKORC1 genetic testing. Clin Pharmacol Ther. 2016 Nov;100(5):427-30. doi: 10.1002/cpt.425. PMID: 27393744.

Osei-Anto HA, Greising CH. Using patient data to provide equitable care. Hosp Health Netw. 2011 May;85(5):63. https://www.researchgate.net/publication/51227748_Using_patient_data_to_provide_equitable_care. PMID: 21682247.

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):1429-30. doi: 10.1007/s11606-018-4475-x. PMID: 29766380.

Price ET. Warfarin pharmacogenomics and African ancestry. Blood. 2015 Jul;126(4):434-6. doi: 10.1182/blood-2015-06-649509. PMID: 26206946.

Ridker PM, Cook NR. The pooled cohort equations 3 years on: building a stronger foundation. Circulation. 2016 Dec;134(23):1789-91. doi: 10.1161/CIRCULATIONAHA.116.024246. PMID: 27920070.

Sarkar R, Martin C, Mattie H, et al. Performance of intensive care unit severity scoring systems across different ethnicities [Preprint]. medRxiv. 2021 Jan. doi: 10.1101/2021.01.19.21249222. PMID: 33501459.

Straw I, Callison-Burch C. Artificial intelligence in mental health and the biases of language based models. PLoS ONE. 2020 Dec;15(12):e0240376. doi: 10.1371/journal.pone.0240376. PMID: 33332380.

Suarez-Kurtz G, Botton MR. Pharmacogenomics of warfarin in populations of African descent. Br J Clin Pharmacol. 2013 Feb;75(2):334-46. doi: 10.1111/j.1365-2125.2012.04354.x. PMID: 22676711.

Yoon G, Matulewicz RS, Major VJ. Generalizing an antibiotic recommendation algorithm for treatment of urinary tract infections to an urban academic medical center. J Urol. 2022 Aug;208(2):234-6. doi: 10.1097/JU.0000000000002672. PMID: 35344386.

Zhao Y, Wood EP, Mirin N, et al. Social determinants in machine learning cardiovascular disease prediction models: a systematic review. Am J Prev Med. 2021 Oct;61(4):596-605. doi: 10.1016/j.amepre.2021.04.016. PMID: 34544559.

Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine. 2021 May;67:103358. doi: 10.1016/j.ebiom.2021.103358. PMID: 33962897.

Key Question 2 Exclusion Reasons

Does not examine the effect of an intervention to mitigate bias of a clinical algorithm or algorithm-based tool

Afrose S, Song W, Nemeroff CB, et al. Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction. Commun Med (Lond). 2022 Sep;2:111. doi: 10.1038/s43856-022-00165-w. PMID: 36059892.

Afrose S, Song W, Nemeroff CB, et al. Subpopulation-specific machine learning prognosis for underrepresented patients with double prioritized bias correction [Preprint]. medRxiv. 2021 Apr:1-48. doi: 10.1101/2021.03.26.21254401.

Agniel D, Martino SC, Burkhart Q, et al. Incentivizing excellent care to at-risk groups with a health equity summary score. J Gen Intern Med. 2021 Jul;36(7):1847-57. doi: 10.1007/s11606-019-05473-x. PMID: 31713030.

Akintoye E, Mahmoud K, Shokr M, et al. Racial/ethnic differences in the prognostic utility of left ventricular mass index for incident cardiovascular disease. Clin Cardiol. 2018 Apr;41(4):502-9. doi: 10.1002/clc.22914. PMID: 29663526.

Al-Mallah MH, Qureshi WT, Keteyian SJ, et al. Racial differences in the prognostic value of cardiorespiratory fitness (results from the Henry Ford Exercise Testing Project). Am J Cardiol. 2016 May;117(9):1449-54. doi: 10.1016/j.amjcard.2016.02.013. PMID: 26976790.

Aminsharifi A, Schulman A, Anderson J, et al. Primary care perspective and implementation of a multidisciplinary, institutional prostate cancer screening algorithm embedded in the electronic health record. Urol Oncol. 2018 Nov;36(11):502.e1-.e6. doi: 10.1016/j.urolonc.2018.07.016. PMID: 30170982

Arnold SV, Spertus JA, Jones PG, et al. Predicting adverse outcomes after myocardial infarction among patients with diabetes mellitus. Circ Cardiovasc Qual Outcomes. 2016;9(4):372-9. doi: 10.1161/CIRCOUTCOMES.115.002365. PMID: 27220369.

Awasthi S, Mahal BA, Park JY, et al. Substantial Gleason reclassification in Black men with national comprehensive cancer network low-risk prostate cancer – a propensity score analysis. Prostate Cancer Prostatic Dis. 2022 Sep;25(3):547-52. doi: 10.1038/s41391-022-00510-z. PMID: 35194179.

Bernardez-Pereira S, Gioli-Pereira L, Marcondes-Braga FG, et al. Genomic ancestry as a predictor of haemodynamic profile in heart failure. Open Heart. 2016;3(2):e000434. doi: 10.1136/openhrt-2016-000434. PMID: 27547430.

Blaga V, Seth K, Valentim C, et al. Opioid prescription in ophthalmology and the impact of a decision support tool in reducing excess dosing. Am J Ophthalmol. 2022 Nov;243:34-41. doi: 10.1016/j.ajo.2022.06.019. PMID: 35809659.

Borgese M, Joyce C, Anderson EE, et al. Bias assessment and correction in machine learning algorithms: a use-case in a natural language processing algorithm to identify hospitalized patients with unhealthy alcohol use. AMIA Annu Symp Proc. 2022 Feb;2021:247-54. https://knowledge.amia.org/74229-amia-1.4622266/t003-1.4626466/t003-1.4626467/3575666-1.4626789/3575666-1.4626790?qr=1. PMID: 35308909.

Bose A, O'Neal WT, Bennett A, et al. Relation between estimated cardiorespiratory fitness and atrial fibrillation (from the reasons for geographic and racial differences in stroke study). Am J Cardiol. 2017 Jun;119(11):1776-80. doi: 10.1016/j.amjcard.2017.03.008. PMID: 28390681.

Brewster LM, Van Valkengoed I, van Montfrans GA. African ancestry vs. creatine kinase to predict hypertension control: time for a change? Am J Hypertens. 2021 Dec;34(12):1264-8. doi: 10.1093/ajh/hpab114. PMID: 34272843.

Burton MJ, Sunesara I, Penman A, et al. Comparing the aspartate aminotransferase (AST) to platelet ratio index (APRI) between African American and White veterans with chronic hepatitis C. South Med J. 2011 May;104(5):309-14. doi: 10.1097/SMJ.0b013e318213cf52. PMID: 21606706.

Butt JH, Adamson C, Docherty KF, et al. Eligibility for pharmacological therapies in heart failure with reduced ejection fraction: implications of the new chronic kidney disease Epidemiology Collaboration creatinine equation for estimating glomerular filtration rate. Eur J Heart Fail. 2022 May;24(5):861-6. doi: 10.1002/ejhf.2460. PMID: 35199418.

Castro VM, Apperson WK, Gainer VS, et al. Evaluation of matched control algorithms in EHR-based phenotyping studies: a case study of inflammatory bowel disease comorbidities. J Biomed Inform. 2014 Dec;52:105-11. doi: 10.1016/j.jbi.2014.08.012. PMID: 25196084.

Castro Y, Kendzor DE, Businelle MS, et al. Structural and predictive equivalency of the Wisconsin smoking withdrawal scale across three racial/ethnic groups. Nicotine Tob Res. 2011 Dec;13(7):548-55. doi: 10.1093/ntr/ntr039. PMID: 21454912.

Chapman CH, Schechter CB, Cadham CJ, et al. Identifying equitable screening mammography strategies for Black women in the United States using simulation modeling. Ann Intern Med. 2021 Dec;174(12):1637-46. doi: 10.7326/M20-6506. PMID: 34662151.

Chase EC, Bryant AK, Sun Y, et al. Development and validation of a life expectancy calculator for US patients with prostate cancer. BJU Int. 2022 Oct;130(4):496-506. doi: 10.1111/bju.15740. PMID: 35373440.

Chen Z, Cao B, Edwards A, et al. A deep imputation and inference framework for estimating personalized and race-specific causal effects of genomic alterations on PSA. J Bioinform Comput Biol. 2021 Aug;19(4):2150016. doi: 10.1142/S0219720021500165. PMID: 34225568.

Christine PJ, Young R, Adar SD, et al. Individual- and area-level SES in diabetes risk prediction: the multi-ethnic study of atherosclerosis. Am J Prev Med. 2017 Aug;53(2):201-9. doi: 10.1016/j.amepre.2017.04.019. PMID: 28625713.

Clark CR, Ommerborn MJ, Moran K, et al. Predicting self-rated health across the life course: health equity insights from machine learning models. J Gen Intern Med. 2021 May;36(5):1181-8. doi: 10.1007/s11606-020-06438-1. PMID: 33620624.

Cohn T, Miller A, Fogg L, et al. Impact of individual and neighborhood factors on cardiovascular risk in White Hispanic and non-Hispanic women and men. Res Nurs Health. 2017 Apr;40(2):120-31. doi: 10.1002/nur.21778. PMID: 27862050.

Coley RY, Johnson E, Simon GE, et al. Racial/ethnic disparities in the performance of prediction models for death by suicide after mental health visits. JAMA Psych. 2021 Jul;78(7):726-34. doi: 10.1001/jamapsychiatry.2021.0493. PMID: 33909019.

Collin LJ, Yan M, Jiang R, et al. Oncotype DX recurrence score implications for disparities in chemotherapy and breast cancer mortality in Georgia. npj Breast Cancer. 2019 Sep;5:32. doi: 10.1038/s41523-019-0129-3. PMID: 31583272.

Congdon HB, Eldridge BH, Truong HA. Development and implementation of a navigator- facilitated care coordination algorithm to improve clinical outcomes of underserved Latino patients with uncontrolled diabetes. J Health Care Poor Underserved. 2013 Nov;24(4):1604-13. doi: 10.1353/hpu.2013.0181. PMID: 24185156.

Conran CA, Shi Z, Resurreccion WK, et al. Assessing the clinical utility of genetic risk scores for targeted cancer screening. J Transl Med. 2021 Jan;19(1):41. doi: 10.1186/s12967-020-02699-w. PMID: 33482857.

Cooper DL, Manago J, Patel V, et al. Universal posttransplant cyclophosphamide after allogeneic transplant, a retrospective single institution study. Leuk Res. 2022 Nov;122:10693. doi: 10.1016/j.leukres.2022.106934. PMID: 36084368.

Coram MA, Candille SI, Duan Q, et al. Leveraging multi-ethnic evidence for mapping complex traits in minority populations: an empirical Bayes approach. Am J Hum Genet. 2015 May;96(5):740-52. doi: 10.1016/j.ajhg.2015.03.008. PMID: 25892113.

Cullen J, Lynch JA, Klein EA, et al. Multicenter comparison of 17-gene genomic prostate score as a predictor of outcomes in African American and Caucasian American men with clinically localized prostate cancer. J Urol. 2021 Apr;205(4):1047-54. doi: 10.1097/JU.0000000000001484. PMID: 33493001.

Davoudi A, Sajdeya R, Ison R, et al. Fairness in the prediction of acute postoperative pain using machine learning models. Front Digit Health. 2023 Jan;4:970281. doi: 10.3389/fdgth.2022.970281. PMID: 36714611.

Dennis JA. Racial/ethnic disparities in triage scores among pediatric emergency department fever patients. Pediatr Emerg Care. 2021 Dec;37(12):e1457-e61. doi: 10.1097/PEC.0000000000002072. PMID: 32150002.

Devick KL, Valeri L, Chen J, et al. The role of body mass index at diagnosis of colorectal cancer on Black-White disparities in survival: a density regression mediation approach. Biostatistics. 2022 Apr;23(2):449-66. doi: 10.1093/biostatistics/kxaa034. PMID: 32968805.

Dipnall JF, Pasco JA, Berk M, et al. Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. PLoS ONE. 2016 Feb;11(2):e0148195. doi: 10.1371/journal.pone.0148195. PMID: 26848571.

Dixon SE, Haas SA, Klopp A, et al. A quality improvement project: using the STOP-BANG tool in a military population to improve equity in preoperative screening. J Perianesth Nurs. 2016 Oct;31(5):371-80. doi: 10.1016/j.jopan.2014.12.002. PMID: 27667343.

Dutta A, Ikwuezunma A, Castellanos MI, et al. An evidence-based, risk-adapted algorithm for antifungal prophylaxis reduces risk for invasive mold infections in children with hematologic malignancies. Pediatr Blood Cancer. 2021 Dec;68(12):e29228. doi: 10.1002/pbc.29228. PMID: 34268879.

Eapen ZJ, McCoy LA, Fonarow GC, et al. Utility of socioeconomic status in predicting 30-day outcomes after heart failure hospitalization. Circ Heart Fail. 2015 May;8(3):473-80. doi: 10.1161/CIRCHEARTFAILURE.114.001879. PMID: 25747700.

Echouffo-Tcheugui JB, Greene SJ, Papadimitriou L, et al. Population risk prediction models for incident heart failure: a systematic review. Circ Heart Fail. 2015 May;8(3):438-47. doi: 10.1161/CIRCHEARTFAILURE.114.001896. PMID: 25737496.

Ekström M, Mannino D. Research race-specific reference values and lung function impairment, breathlessness and prognosis: analysis of NHANES 2007–2012. Respir Res. 2022 Oct;23(1):271. doi: 10.1186/s12931-022-02194-4. PMID: 36182912.

Feng Y, Min Y, Chen H, et al. Construction and validation of a nomogram for predicting cervical lymph node metastasis in classic papillary thyroid carcinoma. J Endocrinol Invest. 2021 Oct;44(10):2203-11. doi: 10.1007/s40618-021-01524-5. PMID: 33586026.

Fiscella K, Winters P, Farah S, et al. Do lung cancer eligibility criteria align with risk among blacks and hispanics? PLoS ONE. 2015 Nov;10(11):e0143789. doi: 10.1371/journal.pone.0143789. PMID: 26618478.

Fremont A, Weissman JS, Hoch E, et al. When race/ethnicity data are lacking: using advanced indirect estimation methods to measure disparities. Rand Health Q. 2016 Jun;6(1):16. doi: 10.7249/RR1162. PMID: 28083444.

Fudim M, Zalawadiya S, Patel DK, et al. Data on coronary artery calcium score performance and cardiovascular risk reclassification across gender and ethnicities. Data Brief. 2016 Mar;6:578-81. doi: 10.1016/j.dib.2016.01.002. PMID: 26909370.

Gadrey SM, Mohanty P, Haughey SP, et al. Overt and occult hypoxemia in patients hospitalized with novel coronavirus disease 2019 [Preprint]. medRxiv. 2022 Jun. doi: 10.1101/2022.06.14.22276166. PMID: 35734082.

Gallifant J, Zhang J, Del Pilar Arias Lopez M, et al. Artificial intelligence for mechanical ventilation: systematic review of design, reporting standards, and bias. Br J Anaesth. 2022 Feb;128(2):343-51. doi: 10.1016/j.bja.2021.09.025. PMID: 34772497.

Galvez JM, Restrepo CM, Contreras NC, et al. Creating and validating a warfarin pharmacogenetic dosing algorithm for Colombian patients. Pharmacogenomics Person Med. 2018;11:169-78. doi: 10.2147/PGPM.S170515. PMID: 30410385.

Gama RM, Clery A, Griffiths K, et al. Estimated glomerular filtration rate equations in people of self-reported black ethnicity in the United Kingdom: inappropriate adjustment for ethnicity may lead to reduced access to care. PLoS ONE. 2021 Aug;16(8):e0255869. doi: 10.1371/journal.pone.0255869. PMID: 34383841.

Gao Y, Cui Y. Deep transfer learning provides a Pareto improvement for multi-ancestral clinico-genomic prediction of diseases [Preprint]. bioRxiv. 2022 Nov. doi: 10.1101/2022.09.22.509055.

Gao Y, Cui Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nat Commun. 2020 Oct;11(1):5131. doi: 10.1038/s41467-020-18918-3. PMID: 33046699.

Gershengorn HB, Patel S, Shukla B, et al. Predictive value of sequential organ failure assessment score across patients with and without COVID-19 infection. Ann Am Thorac Soc. 2022 May;19(5):790-8. doi: 10.1513/AnnalsATS.202106-680OC. PMID: 34784497.

Gianattasio KZ, Ciarleglio A, Power MC. Development of algorithmic dementia ascertainment for racial/ethnic disparities research in the US Health and Retirement Study. Epidemiology. 2020 Jan;31(1):126-33. doi: 10.1097/EDE.0000000000001101. PMID: 31567393.

Gim J, An J, Sung J, et al. A between ethnicities comparison of chronic obstructive pulmonary disease genetic risk. Front Genet. 2020;11:329. doi: 10.3389/fgene.2020.00329. PMID: 32373161.

Gobardhan SN, Dimitriu-Leen AC, van Rosendael AR, et al. Prevalence by computed tomographic angiography of coronary plaques in South Asian and White patients with type 2 diabetes mellitus at low and high risk using four cardiovascular risk scores (UKPDS, FRS, ASCVD, and JBS3). Am J Cardiol. 2017 Mar;119(5):705-11. doi: 10.1016/j.amjcard.2016.11.029. PMID: 28024655.

Goldie SJ, Daniels N. Model-based analyses to compare health and economic outcomes of cancer control: inclusion of disparities. J Natl Cancer Inst. 2011 Sep;103(18):1373-86. doi: 10.1093/jnci/djr303. PMID: 21900120.

Goldstein JR, Atherwood S. Improved measurement of racial/ethnic disparities in COVID-19 mortality in the United States [Preprint]. medRxiv. 2020 Jun. doi: 10.1101/2020.05.21.20109116. PMID: 32511557.

Gu T, Han Y, Duan R. A transfer learning approach based on random forest with application to breast cancer prediction in underrepresented populations. Pac Symp Biocomput. 2023;28:186-97. doi: 10.1142/9789811270611_0018. PMID: 36540976.

Gulbahce HE, White S, Herget KA, et al. 21-gene recurrence score testing utilization among older women from different races: a population-based study. J Geriatr Oncol. 2021 Mar;12(2):206-11. doi: 10.1016/j.jgo.2020.06.004. PMID: 32646620.

Gurinovich A, Bae H, Farrell JJ, et al. PopCluster: an algorithm to identify genetic variants with ethnicity-dependent effects. Bioinformatics. 2019 Sep;35(17):3046-54. doi: 10.1093/bioinformatics/btz017. PMID: 30624692.

Gurka MJ, Lilly CL, Oliver MN, et al. An examination of sex and racial/ethnic differences in the metabolic syndrome among adults: a confirmatory factor analysis and a resulting continuous severity score. Metabolism. 2014 Feb;63(2):218-25. doi: 10.1016/j.metabol.2013.10.006. PMID: 24290837.

Hall EC, Massie AB, James NT, et al. Effect of eliminating priority points for HLA-B matching on racial disparities in kidney transplant rates. Am J Kidney Dis. 2011 Nov;58(5):813-6. doi: 10.1053/j.ajkd.2011.05.023. PMID: 21802805.

Haas Pizarro M, Conte Santos D, Gomes Nunes Melo L, et al. Glomerular filtration rate estimated by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation in type 1 diabetes based on genomic ancestry. Diabetol Metab Syndr. 2020 Aug;12:71. doi: 10.1186/s13098-020-00578-4. PMID: 32821292.

Han JH, Bilker WB, Edelstein PH, et al. Derivation and validation of clinical prediction rules for reduced vancomycin susceptibility in staphylococcus aureus bacteraemia. Epidemiol Infect. 2013 Jan;141(1):165-73. doi: 10.1017/S0950268812000295. PMID: 22490228.

Han Y, Miao ZF, Lian M, et al. Racial and ethnic disparities in 21-gene recurrence scores, chemotherapy, and survival among women with hormone receptor-positive, node-negative breast cancer. Breast Cancer Res Treat. 2020 Dec;184(3):915-25. doi: 10.1007/s10549-020-05902-0. PMID: 32929567.

Harari Y, Harari Y, O'Brien MK, et al. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. J Neuroengineering Rehabil. 2020 Jun;17(1):71. doi: 10.1186/s12984-020-00704-3. PMID: 32522242.

Hernandez SE, Sylling PW, Mor MK, et al. Developing an algorithm for combining race and ethnicity data sources in the Veterans Health Administration. Mil Med. 2020 Apr;185(3):e495-e500. doi: 10.1093/milmed/usz322. PMID: 31603222.

Hernandez W, Gamazon ER, Aquino-Michaels K, et al. Integrated analysis of genetic variation and gene expression reveals novel variant for increased warfarin dose requirement in African Americans. J Thromb Haemost. 2017 Apr;15(4):735-43. doi: 10.1111/jth.13639. PMID: 28135054.

Hivert MF, Christophi CA, Jablonski KA, et al. Genetic ancestry markers and difference in A1c between African American and White in the diabetes prevention program. J Clin Endocrinol Metab. 2018 Nov;104(2):328-36. doi: 10.1210/jc.2018-01416. PMID: 30358859.

Homkrailas P, Bunnapradist S. Association between ethnicity and kidney transplant waitlist outcomes beyond estimated post-transplant survival score. Transpl Int. 2021 Oct;34(10):1837-44. doi: 10.1111/tri.13965. PMID: 34192375.

Hoppmann AL, Chen Y, Landier W, et al. Individual prediction of nonadherence to oral mercaptopurine in children with acute lymphoblastic leukemia: results from COG AALL03N1. Cancer. 2021 Oct;127(20):3832-9. doi: 10.1002/cncr.33760. PMID: 34161608.

Hosein A, Stoute V, Chadee S, et al. Evaluating cardiovascular disease (CVD) risk scores for participants with known CVD and non-CVD in a multiracial/ethnic Caribbean sample. PeerJ. 2020 Mar;8:e8232. doi: 10.7717/peerj.8232. PMID: 32195041.

Hsia DS, Rasouli N, Pittas AG, et al. Implications of the hemoglobin glycation index on the diagnosis of prediabetes and diabetes. J Clin Endocrinol Metab. 2020 Mar;105(3):e130-e8. doi: 10.1210/clinem/dgaa029. PMID: 31965161.

Hu H, Huff CD, Yamamura Y, et al. The relationship between Native American ancestry, body mass index and diabetes risk among Mexican-Americans. PLoS ONE. 2015 Oct;10(10):e0141260. doi: 10.1371/journal.pone.0141260. PMID: 26501420.

Hu T, Mortensen K, Chen J. Medicaid managed care in Florida and racial and ethnic disparities in preventable emergency department visits. Med Care. 2018;56(6):477-83. doi: 10.1097/MLR.0000000000000909. PMID: 29629922.

Hussain A, Virani SS, Zheng L, et al. Potential impact of 2017 American College of Cardiology/American Heart Association hypertension guideline on contemporary practice: a cross-sectional analysis from NCDR PINNACLE registry. J Am Heart Assoc. 2022 Jun;11(11):e024107. doi: 10.1161/JAHA.121.024107. PMID: 35656989.

Inker LA, Levey AS, Tighiouart H, et al. Performance of glomerular filtration rate estimating equations in a community-based sample of Blacks and Whites: the multiethnic study of atherosclerosis. Nephrol Dial Transplant. 2018 Mar;33(3):417-25. doi: 10.1093/ndt/gfx042. PMID: 28505377.

Inker LA, Shafi T, Okparavero A, et al. Effects of race and sex on measured GFR: the multi-ethnic study of atherosclerosis. Am J Kidney Dis. 2016 Nov;68(5):743-51. doi: 10.1053/j.ajkd.2016.06.021. PMID: 20160831039, 27555103.

Inoue Y, Howard AG, Stickley A, et al. Sex and racial/ethnic differences in the association between childhood attention-deficit/hyperactivity disorder symptom subtypes and body mass index in the transition from adolescence to adulthood in the United States. Pediatr Obes. 2019 May;14(5):e12498. doi: 10.1111/ijpo.12498. PMID: 30629806.

Jasem J, Amini A, Rabinovitch R, et al. 21-gene recurrence score assay as a predictor of adjuvant chemotherapy administration for early-stage breast cancer: an analysis of use, therapeutic implications, and disparity profile. J Clin Oncol. 2016 Jun;34(17):1995-2002. doi: 10.1200/JCO.2015.65.0887. PMID: 27001563.

Jayakumar T, Savithri SR. Effect of geographical and ethnic variation on Dysphonia Severity Index: a study of Indian population. J Voice. 2012 Jan;26(1):e11-e6. doi: 10.1016/j.jvoice.2010.05.008. PMID: 20951546.

Jialal I, Remaley AT, Adams-Huet B. The triglyceride-waist circumference index is a valid biomarker of metabolic syndrome in African Americans. Am J Med Sci. 2023;365(2):184-8. doi: 10.1016/j.amjms.2022.11.003. PMID: 36435217.

Jiang X, Morgenstern LB, Cigolle CT, et al. Multiple chronic conditions explain ethnic differences in functional outcome among patients with ischemic stroke. Stroke. 2022 Jan;53(1):120-7. doi: 10.1161/STROKEAHA.120.032595. PMID: 34517767.

Johnston V, Bao Y. Race/ethnicity-related and payer-related disparities in the timeliness of emergency care in U.S. emergency departments. J Health Care Poor Underserved. 2011 May;22(2):606-20. doi: 10.1353/hpu.2011.0050. PMID: 21551937.

Karmali KN, Goff DC, Ning H, et al. A systematic examination of the 2013 ACC/AHA pooled cohort risk assessment tool for atherosclerotic cardiovascular disease. J Am Coll Cardiol. 2014;64(10):959-68. doi: 10.1016/j.jacc.2014.06.1186. PMID: 25190228.

Kaspers M, Llamocca E, Quick A, et al. Black and Hispanic women are less likely than white women to receive guideline-concordant endometrial cancer treatment. Am J Obstet Gynecol. 2020 Sep;223(3):398.e1-.e18. doi: 10.1016/j.ajog.2020.02.041. PMID: 32142825.

Kasturi SN, Park J, Wild D, et al. Predicting COVID-19-related health care resource utilization across a statewide patient population: model development study. J Med Internet Res. 2021 Nov;23(11):e31337. doi: 10.2196/31337. PMID: 34581671.

Kay DM, Langfelder-Schwind E, Decelie-Germana J, et al. Utility of a very high IRT/No mutation referral category in cystic fibrosis newborn screening. Pediatr Pulmonol. 2015 Aug;50(8):771-80. doi: 10.1002/ppul.23222. PMID: 26098992.

Keeney BJ, Koenig KM, Paddock NG, et al. Do aggregate socioeconomic status factors predict outcomes for total knee arthroplasty in a rural population? J Arthroplasty. 2017 Dec;32(12):3583-90. doi: 10.1016/j.arth.2017.07.002. PMID: 28781014.

Kim JS, Gao X, Rzhetsky A. RIDDLE: Race and ethnicity Imputation from Disease history with Deep LEarning. PLoS Comput Biol. 2018;14(4):e1006106. doi: 10.1371/journal.pcbi.1006106. PMID: 29698408.

Kostick-Quenet KM, Cohen IG, Gerke S, et al. Mitigating racial bias in machine learning. J Law Med Ethics. 2022;50(1):92-100. doi: 10.1017/jme.2022.13. PMID: 35243993.

Kowlgi GN, Gunda S, Padala SK, et al. Comparison of frequency of atrial fibrillation in Blacks versus Whites and the utilization of race in a novel risk score. Am J Cardiol. 2020 Nov;135:68-76. doi: 10.1016/j.amjcard.2020.08.029. PMID: 32866451.

Krissberg J, Kaufmann M, Gupta A, et al. Racial disparities in pediatric kidney transplantation under the new kidney allocation system in the United States. Clin J Am Soc Nephrol. 2021 Oct;16(12):1862-71. doi: 10.2215/CJN.06740521. PMID: 34670797.

Ku KC, Li J, Ha NB, et al. Chronic hepatitis B management based on standard guidelines in community primary care and specialty clinics. Dig Dis Sc. 2013 Dec;58(12):3626-33. doi: 10.1007/s10620-013-2889-1. PMID: 24122622.

Kurian AW, Hughes E, Simmons T, et al. Performance of the IBIS/Tyrer-Cuzick model of breast cancer risk by race and ethnicity in the women's health initiative. Cancer. 2021 Oct;127(20):3742-50. doi: 10.1002/cncr.33767. PMID: 34228814.

Lakoski SG, Mallick H, McClure LA, et al. A risk algorithm for assessing short-term mortality for obese black and white men and women. Obesity. 2014 Apr;22(4):1142-8. doi: 10.1002/oby.20622. PMID: 24115735.

Leonard SA, Main EK, Lyell DJ, et al. Obstetric comorbidity scores and disparities in severe maternal morbidity across marginalized groups. Am J Obstet Gynecol MFM. 2021 Nov;4(2):100530. doi: 10.1016/j.ajogmf.2021.100530. PMID: 34798329.

Li J, Bzdok D, Chen J, et al. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. Sci Adv. 2022 Mar;8(11):eabj1812. doi: 10.1126/sciadv.abj1812. PMID: 35294251.

Liu R, Li X, Zhang W, et al. Comparison of nine statistical model based warfarin pharmacogenetic dosing algorithms using the racially diverse international warfarin pharmacogenetic consortium cohort database. PLoS ONE. 2015 Aug;10(8):e0135784. doi: 10.1371/journal.pone.0135784. PMID: 26305568.

Liu S, Pan H, Xia J, et al. Bridging continual reassessment method for phase I clinical trials in different ethnic populations. Stat Med. 2015 May;34(10):1681-94. doi: 10.1002/sim.6442. PMID: 25626429.

Loomba RS, Raskin A, Gudausky TM, et al. Role of the Egami score in predicting intravenous immunoglobulin resistance in Kawasaki disease among different ethnicities. Am J Ther. 2016 Nov;23(6):e1293-e9. doi: 10.1097/MJT.0000000000000045. PMID: 25611359.

Lubelski D, Alentado V, Nowacki AS, et al. Preoperative nomograms predict patient-specific cervical spine surgery clinical and quality of life outcomes. Clin Neurosurg. 2018 Jul;83(1):104-13. doi: 10.1093/neuros/nyx343.

Lynch JA, Berse B, Coomer N, et al. 21-gene recurrence score testing among Medicare beneficiaries with breast cancer in 2010-2013. Genet Med. 2017 Oct;19(10):1134-43. doi: 10.1038/gim.2017.19. PMID: 28333918.

Mahle WT, Simpson SA, Fye P, et al. Management of warfarin in children with heart disease. Pediatr Cardiol. 2011 Dec;32(8):1115-9. doi: 10.1007/s00246-011-9984-x. PMID: 21499856.

Mancini GBJ, Ryomoto A. Comparison of cardiovascular risk assessment algorithms to determine eligibility for statin therapy: implications for practice in Canada. Can J Cardiol. 2014 Jun;30(6):661-6. doi: 10.1016/j.cjca.2014.04.001. PMID: 24882538.

Marini S, Lena UK, Crawford KM, et al. Comparison of genetic and self-identified ancestry in modeling intracerebral hemorrhage risk. Front Neurol. 2018;9:514. doi: doi.org/10.3389/fneur.2018.00514. PMID: 30034361.

Matsushita K, Mahmoodi BK, Woodward M, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. JAMA. 2012 May;307(18):1941-51. doi: 10.1001/jama.2012.3954. PMID: 22570462.

Maziarz M, Black RA, Fong CT, et al. Evaluating risk of ESRD in the urban poor. J Am Soc Nephrol. 2015 Jun;26(6):1434-42. doi: 10.1681/ASN.2014060546. PMID: 25475746.

McCall CJ, DeCaprio D, Gartner J. The measurement and mitigation of algorithmic bias and unfairness in healthcare AI models developed for the CMS AI health outcomes challenge [Preprint]. medRxiv. 2022 Oct. doi: 10.1101/2022.09.29.22280537.

McCarthy AM, Liu Y, Ehsan S, et al. Validation of breast cancer risk models by race/ethnicity, family history and molecular subtypes. Cancers. 2021 Dec;14(1):45. doi: 10.3390/cancers14010045. PMID: 35008209.

McCoy D, Mgbara W, Horvitz N, et al. Ensemble machine learning of factors influencing COVID-19 across US counties. Sci Rep. 2021 Jun;11(1):11777. doi: 10.1038/s41598-021-90827-x. PMID: 34083563.

Mehta A, Pandey A, Ayers CR, et al. Predictive value of coronary artery calcium score categories for coronary events versus strokes: impact of sex and race: MESA and DHS. Circ Cardiovasc Imaging. 2020 Aug;13(8):e010153. doi: 10.1161/CIRCIMAGING.119.010153. PMID: 32806939.

Mercado CI, Cogswell ME, Loria CM, et al. Validity of predictive equations for 24-h urinary potassium excretion based on timing of spot urine collection among adults: the MESA and CARDIA Urinary Sodium Study and NHANES Urinary Sodium Calibration Study. Am J Clin Nutr. 2018 Sep;108(3):532-47. doi: 10.1093/ajcn/nqy138. PMID: 30535091.

Meredith ME, Steimle LN, Stanhope KK, et al. Racial/ethnic differences in pre-pregnancy conditions and adverse maternal outcomes in the nuMoM2b Cohort [Preprint]. medRxiv. 2022 Nov. doi: 10.1101/2022.11.02.22281812.

Miller V. COVID-19 surveillance and Black American substance use disorder: an examination of data and policy. J Subst Abuse Treat. 2021 Apr;123:108243. doi: 10.1016/j.jsat.2020.108243. PMID: 33612203.

Moss HE, An R, Nelson T, et al. Risk of atherosclerotic cardiovascular disease among US adults: use of 1999-2014 NHANES data. J Prim Prev. 2019 Oct;40(5):569-73. doi: 10.1007/s10935-019-00564-1. PMID: 31571032.

Nakash O, Gerber Y, Goldbourt U, et al. Ethnicity and long-term prognosis after myocardial infarction: a population-based cohort study. Med Care. 2013 Feb;51(2):137-43. doi: 10.1097/MLR.0b013e318270bab5. PMID: 23032353.

Nayan M, Salari K, Bozzo A, et al. Predicting survival after radical prostatectomy: variation of machine learning performance by race. Prostate. 2021 Dec;81(16):1355-64. doi: 10.1002/pros.24233. PMID: 34529282.

Noseworthy PA, Attia ZI, Brewer LC, et al. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. Circ Arrhythm Electrophysiol. 2020 Mar;13(3):e007988. doi: 10.1161/CIRCEP.119.007988. PMID: 32064914.

Okwuosa TM, Greenland P, Ning H, et al. Distribution of coronary artery calcium scores by Framingham 10-year risk strata in the MESA (multi-ethnic study of atherosclerosis): potential implications for coronary risk assessment. J Am Coll Cardiol. 2011 May;57(18):1838-45. doi: 10.1016/j.jacc.2010.11.053. PMID: 21527159.

Oliveira N, Poon LC, Nicolaides KH, et al. First trimester prediction of HELLP syndrome. Prenat Diagn. 2016 Jan;36(1):29-33. doi: 10.1002/pd.4694. PMID: 26402854.

Oliveira N, Doyle LE, Atlas RO, et al. External validity of first-trimester algorithms in the prediction of pre-eclampsia disease severity. Ultrasound Obstet Gynecol. 2014 Sep;44(3):286-92. doi: 10.1002/uog.13433. PMID: 24912952.

Omuse G, Maina D, Mwangi J, et al. Comparison of equations for estimating glomerular filtration rate in screening for chronic kidney disease in asymptomatic black Africans: a cross sectional study. BMC Nephrol. 2017 Dec;18:369. doi: 10.1186/s12882-017-0788-y. PMID: 29262800.

Orandi BJ, Kumar V, Reed RD, et al. Reclassification of CKD in living kidney donors with the refitted race-free eGFR formula. Am J Surg. 2023 Feb;225(2):425-8. doi: 10.1016/j.amjsurg.2022.09.024. PMID: 36167624.

Panchal HJ, Durinka JB, Patterson J, et al. Survival outcomes in liver transplant recipients with model for end-stage liver disease scores of 40 or higher: a decade-long experience. HPB (Oxford). 2015 Dec;17(12):1074-84. doi: 10.1111/hpb.12485. PMID: 26373873.

Park J, Artin MG, Lee KE, et al. Deep learning on time series laboratory test results from electronic health records for early detection of pancreatic cancer. J Biomed Inform. 2022 Jul;131:104095. doi: 10.1016/j.jbi.2022.104095. PMID: 35598881.

Pastor-Barriuso R, Ascunce N, Ederra M, et al. Recalibration of the Gail model for predicting invasive breast cancer risk in Spanish women: a population-based cohort study. Breast Cancer Res Treat. 2013 Feb;138(1):249-59. doi: 10.1007/s10549-013-2428-y. PMID: 23378108.

Patel A, García-Closas M, Olshan AF, et al. Gene-level germline contributions to clinical risk of recurrence scores in black and white patients with breast cancer. Cancer Res. 2022 Jan;82(1):25-35. doi: 10.1158/0008-5472.CAN-21-1207. PMID: 34711612.

Pathak P, Panday SB, Ahn J. Artificial neural network model effectively estimates muscle and fat mass using simple demographic and anthropometric measures. Clin Nutr. 2022 Jan;41(1):144-52. doi: 10.1016/j.clnu.2021.11.027. PMID: 34879301.

Payne NR, Puumala SE. Racial disparities in ordering laboratory and radiology tests for pediatric patients in the emergency department. Pediatr Emerg Care. 2013 May;29(5):598-606. doi: 10.1097/PEC.0b013e31828e6489. PMID: 23603649.

Perini W, Snijder MB, Agyemang C, et al. Eligibility for cardiovascular risk screening among different ethnic groups: the HELIUS study. Eur J Prev Cardiol. 2020 Jul;27(11):1204-11. doi: 10.1177/2047487319866284. PMID: 31345055.

Philibert R, Long JD, Mills JA, et al. A simple, rapid, interpretable, actionable and implementable digital PCR based mortality index. Epigenetics. 2021 Oct;16(10):1135-49. doi: 10.1080/15592294.2020.1841874. PMID: 33138668.

Pierson E, Cutler DM, Leskovec J, et al. An algorithmic approach to reducing unexplained pain disparities in underserved populations. Nat Med. 2021 Jan;27(1):136-40. doi: 10.1038/s41591-020-01192-7. PMID: 33442014.

Pollock BD, Hu T, Chen W, et al. Utility of existing diabetes risk prediction tools for young Black and White adults: evidence from the Bogalusa heart study. J Diabetes Complications. 2017 Jan;31(1):86-93. doi: 10.1016/j.jdiacomp.2016.07.025. PMID: 27503406.

Porterhouse MD, Paul S, Lieberenz JL, et al. Black women are less likely to be classified as high-risk for breast cancer using the Tyrer-Cuzick 8 model. Ann Surg Oncol. 2022 Oct;29(10):6419-25. doi: 10.1245/s10434-022-12140-9. PMID: 35790586.

Pressman A, Jacobson A, Eguilos R, et al. Prevalence of migraine in a diverse community - electronic methods for migraine ascertainment in a large integrated health plan. Cephalalgia. 2016 Jul;36(4):325-34. doi: 10.1177/0333102415590242. PMID: 26069243.

