



OPEN ACCESS

EDITED BY

Daisuke Akiba,
The City University of New York, United States

REVIEWED BY

Georg Starke,
Institute for History and Ethics of Medicine,
Germany
Boluwaji Ade Akinuwaesi,
University of Eswatini, Eswatini

*CORRESPONDENCE

Christopher T. Fields
✉ christopher.fields@yale.edu

RECEIVED 11 November 2024

ACCEPTED 22 April 2025

PUBLISHED 15 May 2025

CITATION

Fields CT, Black C, Thind JK, Jegede O,
Aksen D, Rosenblatt M, Assari S, Bellamy C,
Anderson E, Holmes A and Scheinost D (2025)
Governance for anti-racist AI in healthcare:
integrating racism-related stress in psychiatric
algorithms for Black Americans.
Front. Digit. Health 7:1492736.
doi: 10.3389/fdgth.2025.1492736

COPYRIGHT

© 2025 Fields, Black, Thind, Jegede, Aksen,
Rosenblatt, Assari, Bellamy, Anderson, Holmes
and Scheinost. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Governance for anti-racist AI in healthcare: integrating racism-related stress in psychiatric algorithms for Black Americans

Christopher T. Fields^{1,2*}, Carmen Black^{3,4}, Jannat K. Thind¹,
Oluwole Jegede^{4,5}, Damla Aksen⁴, Matthew Rosenblatt²,
Shervin Assari⁶, Chyrell Bellamy¹, Elijah Anderson⁷,
Avram Holmes⁸ and Dustin Scheinost²

¹Program for Recovery and Community Health, Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States, ²Magnetic Resonance Research Center, Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, United States, ³Institute of Living at Hartford Hospital, Hartford, CT, United States, ⁴Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States, ⁵Equity Research and Innovation Center (ERIC), Yale School of Medicine, New Haven, CT, United States, ⁶Department of Family Medicine, Charles R Drew University of Medicine and Science, Los Angeles, CA, United States, ⁷Department of Sociology, Yale University, New Haven, CT, United States, ⁸Department of Psychiatry, Brain Health Institute, Rutgers University, Piscataway, NJ, United States

While the world is aware of America's history of enslavement, the ongoing impact of anti-Black racism in the United States remains underemphasized in health intervention modeling. This Perspective argues that algorithmic bias—manifested in the worsened performance of clinical algorithms for Black vs. white patients—is significantly driven by the failure to model the cumulative impacts of racism-related stress, particularly racial heteroscedasticity. Racial heteroscedasticity refers to the unequal variance in health outcomes and algorithmic predictions across racial groups, driven by differential exposure to racism-related stress. This may be particularly salient for Black Americans, where anti-Black bias has wide-ranging impacts that interact with differing backgrounds of generational trauma, socioeconomic status, and other social factors, promoting unaccounted for sources of variance that are not easily captured with a blanket “race” factor. Not accounting for these factors deteriorates performance for these clinical algorithms for all Black patients. We outline key principles for anti-racist AI governance in healthcare, including: (1) mandating the inclusion of Black researchers and community members in AI development; (2) implementing rigorous audits to assess anti-Black bias; (3) requiring transparency in how algorithms process race-related data; and (4) establishing accountability measures that prioritize equitable outcomes for Black patients. By integrating these principles, AI can be developed to produce more equitable and culturally responsive healthcare interventions. This anti-racist approach challenges policymakers, researchers, clinicians, and AI developers to fundamentally rethink how AI is created, used, and regulated in healthcare, with profound implications for health policy, clinical practice, and patient outcomes across all medical domains.

KEYWORDS

anti-racist AI, racism-related stress, clinical algorithms, algorithmic bias, community-based participatory research

1 Introduction

Artificial Intelligence (AI) and machine learning (ML) technologies are rapidly transforming psychiatric care, offering unprecedented opportunities for early diagnosis, personalized treatment, and improved patient outcomes (1, 2). However, as these technologies become increasingly integrated into mental health services, there is a growing concern that they may inadvertently perpetuate or even exacerbate existing racial disparities, particularly for Black Americans (3–6). This Perspective argues for the urgent need to develop and implement anti-racist AI in healthcare, using psychiatric care as a lens to examine how systemic and multigenerational American anti-Black racism affects mental health outcomes and healthcare delivery.

Recent studies have highlighted the potential for AI systems to exhibit racial bias, even when race is not explicitly included as a variable (7–9). This phenomenon, often referred to as “algorithmic bias,” encompasses a range of deviations from normative standards, including statistical inaccuracies and ethical concerns (10). Algorithmic bias can emerge from multiple sources, such as biased training data, inadequate representation of minority groups in AI development teams, and failure to account for the broader sociocultural context in which these technologies are deployed (5, 11). While some biases may be technical artifacts, others reflect deeply entrenched structural inequities that AI systems inadvertently reproduce.

In the field of psychiatry, where diagnosis and treatment often rely heavily on subjective assessments and cultural nuances, the risk of perpetuating racial biases through AI is particularly acute (6, 12). The poorer performance of current clinical algorithms perpetuates the misdiagnosis and underdiagnosis of mental health conditions in Black Americans, largely due to their failure to account for the pervasive impacts of racism on mental health. This algorithmic bias reflects longstanding clinical tendencies to underestimate the psychological toll of systemic racism, leading to inadequate detection and treatment of mental health issues in Black American populations (13, 14).