Pressman AR, Lockhart SH, Shen Z, et al. Measuring and promoting SARS-CoV-2 vaccine equity: development of a COVID-19 vaccine equity index. Health Equity. 2021 Jul;5(1):476-83. doi: 10.1089/heq.2021.0047. PMID: 34316531.

Pylypchuk R, Wells S, Kerr A, et al. Cardiovascular disease risk prediction equations in 400?000 primary care patients in New Zealand: a derivation and validation study. Lancet. 2018 May;391(10133):1897-907. doi: 10.1016/S0140-6736(18)30664-0. PMID: 29735391.

Quanjer PH, Stanojevic S, Cole TJ, et al. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012 Dec;40(6):1324-43. doi: 10.1183/09031936.00080312. PMID: 22743675.

jef R, Peterson EL, Michaels A, et al. Performance of the meta-analysis global group in chronic heart failure score in black patients compared with whites. Circ Cardiovasc Qual Outcomes. 2019 Jul;12(7):e004714. doi: 10.1161/CIRCOUTCOMES.118.004714. PMID: 31266369.

Raghu VK, Walia AS, Zinzuwadia AN, et al. Validation of a deep learning-based model to predict lung cancer risk using chest radiographs and electronic medical record data. JAMA Netw Open. 2022 Dec;5(12):e2248793. doi: 10.1001/jamanetworkopen.2022.48793. PMID: 36576736.

Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. J Arthroplasty. 2019 Apr;34(4):632-7. doi: 10.1016/j.arth.2018.12.030. PMID: 30665831.

Rappaport D, Chuu A, Hullett C, et al. Assessment of alcohol withdrawal in native American patients utilizing the clinical institute withdrawal assessment of alcohol revised scale. J Addict Med. 2013 May-Jun;7(3):196-9. doi: 10.1097/ADM.0b013e31828b3cc3. PMID: 23579238.

Rayford W, Greenberger M, Bradley RV. Improving risk stratification in a community-based African American population using cell cycle progression score. Transl Androl Urol. 2018 Sep;7(Suppl 4):S384-S91. doi: 10.21037/tau.2018.03.09. PMID: 30363476.

Razavi AC, Potts KS, Kelly TN, et al. Pooled cohort equations heart failure risk score predicts cardiovascular disease and all-cause mortality in a nationally representative sample of US adults. BMC Cardiovasc Disord. 2020 Apr;20(1):202. doi: 10.1186/s12872-020-01485-2. PMID: 32334524.

Reuland CP, Collins J, Chiang L, et al. Oregon's approach to leveraging system-level data to guide a social determinants of health-informed approach to children's healthcare. BMJ Innovations. 2021 Jan;7(1):18-25. doi: 10.1136/bmjinnov-2020-000452.

Ricci BC, Sachs J, Dobbertin K, et al. Risk stratification in primary care: value-based contributions of provider adjudication. J Gen Intern Med. 2022 Feb;37(3):601-7. doi: 10.1007/s11606-021-06896-1. PMID: 34100237.

Robinson A, Hirode G, Wong RJ. Ethnicity and insurance-specific disparities in the model for end-stage liver disease score at time of liver transplant waitlist registration and its impact on mortality. J Clin Exp Hepatol. 2021 Mar;11(2):188-94. doi: 10.1016/j.jceh.2020.07.011. PMID: 33746443.

Rocha AD, Garcia S, Santos AB, et al. No race-ethnicity adjustment in CKD-EPI equations is required for estimating glomerular filtration rate in the Brazilian population. Int J Nephrol. 2020 Jul;2020:2141038. doi: 10.1155/2020/2141038. PMID: 32733708.

Roche C, Kumar V, Overman S, et al. Validation of a machine learning-derived clinical metric to quantify outcomes after total shoulder arthroplasty. J Shoulder Elbow Surg. 2021 Oct;30(10):2211-24. doi: 10.1016/j.jse.2021.01.021. PMID: 33607333.

Röösli E, Bozkurt S, Hernandez-Boussard T. Peeking into a black box, the fairness and generalizability of a MIMIC-III benchmarking model. Sci Data. 2022 Jan;9(1):24. doi: 10.1038/s41597-021-01110-7. PMID: 35075160.

Rutkowski R, Salemi J, Tanner JP, et al. Assessing the impact of different race-bridging algorithms on the reported rate of birth defects. J Registry Manag. 2017 Dec;44(4):146-56. PMID: 30133431.

Salahuddin M, Mandell DJ, Lakey DL, et al. Maternal comorbidity index and severe maternal morbidity during delivery hospitalizations in Texas, 2011-2014. Birth. 2020 Mar;47(1):89-97. doi: 10.1111/birt.12465. PMID: 31659788.

Santamaria-Barria JA, Graff-Baker AN, Chang SC, et al. Disparities in the impact of the AJCC 8th edition staging system on differentiated thyroid cancer outcomes. Head Neck. 2022 Oct;44(10):2129-41. doi: 10.1002/hed.27122. PMID: 35766292.

asano M, Ohno M, Fukuda Y, et al. Verification of pharmacogenomics-based algorithms to predict warfarin maintenance dose using registered data of Japanese patients. Eur J Clin Pharmacol. 2019 Jul;75(7):901-11. doi: 10.1007/s00228-019-02656-7. PMID: 30852642.

Schaich CL, Yeboah J, Espeland MA, et al. Association of vascular risk scores and cognitive performance in a diverse cohort: the multi-ethnic study of atherosclerosis. J Gerontol A Biol Sci Med Sci. 2022 Jun;77(6):1208-15. doi: 10.1093/gerona/glab189. PMID: 34216214.

Schindler SE, Karikari TK, Ashton NJ, et al. Effect of race on prediction of brain amyloidosis by plasma Aß42/Aß40, phosphorylated tau, and neurofilament light. Neurology. 2022 Jul;99(3):e245-e57. doi: 10.1212/WNL.0000000000200358. PMID: 35450967.

Schultz D, Lovejoy S, Peet E. Tackling persistent and large disparities in birth outcomes in Allegheny County, Pennsylvania. Matern Child Health J. 2022 May;26(5):978-84. doi: 10.1007/s10995-021-03289-y. PMID: 34982343.

Schwartz JB, Lai J, Lizaola B, et al. A comparison of measured and calculated free 25(OH) vitamin D levels in clinical populations. J Clin Endocrinol Metab. 2014 May;99(5):1631-7. doi: 10.1210/jc.2013-3874. PMID: 24483159.

Seape T, Gounden V, van Deventer HE, et al. Cystatin C- and creatinine-based equations in the assessment of renal function in HIV-positive patients prior to commencing highly active antiretroviral therapy. Ann Clin Biochem. 2016 Jan;53(1):58-66. doi: 10.1177/0004563215579695. PMID: 25766385.

Segar MW, Jaeger BC, Patel KV, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multi-cohort analysis. Circulation. 2021 Jun;143(24):2370-83. doi: 10.1161/CIRCULATIONAHA.120.053134. PMID: 33845593.

Segarra A, de la Torre J, Ramos N, et al. Assessing glomerular filtration rate in hospitalized patients: a comparison between CKD-EPI and four cystatin c-based equations. Clin J Am Soc Nephrol. 2011 Oct;6(10):2411-20. doi: 10.2215/CJN.01150211. PMID: 21852668.

Shachar BZ, Mayo JA, Lee HC, et al. Effects of race/ethnicity and BMI on the association between height and risk for spontaneous preterm birth. Am J Obstet Gynecol. 2015 Nov;213(5):700e1-e9. doi: 10.1016/j.ajog.2015.07.005. PMID: 26187451.

Shah A, Polascik TJ, George DJ, et al. Implementation and impact of a risk-stratified prostate cancer screening algorithm as a clinical decision support tool in a primary care network. J Gen Intern Med. 2021 Jan;36(1):92-9. doi: 10.1007/s11606-020-06124-2. PMID: 32875501.

Shin J, Cao D. Comparison of warfarin pharmacogenetic dosing algorithms in a racially diverse large cohort. Pharmacogenomics. 2011 Jan;12(1):125-34. doi: 10.2217/pgs.10.168. PMID: 21174627.

Shulman E, Aagaard P, Kargoli F, et al. Validation of PR interval length as a criterion for development of atrial fibrillation in non-Hispanic whites, African Americans and Hispanics. J Electrocardiol. 2015 Jul;48(4):703-9. doi: 10.1016/j.jelectrocard.2015.04.015. PMID: 26025203.

Singh K, Valley TS, Tang S, et al. Evaluating a widely implemented proprietary deterioration index model among hospitalized patients with COVID-19. Ann Am Thorac Soc. 2021 Jul;18(7):1129-37. doi: 10.1513/AnnalsATS.202006-698OC. PMID: 33357088.

Srivatsan S, Guduguntla V, Young KZ, et al. Clinical versus patient-reported measures of depression in bariatric surgery. Surg Endosc. 2018 Aug;32(8):3683-90. doi: 10.1007/s00464-018-6101-8. PMID: 29435747.

Steiner HE, Giles JB, Patterson HK, et al. Machine learning for prediction of stable warfarin dose in US Latinos and Latin Americans. Front Pharmacol. 2021 Oct;12:749786. doi: 10.3389/fphar.2021.749786. PMID: 34776967.

Stevens J, Erber-Oakkar E, Cui Z, et al. Cardiovascular disease risk by assigned treatment using the 2013 and 1998 obesity guidelines. Obesity. 2016 Jul;24(7):1554-60. doi: 10.1002/oby.21496. PMID: 27184463.

Stevens J, Ou FS, Cai J, et al. Prediction of percent body fat measurements in Americans 8 years and older. Int J Obes (Lond). 2016 Apr;40(4):587-94. doi: 10.1038/ijo.2015.231. PMID: 26538187.

Sughayer M, Alaaraj R, Alsughayer A. Applying new Magee equations for predicting the Oncotype Dx recurrence score. Breast Cancer. 2018 Sep;25(5):597-604. doi: 10.1007/s12282-018-0860-x. PMID: 29691722.

Sullivan BA, Hochheimer CJ, Chernyavskiy P, et al. Impact of race on heart rate characteristics monitoring in very low birth weight infants. Pediatr Res. 2023 Jan. Online ahead of print. doi: 10.1038/s41390-023-02470-z. PMID: 36650306.

Sung JH, Yeboah J, Lee JE, et al. Diagnostic value of coronary artery calcium score for cardiovascular disease in African Americans: the Jackson Heart Study. Br J Med Med Res. 2016;11(2):BJMMR/2016/21449. doi: 10.9734/BJMMR/2016/21449. PMID: 26949662.

Taber DJ, Hamedi M, Rodrigue JR, et al. Quantifying the race stratified impact of socioeconomics on graft outcomes in kidney transplant recipients. Transplantation. 2016 Jun;100(7):1550-7. doi: 10.1097/TP.0000000000000931. PMID: 26425875.

Tallaj JA, Pamboukian SV, George JF, et al. Have risk factors for mortality after heart transplantation changed over time? Insights from 19 years of cardiac transplant research database study. J Heart Lung Transplant. 2014 Dec;33(12):1304-11. doi: 10.1016/j.healun.2014.08.014. PMID: 25443871.

Tammemagi CM, Pinsky PF, Caporaso NE, et al. Lung cancer risk prediction: prostate, lung, colorectal and ovarian cancer screening trial models and validation. J Natl Cancer Inst. 2011 Jul;103(13):1058-68. doi: 10.1093/jnci/djr173. PMID: 21606442.

Tan-McGrory A, Bennett-AbuAyyash C, Gee S, et al. A patient and family data domain collection framework for identifying disparities in pediatrics: results from the pediatric health equity collaborative. BMC Pediatr. 2018 Jan;18(1):18. doi: 10.1186/s12887-018-0993-2. PMID: 29385988.

Teo BW, Xu H, Wang D, et al. GFR estimating equations in a multiethnic Asian population. Am J Kidney Dis. 2011 Jul;58(1):56-63. doi: 10.1053/j.ajkd.2011.02.393. PMID: 21601325.

Teo BW, Xu H, Wang D, et al. Estimating glomerular filtration rates by use of both cystatin C and standardized serum creatinine avoids ethnicity coefficients in Asian patients with chronic kidney disease. Clin Chem. 2012 Feb;58(2):450-7. doi: 10.1373/clinchem.2011.172346. PMID: 22205693.

Thurlapati A, Velez-Martinez CS, Hirani S, et al. Do the 2013 United States Preventive Services Task Force guidelines for lung cancer screening fail high-risk African American smokers? An institutional retrospective observational cohort study. Eur J Cancer Prev. 2021 Sep;30(5):375-81. doi: 10.1097/CEJ.0000000000000652. PMID: 34010237.

Tiwari A, Dadhania AV, Ragunathrao VAB, et al. Using machine learning to develop a novel COVID-19 vulnerability index (C19VI). Sci Total Environ. 2021 Jun;773:145650. doi: 10.1016/j.scitotenv.2021.145650. PMID: 33940747.

Tolksdorf J, Kattan MW, Boorjian SA, et al. Multi-cohort modeling strategies for scalable globally accessible prostate cancer risk tools. BMC Med Res Methodol. 2019 Oct;19(1):191. doi: 10.1186/s12874-019-0839-0. PMID: 31615451.

Toseef M, Li X, Wong KC. Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning. Brief Bioinform. 2022 May;23(3):bbac078. doi: 10.1093/bib/bbac078. PMID: 35323862.

Valera P, McClernon FJ, Burkholder G, et al. A pilot trial examining African American and White responses to algorithm-guided smoking cessation medication selection in persons living with HIV. AIDS Behav. 2017 Jul;21(7):1975-84. doi: 10.1007/s10461-016-1634-0. PMID: 27942999.

Vasan RS, van den Heuvel E. Differences in estimates for 10-year risk of cardiovascular disease in Black versus White individuals with identical risk factor profiles using pooled cohort equations: an in silico cohort study. Lancet Digital Health. 2022 Jan;4(1):e55-e63. doi: doi.org/10.1016/S2589-7500(21)00236-3. PMID: 34952676.

Vigil JM, Alcock J, Coulombe P, et al. Ethnic disparities in emergency severity index scores among U.S. Veteran's Affairs Emergency Department patients. PLoS ONE. 2015 May;10(5):e0126792. doi: 10.1371/journal.pone.0126792. PMID: 2015154600, 26024515.

Wang AY, Wong MS, Humbyrd CJ. Eligibility criteria for lower extremity joint replacement may worsen racial and socioeconomic disparities. Clin Orthop Relat Res. 2018;476(12):2301-8. doi: 10.1097/CORR.0000000000000511. PMID: 30303879.

Wang J, Qiao Y, Tina Shih YC, et al. Potential health implications of racial and ethnic disparities in meeting MTM eligibility criteria. Res Social Adm Pharm. 2014 Jan;10(1):106-25. doi: 10.1016/j.sapharm.2013.03.007. PMID: 23759673.

Wang Y, Katzmarzyk PT, Horswell R, et al. Comparison of the heart failure risk stratification performance of the CKD-EPI equation and the MDRD equation for estimated glomerular filtration rate in patients with type 2 diabetes. Diabet Med. 2016 May;33(5):609-20. doi: 10.1111/dme.12859. PMID: 26202081.

Warsch JRL, Rundek T, Paik MC, et al. Association between northern Manhattan study global vascular risk score and successful aging. J Am Geriatr Soc. 2013 Apr;61(4):519-24. doi: 10.1111/jgs.12166. PMID: 23527874.

Wong RJ, Devaki P, Nguyen L, et al. Increased long-term survival among patients with hepatocellular carcinoma after implementation of model for end-stage liver disease score. Clin Gastroenterol Hepatol. 2014 Sep;12(9):1534-40.e1. doi: 10.1016/j.cgh.2013.12.008. PMID: 24361414.

Worrell FC, Mendoza-Denton R, Wang A. Introducing a new assessment tool for measuring ethnic-racial identity: the cross ethnic-racial identity scale-adult (CERIS-A). Assessment. 2019 Apr;26(3):404-18. doi: 10.1177/1073191117698756. PMID: 29214847.

Wu Q, Xiao X, Xu Y. Performance of FRAX in predicting fractures in US postmenopausal women with varied race and genetic profiles. J Clin Med. 2020 Jan;9(1):285. doi: 10.3390/jcm9010285. PMID: 31968614.

Wuerz TH, Kent DM, Malchau H, et al. A nomogram to predict major complications after hip and knee arthroplasty. J Arthroplasty. 2014 Jul;29(7):1457-62. doi: 10.1016/j.arth.2013.09.007. PMID: 24793891.

Yeboah J, McClelland RL, Polonsky TS, et al. Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals. JAMA. 2012 Aug;308(8):788-95. doi: 10.1001/jama.2012.9624. PMID: 22910756.

.

Yoon S, Davis N, Odlum M, et al. Applying artificial intelligence to predict self-reported poor health among Black and Hispanic caregivers with mild cognitive impairment. Stud Health Technol Inform. 2020 Jun;272:433-6. doi: 10.3233/SHTI200588. PMID: 32604695.

Zavorsky GS, Almamary AS, Alqahtani MK, et al. The need for race-specific reference equations for pulmonary diffusing capacity for nitric oxide. BMC Pulm Med. 2021 Jul;21(1):232. doi: 10.1186/s12890-021-01591-7. PMID: 34256739.

Zhang L, Zhang Z, Zhang Y, et al. Evaluation of Finnish diabetes risk score in screening undiagnosed diabetes and prediabetes among U.S. adults by gender and race: NHANES 1999-2010. PLoS ONE. 2014 May;9(5):e97865. doi: 10.1371/journal.pone.0097865. PMID: 24852786.

Zhou X, Ning Q, Jin K, et al. Development and validation of a preoperative nomogram for predicting survival of patients with locally advanced prostate cancer after radical prostatectomy. BMC Cancer. 2020 Feb;20(1):97. doi: 10.1186/s12885-020-6565-5. PMID: 32019501.

Zilberberg MD, Chaudhari P, Nathanson BH, et al. Development and validation of a bedside risk score for MRSA among patients hospitalized with complicated skin and skin structure infections. BMC Infect Dis. 2012 Jul;12:154. doi: 10.1186/1471-2334-12-154. PMID: 22784260.

Zook HG, Kharbanda AB, Flood A, et al. Racial differences in pediatric emergency department triage scores. J Emerg Med. 2016 May;50(5):720-6. doi: 10.1016/j.jemermed.2015.02.056. PMID: 26899520.

Does not report an outcome of interest (i.e., access to care, quality of care, or health outcomes)

Abate M, Jandovitz N, Hirsch JS, et al. The effect of race coefficients on preemptive listing for kidney transplantation. Clin Kidney J. 2022 May;15(5):942-50. doi: 10.1093/ckj/sfab287. PMID: 35498880.

Allen A, Mataraso S, Siefkas A, et al. A racially unbiased, machine learning approach to prediction of mortality: algorithm development study. JMIR Public Health Surveill. 2020 Oct;6(4):e22400. doi: 10.2196/22400. PMID: 33090117.

Arnall DA, Nelson AG, Hearon CM. Maximal respiratory pressure reference values for Hopi children ages 4 to 13. Cardiopulm Phys Ther J. 2022 Jul;33(3):123-9. doi: 10.1097/CPT.0000000000000195. PMID: 36090687.

Awasthi S, Grass GD, Torres-Roca J, et al. Genomic testing in localized prostate cancer can identify subsets of African-Americans with aggressive disease. J Natl Cancer Inst. 2022 Dec;114(12):1656-64. doi: 10.1093/jnci/djac162. PMID: 36053178.

Belouali A, Bai H, Raja K, et al. Impact of social determinants of health on improving the LACE index for 30-day unplanned readmission prediction. JAMIA Open. 2022 Jun;5(2):ooac046. doi: 10.1093/jamiaopen/ooac046. PMID: 35702627.

Bennett R, Mulla ZD, Parikh P, et al. An imbalance-aware deep neural network for early prediction of preeclampsia. PLoS ONE. 2022 Apr;17(4):e0266042. doi: 10.1371/journal.pone.0266042. PMID: 35385525.

Bhattacharya A, García-Closas M, Olshan AF, et al. A framework for transcriptome-wide association studies in breast cancer in diverse study populations. Genome Biol. 2020 Feb;21(1):42. doi: 10.1186/s13059-020-1942-6. PMID: 32079541.

Bowerman C, Bhatka NR, Brazzale D, et al. A race-neutral approach to the interpretation of lung function measurements. Am J Respir Crit Care Med. 2023 Mar;207(6):768-74. doi: 10.1164/rccm.202205-0963OC. PMID: 36383197.

Buckley A, Sestito S, Ogundipe T, et al. Racial and ethnic disparities among women undergoing a trial of labor after cesarean delivery: performance of the VBAC calculator with and without patients' race/ethnicity. Reprod Sci. 2022 Jul;29(7):2030-8. doi: 10.1007/s43032-022-00959-2. PMID: 35534768.

Bukabau JB, Sumaili EK, Cavalier E, et al. Performance of glomerular filtration rate estimation equations in Congolese healthy adults: the inopportunity of the ethnic correction. PLoS ONE. 2018 Mar;13(3):e0193384. doi: 10.1371/journal.pone.0193384. PMID: 29499039.

Chen LY, Norby FL, Chamberlain AM, et al. CHA_2DS_2-VASc score and stroke prediction in atrial fibrillation in whites, blacks, and hispanics. Stroke. 2019;50(1):28-33. doi: 10.1161/STROKEAHA.118.021453. PMID: 30580712.

Chen Q, Ayer T, Nastoupil LJ, et al. Population-specific prognostic models are needed to stratify outcomes for African-Americans with diffuse large B-cell lymphoma. Leuk Lymphoma. 2016 Apr;57(4):842-51. doi: 10.3109/10428194.2015.1083098. PMID: 26415108.

Cohen JL, Thompson E, Sinvani L, et al. Assessment of warfarin algorithms for hospitalized adults: searching for a safe dosing strategy. J Thromb Thrombolysis. 2019 Nov;48(4):570-9. doi: 10.1007/s11239-019-01902-0. PMID: 31228039.

Delanaye P, Vidal-Petiot E, Björk J, et al. Performance of creatinine-based equations to estimate glomerular filtration rate in White and Black populations in Europe, Brazil, and Africa. Nephrol Dial Transplant. 2023 Jan;38(1):106-18. doi: 10.1093/ndt/gfac241. PMID: 36002032.

Deng J, Vozmediano V, Rodriguez M, et al. Genotype-guided dosing of warfarin through modeling and simulation. Eur J Pharm Sci. 2017 Nov;109S:S9-S14. doi: 10.1016/j.ejps.2017.05.017. PMID: 28502675.

Goodson DA, Chalupsky MR, Wiegley N, et al. GFR estimation in potential living kidney donors: race- and nonrace-based equations and measured GFR. Kidney Medicine. 2022 Dec;4(12):100558. doi: 10.1016/j.xkme.2022.100558. PMID: 36471819.

Guo D, Wang C, Wang B, et al. Learning fair representations via distance correlation minimization. IEEE Trans Neural Netw Learn Syst. 2022 Aug. Online ahead of print. doi: 10.1109/TNNLS.2022.3187165. PMID: 35969542.

Hong C, Pencina MJ, Wojdyla DM, et al. Predictive accuracy of stroke risk prediction models across Black and White race, sex, and age groups. JAMA. 2023 Jan;329(4):306-17. doi: 10.1001/jama.2022.24683. PMID: 36692561.

Howell CR, Zhang L, Yi N, et al. Race versus social determinants of health in COVID-19 hospitalization prediction. Am J Prev Med. 2022 Jul;63(1 Suppl 1):S103-S8. doi: 10.1016/j.amepre.2022.01.034. PMID: 35725136.

Im EO, Chee W. The DSCP-CA: a decision support computer program-cancer pain management. Comput Inform Nurs. 2011 May;29(5):289-96. doi: 10.1097/NCN.0b013e3181f9dd23. PMID: 20975538.

Ku E, Amaral S, McCulloch CE, et al. Comparison of 2021 CKD-EPI equations for estimating racial differences in preemptive waitlisting for kidney transplantation. Clin J Am Soc Nephrol. 2022 Oct;17(10):1515-21. doi: 10.2215/CJN.04850422. PMID: 36122938.

Li F, Wu P, Ong HH, et al. Evaluating and mitigating bias in machine learning models for cardiovascular disease prediction. J Biomed Inform. 2023 Feb;138:104294. doi: 10.1016/j.jbi.2023.104294. PMID: 36706849.

Lin YC, Mallia D, Clark-Sevilla AO, et al. Preeclampsia predictor with machine learning: a comprehensive and bias-free machine learning pipeline [Preprint]. medRxiv. 2022 Jun. doi: 10.1101/2022.06.08.22276107.

Márquez-Luna C, Loh PR, Price AL, et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol. 2017 Dec;41(8):811-23. doi: 10.1002/gepi.22083. PMID: 29110330.

Masekela R, Hall GL, Stanojevic S, et al. An urgent need for African spirometry reference equations: the paediatric and adult African spirometry study. Int J Tuberc Lung Dis. 2019 Aug;23(8):952-8. doi: 10.5588/ijtld.18.0442. PMID: 31533886.

Mifsud F, Saint-Martin C, Dubois-Laforgue D, et al. Monogenic diabetes in adults: a multi-ancestry study reveals strong disparities in diagnosis rates and clinical presentation. Diabetes Res Clin Pract. 2022 May;188:109908. doi: 10.1016/j.diabres.2022.109908. PMID: 35533745.

Ng DK, Furth SL, Warady BA, et al. Self-reported race, serum creatinine, cystatin C, and GFR in children and young adults with pediatric kidney diseases: a report from the chronic kidney disease in children (CKiD) study. Am J Kidney Dis. 2022 Aug;80(2):174-85.e1. doi: 10.1053/j.ajkd.2021.10.013. PMID: 34974031.

Park JI, Bozkurt S, Park JW, et al. Evaluation of race/ethnicity-specific survival machine learning models for Hispanic and Black patients with breast cancer. BMJ Health Care Inform. 2023 Jan;30(1):e100666. doi: 10.1136/bmjhci-2022-100666. PMID: 36653067.

Pasquinelli MM, Tammemägi MC, Kovitz KL, et al. Risk prediction model versus United States Preventive Services Task Force lung cancer screening eligibility criteria: reducing race disparities. J Thorac Oncol. 2020 Nov;15(11):1738-47. doi: 10.1016/j.jtho.2020.08.006. PMID: 32822843.

Pei L, Tian X, Long Y, et al. Establishment of a Han Chinese-specific pharmacogenetic-guided warfarin dosing algorithm. Medicine (Baltimore). 2018 Sep;97(36):e12178. doi: 10.1097/MD.0000000000012178. PMID: 30200121.

Perez Alday EA, Rad AB, Reyna MA, et al. Age, sex and race bias in automated arrhythmia detectors. J Electrocardiol. 2022 Sep-Oct;74:5-9. doi: 10.1016/j.jelectrocard.2022.07.007. PMID: 35878534.

Poventud-Fuentes I, Garnett E, Akcan-Arikan A, et al. Comparison of cystatin C and creatinine-based equations with measured glomerular filtration rate in a diverse pediatric population. J Appl Lab Med. 2022 Sep;7(5):1016-24. doi: 10.1093/jalm/jfac043. PMID: 35671191.

Reeves M, Bhat HS, Goldman-Mellor S. Resampling to address inequities in predictive modeling of suicide deaths. BMJ Health Care Inform. 2022 Apr;29(1):e100456. doi: 10.1136/bmjhci-2021-100456. PMID: 35396246.

Rocco MV, Chapman A, Chertow GM, et al. Chronic kidney disease classification in systolic blood pressure intervention trial: comparison using modification of diet in renal disease and CKD-epidemiology collaboration definitions. Am J Nephrol. 2016 Sep;44(2):130-40. doi: 10.1159/000448722. PMID: 27513312.

Sajal IH, Chowdhury M, Wang T, et al. CBCRisk-Black: a personalized contralateral breast cancer risk prediction model for black women. Breast Cancer Res Treat. 2022 Jul;194(1):179-86. doi: 10.1007/s10549-022-06612-5. PMID: 35562619.

Segar MW, Hall JL, Jhund PS, et al. Machine learning-based models incorporating social determinants of health vs traditional models for predicting in-hospital mortality in patients with heart failure. JAMA Cardiol. 2022 Aug;7(8):844-54. doi: 10.1001/jamacardio.2022.1900. PMID: 35793094.

Shaikh N, Lee M, 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 Jun;176(6):569-75. doi: 10.1001/jamapediatrics.2022.0700. PMID: 35435935.

Shin J, Kayser SR. Accuracy of the pharmacogenetic dosing table in the warfarin label in predicting initial therapeutic warfarin doses in a large, racially diverse cohort. Pharmacotherapy. 2011 Sep;31(9):863-70. doi: 10.1592/phco.31.9.863. PMID: 21923587.

Starlard-Davenport A, Allman R, Dite GS, et al. Validation of a genetic risk score for Arkansas women of color. PLoS ONE. 2018 Oct;13(10):e0204834. doi: 10.1371/journal.pone.0204834. PMID: 30281645.

Stevens LA, Claybon MA, Schmid CH, et al. Evaluation of the chronic kidney disease epidemiology collaboration equation for estimating the glomerular filtration rate in multiple ethnicities. Kidney Int. 2011 Mar;79(5):555-62. doi: 10.1038/ki.2010.462. PMID: 21107446.

Syn NLX, Lee SC, Brunham LR, et al. Pharmacogenetic versus clinical dosing of warfarin in individuals of Chinese and African-American ancestry: assessment using data simulation. Pharmacogenet Genomics. 2015 Sep;25(10):491-500. doi: 10.1097/FPC.0000000000000165. PMID: 26230382.

Teo BW, Koh YY, Toh QC, et al. Performance of the CKD-EPI creatinine-cystatin C glomerular filtration rate estimation equations in a multiethnic Asian population. Singapore Med J. 2014 Dec;55(12):656-9. doi: 10.11622/smedj.2014181. PMID: 25630321.

Wang D, Willis DR, Yih Y. The pneumonia severity index: assessment and comparison to popular machine learning classifiers [Preprint]. medRxiv. 2021 Dec. doi: 10.1101/2021.12.06.21267390.

Weissman GE, Teeple S, Eneanya ND, et al. Effects of neighborhood-level data on performance and algorithmic equity of a model that predicts 30-day heart failure readmissions at an urban academic medical center. J Card Fail. 2021 Sep;27(9):965-73. doi: 10.1016/j.cardfail.2021.04.021. PMID: 34048918.

Wiley LK, Vanhouten JP, Samuels DC, et al. Strategies for equitable pharmacogenomic-guided warfarin dosing among European and African American individuals in a clinical population. Pac Symp Biocomput. 2017;22:545-56. doi: 10.1142/9789813207813_0050. PMID: 27897005.

Williams RM, Kareff SA, Sackstein P, et al. Race & sex disparities related to low-dose computed tomography lung cancer screening eligibility criteria: a lung cancer cases review. Lung Cancer. 2022 Jul;169:55-60. doi: 10.1016/j.lungcan.2022.05.008. PMID: 35644087.

Witonsky J, Elhawary JR, Eng C, et al. Race- and ethnicity-based spirometry reference equations: are they accurate for genetically admixed children? Chest. 2022 Jul;162(1):184-95. doi: 10.1016/j.chest.2021.12.664. PMID: 35033507.

Does not report outcomes by race/ethnicity

Bundy JD, Rahman M, Matsushita K, et al. Risk prediction models for atherosclerotic cardiovascular disease in patients with chronic kidney disease: the CRIC study. J Am Soc Nephrol. 2022 Mar;33(3):601-11. doi: 10.1681/ASN.2021060747. PMID: 35145041.

DeFilippis AP, Young R, Carrubba CJ, et al. An analysis of calibration and discrimination among multiple cardiovascular risk scores in a modern multiethnic cohort. Ann Intern Med. 2015 Feb;162(4):266-75. doi: 10.7326/M14-1281. PMID: 25686167.

Ekladious SMM, Issac MSM, Sharaf S, et al. Validation of a proposed warfarin dosing algorithm based on the genetic make-up of Egyptian patients. Mol Diagn Ther. 2013 Dec;17(6):381-90. doi: 10.1007/s40291-013-0046-3. PMID: 23839801.

Fan L, Levey AS, Gudnason V, et al. Comparing GFR estimating equations using cystatin C and creatinine in elderly individuals. J Am Soc Nephrol. 2015 Aug;26(8):1982-9. doi: 10.1681/ASN.2014060607. PMID: 25527647.

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-.e7. doi: 10.1016/j.ajog.2021.05.021. PMID: 34043983.

Praditpornsilpa K, Townamchai N, Chaiwatanarat T, et al. The need for robust validation for MDRD-based glomerular filtration rate estimation in various CKD populations. Nephrol Dial Transplant. 2011 Sep;26(9):2780-5. doi: 10.1093/ndt/gfq815. PMID: 21357214.

Not a full-length primary study (e.g., narrative review, commentary)

Benjamin R. Assessing risk, automating racism: a health care algorithm reflects underlying racial bias in society. Science. 2019 Oct;366(6464):421-2. doi: 10.1126/science.aaz3873. PMID: 31649182.

Boulware LE, Purnell TS, Mohottige D. Systemic kidney transplant inequities for Black individuals: examining the contribution of racialized kidney function estimating equations. JAMA Netw Open. 2021 Jan;4(1):e2034630. doi: 10.1001/jamanetworkopen.2020.34630. PMID: 33443577.

Bozkurt S, Cahan EM, Seneviratne MG, et al. Reporting of demographic data and representativeness in machine learning models using electronic health records. J Am Med Inform Assoc. 2020 Dec;27(12):1878-84. doi: 10.1093/jamia/ocaa164. PMID: 32935131.

Bradley P, Ryan D, Craig C, et al. 62 - The use of PLCOm2012 vs PLCOm2012noRace risk prediction models in a UK lung cancer screening programme. Lung Cancer. 2021 Jun;156(Suppl 1):S25. doi: 10.1016/S0169-5002(21)00260-9.

Byrne L, Toland AE. Polygenic risk scores in prostate cancer risk assessment and screening. Urol Clin North Am. 2021 Aug;48(3):387-99. doi: 10.1016/j.ucl.2021.03.007. PMID: 34210493.

Cahan EM, Hernandez-Boussard T, Thadaney-Israni S, et al. Putting the data before the algorithm in big data addressing personalized healthcare. NPJ Digit Med. 2019;2(1):78. doi: 10.1038/s41746-019-0157-2. PMID: 31453373.

Delgado C, Baweja M, Crews DC, et al. A unifying approach for GFR estimation: recommendations of the NKF-ASN task force on reassessing the inclusion of race in diagnosing kidney disease. Am J Kidney Dis. 2022 Feb;79(2):268-88.e1. doi: 10.1053/j.ajkd.2021.08.003. PMID: 34563581.

Han Y, Zhang X, Bragg-Gresham J, et al. 113 The impact of removing the EGFR race coefficient on CKD care practices among Black Veterans receiving care in US. Am J Kidney Dis. 2021 Apr;77(4):602. doi: 10.1053/j.ajkd.2021.02.118.

Hanaway MJ, MacLennan PA, Locke JE. Exacerbating racial disparities in kidney transplant: the consequences of geographic redistribution. JAMA Surg. 2020 Aug;155(8):679-81. doi: 10.1001/jamasurg.2020.1455. PMID: 32492119.

Hauspurg A. Reducing maternal morbidity and racial disparities in hypertension care using an automated care pathway. Obstet Gynecol. 2021 Feb;137(2):209-10. doi: 10.1097/AOG.0000000000004265. PMID: 33416282.

Lennon JC. Machine learning algorithms for suicide risk: a premature arms race? Gen Psychiatr. 2020;33(6):e100269. doi: 10.1136/gpsych-2020-100269. PMID: 33089067.

Lucas A, Wyatt CM, Inker LA. Removing race from GFR estimates: balancing potential benefits and unintended consequences. Kidney Int. 2021 Jul;100(1):11-3. doi: 10.1016/j.kint.2021.02.017. PMID: 33647323.

Medlock S, Ravelli ACJ, Tamminga P, et al. Prediction of mortality in very premature infants: a systematic review of prediction models. PLoS ONE. 2011 Sep;6(9):e23441. doi: 10.1371/journal.pone.0023441. PMID: 21931598.

Obeng AO, Kaszemacher T, Abul-Husn NS, et al. Implementing algorithm-guided warfarin dosing in an ethnically diverse patient population using electronic health records and preemptive CYP2C9 and VKORC1 genetic testing. Clin Pharmacol Ther. 2016 Nov;100(5):427-30. doi: 10.1002/cpt.425. PMID: 27393744.

Osei-Anto HA, Greising CH. Using patient data to provide equitable care. Hosp Health Netw. 2011 May;85(5):63. https://www.researchgate.net/publication/51227748_Using_patient_data_to_provide_equitable_care. PMID: 21682247.

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):1429-30. doi: 10.1007/s11606-018-4475-x. PMID: 29766380.

Price ET. Warfarin pharmacogenomics and African ancestry. Blood. 2015 Jul;126(4):434-6. doi: 10.1182/blood-2015-06-649509. PMID: 26206946.

Ridker PM, Cook NR. The pooled cohort equations 3 years on: building a stronger foundation. Circulation. 2016 Dec;134(23):1789-91. doi: 10.1161/CIRCULATIONAHA.116.024246. PMID: 27920070.

Sarkar R, Martin C, Mattie H, et al. Performance of intensive care unit severity scoring systems across different ethnicities [Preprint]. medRxiv. 2021 Jan. doi: 10.1101/2021.01.19.21249222. PMID: 33501459.

Straw I, Callison-Burch C. Artificial intelligence in mental health and the biases of language based models. PLoS ONE. 2020 Dec;15(12):e0240376. doi: 10.1371/journal.pone.0240376. PMID: 33332380.

Suarez-Kurtz G, Botton MR. Pharmacogenomics of warfarin in populations of African descent. Br J Clin Pharmacol. 2013 Feb;75(2):334-46. doi: 10.1111/j.1365-2125.2012.04354.x. PMID: 22676711.

Yoon G, Matulewicz RS, Major VJ. Generalizing an antibiotic recommendation algorithm for treatment of urinary tract infections to an urban academic medical center. J Urol. 2022 Aug;208(2):234-6. doi: 10.1097/JU.0000000000002672. PMID: 35344386.

Zhao Y, Wood EP, Mirin N, et al. Social determinants in machine learning cardiovascular disease prediction models: a systematic review. Am J Prev Med. 2021 Oct;61(4):596-605. doi: 10.1016/j.amepre.2021.04.016. PMID: 34544559.

Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine. 2021 May;67:103358. doi: 10.1016/j.ebiom.2021.103358. PMID: 33962897.

Appendix C. Characteristics of Key Question 1 and 2 Studies

Table C-1. Characteristics of studies addressing Key Question 1

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Ashana et al. 2021[1] *Also addressed KQ2	Assess the performance of the Sequential Organ Failure Assessment (SOFA) score and LAPS2 among Black and White patients admitted through the emergency department (ED) with sepsis or acute respiratory failure (ARF).	SOFA score Laboratory-based Acute Physiology Score version 2 (LAPS2) Mitigation strategy: modified versions of SOFA and LAPS2	Compared original and modified versions of algorithms	**SOFA score:** Composed of organ function scores from 6 organ systems (cardiovascular, respiratory, hepatic, renal, coagulation, neurological). In this study, the renal subscore was calculated using creatinine alone, and the highest value of each subscore during the patient's ED stay was used to calculate total score (continuous variable 0 to 24 points, with higher values representing greater illness severity). **SOFA score modifications:** 1) divided score into 4 categories (<6, 6 to 8, 9 to 11, ≥ 12) 2) subtracted one-half point from renal subscore for Black patients whose raw renal subscore >0 3) eliminated renal subscore **LAPS2:** 2-stage algorithm in which patients are first stratified into low- and high-mortality-risk groups, and then vital signs and laboratory values are added to the algorithm. Total score is based on risk stratum and most deranged laboratory value and ranges from 0 to 414 (continuous variable, scores > 200 uncommon). **LAPS2 modifications:** 1) continuous LAPS2 divided into 8 categories 2) continuous LAPS2 divided into 4 equal categories	Modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated EDs of academic medical centers	Patients admitted for sepsis or ARF at 27 hospitals (Kaiser Permanente Northern California and Penn Medicine) between 2013 and 2018. Patients were ≥ 18 years with sepsis at all hospitals and ARF at Penn Medicine hospitals and were admitted from the ED to an inpatient location. Study does not report how race/ethnicity was defined (e.g., self-reported)	**Patients (n):** 113,158 **Race/Ethnicity:** 75.6% White; 24.4% Black **Mean Age (SD):** 67.7 (15.2) White; 61.7 (16.6) Black p<0.001 **Sex:** White 54.1% male and 45.9% female; Black 48.2% male and 51.8% female p<0.001 for % female between groups

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Boley et al. 2022[2]	Examine impact of a rapid triage fast-track (FT) model on outcomes in Black non-Hispanic and White non-Hispanic patients presenting to the ED.	Rapid triage fast-track (FT) model Providers assign an emergency severity index (ESI) score, then determine whether patients meet additional criteria for FT or main ED status.	No comparator. Compared triage process outcomes by racial groups.	**Triage process:** Nurse checks in patient, determines chief complaint, and obtains set of vital signs. Nurse applies ESI protocol to give patient a score of 1 (most acute) to 5 (least acute). Nurse then determines whether patient is appropriate for FT based on the following requirements: 1. Patient is able to sit in a recliner. 2. Patient is ambulatory and able to speak. 3. Patient's ESI score is 3, 4, or 5 (lowest acuity). 4. Patient is determined to be not critical based on the triage determination Patients identified for FT are seen in a separate 5-bed area of ED. Patients not eligible for FT wait until an ED bed becomes available. FT status: patients placed in a separate ED section and can receive intravenous fluids, medications, and laboratory and radiology tests. A physician assistant, nurse practitioner, nurse, or ED technician provides care, but an ED physician can be involved if needed (e.g., case's complexity). Main ED status: patients wait until an ED bed is available. Care typically provided by ED physician.	Retrospec-tive matched cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Tertiary care hospital (556-bed)	EHR ED encounters during a 1-year period (2/1/2019 to 1/31/2020) after full implementation of rapid triage system. Black and White patients were exact-matched on potential confounders including presence of abnormal vital signs. Race and ethnicity collected as separate measures in the EHR based on patient self-reporting and patients can report multiple races. Race and ethnicity not included as an input variable.	**Patients (n):** 9704 with 12,330 unique encounters (5151 Black Non-Hispanic encounters; 7170 White Non-Hispanic encounters) **Race/Ethnicity:** 58.2% White Non-Hispanic encounters; 41.7% Black Non-Hispanic encounters **Mean Age (SD):** 37.4 (36.9) White Non-Hispanic; 36.9 (13.2) Black Non-Hispanic **Sex:** 70.4% female; 29.6% male

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Carbuna-ru et al. 2019[3]	Compare the frequency of avoided biopsies and missed clinically significant prostate cancer (csPCa) resulting from use of 2 risk prediction algorithms across racial groups in an urban, multi-racial cohort.	Prostate Cancer Prevention Trial Risk Calculator 2.0 (PCPT RC) Prostate Biopsy Collaborative Group (PBCG) RC	Compared algorithms	Input variables include prostate-specific antigen (PSA) level, digital rectal exam (DRE) result, first-degree family history of PCa (father, brother or son ever diagnosed with PCa) and history of a prior negative prostate biopsy. Both algorithms "take race into consideration".	Retrospec-tive (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Urology clinics	The sample consisted of consecutive ambulatory patients from urology clinics at 2 privately funded and 3 publicly funded institutions who were undergoing their first prostate biopsy for an abnormal PSA level or digital rectal exam (DRE). Data obtained from a prospectively maintained dataset. Race/ethnicity was self-reported.	**Patients (n)**: 954 **Race/Ethnicity**: Black (463, 48.5%), white (355, 37.2%), Other race (136, 14.2%). The Other group included Hispanic (n = 103, 75.7%), Asian (n = 28) and Middle Eastern men (n = 5). **Median Age (IQR):** Black, 61 (57, 67); white, 62 (58, 67); other, 62 (57, 67) **Sex**: 100% male

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Han et al. 2020[4]	Characterize individuals who would be selected for lung cancer screening based on risk factors but would not be recommend-ed for screening based on the current USPSTF guidelines.	PLCOm2012 Model, which predicts 6-year risk of lung cancer based on demographic, environmental, and clinical risk factors.	Current USPSTF guidelines (annual low-dose computed tomogra-phy screening of individuals aged 55–80 years with at least 30 pack-years of smoking and within 15 years since cessation)	Age, race/ethnicity, education, body mass index (BMI), chronic obstructive pulmonary disease (COPD), personal history of cancer, family history of lung cancer, and smoking status	Simulation study (modeling using synthetic data – source data for calculation of algorithmic scores is synthetic data and outcomes that would have resulted from using the algorithm are simulated) 1950 U.S. birth cohort aged 50-90 years	Analyses were performed on a simulated dataset of 100,000 individuals in the 1950 U.S. birth cohort, containing: (a) smoking history data generated by the CISNET Smoking History Generator based on data from the NHIS, Cancer Prevention studies I and II, and the Human Mortality Database, and (b) risk factor data generated by the LC Risk Factor Generator based on data from the NHIS, PLCO trial, U.S. Census Bureau, and NHANES. Sensitivity analyses were performed on a similar dataset representing the 1960 U.S. birth cohort. Race/ethnicity data generated from U.S. Census Bureau data.	1950 birth cohort: **Patients (n)**: 100,000 **Race/Ethnicity**: White (76%), Black (10%), Hispanic (8%), Asian (5%) **Age**: Not reported **Sex**: Not reported

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Metzger et al. 2022[5]	Assess the association of race and language with ED triage scores.	ESI score	No comparator. Compared triage scores by racial groups.	A 5-level triage algorithm ESI 1 (Immediate medical attention). Study excluded visits with an ESI score of 1. ESI 2 (Emergency) ESI 3 (Urgent) ESI 4 (Nonurgent) ESI 5 (Minor)	Retrospec-tive cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Pediatric ED	EHR data from July 2015 to June 2016 for patients aged 0 to 17 years. Study does not indicate whether race and ethnicity was self-reported. Patients were categorized as Non-Hispanic White if White/Caucasian and non-Hispanic ethnicity were reported. Patients were categorized as non-White if any other race was reported (Black/African American, Asian, American Indian/Alaska Native, Native Hawaiian/Pacific Islander, other) or if an ethnicity of Hispanic was reported. Race and ethnicity not included as an input variable.	**Patients (n):** 8928 (3086 Non-Hispanic White; 5842 Non-White) with 10,815 visits (3538 Non-Hispanic White; 7277 Non-White) **Race/Ethnicity:** 34.6% Non-Hispanic White; 65.4% Non- White (1.2% American Indian/Alaska Native, 14.6% Asian, 23.8% Black, 2.5% Native Hawaiian/Pacific Islander, 38.7% other, 11% more than 1 race) **Median Age (Months at Visit; IQR):** 39.2 (14.1 to 88.5) Non-Hispanic White; 33.4 (13.7 to 74.2) Non-White **Sex:** 46.8% female; 53.2% male

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Miller et al. 2021[6]	Investigate whether using the SOFA is associated with deprioritiza-tion of Black patients in currently adopted crisis standards of care (CSC).	SOFA -- continuous variable used to predict in-hospital mortality risk, scored from 0 (lowest risk) to 24 (highest risk). Scores are collapsed into tiers for the purpose of prioritizing resources to patients most likely to survive with appropriate care when resources are overwhelmed. This study examined 3 tiering systems, termed A (4 tiers, with scores <6 forming the highest-priority tier and scores ≥12 forming the lowest), B (3 tiers, scores <8 highest priority, ≥12 lowest), and C (4 tiers, scores <9 highest priority, ≥15 lowest).	The authors quantified how much the SOFA threshold required for inclusion in a priority tier would have to be increased for Black patients so that mortality would be equivalent for Black and White patients eligible for resource allocation.	Blood pressure, hypoxemia, creatinine, bilirubin, platelet count, and the Glasgow Coma Scale.	Modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated ICUs	The SOFA score was developed using a consensus-based process. In this study, the data source is the eICU Collaborative Research Database, a cohort of patients admitted to ICUs in 208 U.S. hospitals from 2014 to 2015. Eligibility criteria: age ≥18 years; Black or white race; at least 1 SOFA variable recorded within 24 hours of ICU admission; in-hospital mortality documented. Records representing the first ICU stay during a hospitalization were included. Study does not report how race/ethnicity was defined (e.g., self-reported).	**Patients (n):** 111,885 patient encounters for 95,549 unique patients **Race/Ethnicity**: 16,688 encounters with Black patients (14.9%) and 95,197 encounters with White patients (85.1%) **Mean (SD) age**: 63.3 (16.9) years **Sex**: 51,464 encounters with women (46.0%)

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Obermeyer et al. 2019[7] *Also addressed KQ2	Quantify racial disparities in health care resource allocation produced by a widely used commercial risk prediction algorithm.	The algorithm is used to predict complex health needs in primary care patients; the goal is to direct additional resources to such patients, based on the assumption that they will benefit the most from them. The original algorithm predicts costs over the following year. In the health system studied here, patients scoring above the 97th percentile are automatically identified for enrollment into the system's care management program. For those above the 55th percentile, their primary care physician is asked whether they would benefit from the program.	The authors developed and internally validated 3 new algorithms, predicting different outcomes or "labels" (total costs, avoidable costs, and health).	Features of raw insurance claims data from the previous year, including age, sex, insurance type, diagnosis and procedure codes, medications, and detailed costs. Race is not an input variable in the original algorithm or in the 3 new algorithms.	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Hospital	Development dataset for original algorithm is not specified. In this study, the data source is all primary care patients enrolled in risk-based contracts at a large academic hospital from 2013 to 2015 and self-identifying as either Black or as white without another race or ethnicity.	**Patients (n):** 49,618 **Race/Ethnicity**: 87.7% White, 12.3% Black **Mean Age**: 51.3 (White), 48.6 (Black) **Sex**: 62% female (White), 69% female (Black)

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Pasquinelli et al. 2021[8]	Compare 2 different lung cancer screening criteria, USPSTF 2013 and PLCOm2012.	PLCOm2012, a validated logistic regression lung cancer risk prediction model	USPSTF 2013: criteria based on findings of the National Lung Screening Trial (NLST). The Task Force recom-mends low-dose computed tomogra-phy for individuals who meet NLST-like eligibility.	**USPSTF 2013:** NLST-like eligibility criteria include age 55 to 80 years, ≥ 30 pack-year cigarette smoking history, and having quit smoking within the past 15 years. **PLCOm2012:** age, highest level of education obtained, BMI, COPD, personal history of cancer, family history of lung cancer, race and ethnicity, smoking status (former or current), average number of cigarettes smoked per day, duration smoked, years of quitting smoking	Retrospective study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Urban academic medical center (13 federally qualified health centers)	Lung cancer cohort at the University of Illinois Hospital and Health Sciences System between 2010 and 2019. Data collected up until March 15, 2020. Study does not report how race/ethnicity was defined (e.g., self-reported)	**Patients (n):** 883 **Race/Ethnicity:** 56.3% African American; 29.2% White; 7.8% Hispanic; 2.7% Asian; 4.0% Other or missing **Mean Age (SD):** 64.8 (9.4) **Sex:** 55.8% male; 44.2% female

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Presti et al. 2021[9]	Externally validate a newly developed prostate cancer risk prediction algorithm, and compare with 2 other calculators.	Prostate cancer risk prediction Kaiser Permanente Prostate Cancer Risk Calculator (KPPC RC) (range: 0% to 100%)	Compared 2 versions of the algorithm	Version A: age, race (patient-reported), BMI, family history of prostate cancer, number of prior biopsies, PSA level, DRE result Version B: Version A variables plus prostate volume (The study also examined a version that did not include DRE result or prostate volume, but did not report results by race or ethnicity for that version.)	Retrospective Large integrated health care system (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Validation (this study): Kaiser Permanente Northern California (KPNC) All men with no prior diagnosis of prostate cancer who underwent prostate biopsy at any of 21 KPNC urology departments between 12/2017 and 8/2019 for either an abnormal DRE and/or an elevated PSA and had complete data on all analysis variables. Prospective data collection methods used to capture biopsies during this time. Study does not report how race/ethnicity was identified.	**Patients (n):** 4178 (5353 men underwent biopsy; 4178 had complete data) **Race/Ethnicity**: 56.7% Caucasian, 11.3% African American, 16.2% Asian/Pacific Islander, 13.5% Hispanic, 2.3% other **Median Age (IQR)**: 63 (57-67) **Sex**: 100% male

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Riviello et al. 2022[10]	Analyze the association of Crisis Standards of Care (CSC) scoring system with resource prioritization and estimated excess mortality by race, ethnicity, and residence in a socially vulnerable area.	CSC scoring system	Compared outcomes by racial and ethnic groups and compared CSC and random allocation lottery in a simulated model to estimate number of excess deaths.	**CSC scoring system:** aggregate score outlined by Commonwealth of Massachusetts for application in individual hospitals. Score based on points derived from SOFA score and a chronic severity of illness score based on comorbidities or a life expectancy score based on physician assessment. SOFA converted into a 4-point scale: 1 for SOFA < 6, 2 for SOFA 6 to 9, 3 for SOFA 10 to 12, and 4 for SOFA>12. Comorbidities based on a 3-level system: 0 points no significant comorbidities, 2 points major comorbid conditions with substantial impact on long-term survival, 4 points severely life-limiting conditions prior to acute illness. Life expectancy based on a 3-level score: 0 points death not likely in 5 years, 2 points death likely within 5 years, and 4 points death likely within 1 year. Points totaled to create raw ordinal priority score from 1 to 8. Highest scores 1 to 2, intermediate 3 to 5, and lowest 6 to 8. Highest-scores first to receive scarce critical care resources, then intermediate scores, then lowest scores. **Simulation:** Simulation of mortality outcomes using CSOC score vs random lottery in a subset of patients receiving ventilation. Created a scenario of scarcity requiring allocation of ventilators using 2 state-recommended cutoff scores of ≤ 2 (highest-priority category patients receive ventilator) and ≤ 5 (both highest and intermediate-priority category patients receive ventilator). Authors ran 10,000 trials randomly assigning individuals to receive a ventilator.	Retrospective cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) ICUs of 6 Boston-area tertiary and community hospitals in Beth Israel and Lahey Hospital systems	EHR data: April 13, 2020, to May 22, 2020. April 28, 2020, hospitals used an estimate of life expectancy instead of comorbidities in response to revised guidelines which required collection of additional data from EHR (discharge dates, vital status at discharge, discharged destination). Attending physicians assessed likelihood that a patient would survive past 1 or 5 years based on baseline health status at time of ICU admission. Race and ethnicity self-reported or reported by patient surrogate and recorded in medical record. Race categorized as other for any self-reported race that was not White, Black, or Asian. Race and ethnicity listed unknown if self-report not recorded. Race and ethnicity not included as an input variable.	**Patients (n):** 498 (79 Black, 298 White, 11 Asian, 46 Other, 64 Unknown) **Race/Ethnicity (%):** 15.8% Black, 59.8% White, 2.2% Asian, 9.2% Other, 12.8% Unknown **Median Age (IQR), years:** 67 (56 to 75) Black 68 (59 to 75), White 69 (57 to 76), Asian 62 (59 to 72), Other 63 (52 to 73), Unknown 59 (50 to 69) **Sex:** 38,4% female; 61.6% male Black (32.9% female, 67,1% male), White (39.3% female, 60.7% male), Asian (36.4% female, 63.6% male), Other (32.6% female, 67.4% male), Unknown (45.3% female, 54.7% male) **Subgroup (n):** 244 (16.8% Black, 49.2% White, 2.9% Asian, 10.7% other, 20.5% unknown) **Race/Ethnicity (%):** 16.8% Black, 49.1% White, 2.8% Asian, 10.6% other, 20.4% unknown)

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Sarkar et al. 2021[11]	Examine the performance of 3 severity scoring models.	APACHE IVa generates a risk score for hospital, ICU mortality, and length of stay. OASIS predicts hospital mortality and ICU mortality of critically ill patients SOFA characterizes severity state in sepsis but has been used to predict patient outcomes	Compared models	**APACHE IVa:** 142 patient variables including 116 admission categories and 17 acute physiologic parameters (65.9% of score and includes age, chronic health condition, underlying diagnosis, ventilation status). **OASIS:** 10 variables collected in first 24 hours of ICU stay (heart rate, mean arterial pressure, temperature, respiratory rate, urine output, pre-ICU admission length of stay, GCS, age, being placed on a mechanical ventilator at any point during day 1 and admission following elective surgery). **SOFA:** composed of organ function scores from 6 organ systems (cardiovascular, respiratory, hepatic, renal, coagulation, neurological). SOFA categories based on proposed categories for COVID-19 ventilator allocation.	Modeling study using real-world data to determine effect of illness severity scores on CSC allocation (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) ICU admissions	**eICU-Collaborative Research Database (eICU-CRD)** includes > 200,000 discharge episodes across 335 ICUs at 208 hospitals between 2014 and 2015. Data available includes age, sex, ethnicity, vital signs, diagnoses, laboratory measurements, clinical history, problem lists, APACHE IVa scores, and treatments. **Medical Information Mart for Intensive Care-III database** consists of >60,000 ICU admissions to Beth Israel Deaconess Medical Center between 2001 and 2012 and includes OASIS as a mortality prediction model.	**eICU-CRD** **Patients (n):** 122,919 **Race/Ethnicity:** 81.9% White; 12.4% African American; 4.1% Hispanic; 1.5% Asian **Median Age (IQR):** 64 (52 to 75) **Sex:** 54% male; 46% female **MIMIC-III** **Patients (n):** 43,823 **Race/Ethnicity:** 82.14% White; 11.07% African American; 4.07% Hispanic; 2.71% Asian **Median Age (IQR):** 64.5 (52 to 76) **Sex:** 57% male; 43% female **Note:** race/ethnicity is based on definitions within each database. Information is typically entered by an administrator. Patients are either asked which group they identify with or the group is entered based on available records.

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Snavely et al. 2021[12]	Compare the safety and effectiveness of the HEART pathway among women vs men, and white vs non-white patients, presenting to the ED with acute chest pain.	HEART Pathway provides test ordering and disposition decision support to clinicians and risk-based care planning.	Compared pre- and post-implementation of the HEART Pathway.	HEART Pathway risk assessment is based on the HEAR score (History, ECG, Age, and Risk factor) and 0- and 3-hour troponin measures. HEAR score ≤ 3 without elevated troponin is classified as low-risk and recommended for discharge without objective cardiac testing. HEAR score ≤ 3 with elevated troponin leads to a cardiology consult and admission and/or further observation or testing. HEAR score ≥ 4 with elevated troponin, known coronary artery disease, or ischemic ECG is classified as non-low risk and designated for further testing. HEAR score ≥ 4 without elevated troponin leads to observation/admission and/or cardiology consult or testing. *Authors focused on low-risk (≤ 3) and non-low risk (≥ 4) groups for analysis.	Preplanned subgroup analysis of a prospective pre-post study (pre-post study). 3 EDs in North Carolina (large urban academic medical center, rural medical center, small community hospital).	EHR (Clarity-EPIC systems) index encounter and claims data. Race/ethnicity was self-reported.	**Patients (n):** 3713 pre-implementation; 4,761 post-implementation **Race/Ethnicity Pre-implementation:** 66.9% (2,484) White; 28.3% (1,052) Black or African American; 0.6% (21) Asian, 0.2% (9) American Indian, 0.03% (1) Hawaiian or Pacific Islander, 3.9% Other (145), 0.03% (1) Refused to provide information or Unknown **Race/Ethnicity Post-implementation:** 65.2% (3,106) White, 28.8% (1,371) Black, 0.6% (27) Asian, 0.3% (16) American Indian, 0.04% (2) Hawaiian or Pacific Islander, 4.9% (234) Other, 0.1% (5) Refused to provide information or Unknown. **Median Age:** 54 years **Sex:** 46.4% male; 53.6% female

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Thompson et al. 2021[13] *Also addressed KQ2	Assess fairness and bias of a previously validated machine-learning opioid misuse classifier.	Natural language opioid misuse classifier using a convolutional neural network. Input is electronic health record (EHR) data from a hospitalization. The algorithm's goal is "to provide point-of-care education, treatment options, and care pathways to patients who misuse opioids." Thus, false negatives (Type II errors) represent failures of the model to recommend appropriate resources.	2 post-hoc analyses were performed to mitigate the classifier's bias: (a) the threshold value dividing negative predictions from positive predictions in the subgroup with biased false-negative rate was varied to improve sensitivity without losing specificity, and (b) the classifier was recalibrated by subgroup.	Clinical notes	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Hospital and tertiary care academic center	Development dataset: adult hospital encounters from EHR between 2007 and 2017 at a U.S. hospital and tertiary academic center. Opioid-related hospitalizations were oversampled. "The final dataset …consisted of 367 manually labeled cases, age- and sex-matched with controls that had no indications of opioid misuse." External validation dataset: EHR at a different tertiary care academic center. Dataset included "all unplanned adult inpatient encounters …who were screened between October 23, 2017, and December 31, 2019 (n = 53,974)." Appears race/ethnicity was self-reported.	The analyses reported in this article were carried out on the external validation dataset. **Patients (n)**: 53,794 **Race/Ethnicity**: White (n = 23,345), Black (n = 17,541), Hispanic/Latinx (n = 9252), Other (n = 3836) **Age**: not reported **Sex**: not reported

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Wille et al. 2013[14]	Compare ethnic disparities in lung transplantation rates and time to death on the wait list, before vs after introduction of the Lung Allocation Score (LAS).	In 2005, LAS became the main method for determining allocation of deceased donor lungs for transplantation in the United States. The LAS is a numerical score based on survival models that estimate likelihood of survival both while on the wait list and post-transplant; thus, it reflects the net benefit of transplantation.	In the pre-LAS period, time on the wait list was the sole basis for allocation.	Diagnosis (4 categories), age, height, weight, cardiac index at rest, bilirubin, functional status, PA systolic pressure, O_2 required at rest, six-minute walk distance, continuous mechanical ventilation, PCO_2, increase in PCO_2, creatinine (from https://optn.transplant.hrsa.gov/data/allocation-calculators/las-calculator/)	Retrospective pre/post-implementation study (pre-post study) U.S. health care system	The study population consisted of all White and Black non-Hispanic adults who were listed for lung transplantation during 2 time periods: pre-LAS (January 1, 2000–May 3, 2005) and LAS (May 4, 2005–September 4, 2010). Race/ethnicity was self-reported.	**Patients (n):** 8765 (pre-LAS), 8806 (LAS). **Race/Ethnicity**: White (89.9%), Black (10.1%) **Mean (SD) Age**: Pre-LAS: White, 49.3 (12.6); Black, 47.2 (9.6) LAS: White, 54.0 (13.0); Black, 50.4 (10.5) **Sex**: Pre-LAS: 51.3% female LAS: 45.5% female

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Williams 2022[15]	Compare number eligible for lung cancer screening between USPSTF criteria in 2013 with revised criteria in 2021, and with more detailed criteria from the PLCOm2012 model	USPSTF-2021	USPSTF-2013 and PCLOm2012	Eligibility criteria for USPSTF 2021: Age (50-80), 20+ pack year smoking history, for people who currently smoke or who had quit within the past 15 years	Retrospective application of USPSTF 2013, USPSTF 2021, and PLCOm2012 criteria on 2019 cohort data	The Center for Disease Control and Prevention Behavioral Risk Factor Surveillance System (BRFSS), which is a health-related telephone survey that collects data from more than 400,000 adults annually in 50 states, the District of Columbia, and 3 U.S. territories.	**Patients (n):** 41,544 **Race/Ethnicity:** White-Non-Hispanic (n=36,787); Black Non-Hispanic (n=66); Hispanic (n=786); Other Non-Hispanic (n=1905)

Author/ Year	Study Objective	Algorithm	Control	Component Variables	Study Design Setting	Source of Study Data	Patient Characteristics
Yoo et al. 2023[16]	Evaluate how predictive performance of a clinical calculator affects downstream health outcomes.	CHA_2DS_2-VASc	No comparator. Compared outcomes by racial and ethnic groups *Calculator outputs are used to guide clinical guideline-based care	**CHA_2DS_2-VASc (start at 0):** age 65 to 74 (+1) or > 75 (+2), female (+1), CHF history (+1), HTN history (+1), stroke / TIA / thromboembolism history (+2), vascular disease history (+1), diabetes history (+1). The algorithm informs the American College of Cardiology (ACC)/American Heart Association (AHA) atrial fibrillation treatment guideline. **2014 ACC/AHA guideline recommendation**: *do not* recommend antithrombotic therapy for male patients with a score of 0 or female patients with a score of 1. **2020 ACC/AHA guideline recommendation**: *recommend* antithrombotic therapy for male patients with a score $\geq$ 2 and female patients with a score $\geq$ 3. *Consider* antithrombotic therapy for male patients with a score of 1 and female patients with a score of 2.	Retrospective cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated) Stanford Health Care and Lucile Packard Children's Hospital	Stanford Medicine Research Data Repository (STARR). STARR was linked with the Social Security Administration's Death Master File to determine out-of-hospital deaths. Race and ethnicity self-reported. Study used the 5 U.S. Census Bureau categories and used Hispanic as a dedicated ethnicity. Race and ethnicity are not included as input variables in MELD, CHA_2DS_2-VASc, or sPESI.	**CHA_2DS_2-VASc** **Patients (n):** 233,129 **Race/Ethnicity:** Asian 15% (n=33,927), Black 3% (n=7323), White 76% (n=176,278), Hispanic 6% (n=13,578), Other 1% (n=2,023) **Median Age (25^{th} to 75^{th} percentiles):** 77 years (71 to 83) **Sex:** 56% male (n=129,621), 44% female (n=103,508)

Zhang et al. 2018[17]	Assess impact of the 2014 Kidney Allocation System (KAS) policy change on waitlisting overall and evaluate whether racial/ethnic disparities in waitlisting in the United States changed following implementation.	Kidney allocation KAS was developed to improve equity related to dialysis time and to patients with high panel reactive antibody. Specific changes to the system include a change in the calculation of waiting time and prioritization of the most sensitized patients. Waiting time starts at dialysis start instead of at the time of waitlist.	Pre-post comparison	KAS uses Kidney Donor Profile Index (KDPI) and Expected Post Transplant Survival (EPTS) score for longevity matching between donors and recipients. **KDPI (donor variables):** age, height, weight, ethnicity, history of hypertension, history of diabetes, cause of death, serum creatinine, hepatitis C Virus status, donation after Circulatory Death status (range: 0% to 100%). Options for ethnicity variable: American Indian or Alaska Native, Asian, Black or African American, Hispanic/Latino, Native Hawaiian or Other Pacific Islander, White, or Multi Racial. **EPTS score:** age, time on dialysis, current diabetes status, and if candidate had a previous solid organ transplant (range: 0% to 100%). KAS also incorporates a points system to increase priority for patients with high panel reactive antibody (i.e., patients less likely to find a compatible donor), includes pre-registration dialysis time as part of a candidate's waiting time, provides increased access for candidates with blood type B, uses KDPI scores to inform pediatric priority, and eliminates the payback system (i.e., if an organ was received from another organization, the receiving organization had to pay back an organ to the national pool). *Information from the Organ Procurement & Transplantation Network New KAS FAQs.	Retrospective cohort study (pre-post study) U.S. medical centers	New patients on dialysis and existing patients on dialysis with end-stage renal disease (ESRD) from the United States Renal Data System (USRDS). **Pre-KAS group:** beginning dialysis between 1/1/2005 and 12/03/2014 **Post-KAS:** beginning dialysis between 12/4/2014 and 12/31/2015 The United Network for Organ Sharing system used to collect information about active and inactive status of newly waitlisted patients from 2005 to 2015. Study does not report how race/ethnicity was defined (e.g., self-reported)	**Pre-KAS (incident patients)** **Patients (n):** 1,120,655 **Race/Ethnicity:** 52.1% White; 26.5% Black; 13.7% Hispanic; 4.2% Asian **Age group, N (%):** 18 to 39: 7.5% 40 to 49: 10.7% 50 to 59: 19.9% 60 to 69: 24.9% ≥ 70: 47% **Sex:** 56.8% male; 43.2% female **Post-KAS (incident patients)** **Patients (n):** 132,445 **Race/Ethnicity:** 51% White; 25.1% Black; 13.7% Hispanic; 4.8% Asian **Age group, N (%):** 18 to 39: 7.3% 40 to 49: 10.4% 50 to 59: 19.7% 60 to 69: 27.2% ≥ 70: (35.4%) **Sex:** 58.2% male; 41.8% female **Prevalent Dialysis Cohort** *Patients eligible for first time waitlisting anytime during period (1/1/2005 to 12/31/2015). Baseline characteristics NR for race/ethnicity, age, and sex. **Patients (n)**: 1,556,954

Abbreviations: APACHE IVa = Acute physiology and chronic health evaluation; ARF = acute respiratory failure; BMI = body mass index; CAD=coronary artery disease; CG CrCl = Cockcroft-Gault Creatinine Clearance; CISNET = Cancer intervention and surveillance modeling network; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; CLRD = chronic lower respiratory disease; COPD = chronic obstructive pulmonary disease; CSC = crisis standards of care; csPCa = clinically significant prostate cancer; DRE = digital rectal exam; ECG = electrocardiogram; ED = emergency department; eGFR = estimated glomerular filtration rate; EHR = electronic health record; eICU-CRD = eICU-Collaborative Research Database; ESRD = end-stage renal disease; FAQ = frequently asked question; ICU = intensive care unit; IQR = interquartile range; KAS = kidney allocation system; KDPI = Kidney Donor Profile Index; KPPC RC = Kaiser Permanente prostate cancer risk calculator; LAPS2 = Laboratory-based Acute Physiology Score version 2; LAS = Lung Allocation Score; LYFS-CT = life-years from screening-computed tomography; MIMIC-III = Medical Information Mart for Intensive Care III; NHANES = National health and nutrition examination survey; NHIS = National health interview survey; NSLT = National lung screening trial; OASIS = Oxford Acute Severity of Illness Score; PBCG = Prostate Biopsy Collaborative Group; PCPT RC = Prostate Cancer Prevention Trial risk calculator; PSA = prostate-specific antigen; SD = standard deviation; SOFA = Sequential Organ Failure Assessment; USPSTF = United States preventive services taskforce; USRDS = United States Renal Data System

Table C-2. Characteristics of studies addressing Key Question 2

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Ahmed et al. 2021[18]	Examine the impact of the race coefficient in the CKD-EPI eGFR equation on CKD classification and care delivery.	eGFR (CKD-EPI)	Removed race	Modeling study using cross-sectional data (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Partners Health Care Chronic Kidney Disease Registry data obtained June 2019	**Patients (n):** 56,485 **Race/Ethnicity:** 87% White; 3.9% African American; 2.3% Asian; 1.0% Hispanic; 0.1% Native American; 6.1% Other **Median Age**: White 77; African American 73; Asian 77; 74 Hispanic; Native American 73; Other 76 **Sex:** 56.5% female; 43.5% male

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Ashana et al. 2021[1] *Also addressed KQ1	Assess the performance of the SOFA score and LAPS2 among Black and White patients admitted through the emergency department with sepsis or acute respiratory failure (ARF).	Illness severity prediction models SOFA score Laboratory-based Acute Physiology Score version 2 (LAPS2)	Simulation analysis adjusted category thresholds	Modeling study to evaluate potential effect on severity of illness scores using real-world data (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Patients admitted for sepsis or ARF at 27 hospitals (Kaiser Permanente Northern California and Penn Medicine) between 2013 and 2018 Patients were ≥ 18 years with sepsis at all hospitals and ARF at Penn Medicine hospitals. Study does not report how race/ethnicity was defined (e.g., self-reported)	**Patients (n):** 113,158 **Race/Ethnicity:** 75.6% White; 24.4% Black **Mean Age (SD):** 67.7 (15.2) White; 61.7 (16.6) Black p<0.001 **Sex:** White 54.1% male and 45.9% female; Black 48.2% male and 51.8% female p<0.001 for % female between groups
Baugh et al. 2022[19]	Evaluate how removal of race from algorithms affects measurement of lung function in patients with COPD	Percent predicted forced expiratory volume	Removed race	Modeled potential effect of lung function formulas with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Sub-Populations and Intermediate Outcome Measures In COPD Study (SPIROMICS)	**Patients (n):** 2652 **Race/Ethnicity:** 20% Black **Mean age (SD):** 65 (8.4) White; 58 (8.9) Black **Sex:** 46% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Bundy et al. 2022[20]	Evaluate how removal of race coefficient from eGFR affects the Kidney Failure Risk Equation and prediction of 2-year risk of end-stage kidney disease	eGFR (CKD-EPI)	Removed race	Prospective cohort	Chronic Renal Insufficiency Cohort study (CRIC)	**Patients (n):** 3873 **Race/Ethnicity:** 42.1% Black; 57.9% **Mean age (SD):** 57.8 (10.9) **Sex:** Black 51.2% female; Non-Black 40.9% female
Casal et al. 2021[21]	Evaluate how removal of race coefficient from eGFR affects use of anticancer drugs with kidney function cutoffs.	eGFR (CKD-EPI)	Removed race	Modeled potential effect of CKD-EPI formula with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from National Cancer Institute database of patients enrolled in clinical trials between 1995-2010	**Patients (n):** 340 **Race/Ethnicity:** 100% Black **Median age:** 57 **Sex:** 49% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Coresh et al. 2019[22]	Examine an alternative approach to estimating GFR without using race or measuring creatinine through use of a metabolic panel.	eGFR	GFR estimated from metabolic panel without adjustment for race or use of creatinine, as follows: eGFR=exp(3.04584 - 0.450817*ln(Acetyl-L-Threonine) - 0.214876*ln(Beta-pseudouridine) - 0.253004*ln(Myo-inositol) + 0.2265693*ln(Tryptophan))	Modeling of a cross-sectional comparison of eGFR based on metabolic panel without race to eGFR based on creatinine and race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Derived from African American Study of Kidney (AASK) participants; validated in Multi-Ethnic Study of Atherosclerosis (MESA) participants	**Patients (n):** 265 **Race/Ethnicity:** 46% Black **Mean age (SD):** 71 years (9) **Sex:** 47% female
Diao et al. 2023[23]	Evaluate how removal of race coefficient from eGFR affects diagnosis and staging of kidney disease, eligibility for kidney donation and transplantation, medication dosing, and eligibility for medical services.	eGFR (CKD-EPI 2021 and 2009; MDRD 2006)	Removed race	Modeled potential effect of 2021 CKD-EPI formula without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from NHANES, 2001-2018 The NHANES data were extrapolated to the US population at large.	**Patients (n):** 44,360 after removing patients with age <18 years or censored age; pregnant patients; and patients without serum creatinine reported. **Race/Ethnicity:** 21.5% non-Hispanic Black; 41.9% non-Hispanic White; 26.6% Mexican American or Other Hispanic; 9.98% Other Race – Including Multi-Racial **Mean age (IQR):** 45 (26) **Sex:** 50.7% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Doshi et al 2022[24]	Evaluate impact of removing Black donor race indicator from the original KDRI formula (without refitting) on perceived GF risk (as implied by KDPI categorization), GF risk discrimination and predictive accuracy, and organ discard probability.	Race-free KDRI from the donor-only version of KDRI formula in DonorNet	Removed Black race	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Scientific Registry for Transplant Recipients	**Patients (n):** 66,987 9,945 from a Black donor
Drawz et al. 2012[25]	Evaluate if a modified Framingham Risk Score improves prediction of cardiovascular risk in patients with hypertension.	Framingham Risk Score for cardiovascular risk	Added race and chronic kidney disease to Framingham Risk Score	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT)	**Patients (n):** 6,604 **Race/Ethnicity:** 40% Black **Mean age:** 64 **Sex:** 50% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Duggal et al. 2021[26]	Evaluate how removal of race coefficient from eGFR affects medication dosing and risk of kidney failure.	eGFR (CKD-EPI)	Removed race	Modeled potential effect of CKD-EPI formula with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from NHANES 2015-16, and Veterans' Affairs Corporate Data Warehouse	NHANES cohort **Patients (n):** 227,613,357 **Race/Ethnicity:** 11% Black **Mean age (SD):** 47.4 (17.5) **Sex:** 52% female VA cohort **Patients (n):** 4,477,675 **Race/Ethnicity:** 17% Black **Mean age (SD):** 62.9 (15.8) **Sex:** 8% female
Elmaleh-Sachs et al. 2021[27]	Examine whether race and ethnicity–based spirometry reference equations improve the prediction of incident chronic lower respiratory disease (CLRD) events and mortality compared with race and ethnicity–neutral equations.	Spirometry reference equations Global Lung Function Initiative (GLI) reference equation	Removed race	Retrospective study of a prospective cohort modeling the effect of race and ethnicity-based equations compared with race and ethnicity-neutral equations (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	MESA Lung Study cohort (2004-2006)	**Patients (n):** 3344 **Race/Ethnicity:** 36% White; 25% Black; 23% Hispanic; 17% Asian **Mean age (SD):** 65.3 (9.6) **Sex:** 50% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Fairman et al. 2020[28]	Evaluate effect of updated pooled cohort equations compared to original equations for cardiovascular risk prediction.	ASCVD pooled cohort equations (PCE)	Updated PCE were derived from more diverse population and improved statistical techniques	Modeled potential effect of updated equations (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	National Ambulatory Medical Care Survey, 2011-14	**Patients (n):** 12,556 **Race/Ethnicity:** 10% Black **Sex:** 56% female
Foryciarz et al. 2022[29]	Evaluate 2 algorithmic fairness approaches to adjust the risk estimators (group recalibration and equalized odds) for their compatibility with the assumptions underpinning the ACC/AHA primary prevention of ASCVD guidelines' decision rules.	10-year ASCVD risk prediction using PCE (pooled cohort equations) for statin initiation	1) Group recalibrated algorithm model 2) Equalized odds algorithm model	Modeled 1) group recalibrated and 2) equalized odds models using original PCE cohorts compared with revised PCE model previously published and original PCE model (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	ARIC (Atherosclerosis Risk in Communities Study), CARDIA (Coronary Artery Risk Development in Young Adults Study), CHS (Cardiovascular Health Study), FHS OS (Framingham Heart Study Offspring Cohort), MESA (Multi-Ethnic Study of Atherosclerosis) and JHS (Jackson Heart Study)	**Patients (n):** 25,619 **Race/Ethnicity and Sex:** 17.3% Black women; 11.4% Black men; 37.8% Non-Black women; 33.4% Non-Black men **Mean age:** 56.5 ASCVD event incidence: 7.54%