While social and cultural factors impact health outcomes across medical domains, psychiatry particularly grapples with diagnostic constructs that are highly influenced by these factors. The absence of clear biological markers for most psychiatric disorders complicates AI-driven diagnostic predictions, necessitating careful consideration of how racial disparities manifest uniquely in psychiatric diagnoses (15, 16). For instance, the reliance on subjective clinical assessments makes psychiatric AI models particularly vulnerable to encoding racialized diagnostic tendencies rather than actual disease pathology (17). While structural racism pervades all areas of medicine, psychiatry’s classification challenges introduce an additional layer of complexity, as the boundaries between pathology, cultural expression, and systemic bias remain highly contested (17–19). This means that anti-racist AI governance in psychiatry cannot simply borrow strategies from other fields, it must actively account for the ways in which psychiatric diagnoses themselves are shaped by racialized assumptions and sociopolitical context (20).

One critical yet underexplored factor in these disparities is racial heteroscedasticity, which refers to the unequal variance in health

outcomes and algorithmic predictions across racial groups, which may be driven by differential exposure to multigenerational American anti-Black racism-related stress (21). Because Black Americans are subject to systemic racism across multiple domains of life, their health-related experiences may demonstrate higher variability, leading to greater inconsistency in clinical algorithm performance (21). Moreover, beyond present-bound notions of racially-mediated stress, socially-mediated health detriments may accrue across generations, as the poor physical and mental wellbeing of parents has been shown to negatively impact the wellbeing of their children through possible mechanisms such as allostatic load and weathering (22–24). Failure to account for this variability reinforces disparities in psychiatric AI models.

Intersectionality and intergenerationality help explain this heteroscedasticity by identifying which minoritized individuals are more or less vulnerable to racism-related stress (21, 24, 25). Some subgroups, due to factors such as class privilege, skin tone, or generational status, may have stress exposures that resemble the lower variance patterns seen in majority group members, while others experience compounded disadvantages that amplify variability in health outcomes. Thus, intersectionality and intergenerationality provide a framework for understanding the social mechanisms behind racial heteroscedasticity, while heteroscedasticity itself describes the statistical consequences of these disparities in predictive modeling.

The impact of racism on mental health is well-documented in the research literature, with Black Americans experiencing disproportionately higher rates of psychological distress, chronic stress, and untreated mental illness compared to their white counterparts (26, 27). These disparities stem from a complex interplay of factors, including economic marginalization, limited access to care, cultural stigma, and the persistent effects of historical and ongoing racial discrimination (28, 29). This Perspective aims to address the challenges of modeling and eventually tackling these issues by proposing a framework for developing anti-racist AI in psychiatric care. There is a current wave of interest to remove race entirely from clinical algorithms (30, 31). We argue that effective AI systems in this field must go beyond mere “race-blind” approaches and actively work to counteract the effects of systemic racism on mental health outcomes (32, 33). In that light, it is important to note that contemporarily-identified Black Americans are a diverse racial subgroup comprised of people of African descent with vastly different generational identities within American society. The majority of today’s Black Americans possess ancestry that originates within foundational American history, termed Ethnic Black Americans (EBAs) under the Sociohistorical Justice framework, who are the subgroup of Black Americans whose lineages carry the fullest intergenerational transmission of American anti-Black racism (34). In addition to EBAs, a growing number of today’s Black Americans descend from lineages that voluntarily established within the US many centuries after American slavery and who do not possess familial ties to foundational American history. Despite sharing a present-day racial identity, the generationally-transmitted racism-related risk factors owing to American anti-Black racism are different between

EBAs and other Black Americans. Therefore, as will be argued within this Perspective, our field must also go beyond “race-only” approaches to AI governance that homogenize centuries of differential exposure and generational transmission of health detriments born of American anti-Black racism. In suit, our proposed framework emphasizes the importance of incorporating measures of historical identity, cultural identity, resilience, and empowerment in AI models, recognizing these factors as crucial moderators of the relationship between racial stress and mental health outcomes (35, 36). We outline key principles for developing anti-racist psychiatric AI, including inclusive development processes, rigorous bias audits, and accountability measures that prioritize equitable outcomes for Black patients (37, 38).

By challenging researchers, clinicians, and AI developers to reimagine how we create and implement AI in psychiatry, this Perspective aims to contribute to a more equitable and historically informed approach to mental health care. In doing so, we hope to spark a broader conversation about the role of AI in addressing—rather than exacerbating—racial disparities in mental health and to pave the way for more just and effective psychiatric care for all.

2 Current limitations in AI models for Black patients

The rapid integration of AI and machine learning models into healthcare, particularly in psychiatry, has brought to light significant limitations in their performance for Black American patients. One of the seminal papers highlighting this issue was the 2019 study by Obermeyer et al., which revealed racial bias in a widely used healthcare algorithm (8). The study revealed that the algorithm rated Black patients as having the same risk level as White patients who were, in reality, in poorer health. This resulted in Black patients with complex health needs receiving reduced access to care (8). This work exposed how seemingly race-blind variables can perpetuate racial disparities in AI-driven healthcare decisions. While the Obermeyer study focused on a specific algorithm, it underscored a broader problem of racial bias in AI healthcare applications.

These biases may stem from various factors deeply rooted in generationally-compounded and presently-mediated systemic racism, as well as methodological shortcomings in AI development and deployment. One of the primary issues is the underrepresentation of Black individuals in the datasets used to train AI models. For example, in dermatology, a field where skin color plays a crucial role in diagnosis, studies have shown that out of 136 analyzed papers, only one explicitly included Black patients in their datasets, yielding worsened outcomes for Black patients in the AI detection of melanoma (39). This underrepresentation extends to psychiatry, where the lack of diverse representation in mental health research has significant implications for machine learning models' accuracy and equity. Studies show that psychiatric research often fails to adequately include Black participants, leading to unrepresentative datasets (40, 41). Black American youth are less likely to receive mental health treatment across

various sources (42), resulting in skewed clinical datasets. The shortage of Black mental health professionals further compounds these issues, reducing culturally competent data collection and interpretation (43). Addressing these foundational issues of data representation is crucial for developing anti-racist AI in psychiatry. Beyond assuring race-based representation of Black American researchers and participants, due to the intergenerational impacts of American anti-Black racism, the Sociohistorical Justice framework equally calls attention for healthcare systems to attune to the differential historical identities of Black Americans, with specific attention called to ensuring commensurate representation of EBAs within healthcare practice (34).