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Fox et al. 2016[30]	Develop and validate risk prediction models for cardiovascular disease (CVD) incidence in Black adults.	CVD risk prediction	Added 10 biomarkers to standard CVD risk models. 10 biomarkers: adiposity (adiponectin and leptin), neurohormonal activation (aldosterone, B-type natriuretic peptide [BNP], and cortisol), inflammation (high-sensitivity C-reactive protein [hs-CRP]), endothelial function (endothelin and homocysteine), glycemic control (glycated hemoglobin), and insulin resistance (homeostasis model assessment of insulin)	Modeled comparison of models (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from participants in the Jackson Heart Study who had their first examination between September 2000 and March 2004	**Patients (n):** 3689 **Race/Ethnicity:** 100% Black **Mean age (SD):** 53 (11) **Sex:** 65% female
Gutiérrez et al. 2022[31]	Evaluate how removal of race coefficient from eGFR affects prediction of risk of kidney failure with replacement therapy and mortality.	eGFR (CKD-EPI)	Removed race	Retrospective cohort	Data drawn from the Chronic Kidney Disease Prognosis Consortium, 1998 to 2018	**Patients (n):** 62,011 **Race/Ethnicity:** 33.5% Black; 66.5% Non-Black **Mean age:** 63 **Sex:** 53% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Hammond et al. 2020[32]	Evaluate effect of including social determinants of health (SDOH) in algorithm to predict health care use, costs, and death.	Regression models using sex, age, comorbid conditions, and 7 SDOH domains: rural vs urban; alcohol abuse; access to care; economic status; financial strain; marital status; and education	Use of SDOH in addition to or to replace other variables. Race was not a specific component of any algorithm.	Modeled potential effect of algorithms based on sex + age; sex + age + comorbid conditions; all of these + SDOH; or SDOH alone (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Medicare Current Beneficiary Survey, 2016-17	**Patients (n):** 3614 **Race/Ethnicity:** 9.4% Black or Hispanic **Mean age:** 73 **Sex:** 56% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Hoenig et al. 2021[33]	Examine whether the change in use of eGFR without the Black coefficient changed access to transplant listing at our center (quality improvement project).	eGFR (MDRD)	Removed race	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Patients at Beth Israel Deaconess Medical Center registered on the national UNOS waiting list for preemptive kidney transplant from January 1, 2010, to December 31, 2020.	**Patients (n):** 567 **Race/Ethnicity:** 67.7% White; 32.3% Black **Mean age** of patients listed preemptively: White 54.7; Black 52.1 **Mean age** of patients listed on dialysis: White 53.1; Black 51.4 **Sex** of patients listed preemptively: White 66.3% male, 33.7% female; Black 54.3% male, 45.7% female **Sex** of patients listed on dialysis: White 66.2% male, 33.8% female; Black 75.5% male, 24.5% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Huang et al. 2022[34]	Evaluate how removal of race from eGFR affects prediction of risk of acute kidney injury after percutaneous coronary intervention.	eGFR (MDRD)	Removed race	Modeled potential effect of MDRD formula with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	American College of Cardiology's National Cardiovascular Data Registry; data collected between 2009 and 2017	Test cohort **Patients (n):** 947,091 procedures **Race/Ethnicity:** 7.9% Black **Mean age:** 64.8 **Sex:** 32.8% female Validation cohort **Patients (n):** 3,063,853 procedures **Race/Ethnicity:** 8.5% Black **Mean age:** 65.3 **Sex:** 31.9% female
Inker et al. 2021[35]	Examine an alternative approach to estimating GFR without using race through inclusion of 2 biomarkers.	eGFR	Replaced race with 4 potential components: creatinine, cystatin-C, beta-trace protein, beta2-microglobulin	Modeling of a cross-sectional comparison of eGFR based on 4 components without race with variations of those components with race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Derived from 7 studies, validated in 7 different studies; specific studies not identified	**Patients (n):** 2245 **Race/Ethnicity:** 24% Black **Mean age (SD)**: 52.8 (12.8) **Sex:** 29% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Inker et al. 2021[36]	Evaluate how removal of race coefficient from eGFR, and addition of cystatin C, affects diagnosis of kidney disease.	eGFR	1) Removed race 2) Replace race with cystatin C	Modeled potential effect with and without race and cystatin C (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from 12 studies used to validate CKD-EPI formula	**Patients (n):** 4050 **Race/Ethnicity:** 100% Black
Julian et al. 2017[37]	Replace race with relevant genotype to improve the Kidney Donor Risk Index.	Kidney Donor Risk Index (KDRI)	Replaced race with apolipoprotein L1 genotype	Modeled comparison of models with and without race and genotype (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data acquired from 3 kidney transplant centers (Wake Forest University, Emory University, University of Alabama at Birmingham) and from a published study that evaluated samples from 9 organ procurement organizations	**Patients (n):** 622 kidney donors and 1,149 recipients **Race/Ethnicity:** All donors were Black

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Kabra et al. 2016[38]	Add race to existing algorithm to improve prediction of stroke risk in patients with atrial fibrillation.	CHA_2DS_2-VASc score for stroke risk	Added "African-American ethnicity" to the CHA_2DS_2-VASc score, which previously included the following: congestive heart failure, hypertension, age ≥75 years, diabetes, prior stroke, vascular disease, age 65 to 74, and female sex	Comparison of models with and without race added to the scoring algorithm (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from CMS claims files, 2009-2012	**Patients(n):** 460,417 **Race/Ethnicity:** 7% Black **Mean age:** 79
Kimmel et al 2013[39]	Evaluate effect of using genotype data in warfarin dosing algorithm.	Warfarin dosing	Addition of genotype data to standard warfarin dosing algorithms	RCT	Patients enrolled at 18 U.S. study sites	**Patients (n):** 1015 **Race/Ethnicity:** 27% Black **Median age:** 59 (genotype group); 57 (standard group) **Sex:** 49% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Landy et al. 2021[40]	Examine whether USPSTF-2020 guidelines reduced racial and ethnic disparities compared with USPSTF-2013 guidelines and whether using an individualized prediction model for life-years gained from screening could reduce racial and ethnic disparities by identifying high-benefit individuals ineligible under USPSTF-2020 guidelines. *Only USPSTF-2020 and USPSTF-2020 plus LYFS-CT included for this report.	USPSTF 2020	Added an individualized prediction model, life-years from screening-computed tomography (LYFS-CT)	Modeled potential effect of LYFS_CT as an addition to USPSTF criteria (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	2015 U.S. National Health Interview Survey	USPSTF-2020 **Patients (n):** 14,508,450 **Race/Ethnicity:** 85.4% White; 8.0% African American; 1.9% Asian American; 4.7% Hispanic American **Age range:** 50 to 79 **Sex:** 43.9% female 2020 plus LYFS-CT **Patients (n):** 17,977,980 **Race/Ethnicity:** 82.7% White; 10.6% African American; 1.9% Asian American; 4.8% Hispanic American **Age range:** 50 to 79 **Sex:** 44.2% female
Limdi et al. 2015[41]	Evaluate role of clinical vs genetic factors in warfarin dosing algorithms.	Warfarin dosing	Use of race-stratified analysis of predictive algorithms for warfarin dosing rather than use of race-combined and adjusted algorithms	Prospective cohort	Patients enrolled at study sites at academic medical centers	**Patients (n):** 1357 **Race/Ethnicity:** 44% African-American **Mean age (SD):** 61 (15.8) **Sex:** 49% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Lindley et al. 2022[42]	Evaluate whether including information on the *CYP2C9*5* variant in warfarin dosing algorithms improves warfarin dose prediction.	Warfarin dosing	Addition of genotype data to standard warfarin dosing algorithms	Retrospective cohort	Patients enrolled at 7 U.S. study sites	**Patients (n):** 2298 **Race/Ethnicity:** 63.1% White, 35.2% Black, 1.6% Other, 1.7% Hispanic **Sex:** 48.8% female
Mahmud et al. 2021[43]	Evaluate how removal of race coefficient from eGFR affects association between eGFR and acute kidney injury (AKI) events in patients with cirrhosis.	eGFR (MDRD-4, MDRD-6, CKD-EPI)	Removed race	Retrospective cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Veterans' Health Administration data from the Veterans Outcomes and Costs Associated with Liver Disease, 2008-2015	**Patients (n):** 72,267 patients with cirrhosis **Race/Ethnicity:** 57.8% White; 19.7% Black; 7.2% Hispanic; 1.3% Asian; 13.9% Other **Median Age (IQR):** 61 (57 to 66) **Sex:** 2.7% female *study reports data for Black patients only
Meeusen et al. 2022[44]	Evaluate how removal of race coefficient from eGFR affects diagnosis of kidney failure and chronic kidney disease.	eGFR (CKD-EPI)	Removed race	Retrospective cohort	Outpatients treated at Mayo Clinic (Rochester, MN) between 2006 and 2021 for whom GFR was estimated by serum creatinine and measured by iothalmate renal clearance	**Patients (n):** 25,512 **Race/Ethnicity:** 2.5% Black **Age (SD):*** Black patients 50.6 (13.9); Non-Black patients 55.5 (14.0) *Study does not report whether age is mean or median

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Miller et al 2022[45]	Estimate changes that would occur in donor KDRI and in the proportion of donors classified as high risk (KDPI > 85%) if KDRI and KDPI were calculated from models without vs with the Black race predictor.	Original KDRI	Removed Black race	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Scientific Registry for Transplant Recipients	**Patients (n):** 69 244 adults
Miller et al. 2021[46]	Investigate the impact of removing the race coefficient from the CKD-EPI equation on renal dosage adjustment recommendations in a predominantly Black patient population.	eGFR (CKD-EPI, Deindexed CKD-EPI, CG CrCl)	Removed race	Retrospective cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Patients hospitalized between October 2019 and December 2019 at Einstein Medical Center, Philadelphia, PA.	**Patients (n):** 210 **Race/Ethnicity:** 84.3% Black; 15.7% White **Median Age:** Black 60.2; White 64.9 Median age higher among Whites compared to Blacks, p=0.001) **Sex:** 42.9% female
Muiru et al. 2023[47]	Evaluate how removal of race from eGFR affects progression of chronic kidney disease in patients with human immunodeficiency virus.	eGFR (CKD-EPI 2021 and CKD-EPI 2009)	Removed race	Modeled potential effect of 2021 CKD-EPI formula without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	North American AIDS Cohort Collaboration on Research and Design between 2005 and 2014	**Patients (n):** 69,135 **Race/Ethnicity:** 45% Black; 40% White, 11.6% Hispanic **Mean age (SD):** Black 44.4 (11.6); White 45.4 (11.3) **Sex:** Black 21.8% female; White 8.2% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Obermeyer et al. 2019[7] *Also addressed KQ1	Quantify racial disparities in health care resource allocation produced by a widely used commercial risk prediction algorithm.	The algorithm is used to predict complex health needs in primary care patients.	Replaced outcomes	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Development dataset for original algorithm is not specified. In this study, the data source is all primary care patients enrolled in risk-based contracts at a large academic hospital from 2013 to 2015 and self-identifying as either Black or as white without another race or ethnicity.	**Patients (n):** 49,618 **Race/Ethnicity**: 87.7% white, 12.3% Black **Mean Age**: 51.3 (white), 48.6 (Black) **Sex**: 62% female (white), 69% female (Black)
Panchal et al. 2022[48]	Evaluate how removal of race coefficient from eGFR affects eligibility for simultaneous liver-kidney transplantation and waitlist outcomes.	eGFR (MDRD-4, CKD-EPI)	Removed race	Retrospective cohort (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	United Network for Organ Sharing national transplant registry data, 2002-2019	**Patients (n):** 7937 patients eligible for liver transplantation **Race/Ethnicity:** 100% Black **Median Age (IQR):** No waitlist CKD 55 (46 to 61); Waitlist CKD 58 (53 to 62) *difference in age between groups, p<0.001 **Sex:** No-waitlist CKD 39.6% female; Waitlist CKD 47.8% female *difference in sex between groups, p<0.001

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Park et al. 2021[49]	Compare 3 techniques for mitigating bias in algorithms predicting postpartum depression.	Prediction of postpartum depression diagnosis and treatment	Recalibrated through reweighing key groups during model training 2) Removed race Added a regularization term that adjusts the algorithm to limit the effect of race-based variables	Modeled potential effect using each technique (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from IBM MarketScan Medicaid Database, 2014-18	**Patients (n):** 532,802 **Race/Ethnicity:** 38% Black **Mean age (SD):** 26 (5.4) **Sex:** 100% female
Schmeusser et al. 2022[50]	Evaluate how removal of race coefficient from eGFR affects eligibility for cancer clinical trials.	eGFR (CKD-EPI, MDRD)	Removed race	Modeled potential effect of CKD-EPI and MDRD formulas with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Patients who underwent nephrectomy at Emory University Hospital between 2009 and 2021	**Patients (n):** 459 **Race/Ethnicity:** 100% Black **Median age (SD):** 60 (8) **Sex:** 41% female Stage 3 or 4 cancer: 29.4%

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Shi et al. 2021[51]	Evaluate how removal of race coefficient from eGFR affects diagnosis of kidney disease.	eGFR	Removed race	Modeled potential effect of CKD-EPI and MDRD formulas with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from records of patients treated at University of Washington Medicine	**Patients (n):** 241,760 **Race/Ethnicity:** 69% White; 10% Asian; 9% Black **Median age:** 53 **Sex:** 51% female
Shores et al. 2013[52]	Develop and validate an alternative Donor Risk Index for liver transplant that is specific to Black recipients with Hepatitis C.	Donor Risk Index for liver transplantation	Modified the Donor Risk Index with data drawn from Black patients. The original Index was derived from a diverse population and included Black race as 1 of 7 components indicating higher risk of graft failure. The revised Index was derived from a population of Black recipients with Hepatitis C. The revised Index has 3 components and includes non-Black race.	Modeled comparison of revised Index to original version (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Data drawn from United Network for Organ Sharing (UNOS) registry	**Patients (n):** 294 patients with hepatitis-C **Race/Ethnicity:** 100% Black

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Thompson et al. 2021[13] *Also addressed KQ1	Assess fairness and bias of a previously validated machine learning opioid misuse classifier.	Natural language opioid misuse classifier using a convolutional neural network.		Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Development dataset consisted of adult hospital encounters from the EHR between 2007 and 2017 at a U.S. hospital and tertiary academic center. Opioid-related hospitalizations were oversampled. "The final dataset …consisted of 367 manually labeled cases, age- and sex-matched with controls that had no indications of opioid misuse." The external validation dataset came from EHR at a different tertiary care academic center. The dataset included "all unplanned adult inpatient encounters …who were screened between October 23, 2017, and December 31, 2019 (n = 53,974)." Appears that race/ethnicity was self-reported.	**Patients (n)**: 53,794 **Race/Ethnicity**: White 23,345; Black 17,541; Hispanic/Latinx 9,252); Other 3,836

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Topel et al. 2018[53]	Compare race-specific atherosclerosis in cardiovascular disease (ASCVD) formula to non-race-specific Framingham Risk Score for predicting subclinical vascular disease.	ASCVD	Added race and presence of diabetes to Framingham Risk Score to calculate ASCVD	Modeling of a cross-sectional comparison of ASCVD score to Framingham Risk score (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Derived from participants in 2 studies: Morehouse and Emory Team Up to Eliminate Health Disparities (META-Health) Study; and the Emory-Georgia Tech Center for Health Discovery and Well-Being (CHDWB) Study	**Patients (n):** 1231 **Race/Ethnicity:** 37% Black **Mean age (SD):** 53 (7) **Sex:** 59% female
Tsai et al. 2021[54]	Evaluate how removal of race coefficient from eGFR affects diagnosis and treatment of kidney disease.	eGFR	Removed race	Modeled potential effect of MDRD formula with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	NHANES, 2015-18	**Patients (n):** 2401 **Race/Ethnicity:** 100% Black

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Weale et al. 2021[55]	Evaluate effect of adding polygenic risk scores to ASCVD for estimating risk of cardiovascular disease.	ASCVD	Added polygenic risk scores to ASCVD	Modeled potential effect of ASCVD with and without polygenic risk scores (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	Atherosclerosis Risk in Communities (ARIC) study; Multi-ethnic Study of Atherosclerosis (MESA); United Kingdom Biobank	**Patients (n):** 18,961 **Race/Ethnicity:** 31% African ancestry **Sex:** 55% female
Yadlowsky et al. 2018[56]	Revised the 2013 pooled cohort equations (PCEs) using newer data and statistical methods, to improve the clinical accuracy of cardiovascular risk.	ASCVD PCEs	Added data from Jackson Heart Study and MESA to better reflect racial and ethnic populations; Adjusted statistical methods to reduce model overfitting by using elastic net regularization; removed race-based subgroups	Pre-post comparison of risk predicted by PCE derived from updated data and statistical methods to original PCE-based risk	Derived from 6 cohort studies: Atherosclerosis Risk in Communities Study (ARIC), Cardiovascular Health Study, Coronary Artery Risk Development in Young Adults Study, Framingham Health Study offspring cohort, Jackson Heart Study (JHS), MESA	**Patients (n):** 26,689 **Race/Ethnicity:** 29% Black **Mean age:** 57 **Sex:** 56% female

Author/ Year	Study Objective	Algorithm	Mitigation Strategy	Study Design	Source of Study Data	Patient Characteristics
Yap et al. 2021[57]	Evaluate how removal of race coefficient from eGFR affects classification of disease severity.	eGFR	Removed race	Modeled potential effect of CKD-EPI and MDRD formulas with and without race (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	EHRs of a large, urban academic medical center	**Patients (n):** 327 **Race/Ethnicity:** 100% Black **Mean age (SD):** 62 (14.2) **Sex:** 60% female 90% had diagnosis of hypertension
Zelnick et al. 2021[58]	Evaluate how removal of race coefficient from eGFR affects accuracy of GFR estimation and time to eligibility for kidney transplant.	eGFR	Removed race	Retrospective cohort study (modeling using real-world data – source data for calculation of algorithmic scores is real world data and outcomes that would have resulted from using the algorithm are simulated)	National Institute of Diabetes and Digestive and Kidney Diseases public repository data for Chronic Renal Insufficiency Cohort study.	**Patients (n):** 1658 **Race/Ethnicity:** 100% Black **Mean Age (SD):** 58 (11) **Sex:** 51% female

Abbreviations: ARIC = Atherosclerosis Risk in Communities study; ASCVD = atherosclerosis in cardiovascular disease; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; COPD = chronic obstructive pulmonary disease; eGFR = estimated glomerular filtration rate; FRA = Framingham Risk Score; IQR = interquartile range; JHS = Jackson Heart Study; KDRI = Kidney Donor Risk Index; MESA = Multi-ethnic Study of Atherosclerosis; MDRD = Modification of Diet in Renal Disease study; NHANES = National health and nutrition examination survey; PCE = pooled cohort equations; SD = standard deviation; SDOH = social determinants of health

Appendix D. Key Question 1 and 2 Evidence Tables

Table D-1. Reported outcomes of studies addressing Key Question 1

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Ashana et al. 2021[1]	Algorithm vs different or same algorithm w/ modifications for same clinical purpose	Health	In-hospital mortality (sensitivity analyses adjusted for hospital centers)	**SOFA, Black vs White patients (center-adjusted models)** ORs for in-hospital mortality after adjustment for: Original SOFA Score: OR 0.77; 95% CI 0.73 to 0.82, p<0.001 CSC SOFA score categories: OR 0.78; 95% CI 0.74 to 0.83, p<0.001 SOFA score w/ creatinine modification: OR 0.87; 95% CI 0.82 to 0.92, p<0.001 SOFA score w/out creatinine: OR 0.92; 95% CI 0.87 to 0.98, p<0.05 **LAPS2, Black vs White patients (center-adjusted models)** ORs for in-hospital mortality after adjustment for: Original LAPS2: OR 0.87; 95% CI 0.81 to 0.92, p<0.001 Eight-category LAPS2: OR 0.88; 95% CI 0.82 to 0.93, p<0.001 Four-category LAPS2: OR 0.89; 95% CI 0.84 to 0.94, p<0.001	In center-adjusted analyses, Ashana et al. found that Black race was associated with significantly lower in-hospital mortality than White race after adjustment for each mortality prediction model score (original and modified versions).[1]
Ashana et al. 2021[1]	Algorithm vs different or same algorithm w/ modifications for same clinical purpose	Access	In-hospital mortality	**Black vs White patients** Actual in-hospital mortality rates among patients in the highest-priority SOFA category (<6): Black patients (5.3%), White patients (6.9%) Similar in-hospital mortality rates for Black and White patients (6.7% vs 6.9%) in the highest-priority category could be achieved by reclassifying Black patients with scores between 6 and 8 (n=2611) into that category.	In a simulation using observed mortality, Ashana et al. found that 2611 Black patients (representing 81.6% of Black patients who were included in lower-priority CSC categories and 9.4% of all Black patients) were erroneously excluded from receiving the highest prioritization.[1]

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Boley et al. 2022[2]	No comparator. Compared triage process outcomes by racial groups.	Access to care	Assignment to FT area rather than main ED	**Black-NH patients vs White-NH patients** 22.6% (95% CI 21.4 to 23.7) vs 18.5% (95% CI 16.8 to 20.3) OR: 1.28; 95% CI 1.12 to 1.46, p<0.001 Black-NH patients were more likely to be triaged to the FT workflow than White-NH patients. **Chief Complaint Subgroups (Black-NH vs White-NH)** ORs for assignment to FT (Black-NH vs. White-NH) among patients with: Abdominal pain: OR: 1.50; 95% CI 0.84 to 2.70, p=0.303 Chest pain: OR: 1.76; 95% CI 0.77 to 4.02, p=0.082 Shortness of breath: OR: 1.56; 95% CI 0.62 to 3.94, p=0.343 Headache: OR: 2.10; 95% CI 1.01 to 4.39, p=0.048 Black-NH patients with a chief complaint of headache were significantly more likely to be assigned to the FT area than White NH patients. The difference between racial and ethnic groups for chief complaints of abdominal pain, chest pain, and shortness of breath was not significant. **Interaction of Race and ESI-level** ORs for assignment to FT (Black-NH vs. White-NH) among: Low-acuity patients: OR: 0.99; 95% CI 0.87 to 1.13, p=0.934, no statistically significant difference between Black-NH and White-NH low-acuity patients assigned FT status. High-acuity patients: OR 1.40; 95% CI: 1.05 to 1.87, p=0.024, Black-NH high-acuity patients were significantly more likely to be assigned FT status than White-NH high-acuity patients.	Black-NH patients were significantly more likely to be triaged to the FT area than White-NH patients. Authors evaluated the interaction of race and ESI acuity level and found that Black-NH high-acuity patients were significantly more likely to be assigned to the FT area than White-NH high-acuity patients. The difference between Black-NH and White-NH low-acuity patients was not significant.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Boley et al. 2022[2]	No comparator. Compared triage process outcomes by racial groups.	Access to care	High-acuity designation (ESI score 1 to 3)	**Black-NH patients vs White-NH patients** 59.8% (95% CI 58.4 to 61.1) vs 67.0% (95% CI 65.0 to 69.1) OR 0.73, 95% CI 0.66 to 0.81, p<0.001 Black-NH patients were less likely to be triaged as high acuity than White-NH patients. **Chief Complaint Subgroups** (**Black-NH vs White-NH)** ORs for high-acuity designation (Black-NH vs. White-NH) among patients with: Abdominal pain: OR 0.53, 95% CI 0.24 to 1.15, p=0.107 Chest pain: OR 0.63, 95% CI 0.37 to 1.07, p=0.086 Shortness of breath: OR 0.45, 95% CI 0.22 to 0.94, p=0.034 Headache: OR 0.73, 95% CI 0.43 to 1.24, p=0.238	Black-NH patients were significantly less likely to be triaged as high acuity than White-NH patients. Compared to White-NH patients, Black-NH patients with a chief complaint of abdominal pain, chest pain, headache, or shortness of breath were less likely to be triaged as high acuity. The difference between racial and ethnic groups was significant for the chief complaint of shortness of breath.
Boley et al. 2022[2]	No comparator. Compared triage process outcomes by racial groups.	Quality of care (timeli-ness)	Wait time (i.e., time from ED arrival to rooming time, in minutes)	**Black-NH patients vs White-NH patients** Black-NH patients: Mean 54.0 minutes; 95% CI 52.3 to 55.7 White-NH patients: Mean 57.5 minutes; 95% CI 54.9 to 60.1 MD -3.47 minutes, 95% CI -6.56 to -0.37, p=0.028 Black-NH patients had a significantly shorter total wait time than White-NH patients. **Chief Complaint Subgroups** (**Black-NH vs White-NH)** Abdominal pain: MD -9.52 minutes, 95% CI -20.02 to -0.03, p=0.028 Chest pain: MD -18.82 minutes, 95% CI -28.93 to -8.72, p<0.001 Shortness of breath: MD 0.91 minutes, 95% CI -13.7 to 15.52, p=0.903 Headache: MD 8.07 minutes, 95% CI -5.62 to 21.76, p=0.247	Black-NH patients had a significantly shorter total wait time from ED arrival to rooming time than White-NH patients. In subgroup analyses, compared with White-NH patients, Black-NH patients with chief complaints of abdominal or chest pain had significantly shorter wait times, but differences between racial and ethnic groups were not significant for shortness of breath or headache.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Carbunaru et al. 2019[3]	Algorithm vs different algorithm for same clinical purpose	Quality	Biopsies avoided by using a ≥10% threshold	**PCPT:** 251/355 (71%) (whites), 23/463 (5%) (Blacks) **PBCG:** 24/355 (7%) (whites), 1/463 (0.02%) (Blacks)	If biopsies had been performed only on men with a ≥10% risk on the PCPT algorithm, 71% of the white men in this sample would not have received biopsies compared with only 5% of the Black men. Using the same risk threshold on the PBCG algorithm, relatively few men in this sample would have avoided biopsies (7% of white men and 0.02% of Black men).
Carbunaru et al. 2019[3]	Algorithm vs different algorithm for same clinical purpose	Quality	Clinically significant prostate cancers (csPCas) missed by using a ≥10% threshold	Rates of csPCa among men who would not have been recommended for biopsies using the given algorithm and threshold: **PCPT:** 27% (whites), 13% (Blacks) **PBCG:** 13% (whites), 0% (Blacks)	Rates of csPCa were higher in the white subgroup than in the Black subgroup among patients who would not have received biopsies based on either algorithm using a ≥10% risk threshold. Because all but 1 of the Black patients would have received biopsies based on the PBCG, no csPCa would have been missed.
Carbunaru et al. 2019[3]	Algorithm vs different algorithm for same clinical purpose	Quality	Unnecessary biopsies performed by using a ≥10% threshold	Percentages of men with no PCa or indolent PCa who would have been recommended for biopsies using the given algorithm and threshold. **PCPT:** 38/209 (18%) (whites), 189/205 (92%) (Blacks) **PCBG:** 191/209 (91%) (whites), 204/205 (99.5%) (Blacks)	If biopsies had been performed on all men with a ≥10% risk on the PCPT algorithm, only 18% of white men with no PCa or indolent PCa would have received a biopsy compared with 92% of Black men. Using the same risk threshold on the PBCG algorithm, percentages would be 91% and 99.5% among white and Black patients, respectively.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Carbunaru et al. 2019[3]	Algorithm vs different algorithm for same clinical purpose	Quality	Biopsies avoided by using a ≥30% threshold	**PCPT**: 342/355 (96%) (whites), 279/463 (60%) (Blacks) **PBCG:** 208/355 (59%) (whites), 114/463 (25%) (Blacks)	If biopsies had been performed only on men with a ≥30% risk on the PCPT algorithm, 96% of white men in this sample would not have received biopsies compared with 60% of Black men. Using the same risk threshold on the PBCG algorithm, 59% of the white men and 25% of the Black men would not have received biopsies.
Carbunaru et al. 2019[3]	Algorithm vs different algorithm for same clinical purpose	Quality	Clinically significant prostate cancers (csPCAs) missed by using a ≥30% threshold	Rates of csPCA among men who would not have been recommended for biopsies using the given algorithm and threshold: **PCPT:** 33% (whites), 27% (Blacks) **PBCG:** 25% (whites), 24% (Blacks)	Using a ≥30% risk threshold, rates of csPCa among those patients who would not have received biopsies based on either algorithm were slightly higher for white patients than for Black patients.
Carbunaru et al. 2019[3]	Algorithm vs different algorithm for same clinical purpose	Quality	Unnecessary biopsies performed by using a ≥30% threshold	Percentages of men with no PCa or indolent PCa who would have been recommended for biopsies using the given algorithm and threshold: **PCPT:** 0/209 (0%) (whites), 25/205 (12%) (Blacks) **PCBG:** 57/209 (27%) (whites), 121/205 (59%) (Blacks)	If biopsies had been performed only on men with a ≥30% risk on the PCPT algorithm, no white men with no PCa or indolent PCa would have received a biopsy compared with 12% of the Black men. Using the same risk threshold on the PBCG algorithm, the percentages would be 27% and 59% among white and Black patients, respectively.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Han et al. 2020[4]	Algorithm vs no algorithm	Quality	Recommendation for lung cancer screening	Individuals in the **1950 birth cohort eligible for lung cancer screening by PLCOm2012 (using ≥1.51% as threshold)** but ineligible by USPSTF criteria: Individuals aged 50-54: Whites, 4.8%; Blacks, 15.6% (p <0.001) Individuals aged 71-80: Whites, 10.8%; Blacks, 14.2% (p <0.001) Individuals aged 55-70: Whites, 3.3%; Blacks, 7% (p-value not reported) Results at varying risk thresholds are presented graphically; the proportion "is consistently higher in Blacks compared with whites independently of risk threshold."	A simulation of the 1950 birth cohort showed that, in all age groups, the percentage of individuals who were ineligible for lung cancer screening based on USPSTF criteria, but eligible based on their values on established lung cancer risk factors, was greater among Black individuals than among white individuals.
Han et al. 2020[4]	Algorithm vs no algorithm	Quality	Recommendation for lung cancer screening	Individuals in the **1960 birth cohort eligible for lung cancer screening by PLCOm2012 (using ≥1.51% as threshold)** but ineligible by USPSTF criteria: All age groups: Whites, 2.3%; Blacks, 5.8% (p <0.001) Individuals aged 50-54: Whites, 2.2%; Blacks, 5.8%* Individuals aged 71-80: Whites, 7.6%; Blacks, 11.1%* Individuals aged 55-70: Whites, 2.3%; Blacks, 4.6%* *P-values not reported for the comparisons within age group	In a simulation of the 1960 birth cohort, the differences between Black and white individuals ineligible for screening based on USPSTF criteria but eligible based on risk factors persisted, but were mostly smaller than those seen in the 1950 cohort.
Metzger et al. 2022[5]	No comparator. Compared triage process outcomes by racial groups.	Access to care	ESI score assignment	**Non-White patients vs NH-White patients** ESI 2 (Emergency) vs 4 (Nonurgent): aOR 0.40, 95% CI: 0.33 to 0.49, p<0.001 Non-White patients were significantly less likely to receive an ESI score of 2 than White-NH patients. ESI 3 (Urgent) vs 4 (Nonurgent): aOR 0.50, 95% CI: 0.45 to 0.56, p<0.001 Non-White patients were significantly less likely to receive an ESI score of 3 than White-NH patients. ESI 5 (Minor) vs 4 (Nonurgent): aOR 1.34, 95% CI 1.07 to 1.69, p=0.012 Non-White patients were significantly more likely to receive an ESI score of 5 than White-NH patients.	In analyses adjusting for illness severity (i.e., abnormal vital signs), compared with NH-White patients, non-White patients were significantly less likely to receive an ESI score of 2 or 3 but significantly more likely to receive an ESI score of 5, vs an ESI score of 4.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Metzger et al. 2022[5]	No comparator. Compared triage process outcomes by racial groups.	Access to care	ESI score assignment for chief complaint subgroups	**Non-White patients vs NH-White patients (unadjusted ORs)** <u>Closed head injury</u> **ESI 2 (Emergent) vs 4 (Nonurgent)**: OR 1.46, 95% CI 0.72 to 2.98, p=0.295; **ESI 3 (Urgent) vs 4 (Nonurgent):** OR 0.98, 95% CI 0.73 to 1.33, p=0.909; **ESI 5 (Minor) vs 4 (Nonurgent)**: OR 1.51, 95% CI 0.95 to 2.40, p=0.083 <u>Fever</u> **ESI 2 (Emergent) vs 4 (Nonurgent)**: OR 0.30, 95% I 0.22 to 0.40, p<0.001; **ESI 3 (Urgent) vs 4 (Nonurgent)**: OR 0.38, 95% CI 0.31 to 0.47, p<0.001; **ESI 5 (Minor) vs 4 (Nonurgent)**: OR 1.60, 95% CI 0.98 to 2.62, p=0.061 <u>Gastrointestinal distress</u> **ESI 2 (Emergent) vs 4 (Nonurgent):** 0.21, 95% CI 0.14 to 0.32, p<0.001; **ESI 3 (Urgent) vs 4 (Nonurgent):** OR 0.31, 95% CI 0.26 to 0.37, p<0.001; **ESI 5 (Minor) vs 4 (Nonurgent):** OR 3.31, 95% CI 1.5 to 7.04, p=0.002 <u>Headache</u> **ESI 2 (Emergent) vs 4 (Nonurgent):** OR 0.12, 95% CI 0.03 to 0.56, p=0.006; **ESI 3 (Urgent) vs 4 (Nonurgent)**: OR 0.23, 95% CI 0.11 to 0.48, p<0.001; **ESI 5 (Minor) vs 4 (Nonurgent):** OR 0.61, 95% CI 0.11 to 3.52, p=0.584 <u>Upper respiratory infection</u> **ESI 2 (Emergent) vs 4 (Nonurgent):** OR 0.52, 95% CI 0.32 to 0.85, p=0.010; **ESI 3 (Urgent) vs 4 (Nonurgent):** OR 0.65, 95% CI 0.51 to 0.83, p=0.001; **ESI 5 (Minor) vs 4 (Nonurgent):** OR 1.06, 95% CI 0.71 to 1.58, p=0.788 <u>Other</u> **ESI 2 (Emergent) vs 4 (Nonurgent):** OR 0.28, 95% I 0.21 to 0.37, p<0.001; **ESI 3 (Urgent) vs 4 (Nonurgent)**: OR 0.33, 95% CI 0.27 to 0.39, p<0.001; **ESI 5 (Minor) vs 4 (Nonurgent):** OR 1.37, 95% CI 0.85 to 2.20, p=0.199	The difference between Non-White and NH-White patients with a chief complaint of closed head injury receiving an ESI score of 2, 3 or 5 was not significant. Non-White patients were significantly less likely than NH-White patients to receive an ESI score of 2 or 3 for a chief complaint of fever, gastrointestinal distress, headache, upper respiratory infection, or other illness. Non-White patients with a chief complaint of gastrointestinal distress were significantly more likely than NH-White patients to receive an ESI score of 5. The difference between racial and ethnic groups for chief complaints of fever, headache, upper respiratory infection, and other chief complaints was not significant.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Metzger et al. 2022[5]	No comparator. Compared triage process outcomes by racial groups.	Quality of care	Time to provider	**Non-White patients vs NH-White patients** HR 1.03, 95% CI 0.97 to 1.08, p=0.352, no significant difference between Non-White and NH-White patients in time to see a provider. **Non-White patients** Median (IQR), minutes: 32 (17 to 57) **NH-White patients** Median (IQR), minutes: 30 (16 to 53)	No significant difference between Non-White and NH-White patients in time to see a provider.
Metzger et al. 2022[5]	No comparator. Compared triage process outcomes by racial groups.	Quality of care	ED LOS	**Non-White patients vs NH-White patients** HR 1.08, 95% CI 1.03 to 1.14, p value NR. Non-White patients were discharged at a higher rate than White patients indicating a shorter LOS. **Non-White patients** Median (IQR), minutes: 147 (103 to 212) **NH-White patients** Median (IQR), minutes: 173 (117 to 254)	Non-White patients were discharged at a higher rate than NH-White patients, indicating a shorter LOS.
Metzger et al. 2022[5]	No comparator. Compared triage process outcomes by racial groups.	Quality of care	Hospital admission	**Non-White patients vs NH-White patients** aOR: 0.87, 95% CI 0.73 to 1.03, p=0.941 Unadjusted OR: 0.53, 95% CI 0.46 to 0.61, p<0.0001	In an unadjusted analysis, non-White patients were significantly less likely to be admitted to the hospital than NH-White patients. However, in the adjusted analysis, the difference between patient groups was nonsignificant.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Miller et al. 2021[6]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Access	Prioritization for ICU resources	**Adjusted odds of in-hospital mortality** for Black patients compared with White patients with equivalent SOFA scores: *Model including race, SOFA score, interaction of race by SOFA score, and hospital fixed effects*: OR for interaction, 0.98; 95% CI, 0.97-0.99; p<0.001 *Model including race, SOFA score, interaction of race by SOFA score, hospital, Charlson Comorbidity Index score, age, and acute admission diagnosis*: OR for interaction, 0.99; 95% CI, 0.98-0.996; p=0.004	The SOFA score, which estimates the risk of in-hospital mortality, is used to prioritize resources to patients with low risk (i.e., those most likely to benefit). Adjusted analyses indicated a small but statistically significant tendency for the SOFA score to overestimate the true risk of death among Black patients relative to White patients, thus lowering their eligibility for resources relative to White patients.
Miller et al. 2021[6]	Algorithm vs same algorithm w/modifica- tions for same clinical purpose	Access	Prioritization for ICU resources	**Adjusted odds of in-hospital mortality** for Black patients vs White patients in the highest-priority tier of each system examined: *System A*: OR, 0.65; 95% CI, 0.58-0.74; p<0.001 *System B*: OR, 0.70; 95% CI, 0.64-0.78; p<0.001 *System C*: OR, 0.73; 95% CI, 0.67-0.80; p<0.001	Black patients had to have a lower true risk of death in order to qualify for resources than did White patients.
Miller et al. 2021[6]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Access	Prioritization for ICU resources	**Percent of Black patients inappropriately deprioritized under conditions of severe shortage** (i.e., assigned to a lower-priority tier even though their true risk of death was lower than some patients in the highest-priority tier): *System A*: 15.6% *System B*: 9.0% *System C*: 6.5%	Across 3 tiering systems, in a situation in which only patients in the top tier receive resources, up to 15.6% of Black patients should have qualified (based on their true risk of death) but did not.
Miller et al. 2021[6]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Access	Prioritization for ICU resources	**Increase in SOFA thresholds** for Black patients necessary in order to equalize the adjusted odds of death for Black and White individuals who qualify for high priority: 2 points (across all tiering systems and levels of shortage)	Proposed mitigation (separate thresholds for Black and White patients) would remove bias.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Obermeyer et al. 2019[7]	Algorithm vs different algorithm for same clinical purpose	Access	Eligibility for a care management program	At every level of algorithm-predicted risk, Black and white patients had similar actual costs in the following year. However, at a given level of *health*, Black patients generated lower costs than white patients – on average, $1801 less per year.	The algorithm predicted health care costs equally well for Black and white patients, but the use of costs as a proxy for healthcare needs is flawed because the association between costs and health differs across racial/ethnic groups.
Obermeyer et al. 2019[7]	Algorithm vs different algorithm for same clinical purpose	Access	Eligibility for a care management program	At the cutoff score for automatic enrollment into the care management program (97th percentile), Black patients had 26.3% more chronic conditions than White patients ($p < 0.001$).	The system is biased to accept White patients who have a lower level of actual need than Black patients.
Pasquinelli et al. 2021[59]	Algorithm vs different algorithm for same clinical purpose	Access	Eligibility for lung cancer screening	**Number (percent) of individuals:** **Eligible for screening by USPSTF:** **White:** 167/258 (64.7%) **African American:** 257/497 (51.7%) **Ineligible by USPSTF but eligible by PLCOm2012 >1.7%/6-year risk criteria:** **White:** 24/258 (9.3%) **African American:** 109/497 (21.9%) **Eligible for screening by USPSTF or PLCOm2012:** **White:** 191/258 (74.0%) **African American:** 366/497 (73.6%)	By USPSTF criteria, a greater percentage of White individuals were eligible for screening than African American individuals, but broadening the eligibility criteria to include patients classified as at risk by PLCOm2012 criteria (i.e., >1.7% predicted 6-year risk of lung cancer) eliminated the disparity.[59]