In addition to data representation issues, recent research has uncovered more nuanced challenges in AI model performance for Black patients. Recent studies have demonstrated that even when machine learning models are trained exclusively on data from Black subjects, they still show lower predictive accuracy for this group compared to white subjects (6). This persistent disparity, even when controlling for socioeconomic factors, suggests that there are underlying mechanisms related to the experience of racism that are not being captured by current modeling approaches (6). We hypothesize that this phenomenon may be due to differences in variance across differential historical identities of contemporary racial groups. Specifically, we propose that EBA subgroups bear the synergistic effects of both historically-compounded and presently-mediated American anti-Black racism, which stands in contrast to other Black American subgroups whose lineages do not carry the historically-compounded elements of American anti-Black racism (34). Thus, Black Americans without generationally-transmitted American anti-Black trauma may exhibit variance more similar to that observed among the White patient group due to the absence of intergenerational detriments owing to anti-Black racism, and other historically-mediated stressors within a US social context. This increased variance may stem from varied exposure and psychological responses to historically-compounded racism-related stress among various generational identities of contemporary Black Americans (14, 27, 44). Importantly, racial heteroscedasticity violates fundamental assumptions of linear regression models (45–49), which are still widely used in both health disparities research and psychiatric prediction algorithms. Several high-impact papers have recently been published that demonstrate the failures of these linear models in generalizing to minoritized-population datasets (5, 6, 50); again, even those linear ML models trained solely on Black patient sample data (6). This reflects the need to further disaggregate historically-mediated risk factors of intergenerational American anti-Black racism.

Figure 1 provides a visual framework for how racial heteroscedasticity potentially impacts model accuracy across sociohistorical and racial groups. In the lower left panel, standard linear models fail to capture the complexity of variance in stress exposure, leading to biased predictions. This aligns with empirical findings demonstrating that traditional models tend to misclassify high-variability subgroups, particularly in datasets reflecting racialized health disparities. The lower middle and lower right panels shows how incorporating racism-related stress