Presti et al. 2021[9]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Quality	Total biopsies avoided, cancers missed, and negative biopsies avoided – Version A with ≥7.5% risk threshold	**Total biopsies avoided:** 3% (Whites), 7% (Asians), <1% (Hispanic), 7% (Blacks) **High-grade disease missed:** <1% (Whites), 2% (Asians), 0% (Hispanic), 1% (Blacks) **Low-grade disease missed:** 3% (Whites), 6% (Asians), 0% (Hispanic), 5% (Blacks) **Negative biopsies avoided:** 6% (Whites), 9% (Asians), 1% (Hispanic), 12% (Blacks)	If biopsies had been performed only on men with ≥7.5% risk on Version A, relatively few men in this sample would have avoided biopsies (3% of White men and 7% of Black men). This strategy would have avoided 6% of negative biopsies among White patients while missing <1% of high-grade cancers; among Black patients the percentages would be 12% and 1%. Although this study does not calculate net benefits, it is possible to look at the percentages of biopsies avoided and cancers missed and thereby get an idea of the net benefit that would accrue to each racial/ethnic group from the use of each algorithm at each threshold value. If we simply take the ratio of unnecessary biopsies avoided to high-grade cancers missed, it is 12:1 for Black patients; for White patients, it is at least 6:1, but cannot be calculated exactly. Thus, it is unclear whether the net benefit of the strategy would be greater for White or Black patients. The effect on Hispanic patients would have been very low; <1% would have avoided biopsies. The net benefit for Asians was low compared with others; 2% of high-grade cancers would have been missed and 9% of negative biopsies would have been avoided, a ratio of 4.5:1.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Presti et al. (2021)[9]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Quality	Total biopsies avoided, cancers missed, and negative biopsies avoided – Version A with ≥10% risk threshold	**Total biopsies avoided:** 6% (Whites), 19% (Asians), 1% (Hispanic), 19% (Blacks) **High-grade disease missed:** 1% (Whites), 6% (Asians), 1% (Hispanic), 6% (Blacks) **Low-grade disease missed:** 6% (Whites), 18% (Asians), 0% (Hispanic), 17% (Blacks) **Negative biopsies avoided:** 9% (Whites), 25% (Asians), 1% (Hispanic), 25% (Blacks)	Using a ≥10% cutoff for Version A, 9% of negative biopsies would have been avoided among White patients while missing 1% of high-grade cancers; among Black patients, 25% of negative biopsies would have been avoided, but 6% of high-grade cancers would have been missed, yielding a ratio of 4.2:1. Thus, using a ≥10% cutoff rather than ≥7.5% would result in the net benefit for Black patients decreasing relative to that for White patients. The effect on Hispanic patients would have been very low; only 1% would have avoided biopsies. The net benefit for Asians was low compared with other racial/ethnic groups; 6% of high-grade cancers would have been missed and only 25% of negative biopsies would have been avoided, yielding a ratio of 4.2:1.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Presti et al. (2021)[9]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Quality	Total biopsies avoided, cancers missed, and negative biopsies avoided – Version B with ≥7.5% risk threshold	**Total biopsies avoided:** 16% (Whites), 26% (Asians), 9% (Hispanic), 31% (Blacks) **High-grade disease missed:** 3% (Whites), 4% (Asians), 1% (Hispanic), 4% (Blacks) **Low-grade disease missed:** 10% (Whites), 20% (Asians), 7% (Hispanic), 20% (Blacks) **Negative biopsies avoided:** 28% (Whites), 37% (Asians), 19% (Hispanic), 48% (Blacks)	Overall, Version B results in fewer biopsies being recommended than Version A; thus, compared with Version A, Version B will always have both more unnecessary biopsies avoided and more high-grade cancers missed. Using a ≥7.5% risk cutoff, 28% of negative biopsies would have been avoided among White patients while missing 3% of high-grade cancers, yielding a ratio of 9.3:1; among Black patients the percentages would be 48% and 4%, yielding a ratio of 12:1. This suggests that the net benefit of this strategy would be slightly better for Black patients than for White patients. The effect on Hispanic patients would have been low and relatively positive; 1% of high-grade cancers would have been missed and 19% of negative biopsies would have been avoided. The net benefit for Asians was closer to that for other racial/ethnic groups than under other strategies; 4% of high-grade cancers would have been missed and 37% of negative biopsies would have been avoided, yielding a ratio of 9.25:1.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Presti et al. (2021)[9]	Algorithm vs same algorithm w/modifica-tions for same clinical purpose	Quality	Total biopsies avoided, cancers missed, and negative biopsies avoided – Version B with ≥10% risk threshold	**Total biopsies avoided:** 22% (Whites), 37% (Asians), 11% (Hispanic), 40% (Blacks) **High-grade disease missed:** 4% (Whites), 9% (Asians), 1% (Hispanics), 5% (Blacks) **Low-grade disease missed:** 17% (Whites), 30% (Asians), 8% (Hispanics), 27% (Blacks) **Negative biopsies avoided:** 39% (Whites), 51% (Asians), 24% (Hispanics), 61% (Blacks)	Using a ≥10% cutoff for Version B, the percentages of unnecessary biopsies avoided and high-grade cancers missed would be 39% and 4%, respectively for White patients, yielding a ratio of 9.75:1, and 61% and 5% for Black patients, respectively, yielding a ratio of 12.2:1. As with the ≥7.5% cutoff, this suggests that the net benefit of this strategy would be slightly better for Black patients than for White patients. The effect on Hispanic patients would have been low and relatively positive; 1% of high-grade cancers would have been missed and 24% of negative biopsies would have been avoided. The net benefit for Asians was lower than that for other racial/ethnic groups; 9% of high-grade cancers would have been missed and 51% of negative biopsies would have been avoided, yielding a ratio of 5.7:1.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Riviello et al. 2022[10]	No comparator. Compared outcomes by racial and ethnic groups.	Health	Estimated number of excess deaths *Study modeled the number of excess deaths if a lack of mechanical ventilators required allocating ventilators to a priority score of ≤ 2 (highest priority receives ventilators) or ≤ 5 (highest and intermediate priority groups receive ventilators). Model includes 244 patients who received ventilation.	**Black patients vs White patients (CSOC priority score allocation: ≤ 2 get ventilator, most severe)** 43.9% (18 of 41) vs 33.3% (40 of 120), p=0.22 **Black patients vs Other patients (CSOC priority score allocation: ≤ 2 get ventilator, most severe)** 43.9% (18 of 41) vs 28.6% (58 of 203), p=0.05 **Black patients vs White patients (CSOC priority score allocation: ≤ 5 get ventilator)** 4.9% (2 of 41) vs 4.2% (5 of 120), p=0.85 **Black patients vs Other patients (CSOC priority score allocation: ≤ 5 get ventilator)** 4.9% (2 of 41) vs 3.0% (6 of 203), p=0.53	If only patients in the highest-priority group (≤ 2) received ventilators, there would have been significantly more excess deaths among Black patients (43.9%) than among all other patients (28.6%). However, the difference in estimated excess deaths between Black (43.9%) and White (33.3%) patients was not statistically significant in this scenario. Using only the highest and intermediate priority groups receiving ventilators (≤ 5) to model the outcome, the difference in estimated excess deaths between Black and White patients or Black and all other patients was not statistically significant.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Sarkar et al. 2021[11]	Algorithm vs different algorithm for same clinical purpose	Health	Predicted Mortality	**APACHE IVa scores (eICU-CR database)** **Hispanic:** Mean predicted mortality: 12% (608/5057); Actual mortality: 8.7% (442/5057); Standardized mortality ratio (SMR, actual/predicted): 0.73; 95% CI 0.667 to 0.793 **African American:** Mean predicted mortality: 11.9% (1813/15299); Actual mortality: 8.0% (1219/15299); SMR (actual/predicted): 0.67; 95% CI 0.637 to 0.71 **White:** Mean predicted mortality: 11.4% (11456/100,694); Actual mortality: 8.7% (8732/100,694); SMR (actual/predicted): 0.76; 95% CI 0.748 to 0.777 **Asian:** Mean predicted mortality: 11.8% (220/1869); Actual mortality: 9% (169/1869); SMR (actual/predicted): 0.77; 95% CI 0.669 to 0.882 **OASIS scores (MIMIC-III database)** **Hispanic:** Mean predicted mortality: 11.8% (210/1784); Actual mortality: 7.5% (134/1784); SMR (actual/predicted): 0.64; 95% CI 0.53 to 0.76 **African American:** Mean predicted mortality: 13.5% (657/4853); Actual mortality: 9.1% (443/4853); SMR (actual/predicted): 0.67; 95% CI 0.613 to 0.743 **White:** Mean predicted mortality: 14.1% (5081/35,997); Actual mortality: 11.4% (4114/35,997); SMR (actual/predicted): 0.81; 95% CI 0.787 to 0.833 **Asian:** Mean predicted mortality: 13.9% (165/1189); Actual mortality: 13.1% (156/1189); SMR (actual/predicted): 0.95; 95% CI 0.825 to 1.084 **SOFA (eICU-CR database) -** Ratio of observed mortality to overall mortality (i.e., mortality across ethnic groups) by admission SOFA category and ethnic group **Hispanic:** SOFA score 0 to 7 (0.96); SOFA score 8 to 11 (1.04); SOFA score > 11 (1.08); **African American:** SOFA score 0 to 7 (0.86); SOFA score 8 to 11 (0.95); SOFA score > 11 (0.91); **White:** SOFA score 0 to 7 (1.02); SOFA score 8 to 11 (1.00); SOFA score > 11 (1.01); **Asian:** SOFA score 0 to 7 (1.12); SOFA score 8 to 11 (1.05); SOFA score > 11 (1.06) **SOFA (MIMIC-III database) -** Ratio of observed mortality to overall mortality (i.e., mortality across ethnic groups) by admission SOFA category and ethnic group **Hispanic:** SOFA score 0 to 7 (0.62); SOFA score 8 to 11 (0.66); SOFA score > 11 (1.07); **African American:** SOFA score 0 to 7 (0.74); SOFA score 8 to 11 (0.88); SOFA score > 11 (0.99); **White:** SOFA score 0 to 7 (1.04); SOFA score 8 to 11 (1.03); SOFA score > 11 (1.00); **Asian:** SOFA score 0 to 7 (1.06); SOFA score 8 to 11 (1.07); SOFA score > 11 (0.95)	APACHE IVa and OASIS both overestimated mortality for all race/ethnicities, and these overestimates were worse for African Americans and Hispanics than for Whites and Asians. For SOFA scores 0-7, observed mortality was lower for African-Americans and Hispanics than for Whites and Asians. This was found for both the eICU-CR database and the MIMIC-III database.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Snavely et al. 2021[12]	Algorithm/tool vs no algorithm/tool	Health	30-day death or myocardial infarction (MI)	**Non-White:** Pre HEART-Pathway implementation: 5% (61/1229) Post HEART-Pathway implementation: 4.8% (80/1655) Adjusted OR: 1.19; 95% CI 0.82 to 1.72 **White:** Pre HEART-Pathway implementation: 7.9% (197/2,484) Post HEART-Pathway implementation: 8.8% (273/3,106) Adjusted: 1.28; 95% CI 1.04 to 1.57 **Low-risk Non-White:** Post HEART-Pathway implementation: 0.5% (3/590); 95% CI: 0.1% to 1.5% **Low-risk White:** Post HEART-Pathway implementation: 0.3% (3/871); 95% CI 0.07% to 1.0% Low risk Non-White vs Low-risk White: p=0.69	The race disparity was slightly larger after algorithm implementation (4 percentage points) than before (2.9 percentage points). Snavely et al. found that the HEART Pathway was associated with low 30-day death and MI rate among low-risk patients regardless of sex or race.[12]
Snavely et al. 2021[12]	Algorithm/tool vs no algorithm/tool	Access	Nonadherence rates post HEART-Pathway implementation (low-risk patients receiving stress testing or hospitalization or non-low-risk patients receiving early discharge from the ED)	**Non-White:** 15.6% (213/1,363) **White:** 15.4% (405/2,629)	Rates were similar in non-White and White patients post implementation

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Snavely et al. 2021[12]	Algorithm/tool vs no algorithm/tool	Quality	Risk stratification post HEART-Pathway implementation	**Proportion classified as low risk** *Non-White patients vs White patients* 7.6%; 95% CI 4.8% to 10.5%, p<0.0001, more non-white patients were classified as low risk **Proportion classified as low risk (adjusted for age)** *White females vs White Males:* Adjusted OR (aOR) 1.50; 95% CI 1.24 to 1.80, more likely to be classified as low risk *Non-White females vs White Males:* aOR 1.48; 95% CI 1.20 to 1.82, more likely to be classified as low risk *Non-White males vs White Males:* aOR 1.34; 95% CI 1.06 to 1.68, more likely to be classified as low risk	Compared with White males, White females, non-White females, and non-White males were more likely to be classified as low risk.
Snavely et al. 2021[12]	Algorithm/tool vs no algorithm/tool	Quality	Hospitalizations (inpatient admission, transfer, or observation stay)	**Non-White:** Pre HEART-Pathway implementation: 58.8% (723/1,229) Post HEART-Pathway implementation: 50.2% (831/1,655) 8.6% reduction; aOR: 0.72; 95% CI 0.60 to 0.86 **White:** Pre HEART-Pathway implementation: 63.0% (1,564/2,484) Post HEART-Pathway implementation: 58.5% (1,818/3,106) 4.5% reduction; aOR: 0.83; 95% CI 0.73 to 0.94 **Non-White vs White** 8.3% fewer non-white patients were hospitalized compared to white patients post HEART-Pathway implementation 8.3%; 95% CI 5.3% to 11.1%, p<0.001	Compared with White patients, fewer non-White patients were hospitalized post-implementation.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Snavely et al. 2021[12]	Algorithm/tool vs no algorithm/tool	Quality	Early discharge (proportion of patients discharged from the ED without objective cardiac testing [OCT])	**Non-White:** Pre HEART-Pathway implementation: 40% (492/1,229) Post HEART-Pathway implementation: 49.3% (816/1,655) 9.3% increase; aOR 1.47; 95% CI 1.23 to 1.74 **White:** Pre HEART-Pathway implementation: 36.2% (898/2,484) Post HEART-Pathway implementation: 39.6% (1,230/3,106) 3.4% increase; aOR 1.13; 95% CI 0.99 to 1.28 **Non-White vs White** Early discharge occurred more frequently in non-White patients compared to White patients (a 9.7% difference) 9.7%; 95% CI 6.7 to 12.7, p<0.001post HEART-Pathway implementation.	Compared with White patients, early discharge occurred more frequently in non-White patients post-implementation.
Snavely et al. 2021[12]	Algorithm/tool vs no algorithm/tool	Quality	OCT at 30-days (stress testing, coronary CT angiography, or invasive coronary angiography)	**Non-White:** Pre HEART-Pathway implementation: 32.7% (402/1,229) Post HEART-Pathway implementation: 25.7% (425/1,655) 7.0% decrease; aOR: 0.78; 95% CI 0.65 to 0.93 **White:** Pre HEART-Pathway implementation: 35.4% (879/2,484) Post HEART-Pathway implementation: 33.4% (1,037/3,106) 2% decrease; aOR 0.94; 95% CI 0.83 to 1.06) **Non-White vs White** 7.7% fewer non-White patients completed OCT at 30-days compared with White patients post HEART-Pathway implementation 7.7%; 95% CI 5.0% to 10.4%, p<0.001	Compared with White patients, fewer non-White patients underwent OCT at 30-days post implementation
Thompson et al. 2021[13]	Natural language processing algorithm vs same algorithm w/ modifications	Quality	Referral for education, treatment options, and care pathways	**False-negative Rates** Black patients: 0.32; 95% CI: 0.27–0.37 White patients: 0.17; 95% CI: 0.12–0.23	The algorithm was significantly more likely to falsely classify Black patients as not needing resources than white patients (the 95% CIs did not overlap).

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Wille et al. 2013[14]	Algorithm/tool vs no algorithm/tool	Access	Death or ineligibility for transplantation due to morbidity while on waitlist	**Adjusted odds of death or becoming too sick for transplantation within 3 years of listing** *Pre-LAS cohort:* Blacks (43.8%), whites (30.8%), aOR 1.84; p <0.001 *LAS cohort:* Blacks (14.0%), whites (13.3%); OR 0.93; p = 0.74	Racial disparities in lung transplantation existed before implementation of the LAS but became nonsignificant post-implementation. Analyses were logistic regression models, restricted to patients listed for the first 2 years of each time period to allow for the full 3 years of follow-up.
Wille et al. 2013[14]	Algorithm/tool vs no algorithm/tool	Access	Death while on waitlist	**Adjusted hazard ratio (HR) of waitlist removal for death for Black patients compared with white patients** *Pre-LAS cohort:* 1.18; 95% CI 0.99–1.40; p = 0.06 *LAS cohort:* 0.83; 95% CI 0.62–1.10; p = 0.18	Implementation of the LAS was associated with a decrease in racial disparities in lung transplantation. Analyses were Cox proportional hazards regression models covering the entire pre-LAS and LAS periods.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Williams et al. 2022[15]	Algorithm/tool vs different algorithms for same clinical purpose	Access	Percent eligible for screening by race and ethnicity	Percent (95% CI) eligible: **White non-Hispanic**: USPSTF 2013: 21.9% (21.1, 22.7); USPSTF 2021: 35.8% (34.8, 36.7) PCLOm2012 at 1.5% threshold: 36.2% (27.7, 34.5) PCLOm2012 at 1.0% threshold: 46.3% (45.3, 47.2) **Black non-Hispanic:** USPSTF 2013: 16.0% (13.2, 18.8); USPSTF 2021: 28.5% (25.2, 31.9) PCLOm2012 at 1.5% threshold: 31.1% (27.7, 34.5); PLCOm2012 at 1.0% threshold: 39.3% (35.8, 42.7) **Hispanic:** USPSTF 2013: 9.8% (5.3, 14.3); USPSTF 2021: 18.0% (12.4, 23.7) PLCOm2012 at 1.5% threshold: 15.0% (9.7, 20.2) PLCOm2012 at 1.0% threshold: 20.3% (13.7, 26.9) **Other-non-Hispanic:** USPSTF 2013: 22.1% (18.2, 26.0), USPSTF 2021: 39.3% (34.5, 44.0); PLCOm2012 at 1.5% threshold: 43.4% (38.7, 48.1) PCLOm2012 at 1.5% threshold: 51.4% (46.6, 56.3) **Total**: USPSTF 2013: 21.0% (20.2, 21.8); USPSTF 2021: 34.7% (33.8, 35.6) PLCOm2012 at 1.5% threshold: 35.3% (34.4, 36.2) PLCOm2012 at 1.0% threshold: 45.0% (44.1, 45.9)	USPSTF 2021 criteria will increase the proportion of individuals eligible for lung cancer screening across racial and ethnic groups compared with the 2013 guidelines. However racial differences remain, as White patients would have the highest percentage eligible compared with Black and Hispanic individuals. The PLCOm2012(Race 3L) model using both the 1.0% and 1.5% 6-year risk threshold identified the largest proportion of individuals overall and within racial and ethnic groups, indicating greater sensitivity of this model in identifying more racially diverse groups at risk for lung cancer.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Yoo et al. 2023[16]	No comparator. Compared outcomes by racial and ethnic groups.	Health	Stroke *Percentage of stroke incidents among those who would have not been offered antithrombotic therapy under a specific ACC/AHA guideline)	**CHA_2DS_2-VASc (negative event frequency) under 2020 ACC/AHA guideline** *Asian* 2.15%, 95% CI 1.73% to 2.65% *Black* 2.21% 95% CI 1.03% to 3.44% *White* 2.14%, 95% CI 1.95% to 2.33%, *Hispanic* 1.78%, 95% CI 1.17% to 2.47%, p<0.001 **CHA_2DS_2-VASc (negative event frequency) under 2014ACC/AHA guideline** *Asian* 2.26%, 95% CI 1.70% to 2.90% *Black* 2.19%% 95% CI 0.64% to 3.82% *White* 2.21%, 95% CI 1.98% to 2.45% *Hispanic* 3.30%, 95% CI 2.13% to 4.65%, p<0.001	Under the 2014 ACC/AHA guideline, among those who would not have been offered antithrombotic therapy, Hispanic patients had the highest negative event rate (stroke). Under the 2020 ACC/AHA guideline, Hispanic patients had the lowest negative event rate (stroke).

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Zhang et al. 2018[17]	Algorithm/tool vs no algorithm/tool	Access	Impact of KAS implementation on waitlisting in incident ESRD patients	**Pre to Post KAS** *White* Adjusted HR: 0.89; 95% CI 0.87 to 0.91 Waitlisting rate was 11% lower post KAS implementation *Black* Adjusted HR: 0.96; 95% CI 0.93 to 0.98 Waitlisting rate was 4% lower post KAS implementation *Hispanic* Adjusted HR: 0.90; 95% CI 0.87 to 0.93 Waitlisting rate was 10% lower post KAS implementation *Asian* Adjusted HR: 0.92; 95% CI 0.87 to 0.97 Waitlisting rate was 8% lower post KAS implementation	Zhang et al. observed declines in waitlisting for all racial/ethnic groups post KAS implementation.[17] Sensitivity analyses excluding patients with a living donor transplant within 180 days, patients with a history of cancer, or patients who die within 30 days after starting dialysis were similar to the main findings for each group.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Zhang et al. 2018[17]	Algorithm/tool vs no algorithm/tool	Access	Racial and ethnic differences in waitlisting (time from dialysis start to waitlisting) in incident ESRD patients	**Pre KAS** *Black vs White* Adjusted HR: 0.81; 95% CI 0.80 to 0.82 Waitlisting rate was 19% lower for Black patients with incident ESRD compared with White patients pre-KAS implementation *Hispanic vs White* Adjusted HR: 1.07; 95% CI 1.06 to 1.09 Waitlisting rate was 7% higher for Hispanic patients with incident ESRD compared with White patients pre-KAS implementation *Asian vs White* Adjusted HR: 1.20; 95% CI 1.18 to 1.22 Waitlisting rate was 20% higher for Asian patients with incident ESRD compared with White patients pre-KAS implementation **Post KAS** *Black vs White* Adjusted HR: 0.88; 95% CI 0.85 to 0.90 Waitlisting was 12% lower for Black patients with incident ESRD compared with White patients post KAS implementation. The racial difference in waitlisting declined following KAS implementation (p<0.001). *Hispanic vs White* Adjusted HR: 1.08; 95% CI 1.04 to 1.12 Waitlisting was 8% higher for Hispanic patients with incident ESRD compared with White patients post KAS implementation. The racial difference in waitlisting did not differ between groups following KAS implementation (p=0.62). *Asian vs White* Adjusted HR: 1.23; 95% CI 1.17 to 1.30 Waitlisting was 23% higher for Asian patients with incident ESRD compared with White patients post KAS implementation. The racial difference in waitlisting did not differ between groups following KAS implementation (p=0.27).	Zhang et al. found that the waitlisting rate was lower for Black patients with incident ESRD at pre (19%) and post (12%) KAS implementation compared with White patients. The rate difference between Black and White patients following KAS implementation declined, suggesting a statistically significant difference post KAS. Hispanic and Asian patients had a higher rate of waitlisting compared with White patients at pre and post KAS implementation.[17]

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Zhang et al. 2018[17]	Algorithm/tool vs no algorithm/tool	Access	Monthly waitlisting rate in prevalent ESRD patients (patients on dialysis not already on the waiting list)	**Pre to Post KAS** *White* Adjusted time-series analyses: 3.57/10,000 patients; 95% CI 0.62 to 6.51 Monthly waitlisting rate decreased in White patients with prevalent ESRD from pre to post KAS implementation (p=0.017). *Black* Adjusted time-series analyses: 3.50/10,000 patients; 95% CI 0.79 to 6.21 Monthly waitlisting rate decreased in Black patients with prevalent ESRD from pre to post KAS implementation (p=0.011). *Hispanic* Adjusted time-series analyses: 4.56/10,000 patients; 95% CI 0.38 to 8.74 Monthly waitlisting rate decreased significantly (p=0.03) in Hispanic patients with prevalent ESRD from pre to post KAS implementation. *Asian* Adjusted time-series analyses: 5.43/10,000 patients; 95% CI -1.82 to 12.69 Monthly waitlisting rate decreased in Asian patients with prevalent ESRD from pre to post KAS implementation (p=0.13).	Zhang et al. found that the monthly waitlisting rate decreased in all racial/ethnic groups following KAS implementation. The decrease was statistically significant for White, Black, and Hispanic patients from pre to post KAS implementation.[17]
Zhang et al. 2018[17]	Algorithm/tool vs no algorithm/tool	Access	Newly active waitlisting (patients eligible to be called for transplantation at any time)	**Pre-KAS** Overall: 72.1% 72.3% White, 71.3% Black, 72.2% Hispanic, 72.7% Asian, p<0.001 **Post-KAS** Overall: 73.5% 71.4% White, 76.3% Black, 78.0% Hispanics, 73.5 Asian, p<0.001	The overall proportion of newly active waitlisted patients across all racial/ethnic groups increased from pre- to post-KAS 72.1% vs 73.5%). The individual proportions of Blacks, Hispanics, and Asians increased, while the proportion of newly active White patients decreased from pre- to post-KAS.

Author / Year	Comparison	Outcome Category	Reported Outcome	Results	Summary
Zhang et al. 2018[17]	Algorithm/tool vs no algorithm/tool	Access	Inactive waitlisting counts	Inactive waitlisting counts were lower following KAS implementation (p<0.001) and a greater decline was reported among Black and Hispanic patients (p<0.0001).[17]	Inactive waitlisting counts were lower following KAS implementation (p<0.001) and a greater decline was reported among Black and Hispanic patients (p<0.0001).[17]

Abbreviations: aOR = adjusted odds ratio; APACHE IVa = Acute physiology and chronic health evaluation; ARF = acute respiratory failure; aSHR = adjusted subdistribution hazard ratio; BMI = body mass index; BP = blood pressure; CAD = coronary artery disease; CCI = Charlson Comorbidity Index; CG CrCl = Cockcroft-Gault Creatinine Clearance; CISNET = Cancer intervention and surveillance modeling network; CI = confidence interval; CKD = chronic kidney disease; CKD-EPI = Chronic Kidney Disease Epidemiology Collaboration; CLRD = chronic lower respiratory disease; COPD = chronic obstructive pulmonary disease; CSC = crisis standards of care; csPCa = clinically significant prostate cancer; eGFR = estimated glomerular filtration rate; EHR = electronic health record; eICU-CRD = eICU-Collaborative Research Database; EPTS = estimated post-transplant survival; ESRD = end-stage renal disease; FAQ = frequently asked question; FDR = false detection rate; FEV1 = forced expiratory volume; FNR = false negative rate; FVC = forced vital capacity; GLI = Global Lung Function Initiative; HDL = high-density lipoprotein; HR = hazard ratio; HTN = hypertension; ICU = intensive care unit; IQR = interquartile range; KAS = kidney allocation system; KDPI = Kidney Donor Profile Index; KPPC RC = Kaiser Permanente prostate cancer risk calculator; LAPS2 = Laboratory-based Acute Physiology Score version 2; LAS = Lung Allocation Score; LDL= low-density lipoprotein; LYFS-CT = life-years from screening-computed tomography; MIMIC-III = Medical Information Mart for Intensive Care III; NNS = number needed to screen; NSLT = National lung screening trial; OASIS = Oxford Acute Severity of Illness Score; OCT = objective cardiac testing; OR = odds ratio; PBCG = Prostate Biopsy Collaborative Group; PCPT RC = Prostate Cancer Prevention Trial risk calculator; PPR = predictive positive rate; PSA = prostate-specific antigen; SD = standard deviation; SHR = subdistribution hazard ratio; SMR = standardized mortality ratio; SOFA = Sequential Organ Failure Assessment; UNOS = United Network for Organ Sharing; USPSTF = United States preventive services taskforce; USRDS = United States Renal Data System

Table D-2. Reported outcomes of studies addressing Key Question 2

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Ahmed et al. 2021[18]	Removed race from eGFR	Quality	Reclassification to early-stage CKD	**African American patients** Reclassified to eGFR 52 to 60: 434 patients, a 16% increase from 2225 to 2659 of patients w/ CKD in registry	Removal of race would result in a 16% increase in the number of African Americans with CKD in the registry.
Ahmed et al. 2021[18]	Removed race from eGFR	Quality	Reclassification to more severe CKD Stage	**African American patients** Reclassified to a more severe CKD stage: 33.4% (743/2225) Reclassified from stage 3A to 3B: 47% (549/1167) Reclassified from stage 3B to 4: 24.3% (167/687) Reclassified from stage 4 to 5: 9.2% (27/295)	Removal of race would result in 33.4% of African Americans with CKD being reclassified to a more severe CKD stage.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Ahmed et al. 2021[18]	Removed race from eGFR	Quality	Reclassification from eGFR > 20 to eGFR ≤ 20	**African American patients** CKD-EPI w/race: 7% (156/2225) had an eGFR ≤ 20 and 19.2% (30/156) had a transplant referral, evaluation, or waitlist status CKD-EPI w/out race: 3.1% (64/2069) reclassified and 0% (0/64) had a transplant referral	Removal of race would result in 3.1% of African Americans with an eGFR > 20 being reclassified to an eGFR ≤ 20. None of the reclassified patients had a transplant referral.
Ashana et al. 2021[1]	Simulation analysis adjusted category thresholds	Access	Reclassification of high-priority threshold	**Black vs White patients in the highest-priority category (i.e., SOFA < 6 indicating a higher likelihood of survival)** SOFA <6: 5.3% vs 6.9%, Black patients had lower in-hospital mortality than White patients **Simulation analysis, Black vs White patients** Black patients with SOFA 6 to 8 reclassified into highest-priority category (n=2611): 6.7% vs 6.9%, similar in-hospital mortality for Black and White patients The Black patients thus reclassified represented 9.4% of all Black patients placed in a lower-priority category.	Among patients in the highest-priority category, Black patients had lower in-hospital mortality than White patients. Authors performed a simulation analysis reclassifying Black patients into the highest-priority category until rates of in-hospital mortality were similar between Black and White patients. Findings suggest 9.4% of all Black patients thus reclassified from a lower-priority category would have been ineligible for available resources.
Baugh et al. 2022[19]	Replaced race-based lung function algorithm with non-race-based algorithms	Quality	Estimated lung function	**Black patients (n=530)** *Lung function, $ppFEV_1$* With race (Hankinson): 76.8% (SD: 27.5) vs 71.8% (26.1) in White patients Without race (NHW-H): 64.7% (23.1) vs 71.8 (26.1) in White patients Without race (GLI-O): 70.0% (24.9) vs 77.9 (28.3) in White patients *Lung function, FVC* With race (Hankinson): 92.9% (SD: 20.1) vs 90.6% (17.9) in White patients Without race (NHW-H): 76.6% (16.4) vs 90.6 (17.9) in White patients Without race (GLI-O): 85.5% (18.3) vs 101.8 (20.1) in White patients	Black patients were predicted to have better lung function than White patients when ppFEV1 or FVC were estimated by an algorithm that includes race. Black patients were predicted to have worse lung function than White patients when estimated by 2 different algorithms that do not use race.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Bundy et al. 2022[20]	Removed race from eGFR	Quality	Prediction of 2-year risk of end-stage kidney disease	**Black patients (n=1,631)** *Area under the curve (AUC), creatinine with age, sex, and race* eGFR: 0.924; 95% CI 0.912 to 0.938 Kidney Failure Risk Equation (KFRE): 0.951; 95% CI 0.941 to 0.961 *AUC, creatinine, age, sex without race* eGFR: 0.925; 95% CI 0.913 to 0.938 KFRE: 0.951; 95% CI 0.941 to 0.961 *AUC, creatinine plus cystatin C with age, sex, race* eGFR: 0.926; 95% CI 0.914 to 0.938 KFRE: 0.953; 95% CI 0.943 to 0.962 *AUC, creatinine plus cystatin C, age, sex, without race* eGFR: 0.927; 95% CI 0.916 to 0.939 KFRE: 0.953; 95% CI 0.944 to 0.962 **Non-Black patients (n=2242)** *AUC, creatinine with age, sex, and race* eGFR: 0.922; 95% CI 0.909 to 0.934 KFRE: 0.954; 95% CI 0.944 to 0.962 *AUC, creatinine, age, sex without race* eGFR: 0.923; 95% CI 0.911 to 0.935 KFRE: 0.954; 95% CI 0.944 to 0.962 *AUC, creatinine plus cystatin C with age, sex, race* eGFR: 0.920; 95% CI 0.907 to 0.931 KFRE: 0.953; 95% CI 0.943 to 0.961 *AUC, creatinine plus cystatin C, age, sex, without race* eGFR: 0.921; 95% CI 0.909 to 0.933 KFRE: 0.953; 95% CI 0.943 to 0.961	All versions of the equations with and without race and/or cystatin C were strong predictors of 2-year risk of end-stage kidney disease in Black and non-Black patients. Removing race from creatinine equations that did not include cystatin C improved calibration for Black patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Casal et al. 2021[21]	Removed race from eGFR	Quality	Estimated renal function	**Black patients (n=340)** Median CKD-EPI with race: 103 (IQR 85 to 122) Median CKD-EPI without race*:* 89 (IQR 73 to 105) eGFR <60, CKD-EPI with race: 7% eGFR <60, CKD-EPI without race: 13% Reclassification to more severe disease stage, CKD-EPI without race vs with race: 26%	Removing race from eGFR calculations resulted in a lower median estimate of renal function. More patients were classified as having severe disease when race was removed.
Casal et al. 2021[21]	Removed race from eGFR	Access	Ineligible for drug therapy for cancer	**Black patients (n=340)** The impact of removing race from CKD-EPI: Cisplatin ineligibility increased from 7% to 13% Pemetrexed ineligibility increased from 1% to 3% Bendamustine ineligibility stayed constant at 1% Mitomycin, Capecitabine, and Fludarabine ineligibility increased from 0% to 1% Etoposide and topotecan ineligibility stayed constant at 0%	Cancer drugs often have minimum eGFR thresholds, and patients below these cutoffs are not eligible for therapy. Removing race from eGFR would have led to some Black patients becoming ineligible for some cancer treatments.
Coresh et al. 2019[22]	GFR estimated from metabolic panel without adjustment for race or use of creatinine	Quality	Estimated renal function	**All patients (n=265)** *Accuracy of GFR measurement, 1-P_{30}* Metabolite panel without age, sex, or creatinine: 3.4% Metabolite panel with age and sex, without creatinine: 3.0% Metabolite panel with age,sex, and creatinine: 1.9% CKD-EPI eGFR with creatinine: 18.5% CKD-EPI eGFR with cystatin-C: 9.1%	Estimating GFR using a metabolite panel without race or creatinine was more accurate than the CKD-EPI eGFR algorithm.
Diao et al. 2023[23]	Removed race from eGFR	Quality	Diagnosis of chronic kidney disease (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *New diagnoses:* 433,524 patients; 95% CI 350,081 to 516,966; 1.63% of all Black adults in U.S.; 95 % CI 1.38 to 1.91 **Non-Black patients** *Diagnoses reversed:* 5,511,894 patients; 95% CI 4,860,944 to 6,162,844; 2.71% of all non-Black adults in U.S.; 95% CI 2.44 to 3.00	Estimating GFR using the new equation without the race coefficient would result in new diagnoses of CKD for close to half a million Black patients, and the reversal of CKD diagnoses for 5.5 million non-Black patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Diao et al. 2023[23]	Removed race from eGFR	Quality	Reclassification of chronic kidney disease stage (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *Reclassified to more severe stage:* 584,000 patients; 95% CI 508,000 to 667,000; 14.7% of Black adults already diagnosed; 95 % CI 12.8 to 16.8 **Non-Black patients** *Reclassified to less severe stage:* 4.59 million patients; 95% CI 4.28 to 4.92; 16.4% of all non-Black adults already diagnosed; 95% CI 15.3 to 17.6	Estimating GFR using the new equation without the race coefficient would result in more than half a million Black patients reclassified with more severe CKD, and more than 4.5 million non-Black patients reclassified with less severe CKD.
Diao et al. 2023[23]	Removed race from eGFR	Access	Eligibility for kidney transplantation (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *Newly eligible:* 15,809 patients; 95% CI 3,777 to 27,840; increase of 9.36%; 95% CI 2.16 to 16.57 **Non-Black patients** *Newly ineligible:* 24,958 patients; 95% CI 6,172 to 43,745; eligibility decreased by 8.14%; 95% CI 2.0 to 14.28	Estimating GFR using the new equation without the race coefficient would result in a 9% increase in Black patients who are eligible for kidney transplant, and a decrease in eligibility for 8% of non-Black patients.
Diao et al. 2023[23]	Removed race from eGFR	Access	Eligibility for kidney donation (CKD-EPI 2021 versus MDRD 2006)	**Black patients** *Newly ineligible:* 246,144 patients; 95% CI 189,465 to 302,823; eligibility decreased by 2.4%; 95% CI 1.92 to 2.88 **Non-Black patients** *Newly eligible:* 3,963,346 patients; 95% CI 3,463,585 to 4,463,108; eligibility increased by 6.76%; 95% CI 6.03 to 7.49	Estimating GFR using the new equation without the race coefficient would result in a 2% decrease in Black patients eligible to donate a kidney and a 7% increase in non-Black patients eligible to donate a kidney.
Diao et al. 2023[23]	Removed race from eGFR	Access	Eligibility for nephrologist referral (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *Newly eligible:* 41,769 patients; 95% CI 19,787 to 63,750; eligibility increased by 4.93%; 95% CI 2.29 to 7.57 **Non-Black patients** *No longer eligible:* 75,796 patients; 95% CI 35,406 to 116,186; eligibility decreased by 1.32%; 95% CI 0.61 to 2.04	Estimating GFR using the new equation without the race coefficient would result in an increase of 5% of Black patients eligible for referral to nephrology care, and a 1% decrease in non-Black patients eligible for referral.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Diao et al. 2023[23]	Removed race from eGFR	Access	Eligibility for preemptive arteriovenous fistula (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *Newly eligible:* 17,207 patients; 95% CI 4,103 to 30,312; eligibility increased by 10.92%; 95% CI 2.72 to 19.12 **Non-Black patients** *No longer eligible:* 16,190 patients; 95% CI 2,262 to 34,642; eligibility decreased by 6.35%; 95% CI 0.89 to 13.59	Estimating GFR using the new equation without the race coefficient would result in an 11% increase in Black patients eligible for preemptive arteriovenous fistula and a 6% decrease in non-Black patient eligibility.
Diao et al. 2023[23]	Removed race from eGFR	Access	Kidney disease education covered by Medicare (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *Newly covered:* 22,213 patients; 95% CI 2,915 to 41,511; coverage increased by 45.25%; 95% CI -3.37 to 93.87 **Non-Black patients** *No longer covered:* 86,959 patients; 95% CI 43,000 to 130,918; coverage decreased by 25.32%; 95% CI 15.3 to 35.35	Estimating GFR using the new equation without the race coefficient would result in a 45% increase in Black patients eligible for Medicare coverage of kidney disease education and a 25% decrease in non-Black patients covered.
Diao et al. 2023[23]	Removed race from eGFR	Access	Medical nutrition therapy covered by Medicare (CKD-EPI 2021 vs MDRD 2006)	**Black patients** *Newly covered:* 183,653 patients; 95% CI 144,724 to 222,582; coverage increased by 48.49%; 95% CI 36.08 to 60.9 **Non-Black patients** *No longer covered:* 1,042,060 patients; 95% CI 850,782 to 1,233,338; coverage decreased by 29.38%; 95% CI 25.2 to 33.57	Estimating GFR using the new equation without the race coefficient would result in a 48% increase in Black patients eligible for Medicare coverage of nutrition therapy and a 29% decrease in non-Black patients covered.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Diao et al. 2023[23]	Removed race from eGFR	Access	Dose reduction for any of 10 common medications* (CKD-EPI 2021 vs MDRD 2006) *ACE inhibitors, opioids, ARBs, antihyperglycemics, beta blockers, anticonvulsants, antibacterials, anticoagulants, diuretics, statins	**Black patients** *Dose reduction:* 222,336 patients; 95% CI 169,319 to 275,352; change of 38.12%; 95% CI 28.58 to 47.67 **Non-Black patients** *Reversal of dose reduction:* 1,473,681 patients; 95% CI 1,215,272 to 1,732,091; change of 34.02%; 95% CI 30.04 to 38.00	Estimating GFR using the new equation without the race coefficient would result in a 38% increase in the number of Black patients having the dose of at least 1 medication reduced and a 34% decrease in the number of non-Black patients with dose reductions.
Doshi et al 2022[24]	Removed Black race from KDRI	Access	Kidney Donor Profile Index/Kidney Donor Risk Index	**Black patients** *KDPI score 86-100*: 18.0% with race, 7.3% without race *KDPI 35-85*: 49.5% with race, 42.1% without race *KDPI 21-34*: 19.6% with race, 13.7% without race *KDPI 0-20*: 12.8% with race, 36.9% without race Overall, 49% of donors were reclassified to lower risk categories *Discard probability*: based on KDRI, 4,718 kidneys from Black donors were discarded. If the discard probability for non-Black donors were applied to Black donors, 4,231 kidneys would have been discarded.	Removing race from KDPI would result in 49% of Black kidney donors reclassified into lower risk categories for graft failure. If the probability of discarding Black donor kidneys was equivalent to non-Black donors, approximately 70 more kidneys from Black donors would be available for transplant each year.
Drawz et al. 2012[25]	Added race and chronic kidney disease to Framingham Risk Score	Quality	Net reclassification index (NRI) for 5-year risk of coronary heart disease.	**Black men** NRI: -4.1%, p=0.46 **Black women** NRI: 4.4%, p=0.31 **Non-Black men** NRI: 1.3%, p=0.54 **Non-Black women** NRI: -5.5%, p=0.11	Adding race and chronic kidney disease to the FRS did not improve classification of the risk of heart disease in Black or non-Black men or women.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Duggal et al. 2021[26]	Removed race from eGFR	Quality	Estimated renal function	**NHANES, Black individuals (10.9%)** Mean CKD-EPI with race: 104.5 (SD 27.2) Mean CKD-EPI without race: 90.2 (SD 23.5) CKD-EPI with race: 5.2% had eGFR <60 CKD-EPI without race: 10.6% had eGFR <60 Removing race reclassified 5.6% of individuals with eGFR ≥60 to stage 3 disease (30≤ eGFR <60) Removing race reclassified 6.3% of individuals with stage 3 disease to stage 4 disease (15≤ eGFR <30) Removing race reclassified 30.7% of individuals with stage 4 disease to stage 5 disease (eGFR <30) **VA cohort, Black patients (17.5%)** Mean CKD-EPI with race: 87.3 (SD 23.3) Mean CKD-EPI without race: 75.7 (SD 20.8) CKD-EPI with race: 12.4% had eGFR <60 CKD-EPI without race: 21.6% had eGFR <60 Removing race reclassified 10.5% of individuals with eGFR ≥60 to stage 3 disease Removing race reclassified 6.4% of individuals with stage 3 disease to stage 4 disease Removing race reclassified 29.1% of individuals with stage 4 disease to stage 5 disease	Removing race from eGFR resulted in twice as many Black individuals (5.2% vs 10.6%) meeting criteria for a diagnosis of CKD (eGFR <60), based on data from a nationwide, longitudinal U.S. database. Similarly, diagnosis of Black patients increased from 12% to 22% in a large sample of patients treated by the Veterans' Administration.
Duggal et al. 2021[26]	Removed race from eGFR	Quality	Estimated 2-year risk of kidney failure among patients with sustained eGFR <60	**VA cohort, Black patients (17.5%)** CKD-EPI with race: 2.7% (SD 10.3) CKD-EPI without race: 3.4% (SD 11.5) Patients actually progressing to kidney failure: 3.8% Removing race improved the accuracy of predictions only slightly for the full group (C statistic 0.954 with race and 0.957 without race) and for Black patients specifically (C statistic 0.959 with race and 0.967 without race).	Removing race from eGFR resulted in only slightly more accurate prediction of the risk of kidney failure in Black patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Elmaleh-Sachs et al. 2021[27]	Removed race from GLI	Health	Prediction of CLRD-related events by an 80% predicted FEV1 threshold (HRs compare event rates for individuals with scores above the threshold vs. at or below it)	**Secondary analyses adjusted for predictors* of CLRD events or all-cause mortality in MESA** *BMI, educational attainment, smoking status, pack-years, BP, HDL, LDL, total cholesterol, history of HTN and diabetes **White (n=1,187)***:* Race/ethnicity-based GLI: HR 5.75; 95% CI 3.55 to 9.322; Race/ethnicity-neutral GLI: HR 8.12; 95% CI 5.02 to 13.33 **Black (n=844):** Race/ethnicity-based GLI: HR 3.56; 95% CI 2.10 to 6.05; Race/ethnicity-neutral GLI: HR 3.06; 95% CI 1.81 to 5.18 **Hispanic (n=755):** Race/ethnicity-based GLI: HR 3.79; 95% CI 1.94 to 7.42; Race/ethnicity-neutral GLI: HR 7.96; 95% CI 4.03 to 15.73 **Asian (n=558):** Race/ethnicity-based GLI: HR 1.97; 95% CI 0.84 to 4.60; Race/ethnicity-neutral GLI: HR 2.60; 95% CI 1.09 to 6.20 **Subgroup analysis of participants with airflow limitation (FEV1/FVC <0.70)** **White (n=310):** Race/ethnicity-based GLI: HR 6.65; 95% CI 3.03 to 14.60; Race/ethnicity-neutral GLI: HR 10.61; 95% CI 4.98 to 22.61 **Black (n=170):** Race/ethnicity-based GLI: HR 3.10; 95% CI 1.30 to 7.36; Race/ethnicity-neutral GLI: HR 2.53; 95% CI 0.93 to 6.86: **Hispanic (n=93):** Race/ethnicity-based GLI: HR 2.65; 95% CI 0.89 to 7.92; Race/ethnicity-neutral GLI: HR 4.83; 95% CI 1.62 to 14.44 **Asian (n=97):** Race/ethnicity-based GLI: HR 1.52; 95% CI 0.48 to 4.78; Race/ethnicity-neutral GLI: HR 1.64; 95% CI 0.53 to 5.10	Authors found no evidence that percentage predicted FEV1 or FVC calculated using GLI equations with race/ethnicity improved the prediction of CLRD-related events compared with GLI calculations without race/ethnicity. Findings were similar for all-cause mortality. Additionally, hazard ratios were larger for the prediction of CLRD events for Whites, Hispanics, and Asians when using the race/ethnicity-neutral equation compared with the race/ethnicity-based equation. Hazard ratios for this outcome were lower for Blacks when using the race/ethnicity-neutral equation.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Elmaleh-Sachs et al. 2021[27]	Removed race from GLI	Health	Prediction of CLRD-related events by an 80% predicted FVC threshold (HRs compare event rates for individuals with scores above the threshold vs. at or below it)	**White (n=1,187)** Race/ethnicity-based GLI: HR 4.23; 95% CI 2.52 to 7.13 Race/ethnicity-neutral GLI: HR 4.01; 95% CI 2.10 to 7.67 **Black (n=844)** Race/ethnicity-based GLI: HR 2.29; 95% CI 1.27 to 4.14 Race/ethnicity-neutral GLI: HR 1.48; 95% CI 0.85 to 2.57 **Hispanic (n=755)** Race/ethnicity-based GLI: HR 2.41; 95% CI 1.10 to 5.30 Race/ethnicity-neutral GLI: HR 4.67; 95% CI 2.04 to 10.66 **Asian (n=558)** Race/ethnicity-based GLI: HR 0.90; 95% CI 0.31 to 2.66 Race/ethnicity neutral GLI: HR 1.00; 95% CI 0.30 to 3.38	Authors found no evidence that percentage predicted FEV1 or FVC calculated using GLI equations with race/ethnicity improved the prediction of CLRD-related events compared with GLI calculations without race/ethnicity. Findings were similar for all-cause mortality. Additionally, hazard ratios were larger for the prediction of CLRD events for Whites, Hispanics, and Asians when using the race/ethnicity-neutral equation compared with the race/ethnicity-based equation. Hazard ratios for this outcome were lower for Blacks when using the race/ethnicity-neutral equation.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Elmaleh-Sachs et al. 2021[27]	Removed race from GLI	Health	Prediction of all-cause mortality events by an 80% predicted FEV_1 threshold (HRs compare event rates for individuals with scores above the threshold vs. at or below it)	**White (n=1,187):** Race/ethnicity-based GLI: HR 1.80; 95% CI 1.32 to 2.46; Race/ethnicity-neutral GLI: HR 2.16; 95% CI 1.53 to 3.04 **Black (n=844):** Race/ethnicity-based GLI: HR 2.41; 95% CI 1.73 to 3.37; Race/ethnicity-neutral GLI: HR 1.84; 95% CI 1.35 to 2.53 **Hispanic (n=755):** Race/ethnicity-based GLI: HR 1.89; 95% CI 1.22 to 2.91; Race/ethnicity-neutral GLI: HR 2.55; 95% CI 1.54 to 4.22 **Asian (n=558):** Race/ethnicity-based GLI: HR 0.84; 95% CI 0.50 to 1.42; Race/ethnicity-neutral GLI: HR 0.89; 95% CI 0.49 to 1.61 **Subgroup analysis of participants with airflow limitation (FEV1/FVC <0.70)** **White (n=310):** Race/ethnicity-based GLI: HR 1.20; 95% CI 0.78 to 1.83; Race/ethnicity-neutral GLI: HR 1.55; 95% CI 1.00 to 2.41 **Black (n=170):** Race/ethnicity-based GLI: HR 1.28; 95% CI 0.77 to 2.12; Race/ethnicity-neutral GLI: HR 1.26; 95% CI 0.75 to 2.13 **Hispanic (n=93):** Race/ethnicity-based GLI: HR 1.56; 95% CI 0.70 to 3.47; Race/ethnicity-neutral GLI: HR 1.37; 95% CI 0.60 to 3.14 **Asian (n=97):** Race/ethnicity-based GLI: HR 1.01; 95% CI 0.44 to 2.32; Race/ethnicity-neutral GLI: HR 1.02; 95% CI 0.44 to 2.39	Authors found no evidence that percentage predicted FEV1 or FVC calculated using GLI equations with race/ethnicity improved the prediction of CLRD-related events compared with GLI calculations without race/ethnicity. Findings were similar for all-cause mortality. Additionally, hazard ratios were generally larger for the prediction of all-cause mortality for Whites, Hispanics, and Asians when using the race/ethnicity-neutral equation compared with the race/ethnicity-based equation. Hazard ratios for this outcome were lower for Blacks when using the race/ethnicity-neutral equation.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Elmaleh-Sachs et al. 2021[27]	Removed race from GLI	Health	Prediction of all-cause mortality events by an 80% predicted FVC threshold (HRs compare event rates for individuals with scores above the threshold vs. at or below it)	**White (n=1,187)** Race/ethnicity-based GLI: HR 2.19; 95% CI 1.56 to 3.08 Race/ethnicity-neutral GLI: HR 2.34; 95% CI 1.52 to 3.62 **Black (n=844)** Race/ethnicity-based GLI: HR 1.50; 95% CI 1.01 to 2.23 Race/ethnicity-neutral GLI: HR 1.40; 95% CI 1.00 to 1.96 **Hispanic (n=755)** Race/ethnicity-based GLI: HR 1.89; 95% CI 1.16 to 3.07 Race/ethnicity-neutral GLI: HR 2.98; 95% CI 1.73 to 5.14 **Asian (n=558)** Race/ethnicity-based GLI: HR 0.75; 95% CI 0.41 to 1.39 Race/ethnicity-neutral GLI: HR 0.26; 95% CI 0.08 to 0.81	Authors found no evidence that percentage predicted FEV1 or FVC calculated using GLI equations with race/ethnicity improved the prediction of CLRD-related events compared with GLI calculations without race/ethnicity. Findings were similar for all-cause mortality. Additionally, hazard ratios were larger for the prediction of all-cause mortality for Whites and Hispanics when using the race/ethnicity-neutral equation compared with the race/ethnicity-based equation. Hazard ratios for this outcome were lower for Blacks and Asians when using the race/ethnicity-neutral equation.
Fairman et al. 2020[28]	Updated PCEs were derived from more diverse population and improved statistical techniques	Health	Estimated risk of cardiovascular event	**Black patient visits (n=1250)** Risk of cardiovascular event estimated at 5% to 7.5%: original PCE 14.5% (95% CI 11.8 to 17.8) vs revised PCE 16.0% (12.6 to 20.1) Risk > 7.5%: original PCE 58.5% (54.6 to 62.9) vs revised PCE 41.6% (37.6 to 45.7) **White patient visits (n=11,306)** Risk 5% to 7.5%: original PCE 11.1% (10.3 to 12.0) vs revised PCE 13.5% (12.5 to 14.6) Risk > 7.5%: original PCE 52.8% (51.1 to 54.4) vs revised PCE 39.9% (38.2 to 41.5) Difference between Black and White patients in original PCE was significant when risk was estimated at 5% to 7.5% (p=0.017), and when risk was $\geq$ 7.5% (p=0.006). Differences not significant with revised PCE (p=0.178 and 0.448, respectively).	When using original PCE, Black patients were significantly more likely to be identified as at risk for a cardiovascular event than White patients. When using the revised PCE, the difference between Black and White patients was not significantly different.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Fairman et al. 2020[28]	Updated PCEs were derived from more diverse population and improved statistical techniques	Access	Statin prescribing	**Black patient visits (n=1,250)** Statins prescribed: original PCE 35.0% (95% CI 30.5 to 39.9) vs revised PCE 40.6% (35.0 to 46.6) **White patient visits (n=11,306)** Statins prescribed: original PCE 41.8% (39.9 to 44.4) vs revised PCE 43.0% (40.0 to 45.9) Difference between Black and White patients with original PCE was significant (p=0.013) but not with revised PCE (p=0.482).	When using original PCE, Black patients were significantly less likely to be prescribed a statin compared to White patients. When using the revised PCE, the difference between Black and White patients was not significantly different.
Foryciarz et al. 2022[29]	Compared original PCE and previously published rPCE to algorithm fairness mitigation strategies of 1) group recalibration, 2) equalized odds	Quality	Threshold calibration error (TCE), a novel measure defined as the difference between a therapeutic threshold and the implied threshold, measured by the calibration curve.	*Results reported for 2 thresholds: t1=7.5%, representing intermediate risk for ASCVD, and t2=20%, representing high risk for ASCVD.* **All patients in recalibration set (n=2,562)** TCE t1: original model 0.012 (95% CI 0.006 to 0.019) vs. recalibrated model -0.001 (-0.007 to 0.006) TCE t2: original model 0.006 (-0.0013 to 0.0023) vs. recalibrated model 0.000 (-0.019 to 0.016) **Black men (n=2,931)** TCE t2: original model 0.033 (-0.016 to 0.066) vs. recalibrated model -0.071 (-0.196 to 0.018)	Recalibrating the model at the threshold for intermediate risk resulted in better performance overall and for subgroups. Recalibrating the model at the threshold for high risk resulted in better overall performance but worse performance in Black men.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Fox et al. 2016[30]	Alternative algorithm	Access	Risk of CVD events (events classified to a higher risk category)	**Black patients** *Model 1 vs Model 6* Total events = 188 Model 1 reclassified 1 participant from low risk to high-risk (NRI 0.005) *FRS vs Model 1* Total events = 270 Model 1 reclassified 5 participants from a low-risk to high-risk category (NRI 0.019) *ACC/AHA vs Model 1* Total events = 128 Model 1 reclassified 2 participants from low-risk to high-risk category (NRI 0.016) *FRS vs Model 6* Total events = 188 Model 6 reclassified 3 participants from low-risk to high-risk category (NRI 0.016) *ACC/AHA vs, Model 6* Total events = 83 Model 6 reclassified 0 participants from low-risk to high-risk category (NRI 0.000)	Model 1 or model 6 compared with refitted FRS and ACC/AHA models suggests no substantial improvement in reclassification of CVD events from a low- to high-risk category.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Fox et al. 2016[30]	Alternative algorithm	Access	Risk of CVD nonevents (events classified to a lower-risk category)	**Black patients** *Model 1 vs Model 6* Total events = 2484 Model 6 reclassified 134 participants from high-risk to low-risk category (NRI 0.054) *FRS vs Model 1* Total events = 3419 Model 1 reclassified 67 participants from high-risk to low-risk category (NRI 0.020) *ACC/AHA vs Model 1* Total events = 2,818 Model 1 reclassified 20 participants from high-risk to low-risk category (NRI 0.007) *FRS vs Model 6* Total events = 2,482 Model 6 reclassified 190 participants from high-risk to low-risk category (NRI 0.076) *ACC/AHA vs, Model 6* Total events = 2,030 Model 6 reclassified 49 participants from high-risk to low-risk category (NRI 0.024)	Model 1 or model 6 compared with refitted FRS and ACC/AHA models suggests no substantial improvement in reclassification of CVD events from a high- to low-risk category.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Gutiérrez et al. 2022[31]	Removed race from eGFR, with and without addition of cystatin C	Quality	Risk of kidney failure with replacement therapy (KFRT)	**Hazard ratio of KFRT for eGFR=60 vs eGFR=80** *eGFR with creatinine and race* Black: 9.3; 95% CI 4.9 to 17.9 Non-Black: 3.3; 95% CI 2.5 to 4.4 Ratio, Black vs Non-Black: 2.8; 95% CI 1.6 to 4.9 *eGFR with creatinine and without race* Black: 5.1; 95% CI 2.6 to 9.9 Non-Black: 3.8; 95% CI 2.6 to 5.7 Ratio, Black vs Non-Black: 1.3; 95% CI 0.8 to 2.1 *eGFR with cystatin C and without creatinine and race* Black: 8.0; 95% CI 3.6 to 17.8 Non-Black: 2.7; 95% CI 1.7 to 4.3 Ratio, Black vs Non-Black: 3.0; 95% CI 1.5 to 5.8 *eGFR with cystatin C and creatinine, without race* Black: 9.7; 95% CI 5.0 to 18.9 Non-Black: 3.5; 95% CI 2.6 to 4.8 Ratio, Black vs Non-Black: 2.8; 95% CI 1.4 to 5.4	The eGFR equation that excludes race but includes both creatinine and cystatin C is a better predictor of KFRT risk in Black patients compared with equations that include race or exclude cystatin C.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Gutiérrez et al. 2022[31]	Removed race from eGFR, with and without addition of cystatin C	Quality	Risk of all-cause mortality	**Hazard ratio of KFRT for eGFR=60 vs eGFR=80** *eGFR with creatinine and race* Black: 1.3; 95% CI 1.2 to 1.5 Non-Black: 1.1; 95% CI 1.0 to 1.2 Ratio, Black vs Non-Black: 1.2; 95% CI 1.1 to 1.4 *eGFR with creatinine and without race* Black: 1.2; 95% CI 1.1 to 1.3 Non-Black: 1.2; 95% CI 1.1 to 1.3 Ratio, Black vs Non-Black: 1.0; 95% CI 0.9 to 1.1 *eGFR with cystatin C and without creatinine and race* Black: 1.8; 95% CI 1.5 to 2.1 Non-Black: 1.4; 95% CI 1.3 to 1.6 Ratio, Black vs Non-Black: 1.2; 95% CI 1.1 to 1.4 *eGFR with cystatin C and creatinine, without race* Black: 1.6; 95% CI 1.4 to 1.8 Non-Black: 1.4; 95% CI 1.3 to 1.5 Ratio, Black vs Non-Black: 1.1; 95% CI 1.0 to 1.2	The eGFR equation that excludes race but includes both creatinine and cystatin C is a better predictor of risk of all-cause mortality in Black patients compared with equations that include race or exclude cystatin C.
Hammond et al. 2020[32]	Use of SDOH in addition to or to replace other variables. Race was not a specific component of any algorithm.	Health	Annual incidence of all-cause hospitalization per 100 population	**Black/Hispanic patients (n=342)** True rate: 48.0 Predicted by age/sex/comorbidity: 39.9 Predicted by age/sex/comorbidity/SDOH: 47.4 Prediction by SDOH alone: 49.0 **White/Other patients (n=3272)** True rate: 29.2 Predicted by age/sex/comorbidity: 30.7 Predicted by age/sex/comorbidity/SDOH: 29.8 Predicted by SDOH alone: 29.2	Adding SDOH measures to an algorithm that included age, sex, and comorbidity improved the accuracy of predicting risk of hospitalization. Using SDOH measures alone was equal or superior to the combined algorithm. The algorithm without SDOH underpredicted risk in Black/Hispanic patients and overpredicted risk in White/Other patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Hammond et al. 2020[32]	Use of SDOH in addition to or to replace other variables. Race was not a specific component of any algorithm.	Health	Annual incidence of hospitalization for CVD per 100 population	**Black/Hispanic patients (n=342)** True rate: 16.1 Predicted by age/sex/comorbidity: 9.5 Predicted by age/sex/comorbidity/SDOH: 15.6 Prediction by SDOH alone: 16.5 **White/Other patients (n=3272)** True rate: 7.5 Predicted by age/sex/comorbidity: 8.4 Predicted by age/sex/comorbidity/SDOH: 7.6 Predicted by SDOH alone: 7.6	Adding SDOH measures to an algorithm that included age, sex, and comorbidity improved the accuracy of predicting risk of cardiovascular-related hospitalization. Using SDOH measures alone was superior to the combined algorithm for Black/Hispanic patients and equal to the combined algorithm for White/Other patients. The algorithm without SDOH underpredicted risk in Black/Hispanic patients, and overpredicted risk in White/Other patients.
Hammond et al. 2020[32]	Use of SDOH in addition to or to replace other variables. Race was not a specific component of any algorithm.	Health	Risk of death	**Black/Hispanic patients (n=342)** True rate: 3.5% Predicted by age/sex/comorbidity: 4.4% Predicted by age/sex/comorbidity/SDOH: 3.5% Prediction by SDOH alone: 3.5% **White/Other patients (n=3272)** True rate: 3.9% Predicted by age/sex/comorbidity: 3.8% Predicted by age/sex/comorbidity/SDOH: 3.9% Predicted by SDOH alone: 3.9%	Adding SDOH measures to an algorithm that included age, sex, and comorbidity improved the accuracy of predicting risk of death. Using SDOH measures alone was equal to the combined algorithm for all patients. The algorithm without SDOH overpredicted risk in Black/Hispanic patients, and underpredicted risk in White/Other patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Hammond et al. 2020[32]	Use of SDOH in addition to or to replace other variables. Race was not a specific component of any algorithm.	Access	Total annual health care costs	**Black/Hispanic patients (n=342)** True cost: $11,754 Predicted by age/sex/comorbidity: $11,928 Predicted by age/sex/comorbidity/SDOH: $11,754 Prediction by SDOH alone: $11,754 **White/Other patients (n=3272)** True cost: $9,736 Predicted by age/sex/comorbidity: $9,718 Predicted by age/sex/comorbidity/SDOH: $9,736 Predicted by SDOH alone: $9,736	Adding SDOH measures to an algorithm that included age, sex, and comorbidity slightly improved the accuracy of predicting total annual healthcare costs. Using SDOH measures alone was equal to the combined algorithm for all patients. The algorithm without SDOH slightly overpredicted risk in Black/Hispanic patients, while there was no meaningful difference in White/Other patients.
Hoenig et al. 2021[33]	Removed race from eGFR	Access	Patients preemptively listed for kidney transplant after KAS implementation	**Black patients (n=72)** 83.3% (60/72) listed after January 1, 2017, when the policy changed and race was removed from MDRD eGFR. 15% (9/60) of Black patients would not have been listed for kidney transplant if the Black race coefficient was used in MDRD eGFR. *In 2016, MDRD included race and 26% of Black patients were preemptively listed for kidney transplant compared with 70% of White patients. Authors reported that the "proportion of Black patients preemptively listed continued to increase and approached the proportion of White patients."	Authors found that the proportion of Black patients preemptively listed continued to increase since 2017. Of the 60 Black patients preemptively listed after implementation of KAS, 9 (15%) would not have been listed if the race coefficient had been used. At the end of the study period, these 9 patients had gained an average of 475.9 days on the waiting list, none had received a kidney transplant, and 1 had begun peritoneal dialysis.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Huang et al. 2022[34]	Removed race from eGFR	Quality	Prediction of acute kidney injury following percutaneous coronary intervention	**Black patients** Observed risk, mean: 10.2%; 95% CI 9.8 to 10.6 Predicted risk, regression model with race: 7.6%; 95% CI 7.5 to 7.7 Predicted risk, regression model without race: 8.2%; 95% CI 8.1 to 8.3 Predicted risk, machine learning model with race: 8.6%; 95% CI; 8.5 to 8.7 Predicted risk, machine learning model without race: 8.7%; 95% CI 8.6 to 8.8 **Non-Black patients** Observed risk, mean: 7.1%; 95% CI 7.0 to 7.2 Predicted risk, regression model with race: 7.4%; 95% CI 7.4 to 7.4 Predicted risk, regression model without race: 7.3%; 95% CI 7.3 to 7.4 Predicted risk, machine learning model with race: 7.3%; 95% CI 7.3 to 7.3 Predicted risk, machine learning model without race: 7.3%; 95% CI 7.2 to 7.3	Estimating GFR without race resulted in better or equivalent prediction of acute kidney injury following percutaneous coronary intervention in Black and non-Black patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Inker et al. 2021[35]	Replaced race in eGFR with cystatin C, beta-2 microglobulin, and beta-trace protein.	Quality	Estimated renal function	**All patients (n=2245)** *Accuracy as measured by 1-P_{30}* New model (with cystatin-C, beta-2 microglobulin, beta-trace protein, age, and sex) and creatinine but without race: 8.6% (95% CI 7.5 to 9.8) New model and creatinine but with race: 8.4% (7.3 to 9.5) New model without creatinine or race: 15.6% (14.2 to 17.2) New model without creatinine but with race: 14.8% (13.4 to 16.2) 2012 CKD-EPI eGFR with creatinine, cystatin C, age, sex, and race: 9.4% (8.2 to 10.6) 2012 CKD-EPI eGFR with cystatin C, age, and sex but without race: 17.4% (15.9 to 18.9) 2009 CKD-EPI with creatinine, age, sex, and with race: 11.8% (10.5 to 13.2)	Estimating GFR using 4 filtration markers without race was equivalent to the same algorithm with race and more accurate than previous algorithms that used race.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Inker et al. 2021[36]	1) Removed race and 2) Replaced race in eGFR	Quality	Estimated renal function	**All patients (n=4050)** *eGFR with creatinine, age, sex, and race* Overestimated GFR in Black patients (median difference -3.7 ml/minute/1.73 m^2, 95% CI -1.8 to -5.4) and in non-Black patients (median difference -0.5 ml/minute/1.73 m^2; 95% CI 0.0 to -0.9). *eGFR with creatinine, age, and sex* Underestimated GFR in Black patients (median difference 7.1 ml/minute/1.73 m2; 95% CI 5.9 to 8.8) and overestimated in Black patients (median difference -0.5 ml/minute/1.73 m^2; 95% CI 0.0 to -0.9). *Newly derived eGFR with creatinine, age and sex*: Underestimated GFR in Black patients (median difference 3.6 ml/minute/1.73 m^2; 95% CI 1.8 to 5.5) and overestimated in non-Blacks (median difference -3.9 ml/minute/1.73 m2; 95% CI -3.4 to -4.4). *Newly derived eGFR with creatinine, cystatin C, age, and sex*: Underestimated GFR in Black patients (median difference 0.1 ml/minute/1.73 m^2; 95% CI -0.9 to 1.6) and overestimated GFR in non-Blacks (median difference -2.9 ml/minute/1.73 m2; 95% CI -2.5 to -3.3 to 4.4).	New eGFR equations with creatinine and cystatin C and without race are more accurate and resulted in smaller differences between Black and non-Black patients than equations without race with either creatinine or cystatin C alone.
Julian et al. 2017[37]	Replaced race with apolipoprotein L1 genotype in Kidney Donor Risk Index	Quality	Risk of allograft failure	**Black patients** *Revised Kidney Donor Risk Index* 1-year post-transplant: AUC 0.600 (95% CI 0.561 to 0.638) 3-years post-transplant: 0.602 (0.565 to 0.639) 5-years post-transplant: 0.604 (0.556 to 0.653) *Current Index* 1-year post-transplant: 0.589 (0.553 to 0.625) 3-years post-transplant: 0.592 (0.551 to 0.633) 5-years post-transplant: 0.594 (0.554 to 0.634)	Replacing race with the APOL1 genotype improved the predictive ability of the Kidney Donor Risk Index for kidneys from deceased African-American donors.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Kabra et al. 2016[38]	Added "African-American ethnicity" to the CHA_2DS_2-VASc score, which previously included the following: congestive heart failure, hypertension, age ≥75 years, diabetes, previous stroke, vascular disease, age 65 to 74, and female sex	Quality	Estimated risk of stroke	**CHA2DS2-VASc without race** Hazard ratio:* 1.24 (1.23 to 1.25), p<0.001 C-statistic: 0.60 (0.59 to 0.61), p<0.001 **CHA2DS2-VASc with race** Hazard ratio:* 1.25 (95% CI: 1.24 to 1.26), p<0.001 C-statistic: 0.61 (95% CI: 0.60 to 0.62) Improvement in model fit with addition of race (log likelihood ratio statistic): p<0.001 Net reclassification improvement in algorithm with race: 7.6% (p<0.001) Integrated discrimination improvement, African-American patients: 0.12 Integrated discrimination improvement, White patients: <0.001 *Hazard ratios reflect relative hazard associated with 1-point increase in risk score.	Changes in the hazard ratio and C-statistic indicated that adding race improved the algorithm's ability to predict stroke. The net reclassification index demonstrated that patients whose stroke risk is reclassified by the algorithm were usually moved in the correct direction. Integrated discrimination improvement showed that when race was added, the algorithm's prediction of stroke for African Americans was 1.2% closer to the true risk, while the prediction for White patients was almost unchanged.
Kimmel et al 2013[39]	Addition of genotype data to standard warfarin dosing algorithms	Health	Time in therapeutic range	**Black patients (n=255)** % of time in therapeutic range through 4 weeks: Genotype group: 35.2% (SD 26.0) Clinically guided group: 43.5% (SD 26.5) Mean difference: -8.3 (95% CI -15.0 to -2.0), p=0.01 **Nonblack patients (n=700)** % of time in therapeutic range through 4 weeks: Genotype group: 48.8% (SD 25.9) Clinically guided group: 46.1% (SD 25.5) Mean difference: 2.8 (95% CI -1.0 to 6.6), p=0.15	Adding genotype data did not improve prediction of warfarin dosing for all patients. For Black patients, the algorithm led to worse dosing strategy and increased disparity compared with nonblack patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Landy et al. 2021[40]	Added a prediction model to USPSTF 2020	Access	Number of individuals eligible for screening	**USPSTF-2020 plus LYFS-CT (threshold >12 days life-gained from NSLT screening)** White (n): 14,865,167 (an additional 2,472,808) African American (n): 1,911,784 (an additional 750,507) Asian American (n): 334,691 (an additional 54,927) Hispanic American (n): 866,338 (an additional 191,289) **USPSTF-2020** White (n): 12,392,359 African American (n): 1,161,277 Asian American (n): 279,764 Hispanic American (n): 675,050	Adding LYFS-CT to USPSTF-2020 guidelines increased the number of eligible patients in each racial/ethnic group, offering screening to an additional 3.5 million individuals. This finding was similar when comparing the increase in eligible patients between USPSTF-2013 and USPSTF-2020 (data not reported here)
Landy et al. 2021[40]	Added a prediction model to USPSTF 2020	Health	Number needed to screen per lung cancer death prevented	**USPSTF 2020 plus LYFS-CT (threshold: ≥ 12 days life-gained from NLST-like screening)** White (NNS): 279 African American (NNS): 219 Asian American (NNS): 505 Hispanic American (NNS): 442 **USPSTF 2020** White (NNS): 282 African American (NNS): 202 Asian American (NNS): 550 Hispanic American (NNS): 501	Adding LYFS-CT to USPSTF-2020 guidelines maintained screening efficiency (NNS) in some groups and improved efficiency in other groups (Asian and Hispanic Americans).
Landy et al. 2021[40]	Added a prediction model to USPSTF 2020	Health	Disparity in lung cancer deaths (difference in sensitivity of Whites and each racial/ethnic group)	**USPSTF-2020 plus LYFS-CT (threshold >12 days life-gained from NSLT screening)** African American vs White: 0% Asian American vs White: 19% Hispanic American vs White: 23% **USPSTF-2020** African American vs White: 13% Asian American vs White: 19% Hispanic American vs White: 27%	Adding LYFS-CT to USPSTF-2020 guidelines nearly eliminated the disparity between African American and White patients for preventable lung cancer deaths, slightly reduced the disparity for Hispanic Americans, and was unchanged for Asian Americans.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Landy et al. 2021[40]	Added a prediction model to USPSTF 2020	Health	Number needed to screen per 10 life-years gained	**USPSTF 2020 plus LYFS-CT (threshold: ≥ 12 days life-gained from NLST-like screening)** White (NNS): 196 African American (NNS): 168 Asian American (NNS): 270 Hispanic American (NNS): 293 **USPSTF 2020** White (NNS): 195 African American (NNS): 159 Asian American (NNS): 288 Hispanic American (NNS): 322	Adding LYFS-CT to USPSTF-2020 guidelines maintained screening efficiency (NNS) in some groups, and improved efficiency in other groups (Asian and Hispanic Americans).
Landy et al. 2021[40]	Added a prediction model to USPSTF 2020	Health	Disparity in life-years gained (difference in sensitivity of Whites and each racial/ethnic group)	**USPSTF-2020 plus LYFS-CT (threshold: ≥ 12 days life-gained from NLST-like screening)** African American vs White: 1% Asian American vs White: 19% Hispanic American vs White: 24% **USPSTF-2020** African American vs White: 16% Asian American vs White: 19% Hispanic American vs White: 27%	Adding LYFS-CT to USPSTF-2020 guidelines nearly eliminated the disparity between African American and White patients for preventable lung cancer deaths, slightly reduced the disparity for Hispanic Americans, and was unchanged for Asian Americans.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Limdi et al. 2015[41]	Use of race-stratified analysis of predictive algorithms for warfarin dosing, rather than use of race-combined and adjusted algorithms	Quality	Dose variability explained by clinical vs genetic factors	*Model 1 included clinical and genetic factors used in the 2013 Clarification of Oral Anticoagulation through Genetics (COAG) study; model 2 included additional clinical and genetic factors reported to influence warfarin dose, such as chronic kidney disease, use of statins, and others.* **Black patients (n=595)** Total dosing variability explained, model 1: 29.3% Explained by clinical factors: 21.5% Explained by genetic factors: 7.0% Total variability explained, model 2: 33.9% Clinical factors: 22.8%, genetic factors: 10.0% **White patients (n=762)** Total variability explained, model 1: 51.4% Clinical factors: 14.7%, genetic factors: 34.1% Variability explained, model 2: 54.0% Clinical factors: 16.4%, genetic factors: 34.6% **Race-combined analysis (n=1357)** Total variability explained, model 1: 45.7% Clinical factors: 16.1%, genetic factors: 22.1% Variability explained, model 2: 48.3% Clinical factors: 17.4%, genetic factors: 23.5%	Stratifying this warfarin dosing algorithm by race resulted in dosing accuracy for both Black and White patients that was superior to a combined algorithm.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Lindley et al. 2022[42]	Addition of genotype data to standard warfarin dosing algorithms	Health	Prediction of therapeutic warfarin dose	**Black *CYP2C9*5* carriers (n=19), standard dosing algorithm** Mean actual dose (SD, mg/day): 4.47 (1.49), p=0.0095 (vs. Black noncarriers) Mean predicted dose (SD, mg/day): 6.39 (1.31) Mean prediction error (SD, mg/day): +1.92 (1.67), p<0.001 (vs. Black noncarriers) Algorithm overpredicted the warfarin dose by 30%, 95% CI 19% to 39%, p<0.001. **Black noncarriers (n=791), standard dosing algorithm** Mean actual dose (SD, mg/day): 6.34 (3.13) Mean predicted dose (SD, mg/day): 6.05 (1.56) Mean prediction error (SD, mg/day): -0.30 (2.73) Underpredicted the daily dosing requirement by 0.30 mg/day (p<0.001). **CYP2C9*5 carriers (n=19 Black, n=1 White), standard algorithm adjusted for CYP2C9*5 Variant** *CYP2C9*5* heterozygotes adjustment: 0.70, 95% CI 0.60 to 0.81 *CYP2C9*5* homozygotes adjustment: 0.49, 95% CI 0.36 to 0.65 Adjustment reduced the mean prediction error by 51.9% to 0.97 mg/day (median absolute error 0.93 mg/day) in heterozygotes	The standard dosing algorithm overpredicted the warfarin dose by 30% (average of 1.92 mg/day) in Black CYP2C9*5 carriers. Black noncarriers, by contrast, had a slightly greater than predicted daily dosing requirement. Adjusting the standard algorithm by incorporating the CYP2C9*5 correction factor decreased the prediction error by 51.9% to 0.97 mg/day in heterozygous Black carriers.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Mahmud et al. 2021[43]	Removed race from eGFR	Quality	Association between lower baseline eGFR and rate of AKI events	**Black patients (unadjusted analysis; incidence rate ratios [IRRs] per 15 mL/min/1.73 m^2 decrease in eGFR)** MDRD-4: IRR 1.01; 95% CI 1.00 to 1.02 MDRD-4 w/out race: IRR 1.01; 95% CI 1.00 to 1.02 MDRD-6: IRR 1.06; 95% CI 1.05 to 1.07 MDRD-6 w/out race: IRR 1.07; 95% CI 1.06 to 1.08 CKD-EPI: IRR 1.09; 95% CI 1.08 to 1.10 CKD-EPI w/out race: IRR 1.11; 95% CI 1.10 to 1.12 **Black patients (adjusted analysis)** MDRD-4: IRR 0.98; 95% CI 0.98 to 0.99 MDRD-4 w/out race: IRR 0.98; 95% CI 0.98 to 0.99 MDRD-6: IRR 1.00; 95% CI 1.00 to 1.01 MDRD-6 w/out race: IRR 1.01; 95% CI 1.00 to 1.01 CKD-EPI: IRR 1.02; 95% CI 1.01 to 1.03 CKD-EPI w/out race: IRR 1.02; 95% CI 1.01 to 1.04	Unadjusted analyses found that removing race from equations strengthened the association between lower eGFR and rate of AKI. For adjusted analyses, findings suggest there was minimal difference in the association between lower eGFR and rate of AKI with or without the race coefficient. However, in all analyses, CKD-EPI had the strongest inverse associations between baseline eGFR and incident AKI events.
Mahmud et al. 2021[43]	Removed race from eGFR	Quality	Association between lower baseline eGFR and rate of stage 2 or 3 AKI events (sensitivity analysis)	**Black patients (Unadjusted analysis)** MDRD-4: IRR 1.02; 95% CI 1.01 to 1.02 MDRD-4 w/out race: IRR 1.02; 95% CI 1.01 to 1.03 MDRD-6: IRR 1.05; 95% CI 1.04 to 1.06 MDRD-6 w/out race: IRR 1.06; 95% CI 1.05 to 1.07 CKD-EPI: IRR 1.08; 95% CI 1.06 to 1.09 CKD-EPI w/out race: IRR 1.09; 95% CI 1.07 to 1.11 **Black patients (Adjusted analysis)** MDRD-4: IRR 0.99; 95% CI 0.98 to 1.00 MDRD-4 w/out race: IRR 0.99; 95% CI 0.98 to 1.00 MDRD-6: IRR 1.01; 95% CI 1.00 to 1.02 MDRD-6 w/out race: IRR 1.01; 95% CI 1.00 to 1.03 CKD-EPI: IRR 1.02; 95% CI 1.01 to 1.04 CKD-EPI w/out race: IRR 1.03; 95% CI 1.01 to 1.05	In a sensitivity analysis limited to stage 2 or 3 AKI events, findings were similar to the main study findings, suggesting a stronger association between lower eGFR and rate of AKI in unadjusted analyses and minimal difference in adjusted analyses with or without the race coefficient. In all analyses, CKD-EPI had the strongest inverse associations between baseline eGFR and incident AKI events.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Mahmud et al. 2021[43]	Removed race from eGFR	Quality	Association between lower baseline eGFR and rate of AKI events in patients with Child-Turcotte-Pugh (CTP) B/C cirrhosis (sensitivity analysis)	**Black patients (unadjusted analysis)** MDRD-4: IRR 1.00; 95% CI 0.99 to 1.00 MDRD-4 w/out race: IRR 1.00; 95% CI 0.99 to 1.01 MDRD-6: IRR 1.03; 95% CI 1.02 to 1.04 MDRD-6 w/out race: IRR 1.03; 95% CI 1.02 to 1.05 CKD-EPI: IRR 1.10; 95% CI 1.09 to 1.11 CKD-EPI w/out race: IRR 1.12; 95% CI 1.10 to 1.13 **Black patients (adjusted analysis)** MDRD-4: IRR 0.98; 95% CI 0.98 to 0.99 MDRD-4 w/out race: IRR 0.98; 95% CI 0.97 to 0.99 MDRD-6: IRR 1.00; 95% CI 0.99 to 1.01 MDRD-6 w/out race: IRR 1.00; 95% CI 0.99 to 1.01 CKD-EPI: IRR 1.08; 95% CI 1.06 to 1.09 CKD-EPI w/out race: IRR 1.09; 95% CI 1.07 to 1.11	In a sensitivity analysis limited to patients with CTP B/C cirrhosis, authors reported a stronger association between CKD-EPI eGFR and incident AKI in adjusted analyses and minimal difference in AKI rate ratios when race was removed from the equation. In all analyses, CKD-EPI had the strongest inverse associations between baseline eGFR and incident AKI events.
Meeusen et al. 2022[44]	Removed race from eGFR	Quality	Diagnosis of CKD (GFR<60) (CKD-EPI 2021 vs CKD-EPI 2009)	**Reclassification to GFR <60** *Black patients (n=852)* Correctly reclassified from GFR >60 to GFR<60: 4.7%; 95% CI 2.3 to 7.0 Incorrectly reclassified from GFR >60 to GFR<60: 5.9%; 95% CI 3.8 to 8.0 Net reclassification: -1.2%; 95% CI -5.7 to 3.2 *Non-Black patients (N=33,037)* Incorrectly reclassified from GFR <60 to GFR >60: 4.1%; 95% CI 3.8 to 4.4 Correctly reclassified from GFR <60 to GFR >60: 4.4%; 95% CI 4.1 to 4.7 Net reclassification: 0.33%; 95% CI -0.3 to 0.9	Removal of the race coefficient from eGFR resulted in 5% of Black patients correctly receiving new CKD diagnoses, and 6% incorrectly diagnosed with CKD. In non-Black patients, 4% incorrectly received a reversal of CKD diagnosis, while 4% correctly received a reversal of diagnosis.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Meeusen et al. 2022[44]	Removed race from eGFR	Quality	Diagnosis of kidney failure (GFR <20) (CKD-EPI 2021 vs CKD-EPI 2009)	**Reclassification to GFR <20** *Black patients (n=852)* Correctly reclassified from GFR >20 to GFR <20: 6.7%; 95% CI 1.02 to 12.3 Incorrectly reclassified from GFR >20 to GFR <20: 0.26%; 95% CI -0.10 to 0.61 Net reclassification: 6.4%; 95% CI 0.36 to 12.4 *Non-Black patients (n=33,037)* Incorrectly reclassified from GFR <20 to GFR >20: 5.6%; 95% CI 4.5 to 6.6 Correctly reclassified from GFR <20 to GFR >20: 0.49%; 95% CI 0.41 to 0.57 Net reclassification: -5.1%; 95% CI -4.1 to -6.0	Removal of the race coefficient from eGFR resulted in 7% of Black patients correctly receiving new kidney failure diagnoses, and <1% incorrectly diagnosed with kidney failure. In non-Black patients, 6% incorrectly received a reversal of kidney failure diagnosis, while <1% correctly received a reversal of diagnosis.
Miller et al 2022[45]	Removed Black race from original KDRI	Access	Kidney non-use probability from deceased donors	**KDPI with race** Kidney transplants: 112,881 Non-used kidneys: 29,224 Black recipients: 37,174 Black recipients and Black donor: 7,610 **KDPI modeled without race** Overall non-use would increase by 32 total kidneys Black donor kidneys used would increase by 353 non-Black donor kidneys used would decrease by 385	Removal of the race coefficient does not significantly change the overall effect of KDRI and KDPI on kidney non-use. However, removing race could result in more parity between Black and non-Black donors.
Miller et al. 2021[46]	Algorithm vs same algorithm w/out race	Quality	Rate of discordance among all antibiotic orders	**Black patients (n=297)** Deindexed eGFR w/race vs w/out race: 17.9% (n=53); Cohens k 0.670; 95% CI 0.590 to 0.750, moderate level of agreement between equations Deindexed eGFR w/race vs CG CrCl: 18.5% (n=55); Cohens k 0.651; 95% CI 0.572 to 0.730), moderate level of agreement between equations Deindexed eGFR w/out race vs CG CrCl: 12.5% (n=37); Cohens k 0.779; 95% CI 0.713 to 0.844, moderate level of agreement between equations **White patients (n=57)** Deindexed eGFR vs CG CrCl: 33.9% (n=19); Cohens k 0.419; 95% CI 0.248 to 0.590, weak level of agreement between equations	In Black patients, moderate levels of agreement in dosage recommendations were reported when using equations with and without the race coefficient, although less discordance was reported between Deindexed eGFR w/out race and CG CrCl. When comparing Deindexed eGFR with CG CrCl in White patients, a weak level of agreement was reported.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Muiru et al. 2023[47]	Removed race from eGFR	Quality	Progression of CKD in patients with HIV	**Progression from Stage 1 to Stage 2** Hazard ratio, Black vs White patients, eGFR without race: 1.37; 95% CI 1.30 to 1.45 Hazard ratio, eGFR with race: 0.77; 95% CI 0.73 to 0.82 **Progression from Stage 2 to Stage 3** Hazard ratio, Black vs White patients, eGFR without race: 1.07; 95% CI 0.99 to 1.16 Hazard ratio, eGFR with race: 1.00; 95% CI 0.92 to 1.07 **Progression from Stage 3 to Stage 4** Hazard ratio, Black vs White patients, eGFR without race: 1.71; 95% CI 1.45 to 2.02 Hazard ratio, eGFR with race: 3.06; 95% CI 2.60 to 3.62	Removing race from eGFR resulted in better prediction of the progression of CKD for Black patients with HIV.
Ober-meyer et al. 2019[7]	Replaced biased variable	Access	Eligibility for a care management program	In a simulation that removes algorithmic bias (by replacing white patients above the cutoff with Black patients who are sicker but below the cutoff until the mean number of chronic conditions among patients at the cutoff score is equal for Blacks and Whites), the fraction of patients above the cutoff who are Black rises from 17.7% to 46.5%.	If the program were unbiased (i.e., if Black and White patients with scores at the cutoff value had equal actual need), it would have a markedly higher proportion of patients who were Black.
Ober-meyer et al. 2019[7]	Replaced biased variable	Access	Eligibility for a care management program	Fraction of patients at or above the 97th percentile who are Black when the newly developed algorithms are used: Algorithm predicting total costs: 14.1% Algorithm predicting avoidable costs: 21.0% Algorithm predicting health (chronic conditions): 26.7%	A new algorithm that, like the original algorithm, predicts total costs results in a similar racial distribution in the program membership. Algorithms that predict other outcomes, such as chronic conditions, would result in the program including a higher percentage of Black individuals.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Ober-meyer et al. 2019[7]	Replaced biased variable	Access	Eligibility for a care management program	Fraction of patients at or above the 97th percentile who are Black when the newly developed algorithms are used and race is included in them as an additional predictor: Algorithm predicting total costs: 11.2% Algorithm predicting avoidable costs: 24.1% Algorithm predicting health (chronic conditions): 28.5%	Adding race as a predictor in the new algorithms does not have a sizable effect on the resulting racial distribution of the program membership.
Panchal et al. 2022[48]	Removed race from eGFR	Access	Waitlist CKD classification with and without eGFR race adjustment	**Black patients** MDRD-4 with race: 20.3% (1610/7937) MDRD-4 without race: 27.6% (2193/7937) MDRD-4 with race vs MDRD-4 without race: 36% (583/1610) increase in waitlist CKD classification, p<0.05 CKD-EPI with race: 21.5% (1707/7937) CKD-EPI without race: 27.1% (2154/7937) CKD-EPI with race vs CKD-EPI without race: 26.1% (447/1707) increase in waitlist CKD classification, p<0.05	The proportion of patients with waitlist CKD significantly increased when race was removed from each eGFR equation.
Panchal et al. 2022[48]	Removed race from eGFR	Quality	CKD reclassifications without eGFR race adjustment	**Black patients** MDRD-4 without race: 7.4% (583/7937) CKD-EPI without race: 5.6% (447/7937) MDRD-4 without race vs CKD-EPI without race: p<0.05, favors MDRD-4 without race	The overall proportion of patients reclassified as having CKD was significantly higher with MDRD-4 without race compared with CKD-EPI without race. Authors also observed increasing proportions of patients reclassified with equations without race as having CKD in sensitivity analyses limiting the requirement to establish CKD from 90 days to 60 or 30 days.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Panchal et al. 2022[48]	Removed race from eGFR	Access	Waitlist simultaneous liver-kidney transplantation (SLKT) criteria (CKD pathway) Classification with and without eGFR race adjustment	**Black patients** MDRD-4 with race: 15.3% (1217/7937) MDRD-4 without race: 19.0% (1506/7937) MDRD-4 with race vs MDRD-4 without race: 23.7% (289/1217) increase in SLKT candidacy, p<0.05 CKD-EPI with race: 16.0% (1270/7937) CKD-EPI without race: 19.0% (1509/7937) CKD-EPI with race vs CKD-EPI without race: 18.7% (239/1270) increase in SLKT candidacy, p<0.05	The proportion of patients meeting SLKT criteria via the CKD pathway significantly increased when race was removed from each eGFR equation.
Panchal et al. 2022[48]	Removed race from eGFR	Access	Waitlist SLKT criteria (AKI pathway) Classification with and without eGFR race adjustment	**Black patients** MDRD-4 with race: 1.9% (150/7937) MDRD-4 without race: 2.1% (170/7937) MDRD-4 with race vs MDRD-4 without race: p=0.26 CKD-EPI with race: 2.0% (155/7937) CKD-EPI without race: 2.2% (174/7937) CKD-EPI with race vs CKD-EPI without race: p=0.29	Authors reported an increase in SLKT candidacy via the AKI pathway when race was removed from each equation. The increase was not statistically significant.
Panchal et al. 2022[48]	Removed race from eGFR	Access	SLKT reclassifications (CKD pathway) without eGFR race adjustment	**Black patients** MDRD-4 without race: 3.6% (289/7937) CKD-EPI without race: 3.0% (239/7937) MDRD-4 without race vs CKD-EPI without race: p<0.05, favors MDRD-4 without race	The overall proportion of patients reclassified as qualifying for SLKT via the CKD pathway was significantly higher with MDRD-4 without race than with CKD-EPI without race. Authors also observed increasing proportions of patients reclassified with equations without race as qualifying for SLKT in sensitivity analyses limiting the requirement to establish CKD from 90 days to 60 or 30 days.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Panchal et al. 2022[48]	Removed race from eGFR	Health	Waitlist mortality in patients meeting criteria for SLKT listing	**Black patients** MDRD-4 with race: 37.1% (451/1216) MDRD-4 without race: 37.7% (568/1506) MDRD-4 with race vs MDRD-4 without race: p=0.74, no statistically significant difference CKD-EPI with race: 37.4% (475/1269) CKD-EPI without race: 37.9% (572/1509) CKD-EPI with race vs CKD-EPI without race: no statistically significant difference, data not reported	Waitlist mortality in patients meeting criteria for SLKT listing was similar when eGFR was calculated with or without race for MDRD-4 and CKD-EPI.
Panchal et al. 2022[48]	Removed race from eGFR	Health	Liver Transplant Alone (LTA) in patients meeting criteria for SLKT listing	**Black patients** MDRD-4 with race: 8.6% (104/1216) MDRD-4 without race: 14.8% (223/1506) MDRD-4 with race vs MDRD-4 without race: p<0.05, more likely to receive LTA with MDRD-4 without race CKD-EPI with race: 9.9% (125/1269) CKD-EPI without race: 14.6% (220/1509) CKD-EPI with race vs CKD-EPI without race: p<0.05, more likely to receive LTA with CKD-EPI without race	Patients meeting criteria for SLKT listing were significantly more likely to receive LTA when using MDRD-4 and CKD-EPI without race to calculate eGFR.
Panchal et al. 2022[48]	Removed race from eGFR	Health	SLKT in patients meeting criteria for SLKT listing	**Black patients** MDRD-4 with race: 34.7% (422/1216) MDRD-4 without race: 29.2% (439/1506) MDRD-4 with race vs MDRD-4 without race: p<0.05, less likely to undergo SLKT with MDRD-4 without race CKD-EPI with race: 37.4% (475/1269) CKD-EPI without race: 29.0% (438/1509) CKD-EPI with race vs CKD-EPI without race: p<0.05, less likely to undergo SLKT with CKD-EPI without race	Patients meeting criteria for SLKT listing were significantly less likely to undergo SLKT when using MDRD-4 and CKD-EPI without race to calculate eGFR.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Panchal et al. 2022[48]	Removed race from eGFR	Health	Received a kidney transplantation within 1 year of LTA	**Black patients** MDRD-4 without race: 1.7% (2/118) of patients reclassified as meeting SLKT criteria (CKD pathway) received kidney transplantation (both pre-OPTN policy change) MDRD-4 without race: 0.3% (8/3076) of patients not reclassified as meeting SLKT criteria (CKD pathway) received kidney transplantation (one pre-OPTN policy change) CKD-EPI without race: 2.1% (2/94) of patients reclassified as meeting SLKT criteria (CKD pathway) received kidney transplantation (both pre-OPTN policy change) CKD-EPI without race: 0.3% (8/3079) of patients not reclassified as meeting SLKT criteria (CKD pathway) received kidney transplantation (one pre-OPTN policy change)	A larger proportion of patients reclassified as meeting SLKT criteria via the CKD pathway required kidney transplantation within 1 year of LTA than patients not reclassified.
Park et al. 2021[49]	1) Recalibrated through reweighing key groups during model training 2) Removed race 3) Added an adjustment term to limit the effect of race-based variables	Quality	Risk of postpartum depression	**Black patients (n=217,899), White patients (n=314,903)** *Measures used: Disparate impact (DI) is a ratio of means of predicted favorable outcome between unprivileged and privileged groups. Values closer to 1 generally indicate more fairness between groups.* *Equal opportunity difference (EOD) is a sensitivity measure for DI that compares true positive rates. Values closer to 0 generally indicate greater fairness.* Baseline risk of postpartum depression: DI: 0.31; EOD: -0.19 After reweighing: DI: 0.79; EOD: 0.02 After removing race: DI: 0.61; EOD: 0.05 After adding adjustment term: Improvement in model fairness was roughly comparable to that achieved by removing race (data shown in graphic form or not reported in detail)	Removing race improved algorithm fairness regarding prediction of postpartum depression. Reweighing the model with more diverse patient data resulted in greater improvement than removing race. Adding a statistical adjustment improved fairness to a degree roughly comparable to that achieved by removing race.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Park et al. 2021[49]	1) Recalibrated through reweighing key groups during model training 2) Removed race 3) Added an adjustment term to limit the effect of race-based variables	Quality	Likelihood to use mental health services	**Black patients (n=217,899), White patients (n=314,903)** *Measures used: see above* Baseline risk of using mental health services: DI: 0.45; EOD: -0.11 After reweighing: DI: 0.85; EOD: -0.02 After removing race: DI: 0.63; EOD: -0.04 After adding adjustment term: Improvement in model fairness was roughly comparable to that achieved by removing race (data shown in graphic form or not reported in detail)	For predicting likelihood of using mental health services, removing race improved algorithm fairness. Reweighing the model with more diverse patient data resulted in greater improvement than removing race. Adding a statistical adjustment improved fairness to a degree roughly comparable to that achieved by removing race.
Schmeusser et al. 2022[50]	Removed race from eGFR	Access	Eligibility for enrollment in cancer clinical trials (n=459)	**Black patients (n=459)** *Not eligible for enrollment with eGFR <60* CKD-EPI with race: 26.6% CKD-EPI without race: 39.8%, p<0.0001 MDRD with race: 31.6% MDRD without race: 54.1%, p<0.0001 *Not eligible for enrollment with eGFR <45* CKD-EPI with race: 11.6% CKD-EPI without race: 18.0, p=0.0052 MDRD with race: 12.0% MDRD without race: 24.0%, p<0.0001 *Not eligible for enrollment with eGFR <30* CKD-EPI with race: 3.5% CKD-EPI without race: 5.6%, p=NS MDRD with race: 3.7% MDRD without race: 7.0%, p=0.0265	Estimating GFR without a race coefficient may lead to an increase in Black patients becoming ineligible to participate in clinical trials.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Shi et al. 2021[51]	Removed race from eGFR	Quality	Estimated renal function	**Reclassification of patients from MDRD with race to CKD-EPI without race** *All patients (n=241,760)* Baseline stage 1: 5.43% moved to stage 2 Baseline stage 2: 29.60% moved to stage 1, 1.35% moved to stage 2 Baseline stage 3a: 23.49% moved to stage 2, 4.70% moved to stage 3b Baseline stage 3b: 11.11% moved to stage 3a, 4.24% moved to stage 4 Baseline stage 4: 4.68% moved to stage 3b, 4.34% moved to stage 5 Baseline stage 5: 1.34% moved to stage 4 *Black patients (n=21,751)* Baseline stage 1: 17.92% moved to stage 2 Baseline stage 2: 16.28% moved to stage 3a Baseline stage 3a: 38.08% moved to stage 3b Baseline stage 3b: 23.90% moved to stage 4 Baseline stage 4: 18.18% moved to stage 5 No Black patients moved to a lower stage	Removing race from eGFR resulted in frequent reclassification of Black patients to higher levels of disease severity. For all patients combined, removing race resulted in more reclassification to lower rather than higher levels.
Shores et al. 2013[52]	Modified the Donor Risk Index with data drawn specifically from Black patients. The original Index included Black race as 1 of 7 components. The revised Index includes non-Black race as 1 of 3 components.	Quality	NRI of 1-year risk of graft loss	**Net Reclassification Improvement** 27% of patients in validation set, p =0.04 **C-index** Original algorithm: 0.51 Updated algorithm: 0.55	Revising the liver transplant Donor Risk Index with data from Black patients with hepatitis C led to improved prediction of graft failure in these patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Thompson et al. 2021[13]	Race-specific thresholds	Access	Referral for education, treatment options, and care pathways	**Model performance with cut point in the Black subgroup** reduced from 0.3 to 0.2. *False-negative rates*: Black patients: 0.25; 95% CI: 0.20–0.30 (reduced from 0.32; 95% CI: 0.27–0.37) White patients: presented graphically *False-detection rates*: Black patients: 0.46; 95% CI: 0.41–0.50 (increased from 0.41; 95% CI: 0.37–0.47) White patients: 0.36; 95% CI: 0.31–0.43	False-negative rate was reduced among Black patients by 0.07, rendering the difference in underdiagnosis between Black and White patients not significant. The false-detection rate (i.e., the proportion of patients classified as needing resources when they in fact did not) among Black patients was increased by 0.05 but was not significantly different than that among White patients.
Thompson et al. 2021[13]	Recalibration by subgroup	Access	Referral for education, treatment options, and care pathways	**Model performance after recalibration by subgroup** *False-negative rates*: Black patients: 0.24; 95% CI: 0.19–0.29 White patients: 0.21; 95% CI: 0.15–0.27 *False-detection rates*: Black patients: 0.46; 95% CI: 0.41–0.50 White patients: 0.36; 95% CI: 0.31–0.43	The results of recalibration by subgroup were virtually identical to the results above.
Topel et al. 2018[53]	Added race and presence of diabetes to Framingham Risk Score to calculate ASCVD	Health	Central augmentation index	**Low-risk patients** FRS, Black patients: 24.0% (SD 10.0) vs White patients: 22.0 (10.2), p=0.049 ASCVD, Black patients: 23.9% (10.1) vs White patients: 22.0 (10.0), p=0.13 **High-risk patients** FRS, Black patients: 24.1% (SD 10.8) vs White patients: 22.6 (8.0), p=0.22 ASCVD, Black patients: 24.4% (10.7) vs White patients: 22.9 (7.7), p=0.43	When patients were categorized as low risk by FRS, a significant disparity between Black and White patients remained, while no difference between groups was found when risk was predicted by ASCVD. For high-risk patients, no differences were found with either algorithm. This can indicate that adding race and diabetes resulted in a better estimate of 1 measure of subclinical vascular disease in healthier Black patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Topel et al. 2018[53]	Added race and presence of diabetes to Framingham Risk Score to calculate ASCVD	Health	Central pulse pressure	**Low-risk patients** FRS, Black patients: 35.3 mmHg (SD 9.4) vs White patients: 33.6 (8.9), p=0.016 ASCVD, Black patients: 35.2 (9.3) vs White patients: 34.0 (9.20), p=0.08 **High-risk patients** FRS, Black patients: 40.2 (SD 12.6) vs White patients: 38.3 (11.0), p=0.20 ASCVD, Black patients: 40.5 (12.8) vs White patients: 38.8 (11.1), p=0.31	When patients were categorized as low risk by FRS, a significant disparity between Black and White patients remained, while no significant difference between groups was found when risk was predicted by ASCVD. For high-risk patients, no differences were found with either algorithm. This can indicate that adding race and diabetes resulted in a better estimate of one measure of subclinical vascular disease in healthier Black patients.
Topel et al. 2018[53]	Added race and presence of diabetes to Framingham Risk Score to calculate ASCVD	Health	Pulse wave velocity	**Low-risk patients** FRS, Black patients: 7.4 m/s (SD 1.40) vs White patients: 7.1 (SD 1.0), p<0.001 ASCVD, Black patients: 7.4 (1.4) vs White patients: 7.2 (1.2), p=0.003 **High-risk patients** FRS, Black patients: 8.1 (SD 1.6) vs White patients: 8.0 (1.5), p=0.72 ASCVD, Black patients: 8.0 (1.7) vs White patients: 8.1 (1.6), p=0.65	When patients were categorized as low risk by FRS or ASCVD, a significant disparity between Black and White patients remained. For high-risk patients, no differences were found with either algorithm. This indicates that adding race and diabetes made no difference in estimating this measure of subclinical vascular disease for Black or White patients.
Topel et al. 2018[53]	Added race and presence of diabetes to Framingham Risk Score to calculate ASCVD	Health	Carotid intima-median thickness	**Low-risk patients** FRS, Black patients: 0.66 mm (SD 0.09) vs White patients: 0.63 (SD 0.09), p=0.008 ASCVD, Black patients: 0.65 (0.10) vs White patients: 0.64 (0.09), p=0.10 **High-risk patients** FRS, Black patients: 0.73 (SD 0.13) vs White patients: 0.74 (0.10), p=0.88 ASCVD, Black patients: 0.74 (0.12) vs White patients: 0.76 (0.08), p=0.46	When patients were categorized as low risk by FRS, a significant disparity between Black and White patients remained, while no difference between groups was found when risk was predicted by ASCVD. For high-risk patients, no differences were found with either algorithm. This can indicate that adding race and diabetes resulted in a better estimate of one measure of subclinical vascular disease in healthier Black patients.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Tsai et al. 2021[54]	Removed race from eGFR	Quality	Estimated renal function	**Black patients (n=2401)** eGFR >60, MDRD with race: 93.2% (95 CI: 92.2 to 94.4%) eGFR >60, MDRD without race: 83.7% (82.0 to 85.0%) Population estimate after removal of race: 3.3 million additional diagnoses of moderate kidney disease in Black patients	Removing race from eGFR resulted in large increases in Black patients potentially diagnosed with CKD, based on data from a nationwide, longitudinal U.S. database.
Tsai et al. 2021[54]	Removed race from eGFR	Access	Eligibility for referral to nephrologist	**Black patients (n=2401)** Population estimate after removal of race: 300,000 additional Black patients eligible for referral to nephrologist	Removing race from eGFR resulted in large increases in Black patients becoming eligible for treatment by a nephrologist.
Tsai et al. 2021[54]	Removed race from eGFR	Access	Eligibility for transplant evaluation	**Black patients (n=2401)** Population estimate after removal of race: 31,000 additional Black patients eligible for transplant evaluation	Removing race from eGFR resulted in large increases in Black patients becoming eligible for transplant evaluation.
Weale et al. 2021[55]	Added polygenic risk scores to ASCVD	Quality	Estimated 10-year risk for atherosclerotic cardiovascular disease	**Net Reclassification Index** Black/African American/Black Caribbean/Black African: 2.46% (95% CI 0.57 to 4.34, p=0.01) White: 2.65% (1.12 to 4.18, p<0.001)	Adding genetic risk factors to ASCVD equations led to more accurate prediction of risk for both Black and White patients. Improvement in both groups was similar.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Yadlow-sky et al. 2018[56]	Updated PCE model: Added data from Jackson Heart Study and MESA to better reflect racial and ethnic populations. Adjusted statistical methods to reduce model overfitting by using elastic net regularization, and removing race-based subgroups.	Quality	Extreme variation in 10-year risk of atherosclerotic cardiovascular disease	**Underestimation of risk for Black vs White patients** Risk ratios <0.7 in 2013 algorithm vs 2018 update: *Example 1* 2013 White risk 10.9, Black risk 6.7, RR: 0.61 2018 White risk 6.4, Black risk 4.3, RR: 0.67 *Example 2* 2013 White risk 14.6, Black risk 9.7, RR: 0.66 2018 White risk 7.6, Black risk 5.7, RR: 0.75 *Example 3* 2013 White risk 1.9, Black risk 0.5, RR: 0.26 2018 White risk 0.8, Black risk 1.3, RR: 1.63 *Example 4* 2013 White risk 24.4, Black risk 13.2, RR: 0.54 2018 White risk 8.4, Black risk 7.9, RR: 0.94 *Example 5* 2013 White risk 20.4, Black risk 13.5, RR: 0.66 2018 White risk 12.7, Black risk 9.5, RR: 0.75	When more diverse patient data and newer statistical methods were used, extreme disparities between Black and White patients in predicted risk were significantly reduced. In numerous examples, a Black person's risk of cardiovascular disease was significantly underestimated compared with a White patient in the initial algorithm, but the new algorithm resulted in much less variation by race.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Yadlow-sky et al. 2018[56]	Updated PCE model: Added data from Jackson Heart Study and MESA to better reflect racial and ethnic populations. Adjusted statistical methods to reduce model overfitting by using elastic net regularization, and removing race-based subgroups.	Quality	Extreme variation in 10-year risk of atherosclerotic cardiovascular disease	**Overestimation of risk for Black vs White patients** Risk ratios >2.5 in 2013 algorithm vs 2018 update: *Example 1* 2013 White risk 2.4, Black risk 9.3, RR: 3.88 2018 White risk 4, Black risk 8.4, RR: 2.10 *Example 2* 2013 White risk 6.3, Black risk 26.7, RR: 4.24 2018 White risk 5.7, Black risk 11.8, RR: 20.7 *Example 3* 2013 White risk 1.9, Black risk 9.9, RR: 5.21 2018 White risk 3.8, Black risk 9.3, RR: 2.45 *Example 4* 2013 White risk 3.5, Black risk 8.9, RR: 2.54 2018 White risk 4.2, Black risk 6.7, RR: 1.60 *Example 5* 2013 White risk 1.8, Black risk 8.9, RR: 4.94 2018 White risk 1.8, Black risk 3.6, RR: 2.00	When more diverse patient data and newer statistical methods were used, extreme disparities between Black and White patients in predicted risk were significantly reduced. In numerous examples, a Black person's risk of cardiovascular disease was significantly overestimated compared with a White patient in the initial algorithm, but the new algorithm resulted in much less variation by race.
Yap et al. 2021[57]	Removed race from eGFR	Quality	Estimated renal function	**Reclassification of Black patients (n=327)** *MDRD with race to MDRD without race* Baseline stage 1/2: 39.9% moved to stage 3a Baseline stage 3a: 71.8% moved to stage 3b Baseline stage 3b: 54.1% moved to stage 4 Baseline stage 4: 36.4% moved to stage 5 *CDK-EPI with race to CKD-EPI without race* Baseline stage 1/2: 22.6% moved to stage 3a Baseline stage 3a: 46.5% moved to stage 3b Baseline stage 3b: 38.3% moved to stage 4 Baseline stage 4: 17.4% moved to stage 5	Removing race from eGFR resulted in large increases in Black patients potentially diagnosed with CKD and becoming eligible for treatment and transplant evaluation, based on a sample of patients at an urban academic medical center.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Zelnick et al. 2021[58]	Removed race from eGFR	Quality	Difference in eGFR measurements for patients with iGFR 15 to ≤45 mL/min/1.73 m^2 at baseline	**Black Patients** *Creatinine-based CKD-EPI with race vs iGFR* MD: 3.1 mL/min/1.73m^2; 95% CI 2.2 to 3.9; p<0.001; CKD-EPI with race overestimated iGFR *Creatinine-based CKD-EPI without race vs iGFR* MD: -1.7 mL/min/1.73 m^2; 95% CI -2.5 to -0.9; p<0.001; CKD-EPI without race did not overestimate iGFR *Cystatin C-based CKD-EPI vs iGFR* MD: 5.6 mL/min/1.73 m^2; 95% CI 4.6 to 6.6, p<0.001; Cystatin C-based CKD-EPI overestimated iGFR (race is not included in cystatin c-based equation) *Creatinine-based CKD-EPI with race vs Creatinine-based CKD-EPI without race* MD: 4.8 mL/min/1.73 m^2; 95% CI 4.6 to 4.9; p<0.001; CKD-EPI with race was higher than CKD-EPI without race *Creatinine-based CKD-EPI with race vs Cystatin C-based CKD-EPI* MD: -2.5 mL/min/1.73 m^2; 95% CI -3.5 to -1.5; p<0.001; CKD-EPI with race lower than Cystatin C-based CKD-EPI (race is not included in cystatin c-based equation)	Creatinine-based CKD-EPI equation with the race coefficient was associated with a higher eGFR than when the race coefficient was not used.