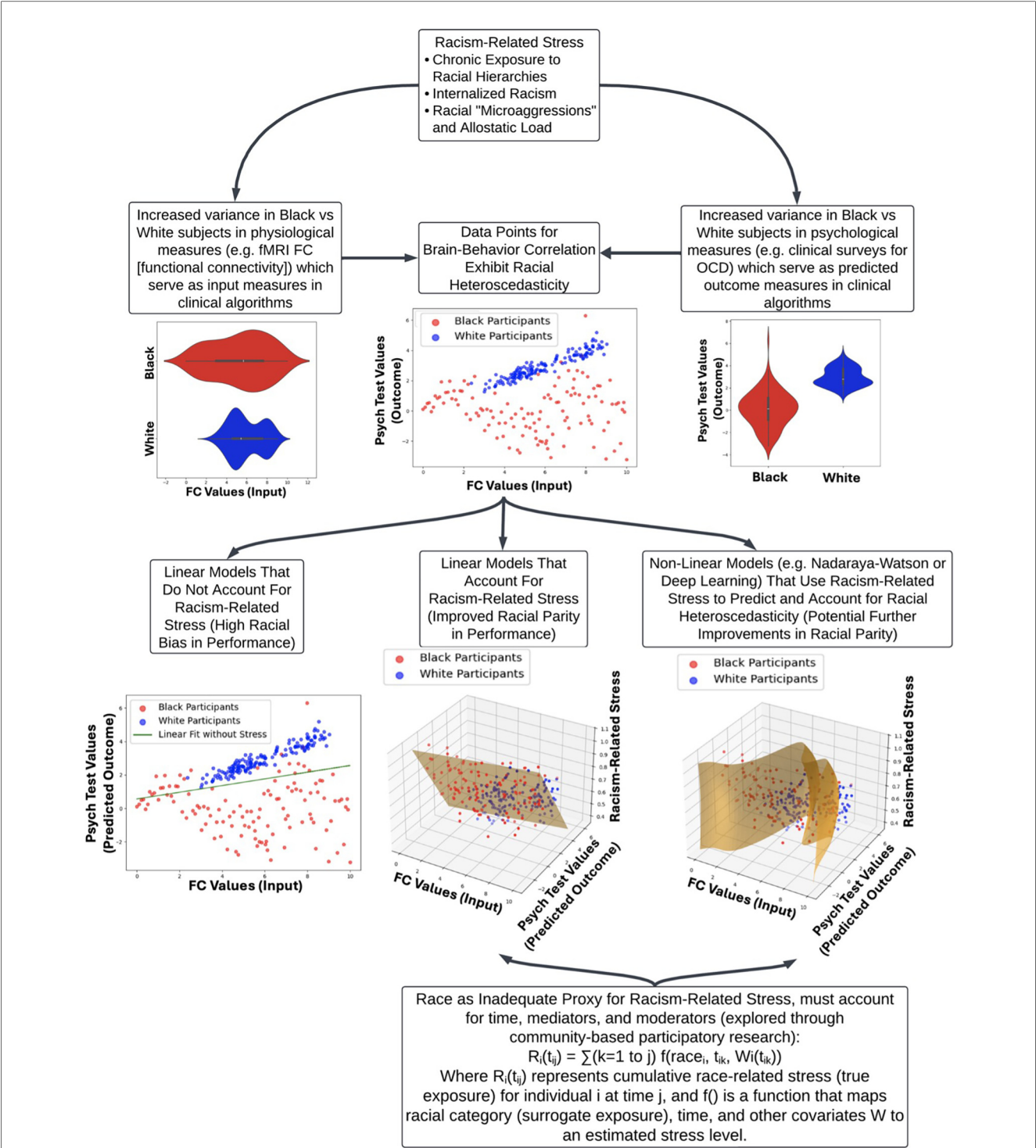


FIGURE 1
Impact of racism-related stress on clinical algorithm performance and racial heteroscedasticity. This flow chart illustrates how racism-related stress contributes to health disparities and affects clinical algorithm performance, focusing on the stress of racialization as a central factor. The top node introduces three key components of racism-related stress: chronic exposure to racial hierarchies, internalized racism, and the cumulative burden of racial microaggressions (allostatic load). Exposure to racism-related stress has wide-ranging impacts on targeted individuals, yielding a range of historically-compounded and non-historically-compounded subpopulations. This is reflected in the increased variance in physiological (e.g., fMRI functional connectivity) and psychological measures (e.g., clinical surveys for OCD) between Black and White subjects, which leads to racial heteroscedasticity in brain-behavior correlations. The lower section compares different modeling approaches: linear models without racism-related stress show significant racial bias, while incorporating stress into linear and non-linear models improves racial parity. The figure underscores the inadequacy of race as a sole proxy for racism-related stress, advocating for a more nuanced approach that includes time, mediators, and moderators, supported by community-based participatory research (CBPR).

as a moderating factor in both linear and non-linear models improves parity in algorithmic outcomes. This suggests that AI models must explicitly account for historically-compounded and directly-experienced American anti-Black racism-related stress to mitigate racial disparities in predictive accuracy. Additionally, the observed heteroscedasticity may indicate the presence of “historically-compounded” and “non-historically-compounded” Black subpopulations, reinforcing the need for tailored modeling approaches that account for intra-group variation rather than treating racialized groups as sociohistorical monolithic entities.

Beyond linear models, non-linear AI architectures such as deep neural networks and ensemble models may also fail to account for racial heteroscedasticity if training data do not sufficiently capture the full range of stress-related variability (21). These models, which optimize performance across the entire population rather than subgroups, often smooth out or obscure race-related and intergenerational-related variance rather than explicitly modeling it. This failure is especially problematic in clinical AI, where the cumulative impact of racism-related stress produces distinct physiological and psychological signatures in different sociohistorical subpopulations within contemporarily racialized groups.

Additionally, there is a concerning trend of AI models amplifying existing biases in clinical practice. For example, if historical data reflect a pattern of underdiagnosis or misdiagnosis of certain mental health conditions in Black populations, AI models trained on these data may perpetuate and even exacerbate these errors (51). Several prior studies reported lower lifetime risk of psychiatric disorders for Black individuals compared to whites, a finding that may reflect underdiagnosis or cultural differences rather than true lower prevalence (52–55). This issue is particularly salient given the complex relationship between race, clinician bias, and mental health diagnoses. Barnes and Bates (2017) highlight a paradox where Black Americans often show lower diagnostic prevalence of major depressive disorder compared to White Americans in epidemiological studies, despite experiencing higher levels of psychological distress and exposure to risk factors (56). Moreover, Jegarl et al. (2023) highlighted mechanisms by which Black Americans may also have their depressive and substance use symptoms misdiagnosed as psychosis (57). These discrepancies suggest potential underdiagnosis or misdiagnosis of depression in Black populations, which may be moderated by clinician racialized bias and lead to downstream disparities in treatment services offered to Black patients (20). Furthermore, Black and Calhoun (2022) argue that biased and carceral responses to racially-minoritized persons with mental illness in acute medical care settings can constitute iatrogenic harms, potentially leading to further misdiagnosis and inadequate treatment (58). These biases, when incorporated into AI models, risk perpetuating and amplifying racial disparities in mental health care, underscoring the critical need for careful consideration of how race and mental health data are used in AI development.

There is currently a great deal of debate about the removal of race as a factor in clinical models (30, 31). The use of race as a simplistic binary variable in many AI models fails to capture the complex, multidimensional nature of racial identity and the differential historical identities within contemporarily racialized groups, and

their synergized impact on health outcomes. This reductionist approach can perpetuate harmful stereotypes and lead to overgeneralized conclusions about Black patients' health needs and risks (33). Moving forward, it is critical that we shift away from using race as an insufficient proxy of history-based or present-day risk and instead focus on identifying and measuring the factors that mediate and modify racism-related stress. This will require a concerted effort to develop anti-racist AI approaches that can capture the nuanced impacts of structural racism, discrimination, and chronic stress on mental health outcomes that may be compounded through generational transmission in a given social context. By substituting more precise measures of intergenerational and race-related stress for crude racial categories, we may be able to build models that more accurately reflect the lived experiences of Black patients and avoid perpetuating harmful biases. Achieving this goal will take dedicated research to elucidate the complex pathways through which racism affects mental health over time and across generations, as well as collaboration between data scientists, clinicians, and communities to ensure new AI tools are equitable and patient-centered. While challenging, this paradigm shift is essential for developing psychiatric AI systems that can help reduce, rather than exacerbate, racial disparities in mental health care.

3 Directly-experienced, American anti-Black racism-related stress as a key contributor to mental health outcomes

The impact of directly-experienced, American anti-Black racism-related stress on mental health outcomes is a critical factor that must be accounted for in the development of anti-racist AI models in psychiatry. Racism-related stress refers to the cumulative psychological and physiological effects of experiencing chronic discrimination, microaggressions, and systemic oppression (27). This stress has been shown to have significant negative impacts on mental health outcomes for Black Americans, contributing to higher rates of depression, anxiety, and post-traumatic stress disorder (PTSD) (28), but has also been linked to a wide range of health outcomes including hypertension, cardiovascular disease, obesity, and accelerated cellular aging as measured by telomere length (59–61). The Everyday Discrimination Scale, a widely used measure of perceived discrimination, has been associated with numerous physical health outcomes, including coronary artery calcification, sleep disturbances, and chronic pain (62–64).

A crucial sociological insight into the importance of centering the impact of racism on mental health predictive models comes from the Anderson (2022) “Black in White Space” framework (12). This comprehensive ethnographic work, spanning over 45 years (12, 65–68), illuminates how structural racism operates in everyday life. Anderson's research underscores the psychological toll on Black Americans navigating predominantly white environments, where they must constantly be prepared for potential discrimination or hostility. The concept of “Black in White Space” emphasizes that the “Black ghetto” is not simply a physical space but has become an icon and a deep source of prejudice, negative stereotypes, and

discrimination. Because of the power of this “iconic ghetto” and the lingering impact of systemic racism, Black people are typically burdened by a negative presumption they must disprove before establishing trusting relations with others. This challenge contributes to a wide range of racial disparities, including in healthcare, employment, education, police contact, incarceration, joblessness, housing, and random insults in public. Burdened with a deficit of credibility, Black people are often required to “dance,” or to perform respectability before a largely unsympathetic audience whose minds are typically already made up about where the Black person belongs — long before they belong in the white space, they are assumed to belong in the Black space. This process of negotiation through social interaction is a constant stressor that AI models must account for when assessing mental health risks and outcomes for Black Americans. The chronic stress and vigilance required in these contexts can contribute significantly to what has been termed “racial weathering,” with deleterious effects on mental health that current AI models often fail to capture.

The concept of “weathering,” introduced by Geronimus (1992), posits that the cumulative impact of racism and socioeconomic disadvantage leads to accelerated biological aging and increased allostatic load (69). This would manifest in higher rates of stress-related physical and mental health conditions among Black Americans compared to their white counterparts. Gee et al. (2019) emphasize the importance of time in understanding racism’s effects on health (13), considering three key dimensions: time as age, exposure, and resource/privilege. Regarding time as age, Black Americans may experience accelerated aging due to chronic stress, evidenced by earlier onset of disease (70), greater morbidity at younger ages (71), shorter telomere lengths (59, 72), and shorter life expectancy compared to White individuals (73). Recent studies have directly linked experiences of racial discrimination to accelerated epigenetic aging and increased depressive symptoms, providing compelling evidence for the weathering hypothesis in the context of mental health outcomes (74). Regarding exposure, the duration, frequency, and timing of racism can significantly impact health outcomes, with potential critical periods during the life course. For instance, perceived racial discrimination during adolescence predicted depressive symptoms in young adulthood, even after controlling for earlier symptoms (75). Furthermore, time itself is a racialized resource, with racially minoritized people often experiencing a “time penalty” in various aspects of life. This inequitable distribution of time can exacerbate stress and contribute to poor health outcomes. Integrating these temporal dimensions provides a more nuanced understanding of how systemic racism affects mental health over the life course.

4 Historically-transmitted, anti-Black racism-related stress as a key contributor to mental health outcomes

The multigenerational impact of racism on mental health is not uniform across Black American sociohistorical subgroups. Factors such as generational exposure to American anti-Black racism, immigration status, and acculturation processes can influence an

individual’s risk owing to historically-transmitted racism-related stress (14). Shervin Assari’s work on Marginalization-related Diminished Returns (MDR) further complicates this picture by demonstrating that socioeconomic status (SES) interacts with racial identity in surprising ways. Contrary to expectations, Assari and colleagues have found that higher SES often fails to protect Black Americans from poor health outcomes to the same degree it does for White Americans (76). For instance, education level has been shown to have a weaker protective effect against depression for Black adults compared to White adults (77). Similarly, income has been found to have a weaker association with self-rated health for Black individuals than for White individuals (78). These findings suggest that racism not only directly impacts health but also reduces the protective effects of socioeconomic resources. Furthermore, research has shown that the health effects of discrimination can vary based on historical subgroup. Namely, Black Americans (being Ethnic Black Americans according to the sociohistorical justice framework) (34) and U.S.-born Black American individuals, being Black Americans who grew up in America’s white dominant society, may report more discrimination and associated health impacts than Caribbean Black individuals or recent African immigrants, who are Black Americans who grew up in predominantly Black societies and/or societies where race is not a dominant sociopolitical construct (79, 80). Additionally, factors such as racial identity strength, coping strategies, and social support have been found to moderate the relationship between perceived discrimination and mental health outcomes (81–83).

In light of the complex and nuanced ways in which racism impacts health outcomes for historically-diverse Black Americans, it is critical that AI healthcare models incorporate a sophisticated understanding of racism-related stress. Furthermore, to truly account for the historically-transmitted vs. directly-experienced nuances of racism, AI healthcare models must be individually tailored to each society’s unique historical, sociopolitical construction of race. That is, the intergenerational health implications of anti-Black racism cannot be meaningfully generalized from an American historical context to a different white dominant social context, like Canada or England, because those countries do not have a foundational population of Black Canadians or Britons since enslavement (84–86). Similarly, intergenerational health implications cannot be automatically generalized to Black-normative social contexts where anti-Blackness is not a dominant sociopolitical construct to determine social privilege, like the Caribbean or West Africa (87, 88). The evidence presented demonstrates that racism affects health through multiple pathways, including direct physiological impacts of chronic stress, reduced returns on protective factors like education and income, and varied effects across different sociohistorical subgroups of Black Americans. Simply including race as a variable in AI models is insufficient and may even perpetuate harmful biases. Instead, developers of AI healthcare models must strive to incorporate measures of racism-related stress, consider the moderating effects of factors like racial identity and coping strategies, and account for the differential impacts of socioeconomic status across racial groups. By doing so, these models can more accurately reflect the historically-diverse lived experiences of Black Americans and provide more

equitable and effective healthcare recommendations. This nuanced approach is essential for improving racial parity in AI healthcare models and, ultimately, for addressing the persistent health disparities that affect Black communities in the United States.