Author/ Year	Mitigation Strategy	Outcome Category	Reported Outcome	Results	Summary
Zelnick et al. 2021[58]	Removed race from eGFR	Quality	Time to an eGFR < 20 mL/min/1.73 m^2 for patients with an eGFR of at least 20 mL/min/1.73 m^2 at baseline (n=1616) Median follow-up 4.4 years (IQR 1.2 to 10.3)	**Black Patients** *Creatinine-based CKD-EPI with race* Achieved eGFR <20: 28.6% (462/1616) Median time to achievement: 13.9 years; 95% CI 13.0 to 13.9 *Creatinine-based CKD-EPI without race* Achieved eGFR <20: 36.4% (589/1616) Median time to achievement: 12.0 years; 95% CI 10.9 to 13.0 *Cystatin C-based CKD-EPI (race not included in equation)* Achieved eGFR <20: 27.7% (448/1616) *Creatinine-based CKD-EPI without race vs Creatinine-based CKD-EPI with race* HR 1.35; 95% CI 1.29 to 1.41; p<0.001 **Cystatin C-based CKD-EPI vs Creatinine-based CKD-EPI with race** HR 1.01; 95% CI 0.94 to 1.08; p=0.85	Creatinine-based CKD-EPI equation without race was associated with a 35% higher risk of achieving eGFR <20 mL/min/1.73 m^2 compared with equation with race. Additionally, the time to achievement of an eGFR less than 20 mL/min/1.73 m^2 did not differ between the Cystatin C-based equation (race not included in equation) and creatinine-based equation with race.
Zelnick et al. 2021[58]	Removed race from eGFR	Quality	Time to an eGFR < 30 mL/min/1.73 m^2 for patients with an eGFR of at least 20 mL/min/1.73 m^2 at baseline (n=1338) Median follow-up time: 2.2 years (IQR 0.0 to 7.9 years)	**Black patients** *Creatinine-based CKD-EPI with race* Achieved eGFR <30: 43.3% (579/1338) *Creatinine-based CKD-EPI without race* Achieved eGFR <30: 55.8% (746/1338) *Cystatin C-based CD-EPI (race not included in equation)* Achieved eGFR <30: 46.1% (617/1338) *Creatinine-based CKD-EPI without race vs Creatinine-based CKD-EPI with race* HR 1.52; 95% CI 1.45 to 1.59; p<0.001 *Cystatin C-based CKD-EPI (race not included in equation) vs Creatinine-based CKD-EPI with race* HR 1.11; 95% CI 1.05 to 1.18; p<0.001	Creatinine-based equation without race was associated with a 52% higher risk of achieving eGFR <30 mL/min/1.73 m^2 compared with equation with race and the difference in median time to event was 3.6 years. Furthermore, the Cystatin C-based CKD-EPI equation (race not included in equation) was associated with an 11% higher risk of achieving an eGFR <30 mL/min/1.73 m^2 compared with creatinine-based CKD-EPI equation with race.