5 Key principles for developing anti-racist psychiatric models

The development of anti-racist psychiatric AI models requires a fundamental shift in approach, prioritizing the inclusion of diverse perspectives throughout the entire process. A critical aspect of this shift is the involvement of community stakeholders in the development of racism-related stress measures that are incorporated into AI models. This inclusion is particularly crucial given the intersectional complexity and nuances of the impacts of racism on health outcomes, as discussed in the previous section. The varied experiences of racism across different Black sociohistorical subgroups, the moderating effects of factors such as socioeconomic status and immigration status, and the temporal dynamics of racism-related stress all contribute to a complex landscape that cannot be adequately captured without direct input from affected communities. As illustrated in [Figure 1](#), the mathematical modeling of historically-compounded race-related stress must account for multiple factors, including time dimensions, intersectional moderating identities, and the use of race as a proxy for directly experienced racism.

By involving community members in the development of these measures, we can ensure that the AI models more accurately reflect the lived experiences of diverse Black Americans and capture the nuanced ways in which racism impacts mental health. This collaborative approach not only improves the validity and reliability of the measures, as has been observed for other patient-centric research (89), but also helps to build trust between researchers and communities, which is essential for the successful implementation and adoption of AI technologies in mental health care. In addition, the following key principles should be incorporated toward the development of anti-racist AI, as illustrated in [Figure 2](#).

5.1 Mandate inclusion of Black researchers and community members in AI development

The development of anti-racist psychiatric AI models necessitates the active inclusion of Black researchers and community members throughout the entire process (37). This principle goes beyond mere tokenism and calls for a fundamental shift towards community-based participatory research (CBPR) approaches in AI development (90). By including researchers with sociohistorically-diverse lived experience of anti-Black racism, along with engaging Black communities as co-researchers, we can ensure that the AI models reflect the lived

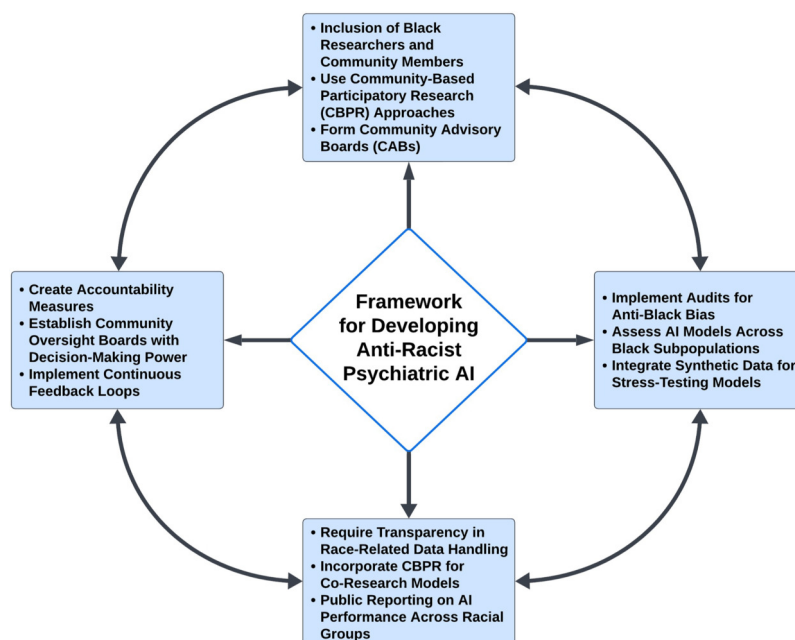


FIGURE 2

Framework for Developing Anti-Racist Psychiatric AI. The framework for developing anti-racist psychiatric AI models emphasizes the inclusion of diverse perspectives and experiences throughout the entire process. It begins with a mandate to include Black researchers and community members, ensuring that AI models accurately reflect the lived experiences and priorities of those they serve. Community-based participatory research (CBPR) plays a central role, facilitating the creation of culturally relevant and responsive AI models. The framework also underscores the importance of rigorous audits to assess for anti-Black bias, transparency in the handling of race-related data by algorithms, and the establishment of robust accountability measures. These components work together to ensure that psychiatric AI systems contribute to equitable mental health outcomes for Black patients, transforming AI from a potential source of bias into a tool for justice and equity in mental healthcare.

experiences and priorities of those they aim to serve. CBPR approaches have shown significant promise in addressing health disparities and developing culturally relevant interventions (37, 91). In the context of psychiatric AI, this could involve forming community advisory boards (CABs) comprising diverse Black community members, mental health professionals, and researchers to guide the development process. These CABs can provide invaluable insights into the nuanced ways racism impacts mental health, help identify culturally specific protective factors, and ensure that the AI models are responsive to community needs and values (37).