Abbreviations: ACC/AHA = American College of Cardiology/American Heart Association; AKI = acute kidney injury; ASCVD = atherosclerotic cardiovascular disease; AUC = area under the curve; CG CrCl = Cockcroft-Gault Creatinine Clearance; CI = confidence interval; CKD = chronic kidney disease; CKD-EPI =

Chronic Kidney Disease Epidemiology Collaboration; CSC = crisis standards of care; CTP-B/C = Child-Turcotte-Pugh B/C cirrhosis; DI = disparate impact; eGFR = estimated glomerular filtration rate; EOD = equal opportunity difference; FEV1 = forced expiratory volume; FNR = false negative rate; FRS = Framingham Risk Score; FVC = forced vital capacity; GLI = Global Lung Function Initiative; iGFR = iothalamate glomerular filtration rate; IQR = interquartile range; IRR = incidence rate ratio; KAS = kidney allocation system; LTA = liver transplant alone; MESA = multi-ethnic study of atherosclerosis; MDRD = Modification of Diet in Renal Disease study; NRI = net reclassification index; OPTN = Organ Procurement and Transplantation Network; OR = odds ratio; PCE = pooled cohort equations; SD = standard deviation; SLKT = simultaneous liver-kidney transplantation; SDOH = social determinants of health; SOFA = Sequential Organ Failure Assessment;

Table D-3. Risk-of-bias assessment for included studies

Author	Bias Due to Confounding	Bias in Selection of Participants Into the Study	Bias in Classification of Interventions	Bias Due to Deviations From Intended Interventions	Bias Due to Missing Data	Bias in Measurement of Outcomes	Bias in Selection of the Reported Result	Overall Risk of Bias
Ahmed et al. 2021[18]	Moderate	Moderate	Low	Moderate	Moderate	Low	Moderate	Moderate
Ashana et al. 2021[1]	Low	Moderate	Low	Moderate	Low	Moderate	Low	Moderate
Baugh et al. 2022[19]	Low	NA	Low	Low	Low	Low	Low	Low
Boley et al. 2022[2]	Low	Low	Low	Moderate	Low	Moderate	Low	Moderate
Bundy et al. 2022[20]	Low	Moderate	Low	Low	Moderate	Low	Moderate	Moderate
Carbunaru et al. 2019[3]	Low	Moderate	Low	Moderate	Low	Low	Moderate	Moderate
Casal et al. 2021[21]	Low	NR	Low	Low	Low	Low	Low	Low
Coresh et al. 2019[22]	Low	Moderate	Moderate	Low	Low	Low	Low	Moderate
Diao et al. 2023[23]	Low	Moderate	Low	Low	Low	Moderate	Low	Moderate
Doshi et al. 2022[24]	Low	Moderate	Low	Low	Low	Low	Low	Moderate
Drawz et al. 2012[25]	Low	NA	Low	Low	Low	Low	Low	Low
Duggal et al. 2021[26]	Low	Low -> Moderate	Low	Moderate	Moderate	Low	Low	Moderate
Elmaleh-Sachs et al. 2021[27]	Low	Low	Low	Moderate	Moderate	Moderate	Low	Moderate
Fairman et al. 2020[28]	NA	Moderate	NA	Moderate	Low	Low -> Moderate	Low	Moderate
Foryciarz et al. 2022[29]	Low	Moderate	Low	Low	Low	Low -> Moderate	Low	Moderate
Fox et al. 2016[30]	Moderate	Low	Low	Moderate	Low	Moderate	Low	Moderate
Gutierez et al. 2022[31]	Low	Moderate	Low	Low	Moderate	Low	Low	Moderate
Hammond et al. 2020[32]	Moderate	Low	Low	Low	High	Low	Low	High
Han et al. 2020[4]	Low	Low	Low	Moderate	Low	Low	Low	Moderate
Hoenig et al. 2021[33]	High	Low -> Moderate	Low	High	Moderate	Low	Low	High
Huang et al. 2022[34]	Low	Low	Low	Low	Low	Low	Low	Low

Author	Bias Due to Confounding	Bias in Selection of Participants Into the Study	Bias in Classification of Interventions	Bias Due to Deviations From Intended Interventions	Bias Due to Missing Data	Bias in Measurement of Outcomes	Bias in Selection of the Reported Result	Overall Risk of Bias
Inker et al. 2021[35]	Low	Low	Moderate	Low	Low	Moderate	Low	Moderate
Inker et al. 2021[36]	Low	Moderate	NA	Moderate	Moderate	Low	Low	Moderate
Julian et al. 2017[37]	Low	Low	Low	Low	Moderate	Moderate	Moderate	Moderate
Kabra et al. 2016[38]	Moderate	Moderate -> High	Low	Moderate	Low	Low	Low	Moderate -> High
Kimmel et al 2013[39]	Low	Low	Low	Low	Low	Low	Low	Low
Landy et al. 2021[40]	Low	Moderate	Low	Moderate	Moderate	Moderate	Low	Moderate
Limdi et al. 2015[41]	Low	NA	Low	Moderate	Low	Low	Low	Moderate
Lindley et al. 2022[42]	Low	Low	Low	Low	Low	Low	Low	Low
Mahmud et al. 2021[43]	Low	Low	Low	Low	Low	Moderate	Moderate	Moderate
Metzger et al. 2022[5]	Low	Low -> Moderate	Low	Low	Low	Moderate	Low	Moderate
Meeusen et al. 2022[44]	Low	Low	Low	Low	Low	Low	Moderate	Moderate
Miller et al. 2022[45]	Low	Moderate	Low	Low	Low	Low	Low	Moderate
Miller et al. 2021[46]	Low	Low	Low	Moderate	Low	Moderate	Low	Moderate
Miller et al. 2021[6]	Low	Moderate	Low	Moderate	Low	Low	Low	Moderate
Muiru et al. 2023[47]	Low	Low	Low	Low	Low	Moderate	Moderate	Moderate
Obermeyer et al. 2019[7]	Low	Moderate	Low	Low	Low	Moderate	Low	Moderate
Panchal et al. 2022[48]	Low	Low	Low	High	Moderate	Low	Low	High
Park et al. 2021[49]	Low	Low -> Moderate*	Low	Moderate	Low	Low	Moderate	Moderate
Pasquinelli et al. 2021[8]	Low	Moderate	Low	Moderate	High	Low	High	High
Presti et al. 2021[9]	Low	Low	Low	Moderate	Low	Moderate	Moderate	Moderate
Riviello et al. 2022[10]	Low	Low	Low	Moderate	Moderate	Moderate	Moderate	Moderate
Sarkar et al. 2021[11]	Moderate	Moderate	Low	Moderate	Moderate	High	Moderate	High
Schmeusser et al. 2022[50]	Low	Low	Low	Low	Low	Low	Low	Low
Shi et al. 2021[51]	Low	Low	NA	NA	Moderate	Low	Low	Moderate
Shores et al. 2013[52]	Moderate	Moderate	Low	Low	Low	Moderate	Low	Moderate
Snavely et al. 2021[12]	Low	Moderate	Low	Low	High	Low	Moderate	High

Author	Bias Due to Confounding	Bias in Selection of Participants Into the Study	Bias in Classification of Interventions	Bias Due to Deviations From Intended Interventions	Bias Due to Missing Data	Bias in Measurement of Outcomes	Bias in Selection of the Reported Result	Overall Risk of Bias
Thompson et al. 2021[13]	Low	Low	Low	Moderate	Low	Moderate	Low	Moderate
Topel et al. 2018[53]	Moderate	Moderate	Low	Low	Low	Moderate	Moderate	Moderate
Tsai et al. 2021[54]	Moderate	Low	Low	Low	Low	Moderate	Low	Moderate
Weale et al. 2021[55]	Moderate	Moderate	Low	Low	Low	Moderate	Low	Moderate
Wille et al. 2013[14]	Moderate	Moderate	Low	Low	Low	Low	Low	Moderate
Williams et al. 2022[15]	Low	Moderate	Low	Moderate	Moderate	Low	Low	Moderate
Yadlowsky et al. 2018[56]	Low	Low	Low	Low	Low	Low	Low	Low
Yap et al. 2021[57]	Low	Low -> Moderate	Low	Low	Low	High	High	High
Yoo et al. 2023[16]	Low	Moderate	Low	Low	Moderate	High	Low	High
Zelnick et al. 2021[58]	Moderate	Moderate	Low	Low	Low	Low	Low	Moderate
Zhang et al. 2018[17]	High	Moderate	Low	Low	Moderate	Low	Moderate	Moderate

Abbreviations: NA = not applicable; NR = not reported. *Changes to risk of bias rating due to the additional equity-based signaling questions are shown. As shown, changes to individual domains were made as a result of these questions, but only impacted the overall risk of bias for Kabra et al.

Appendix E. Contextual Question 4 Detailed Supplement

Get with the Guidelines Heart Failure Risk Score (Peterson et al., 2010)[60]

- **Model variables and scores:** blood pressure (28 points possible; higher points allotted for lower values, BUN (28 points possible; higher points allotted for higher values), sodium (28 points possible, higher points allotted for lower values), age (28 points possible, higher points allowed for higher values), heart rate (8 points possible; higher points allotted for higher values), Black Race (Yes = 0; No = 3 points), COPD (Yes = 2, No = 0). The outcome, "Probability of death" is categorized as > 20-30% for scores 55-70, > 30-40% for scores 71-74, >40-50% for scores 75-78 and >50% for scores over 79.

Development and Validation of Prediction Scores for Early Mortality at Transition to Dialysis[61]

Figure E-1. Cox regression model for predicting mortality among patients with low and high eGFR at the last measurement before dialysis initiation

Variable	Low (<15 mL/min/1.73 m^2)			High (≥15 mL/min/1.73 m^2)		
	Parameter[b]	HR (95% CI)	*P* value	Parameter[b]	HR (95% CI)	*P* value
Age (per year)	0.0311	1.03 (1.03-1.04)	<.001	0.0343	1.03 (1.03-1.04)	<.001
Race						
White		Reference			Reference	
Black	−0.4019	0.67 (0.61-0.73)	<.001	−0.2178	0.80 (0.73-0.89)	<.001
Asian	−0.6334	0.53 (0.29-0.96)	.04	−0.8342	0.43 (0.23-0.81)	.009
Native American	−0.4859	0.61 (0.36-1.04)	.07	−0.6384	0.53 (0.27-1.02)	.06
Other	−0.3692	0.69 (0.55-0.87)	.002	−0.2437	0.78 (0.60-1.02)	.07
Hispanic ethnicity	−0.2926	0.75 (0.64-0.87)	<.001	−0.2547	0.78 (0.64-0.93)	.007

Table 2, partial table with model parameters for race and ethnicity. Obi Y, Nguyen DV, Zhou H, et al. Development and validation of prediction scores for early mortality at transition to dialysis. Mayo Clinic Proceedings. 2018 Sep;93(9):1224-35. Epub 2018 Aug 10. Copyright 2018. Used with permission from Elsevier. doi: 10.1016/j.mayocp.2018.04.017. http://mayoclinicproceedings.org.

Abbreviations: CI = confidence interval; HR = hazard ratio; m^2 = square meters; min = minutes; mL = milliliters

Calculator sample for various race selections when **eGFR < 15 mL/min/1.73m^2**:

- Accessed 08/23/2022; www.dialysisscore.com
- Entered calculator clinical values: Age: 55; Body Mass Index (per Kg/M2): 35; Cause of ESRD: Primary GN; eGFR: 10; White Blood Cells (per x 10^3): 96; Serum albumin (per mg/dl):2; Serum urea nitrogen (per mg/dl): 30; Serum sodium (per mEq/L): 150; Serum alkaline phosphatase (per IU/L): 100. All other entries: No.
- White has the highest risk for mortality compared to all other races and across all timeframes. Risk for mortality decreases for each race group when Hispanic ethnicity is added. White and Hispanic shown as an example.

Table E-1. Results from the Dialysisscore.com Calculator when varying race and holding all other variables constant (eGFR < 15 mL/min/1.73m^2)

Mortality Risk (%) eGFR < 15 mL/min/1.73m^2	3 Months	6 Months	9 Months	12 Months
White	17.0	31.1	40.9	49.7
White + Hispanic	13.1	24.6	32.7	40.4
Black	11.8	22.5	30.0	37.2
Asian	9.6	18.4	24.8	31.0
Native American	11.0	20.9	28.0	34.9
Other Race	12.2	23.1	30.8	38.1

Abbreviations: eGFR = estimated glomerular filtration rate; m^2 = square meters; min = minutes; mL = milliliters

Calculator sample output for various race selections when **eGFR > 15 mL/min/1.73m^2**:

- Accessed 08/23/2022; www.dialysisscore.com
- Entered calculator clinical values: Age: 55; Body Mass Index (per Kg/M2): 35; Cause of ESRD: Primary GN; eGFR: 20; Serum albumin (per mg/dl): 2; Serum urea nitrogen (per mg/dl): 30; Serum sodium (per mEq/L): 150; Serum alkaline phosphatase (per IU/L): 100. All other entries: No.
- White has the highest risk for mortality compared to all other races and across all timeframes. Risk for mortality decreases for each race group when Hispanic ethnicity is added. White and Hispanic shown as an example.

Table E-2. Results from the Dialysisscore.com Calculator when varying race and holding all other variables constant (eGFR > 15 mL/min/1.73m2)

Mortality Risk (%) eGFR > 15 mL/min/1.73m^2	3 Months	6 Months	9 Months	12 Months
White	3.0	5.3	7.3	9.1
White + Hispanic	2.3	4.2	5.7	7.2
Black	2.4	4.3	5.9	7.4
Asian	1.3	2.4	3.3	4.2
Native American	1.6	2.9	4.0	5.0
Other Race	2.4	4.2	5.8	7.2

Abbreviations: eGFR = estimated glomerular filtration rate; m^2 = square meters; min = minutes; mL = milliliters

Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models[62,63]
Calculator sample output for various race selections:

- Accessed 08/23/2022; https://riskcalc.sts.org/stswebriskcalc/calculate
- Entered calculator clinical values: Age: 55; Gender: Male; Weight (kg): 90; Height (cm): 172; Hematocrit: 75; WBC: 96; Platelet: 100,000, Last creatinine level: 25. All other entries: No or not filled.

Table E-3. Results from the Online STS Risk Calculator when varying race and holding all other variables constant

Outcomes for Isolated CABG	White	Black	Asian	American Indian / Alaskan Native	Native Hawaiian / Pacific Islander	Hispanic / Latino / Spanish Ethnicity
Risk of Mortality	1.264	1.503 ↓	1.264	1.264	1.264	1.264
Renal Failure	NA	NA	NA	NA	NA	NA
Premature Stroke	0.538 ↑	0.835 ↓	0.750	0.538 ↑	0.538 ↑	0.538 ↑
Prolonged Ventilation	5.054 ↑	6.923	6.666	6.092	7.221 ↓	5.852
DSW Infection	0.135	0.135	0.135	0.135	0.135	0.168 ↓
Reoperation	2.249 ↑	2.249 ↑	2.861	2.249 ↑	3.514 ↓	2.484
Morbidity + Mortality Composite	44.950 ↑	52.326 ↓	50.701	49.958	52.001	48.596
Short Length of Stay	51.375 ↑	42.913 ↓	45.110	51.375 ↑	48.148	47.839
Long Length of Stay	4.289 ↑	6.202 ↓	5.234	4.289 ↑	5.580	4.610

Abbreviations: CABG = coronary artery bypass graft; DSW = mediastinitis/deep sternal wound infection; NA = not applicable.
Note: Up arrows indicate best outcome risk in each outcome category; down arrows indicate worst outcome risk in each outcome category. Color shading is for emphasis only.

Denver Human Immunodeficiency Virus (HIV) Risk Score for Targeted HIV Screening[64-67]

- The risk score was created by multiplying the final model's regression coefficients by 10 and rounding them to the nearest integer.

Figure E-2. Multivariate logistic regression model for the prediction of newly diagnosed human immunodeficiency virus Infection and translation of regression coefficients into the Denver HIV Risk Score.[64]

Table 2. Multivariate Logistic Regression Model for the Prediction of Newly Diagnosed Human Immunodeficiency Virus Infection and Translation of Regression Coefficients into the Denver HIV Risk Score, Denver, Colorado, 1996–2008

Variable	β	95% Confidence Interval	Score
Age, years			
<22 or >60	0	Referent	0
22–25 or 55–60	0.4	0.3, 0.8	+4
26–32 or 47–54	1.0	0.7, 1.3	+10
33–46	1.2	0.9, 1.5	+12
Gender			
Female	0	Referent	0
Male	2.1	1.8, 2.4	+21
Race/ethnicity			
Black	0.9	0.7, 1.1	+9
Hispanic	0.3	0.1, 0.5	+3
Other[a]	−0.1	0.3, 0.1	0
White	0	Referent	0
Sexual practices			
Sex with a male	2.2	2.0, 2.5	+22
Vaginal intercourse	−1.0	0.8, −1.2	−10
Receptive anal intercourse	0.8	0.6, 1.0	+8
Other risk factors			
Injection drug use	0.9	0.7, 1.1	+9
Past HIV testing	−0.4	0.2, −0.6	−4

Abbreviation: HIV, human immunodeficiency virus.
[a] American Indian or Alaska Native, Native Hawaiian, or non-Hawaiian Pacific Islander.

Haukoos JS and Lyons MS. Derivation and validation of the Denver human immunodeficiency virus (HIV) risk score for targeted HIV screening. American Journal of Epidemiology. 2012;175(8):838-46. Used with permission from Oxford University Press. doi: 10.1093/aje/kwr389. http://academic.oup.com/aje.

Score calculator as implemented in the randomized controlled trial of Denver HIV Tool in 2021.[66] Note that in the original model, "Other" race was protective against HIV risk and received a score of -1, whereas in the 2021 study, "Other" was assigned a score of 0.

Figure E-3. Scoring tool as implemented for the randomized control trial [66]

eTable 1. Denver HIV Risk Score.

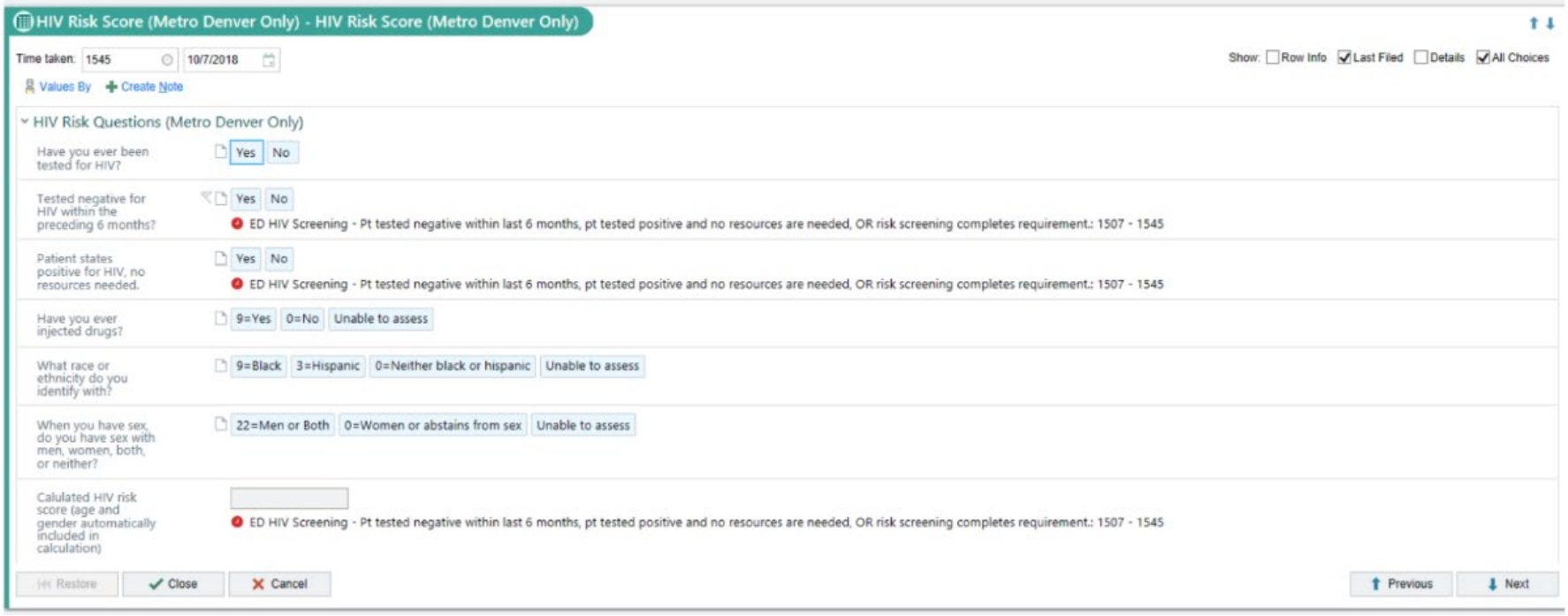
Variable	Score
Age	
22-25 or 55-60 years	+4
26-32 or 47-54 years	+10
33-46 years	+12
Gender	
Male	+21
Race/Ethnicity	
Black	+9
Hispanic	+3
Sexual Practices	
Sex with a male	+22
Other Risks	
Injection drug use	+9
Past HIV test	-4

*Reference groups that score zero are:
Age, <22 or >60 years; Gender, female;
Race/Ethnicity, white or "other", defined as American or Alaskan Native, Native Hawaiian, or non-Hawaiian Pacific Islander.
Dashes represent exclusion of the variable from the refined score. The refined score ranges from -4 to +73 with risk groups stratified as <20 (very low risk), 20 – 29 (low risk), 30 – 39 (moderate risk), 40 – 49 (high risk, and ≥50 (very high risk).

Haukoos JS, Lyons MS, Rothman RE et al. Comparison of HIV screening strategies in the emergency department: a randomized clinical trial. JAMA Network Open. 2021;4(7):e2117763. Used with permission from JAMA Network Open. doi: 10.1001/jamanetworkopen.2021.17763. http://jamanetwork.com/journals/jamanetworkopen.

University of Colorado Denver, Aurora, CO.[67] Nursing driven tool in the ED used to identify people at risk for HIV infection

Figure E-4. Screen shot of the tool as implemented in the electronic medical record [67]



Dunlevy H, Robins M, Ashwood E, et al. Targeted HIV-testing in the emergency department with linkage to care using an HIV risk score. 2018 National Ryan White Conference on HIV Care & Treatment, Aurora, CO; 2018. Used with permission from Hillary Dunlevy. https://targethiv.org/sites/default/files/supporting-files/11016_Dunlevy_508.pdf.

Abbreviations: ED = emergency department; HIV = human immunodeficiency virus

Appendix F. Appendix References

1. Ashana DC, Anesi GL, Liu VX, et al. Equitably allocating resources during crises: racial differences in mortality prediction models. Am J Respir Crit Care Med. 2021 Jul;204(2):178-86. doi: 10.1164/rccm.202012-4383OC. PMID: 33751910.

2. Boley S, Sidebottom A, Vacquier M, et al. Investigating racial disparities within an emergency department rapid-triage system. Am J Emerg Med. 2022 Jul;60:65-72. doi: 10.1016/j.ajem.2022.07.030. PMID: 35907271.

3. Carbunaru S, Nettey OS, Gogana P, et al. A comparative effectiveness analysis of the PBCG vs PCPT risks calculators in a multi-ethnic cohort. BMC Urol. 2019 Nov;19(1):121. doi: 10.1186/s12894-019-0553-6. PMID: 31771578.

4. Han SS, Chow E, Ten Haaf K, et al. Disparities of National Lung Cancer Screening Guidelines in the US population. J Natl Cancer Inst. 2020 Nov;112(11):1136-42. doi: 10.1093/jnci/djaa013. PMID: 32040195.

5. Metzger P, Allum L, Sullivan E, et al. Racial and language disparities in pediatric emergency department triage. Pediatr Emerg Care. 2022 Feb;38(2):e556-e62. doi: 10.1097/PEC.0000000000002439. PMID: 34009885.

6. Miller WD, Han X, Peek ME, et al. Accuracy of the sequential organ failure assessment score for in-hospital mortality by race and relevance to crisis standards of care. JAMA Netw Open. 2021 Jun;4(6):e2113891. doi: 10.1001/jamanetworkopen.2021.13891. PMID: 34143190.

7. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019 Oct;366(6464):447-53. doi: 10.1126/science.aax2342. PMID: 31649194.

8. Pasquinelli MM, Tammemägi MC, Kovitz KL, et al. Brief report: risk prediction model versus United States Preventive Services Task Force 2020 draft lung cancer screening eligibility criteria-reducing race disparities. JTO Clin Res Rep. 2021 Mar;2(3):100137. doi: 10.1016/j.jtocrr.2020.100137. PMID: 34590000.

9. Presti JC, Alexeeff S, Horton B, et al. Prospective validation of the Kaiser Permanente prostate cancer risk calculator in a contemporary, racially diverse, referral population. Urol Oncol. 2021 Nov;39(11):783.e11-.e19. doi: 10.1016/j.urolonc.2021.03.023. PMID: 33962850.

10. Riviello ED, Dechen T, O'Donoghue AL, et al. Assessment of a crisis standards of care scoring system for resource prioritization and estimated excess mortality by race, ethnicity, and socially vulnerable area during a regional surge in COVID-19. JAMA Netw Open. 2022 Mar;5(3):e221744. doi: 10.1001/jamanetworkopen.2022.1744. PMID: 35289860.

11. Sarkar R, Martin C, Mattie H, et al. Performance of intensive care unit severity scoring systems across different ethnicities in the USA: a retrospective observational study. Lancet Digital Health. 2021 Apr;3(4):e241-e9. doi: 10.1016/S2589-7500(21)00022-4. PMID: 33766288.

12. Snavely AC, Hendley N, Stopyra JP, et al. Sex and race differences in safety and effectiveness of the HEART pathway accelerated diagnostic protocol for acute chest pain. Am Heart J. 2021 Feb;232:125-36. doi: 10.1016/j.ahj.2020.11.005. PMID: 33160945.

13. Thompson HM, Sharma B, Bhalla S, et al. Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. J Am Med Inform Assoc. 2021 Nov;28(11):2393-403. doi: 10.1093/jamia/ocab148. PMID: 34383925.

14. Wille KM, Harrington KF, Deandrade JA, et al. Disparities in lung transplantation before and after introduction of the lung allocation score. J Heart Lung Transplant. 2013 Jul;32(7):684-92. doi: 10.1016/j.healun.2013.03.005. PMID: 23582477.

15. Williams RM, Li T, Luta G, et al. Lung cancer screening use and implications of varying eligibility criteria by race and ethnicity: 2019 behavioral risk factor surveillance system data. Cancer. 2022 May;128(9):1812-9. doi: 10.1002/cncr.34098. PMID: 35201610.

16. Yoo RM, Dash D, Lu JH, et al. Investigating real-world consequences of biases in commonly used clinical calculators. Am J Manag Care. 2023 Jan;29(1):e1-e7. doi: 10.37765/ajmc.2023.89306. PMID: 36716157.

17. Zhang X, Melanson TA, Plantinga LC, et al. Racial/ethnic disparities in waitlisting for deceased donor kidney transplantation 1 year after implementation of the new national kidney allocation system. Am J Transplant. 2018 Aug;18(8):1936-46. doi: 10.1111/ajt.14748. PMID: 29603644.

18. Ahmed S, Nutt CT, Eneanya ND, et al. Examining the potential impact of race multiplier utilization in estimated glomerular filtration rate calculation on African-American care outcomes. J Gen Intern Med. 2021 Feb;36(2):464-71. doi: 10.1007/s11606-020-06280-5. PMID: 33063202.

19. Baugh AD, Shiboski S, Hansel NN, et al. Reconsidering the utility of race-specific lung function prediction equations. Am J Respir Crit Care Med. 2022 Apr;205(7):819-29. doi: 10.1164/rccm.202105-1246OC. PMID: 34913855.

20. Bundy JD, Mills KT, Anderson AH, et al. Prediction of end-stage kidney disease using estimated glomerular filtration rate with and without race: a prospective cohort study. Ann Intern Med. 2022 Mar;175(3):305-13. doi: 10.7326/M21-2928. PMID: 35007146.

21. Casal MA, Ivy SP, Beumer JH, et al. Effect of removing race from glomerular filtration rate-estimating equations on anticancer drug dosing and eligibility: a retrospective analysis of National Cancer Institute phase 1 clinical trial participants. Lancet Oncol. 2021 Sep;22(9):1333-40. doi: 10.1016/S1470-2045(21)00377-6. PMID: 34399096.

22. Coresh J, Inker LA, Sang Y, et al. Metabolomic profiling to improve glomerular filtration rate estimation: a proof-of-concept study. Nephrol Dial Transplant. 2019 May;34(5):825-33. doi: 10.1093/ndt/gfy094. PMID: 29718360.

23. Diao JA, Wu GJ, Wang JK, et al. National projections for clinical implications of race-free creatinine-based GFR estimating equations. J Am Soc Nephrol. 2023 Feb;34(2):309-21. doi: 10.1681/ASN.2022070818. PMID: 36368777.

24. Doshi MD, Schaubel DE, Xu Y, et al. Clinical utility in adopting race-free kidney donor risk index. Transplant Direct. 2022 Jul;8(7):e1343. doi: 10.1097/TXD.0000000000001343. PMID: 35747522.

25. Drawz PE, Baraniuk S, Davis BR, et al. Cardiovascular risk assessment: addition of CKD and race to the Framingham equation. Am Heart J. 2012 Dec;164(6):925-31.e2. doi: 10.1016/j.ahj.2012.09.003. PMID: 23194494.

26. Duggal V, Thomas IC, Montez-Rath ME, et al. National estimates of CKD prevalence and potential impact of estimating glomerular filtration rate without race. J Am Soc Nephrol. 2021 Jun;32(6):1454-63. doi: 10.1681/ASN.2020121780. PMID: 33958490.

27. Elmaleh-Sachs A, Balte P, Oelsner EC, et al. Race/ethnicity, spirometry reference equations and prediction of incident clinical events: the multi-ethnic study of atherosclerosis (MESA) lung study. Am J Respir Crit Care Med. 2022 Mar;205(6):700-10. doi: 10.1164/rccm.202107-1612OC. PMID: 34913853.

28. Fairman KA, Romanet D, Early NK, et al. Estimated cardiovascular risk and guideline-concordant primary prevention with statins: retrospective cross-sectional analyses of US ambulatory visits using competing algorithms. J Cardiovasc Pharmacol Ther. 2020 Jan;25(1):27-36. doi: 10.1177/1074248419866153. PMID: 31353942.

29. Foryciarz A, Pfohl SR, Patel B, et al. Evaluating algorithmic fairness in the presence of clinical guidelines: the case of atherosclerotic cardiovascular disease risk estimation. BMJ Health Care Inform. 2022 Apr;29(1):e100460. doi: 10.1136/bmjhci-2021-100460. PMID: 35396247.

30. Fox ER, Samdarshi TE, Musani SK, et al. Development and validation of risk prediction models for cardiovascular events in black adults: the Jackson heart study cohort. JAMA Cardiol. 2016 Apr;1(1):15-25. doi: 10.1001/jamacardio.2015.0300. PMID: 27437649.

31. Gutiérrez OM, Sang Y, Grams ME, et al. Association of estimated GFR calculated using race-free equations with kidney failure and mortality by black vs non-black race. JAMA. 2022 Jun;327(23):2306-16. doi: 10.1001/jama.2022.8801. PMID: 35667006.

32. Hammond G, Johnston K, Huang K, et al. Social determinants of health improve predictive accuracy of clinical risk models for cardiovascular hospitalization, annual cost, and death. Circ Cardiovasc Qual Outcomes. 2020 Jun;13(6):e006752. doi: 10.1161/CIRCOUTCOMES.120.006752. PMID: 32412300.

33. Hoenig MP, Mann A, Pavlakis M. Removal of the Black race coefficient from the estimated glomerular filtration equation improves transplant eligibility for Black patients at a single center. Clin Transplant. 2022 Feb;36(2):e14467. doi: 10.1111/ctr.14467. PMID: 34605076.

34. Huang C, Murugiah K, Li X, et al. Effect of removing race correction factor in glomerular filtration rate estimation on predicting acute kidney injury after percutaneous coronary intervention. medRxiv. 2022 Jan. doi: 10.1101/2022.01.18.22269155.

35. Inker LA, Couture SJ, Tighiouart H, et al. A new panel-estimated GFR, including ß(2)-microglobulin and ß-trace protein and not including race, developed in a diverse population. Am J Kidney Dis. 2021 Nov;77(5):673-83.e1. doi: 10.1053/j.ajkd.2020.11.005. PMID: 33301877.

36. Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. N Engl J Med. 2021 Nov;385(19):1737-49. doi: 10.1056/NEJMoa2102953. PMID: 34554658.

37. Julian BA, Gaston RS, Brown WM, et al. Effect of replacing race with apolipoprotein L1 genotype in calculation of kidney donor risk index. Am J Transplant. 2017 Jun;17(6):1540-8. doi: 10.1111/ajt.14113. PMID: 27862962.

38. Kabra R, Girotra S, Vaughan Sarrazin M. Refining stroke prediction in atrial fibrillation patients by addition of African-American ethnicity to CHA(2)DS(2)-VASc score. J Am Coll Cardiol. 2016 Aug;68(5):461-70. doi: 10.1016/j.jacc.2016.05.044. PMID: 27470453.

39. Kimmel SE, French B, Kasner SE, et al. A pharmacogenetic versus a clinical algorithm for warfarin dosing. N Engl J Med. 2013;369(24):2283-93. doi: 10.1056/NEJMoa1310669. PMID: 24251361.

40. Landy R, Young CD, Skarzynski M, et al. Using prediction-models to reduce persistent racial/ethnic disparities in draft 2020 USPSTF lung-cancer screening guidelines. J Natl Cancer Inst. 2021 Nov;113(11):1590-4. doi: 10.1093/jnci/djaa211. PMID: 33399825.

41. Limdi NA, Brown TM, Yan Q, et al. Race influences warfarin dose changes associated with genetic factors. Blood. 2015 Jul;126(4):539-45. doi: 10.1182/blood-2015-02-627042. PMID: 26024874.

42. Lindley KJ, Limdi NA, Cavallari LH, et al. Warfarin dosing in patients with CYP2C9*5 variant alleles. Clin Pharmacol Ther. 2022;111(4):950-5. doi: 10.1002/cpt.2549. PMID: 35108398.

43. Mahmud N, Asrani SK, Reese PP, et al. Race adjustment in eGFR equations does not improve estimation of acute kidney injury events in patients with cirrhosis. Dig Dis Sc. 2022 Apr;67(4):1399-408. doi: 10.1007/s10620-021-06943-1. PMID: 33761091.

44. Meeusen JW, Kasozi RN, Larson TS, et al. Clinical impact of the refit CKD-EPI 2021 creatinine-based eGFR equation. Clin Chem. 2022;68(4):534-9. doi: 10.1093/clinchem/hvab282. PMID: 35038721.

45. Miller J, Lyden GR, McKinney WT, et al. Impacts of removing race from the calculation of the kidney donor profile index. Am J Transplant. 2023 May;23(5):636-41. doi: 10.1016/j.ajt.2022.12.016. PMID: 36695678.

46. Miller J, Knorr JP. Impact of removing the race coefficient in renal function estimate equations on drug dosage recommendations. Ann Pharmacother. 2022 Jan;56(1):44-51. doi: 10.1177/10600280211010228. PMID: 33866823.

47. Muiru AN, Madden E, Scherzer R, et al. Effect of adopting the new race-free 2021 chronic kidney disease epidemiology collaboration estimated glomerular filtration rate creatinine equation on racial differences in kidney disease progression among people with human immunodeficiency virus: an observational study. Clin Infect Dis. 2023 Feb;76(3):461-8. doi: 10.1093/cid/ciac731. PMID: 36069064.

48. Panchal S, Serper M, Bittermann T, et al. Impact of race-adjusted glomerular filtration rate estimation on eligibility for simultaneous liver-kidney transplantation. Liver Transpl. 2022 Jun;28(6):959-68. doi: 10.1002/lt.26310. PMID: 34558791.

49. Park Y, Hu J, Singh M, et al. Comparison of methods to reduce bias from clinical prediction models of postpartum depression. JAMA Netw Open. 2021 Apr;4(4):e213909. doi: 10.1001/jamanetworkopen.2021.3909. PMID: 33856478.

50. Schmeusser BN, Palacios AR, Midenberg ER, et al. Race-free renal function estimation equations and potential impact on Black patients: implications for cancer clinical trial enrollment. Cancer. 2023 Mar;129(6):920-4. doi: 10.1002/cncr.34637. PMID: 36606692.

51. Shi J, Lindo EG, Baird GS, et al. Calculating estimated glomerular filtration rate without the race correction factor: observations at a large academic medical system. Clin Chim Acta. 2021 Sep;520:16-22. doi: 10.1016/j.cca.2021.05.022. PMID: 34052206.

52. Shores NJ, Dodge JL, Feng S, et al. Donor risk index for African American liver transplant recipients with hepatitis C virus. Hepatology. 2013 Oct;58(4):1263-9. doi: 10.1002/hep.26478. PMID: 23696235.

53. Topel ML, Shen J, Morris AA, et al. Comparisons of the Framingham and pooled cohort equation risk scores for detecting subclinical vascular disease in Blacks versus whites. Am J Cardiol. 2018 Mar;121(5):564-9. doi: 10.1016/j.amjcard.2017.11.031. PMID: 29361288.

54. Tsai JW, Cerdeña JP, Goedel WC, et al. Evaluating the impact and rationale of race-specific estimations of kidney function: estimations from U.S. NHANES, 2015-2018. EClinicalMedicine. 2021 Dec;42:101197. doi: 10.1016/j.eclinm.2021.101197. PMID: 34849475.

55. Weale ME, Riveros-Mckay F, Selzam S, et al. Validation of an integrated risk tool, including polygenic risk score, for atherosclerotic cardiovascular disease in multiple ethnicities and ancestries. Am J Cardiol. 2021 Jun;148:157-64. doi: 10.1016/j.amjcard.2021.02.032. PMID: 33675770.

56. Yadlowsky S, Hayward RA, Sussman JB, et al. Clinical implications of revised pooled cohort equations for estimating atherosclerotic cardiovascular disease risk. Ann Intern Med. 2018 Jul;169(1):20-9. doi: 10.7326/M17-3011. PMID: 29868850.

57. Yap E, Prysyazhnyuk Y, Ouyang J, et al. The implication of dropping race from the MDRD equation to estimate GFR in an African American-only cohort. Int J Nephrol. 2021 Nov;2021:1880499. doi: 10.1155/2021/1880499. PMID: 34824870.

58. Zelnick LR, Leca N, Young B, et al. Association of the estimated glomerular filtration rate with vs without a coefficient for race with time to eligibility for kidney transplant. JAMA Netw Open. 2021 Jan;4(1):e2034004. doi: 10.1001/jamanetworkopen.2020.34004. PMID: 33443583.

59. Pasquinelli MM, Tammemägi MC, Kovitz KL, et al. Risk prediction model versus United States Preventive Services Task Force lung cancer screening eligibility criteria: reducing race disparities. J Thorac Oncol. 2020 Nov;15(11):1738-47. doi: 10.1016/j.jtho.2020.08.006. PMID: 32822843.

60. Peterson PN, Rumsfeld JS, Liang L, et al. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. Circ Cardiovasc Qual Outcomes. 2010 Jan;3(1):25-32. doi: 10.1161/circoutcomes.109.854877. PMID: 20123668.

61. Obi Y, Nguyen DV, Zhou H, et al. Development and validation of prediction scores for early mortality at transition to dialysis. Mayo Clin Proc. 2018 Sep;93(9):1224-35. doi: 10.1016/j.mayocp.2018.04.017. PMID: 30104041.

62. O'Brien SM, Feng L, He X, et al. The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: part 2-statistical methods and results. Ann Thorac Surg. 2018 May;105(5):1419-28. doi: 10.1016/j.athoracsur.2018.03.003. PMID: 29577924.

63. Shahian DM, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons 2018 adult cardiac surgery risk models: part 1-background, design considerations, and model development. Ann Thorac Surg. 2018 May;105(5):1411-8. doi: 10.1016/j.athoracsur.2018.03.002. PMID: 29577925.

64. Haukoos JS, Lyons MS, Lindsell CJ, et al. Derivation and validation of the Denver human immunodeficiency virus (HIV) risk score for targeted HIV screening. Am J Epidemiol. 2012 Apr;175(8):838-46. doi: 10.1093/aje/kwr389. PMID: 22431561.

65. Haukoos JS, Hopkins E, Bucossi MM, et al. Brief report: validation of a quantitative HIV risk prediction tool using a national HIV testing cohort. J Acquir Immune Defic Syndr. 2015 Apr;68(5):599-603. doi: 10.1097/QAI.0000000000000518. PMID: 25585300.

66. Haukoos JS, Lyons MS, Rothman RE, et al. Comparison of HIV screening strategies in the emergency department: a randomized clinical trial. JAMA Netw Open. 2021 Jul;4(7):e2117763. doi: 10.1001/jamanetworkopen.2021.17763. PMID: 34309668.

67. Dunlevy H, Robins M, Ashwood E, et al. Targeted HIV-testing in the emergency department with linkage to care using an HIV risk score. Aurora (CO): 2018 National Ryan White Conference on HIV Care & Treatment; 2018. https://targethiv.org/sites/default/files/supporting-files/11016_Dunlevy_508.pdf. Accessed on August 1, 2022.