Moreover, it is crucial to recognize and explore psychological protective factors that may vary across different Black American sociohistorical subgroups (92, 93). This exploration necessitates the involvement of sociohistorically-diverse Black researchers and community members who can provide invaluable insights into these nuanced experiences. For instance, research has shown that first-generation Black immigrants from Black-normative societies often demonstrate better mental health outcomes compared to their U.S.-born counterparts (79). This “immigrant paradox” may be attributed to stronger cultural identity, community connectedness, and different experiences with racism (94, 95). However, these protective effects tend to erode over time and generations as immigrants acculturate to being “Black in White Space” (12). The involvement of diverse historical subgroups within the Black American community is necessary to capture these nuances. AI models must be designed to account for these temporal dynamics and the varying impacts of racism across generations, which can only be achieved through the active participation of sociohistorically-representative Black researchers and community members who have lived these experiences.

The IMANI Breakthrough project provides an instructive example of how CBPR principles can transform healthcare interventions for marginalized communities (96). This substance use intervention for Black and Latinx communities began with nine months of relationship-building through community meetings to develop research questions and implementation strategies. Through learning conversations in churches across Connecticut, researchers discovered how structural racism and discriminatory policies shaped community needs around treatment. This deep community engagement led to substantial outcomes: a 42% retention rate compared to reported outpatient treatment completion rates of 20.4% for Black individuals and 14.7% for Latinx individuals in traditional programs (97). IMANI researchers demonstrated, working directly with community members, how standardized health measures needed to be adapted to capture community-specific experiences of social determinants of health (96).

Similar principles should guide AI development in psychiatry. Just as IMANI spent months building community relationships before intervention design, AI developers should engage with affected communities well before algorithm development begins. Like IMANI’s community-driven adaptation of wellness measures, AI researchers need to work with communities to develop more culturally responsive ways of measuring and modeling mental health outcomes that capture the impacts of racism-related

stress. While this approach requires greater upfront investment in relationship-building, the IMANI project demonstrates how such engagement ultimately produces more effective and equitable interventions (96).

5.2 Implement audits that specifically assess for anti-Black bias

To ensure that psychiatric AI models do not perpetuate racial disparities, it is crucial to implement rigorous and ongoing audits specifically assessing for anti-Black bias that is sociohistorically-nuanced. These audits should transcend simple measures of overall model performance and probe the nuanced ways AI systems may discriminate against Black patients. Regular assessments of model performance and impact across different racial groups, particularly within various Black subpopulations, are essential. This includes evaluating the model’s accuracy, fairness, and potential for harm across intersections of race, ethnicity, gender, age, and socioeconomic status within Black communities. Such comprehensive audits can uncover disparities in model performance that might be masked by aggregated data, which is a common issue when broad metrics are used without considering subgroup variabilities (26, 98–100).

The potential for AI models to amplify or reinforce existing biases in clinical practice should be another key focus of these audits. For example, these audits should determine whether model recommendations align with known patterns of under-diagnosis or misdiagnosis of mental health conditions in Black American populations. This involves evaluating how models handle culturally specific expressions of mental distress that may not conform to traditional diagnostic criteria (56).

Finally, it is imperative that audits for American anti-Black bias in psychiatric AI models may be both retrospective and predictive. This means that the audits should actively seek to forecast and preempt potential biases before they manifest in real-world clinical settings. One effective approach could be the integration of synthetic data representing under-represented Black subpopulations into the training and validation processes of these models (101, 102). By simulating scenarios where the AI might encounter diverse Black patients with complex intersecting identities (e.g., a young Ethnic Black American woman with a low socioeconomic status suffering from culturally specific expressions of distress), the audit can stress-test the model’s fairness and accuracy across various contexts. This proactive strategy ensures that models are equipped to handle the rich diversity within Black American communities and reduces the risk of perpetuating existing biases, ultimately fostering more equitable mental health outcomes.

5.3 Require transparency in how algorithms handle race-related data

The lack of transparency in many AI algorithms, often protected as proprietary information, makes it difficult to identify

and address potential biases. This “black box” nature of AI systems can obscure discriminatory practices and hinder efforts to improve model performance for Black American patients (103). To counteract this, it is essential to mandate transparency throughout the entire AI development lifecycle, particularly concerning the treatment of race-related data. Given the historical context, where Black American communities have been exploited and misled in research, such as during the Tuskegee Syphilis Study (104), it is imperative that the development and employment of the most explainable models be prioritized for use in clinical settings (105). This step is vital in fostering trust, especially within Black communities that have historically been marginalized in healthcare systems, such as Ethnic Black American communities.

Central to achieving this transparency is the incorporation of CBPR principles and co-research models in AI development. CBPR emphasizes the involvement of community members as equal partners in the research process, from the initial stages of algorithm design to the final deployment in clinical settings (37). By engaging Black American communities directly in the development and evaluation of AI models, researchers can ensure that the models are not only explainable but also culturally relevant and aligned with the needs and concerns of those they are designed to serve. This collaborative approach not only enhances the transparency and explainability of AI but also empowers Black American communities, giving them agency in the tools that affect their healthcare. Furthermore, co-research fosters a deeper understanding of the social determinants that influence health disparities, allowing AI models to account for these factors more effectively (37, 106). This engagement is crucial for transforming AI from a potential source of bias into a tool for equity and justice in mental healthcare.

Finally, routine public reporting on the performance of AI systems across different racial and historical groups is essential for building trust and accountability. These reports should be accessible to the general public and presented in a format that is easy for non-experts to understand. Engaging the community in discussions about AI transparency and equity, grounded in CBPR principles, will further strengthen the relationship between healthcare providers and Black American communities. This approach ensures that AI tools are used ethically and effectively to improve mental health outcomes for all patients, particularly those from historically marginalized groups.

5.4 Create accountability measures that prioritize equitable outcomes for Black American patients

While transparency is crucial, it must be coupled with robust accountability measures to ensure that psychiatric AI models actively work towards achieving equitable outcomes for Black American patients. This involves moving beyond mere oversight to implementing concrete mechanisms that hold developers, healthcare providers, and institutions responsible for the performance and impact of AI systems on Black mental health.

In addition to using CBPR principles in the development of clinical models, another critical accountability measure is the establishment of community oversight boards with decision-making power. These boards, comprised primarily of Black American community members and mental health professionals, would have the authority to approve or reject the use of AI models based on their potential impact on Black American mental health outcomes. This approach ensures that the community most affected by these technologies has a direct say in their implementation (107).

Moreover, to deepen accountability, it is essential to implement a system of continuous feedback loops between the AI developers and the communities they serve. This could involve the use of real-time data sharing where communities are regularly updated on the performance of AI models, and their feedback is actively sought and incorporated into ongoing model adjustments. Such a system not only places community voices at the center of AI development but also ensures that the models evolve in response to the lived experiences and needs of Black American patients. By making these feedback loops a mandatory aspect of AI model deployment, we can guarantee that these systems remain responsive and accountable to those they are designed to serve.

6 Conclusion

The development of anti-racist AI in psychiatry would represent a pivotal advancement for mental health care and serve as a model for addressing racial disparities across all healthcare domains. This approach has the potential to enable more accurate diagnoses and timely interventions tailored to the unique experiences of Black American patients and other marginalized groups throughout the healthcare system.

However, realizing the full potential of anti-racist AI will require overcoming significant challenges and enacting major culture shifts among American healthcare research. Key among these is the need for inclusive development teams, comprehensive data that captures the historically nuanced experiences of racism, and the application of advanced statistical methods that address racial heteroscedasticity. Furthermore, ongoing community engagement, rigorous ethical standards, and robust accountability measures must be central to the development process. By prioritizing these elements, we can ensure that AI not only serves as a tool for reducing disparities but also contributes to the broader goal of dismantling systemic inequities in mental health care.

While this paper focuses on psychiatric AI, its challenges and governance principles apply across medicine. In psychiatric AI, bias stems not just from underrepresentation in training datasets but also from failing to account for intergenerational racism-related stress as a determinant of brain function and behavior. This issue extends beyond psychiatry to numerous clinical algorithms that have been widely criticized for their poor performance in Black American patients, regardless of whether they currently use “race-correction” factors. The eGFR (Estimated Glomerular Filtration Rate), which includes a race-based correction factor, misestimates kidney function in Black patients, leading to delayed

referrals for dialysis and transplantation (33, 108, 109). However, removing the race factor does not fully eliminate the racial disparities in CKD classification, as race-blind equations still result in lower eGFR estimates for Black American patients, increasing their likelihood of being diagnosed with CKD and reclassified into more severe disease stages compared to White patients (109). The Pooled Cohort Equations (PCE), used for cardiovascular risk assessment, differentially overestimates or underestimates risk based on race, contributing to inequitable allocation of preventive treatments (110, 111). Maternal health risk models consistently underestimate risks for Black American women, including for conditions like preeclampsia and postpartum hemorrhage, resulting in preventable maternal health disparities (31, 112, 113). Similarly, cancer risk prediction models, such as those for breast cancer, demonstrate lower accuracy in Black American women, exacerbating disparities in screening and early detection (114).

Beyond dataset underrepresentation, current clinical AI models fail to account for the broad health consequences of racism and racial socialization. Future clinical AI models must move beyond simplistic race adjustments and instead develop robust racism indices that leverage existing bodies of work that explicitly capture the historically-compounded biological and psychological consequences of American anti-Black racism. These include extensive research showing profound health correlates of David Williams' Everyday Discrimination Scale, Arline Geronimus' Weathering Hypothesis, Nancy Krieger's Ecosocial Theory, and Camara Jones' framework on levels of racism (69, 115–118). Relevant to a wide range of clinical models, studies have demonstrated that racial stress is embodied physiologically, accelerating allostatic load, neurobiological aging, and disease risk, particularly in Black Americans; however, these findings have yet to be fully integrated into clinical models (72, 119–121). These indices should complement existing social determinants of health (SDOH) measures, such as the Area Deprivation Index (ADI), and be further developed to account for sociohistorical context, ensuring that structural racism is formally integrated into risk assessment rather than erased as solely the effects of socioeconomic status (76, 122–124).

While clinical models fail to account for the impacts of American anti-Black racism within the context of US healthcare, similar biases may be present in other underrepresented groups, including Latinx populations and non-US contexts. Future work should explore how the concepts discussed here generalize across different healthcare systems and populations. In particular, the mechanisms underlying racial heteroscedasticity may vary depending on historical, economic, and policy-driven factors shaping racialization in different nations. For example, while US-based studies highlight the cumulative impacts of American anti-Black racism on psychiatric diagnosis and treatment, similar disparities exist in European or Latin American contexts through distinct but functionally equivalent pathways of medical neglect and epistemic injustice (18, 125–127). Caution must be exercised against generalization of AI healthcare models that have been calibrated for one sociopolitical and historical context to another. Nevertheless, expanding this work to international settings would help clarify how the AI governance principles

outlined here can be optimally adapted for different sociopolitical and national landscapes.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Author contributions

CF: Writing – original draft. CB: Writing – original draft, Writing – review & editing. JT: Writing – review & editing. OJ: Writing – review & editing. DA: Writing – review & editing. MR: Conceptualization, Writing – review & editing. SA: Writing – review & editing. CB: Funding acquisition, Writing – review & editing. EA: Writing – review & editing. AH: Writing – review & editing. DS: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by grant support to the authors: Yale ASCEND grant to DS, NIMH grant 7R01MH120080-06 to AH, and National Institute of Drug Abuse (NIDA) grant 1U01OD033241-01 to CB. DA is supported through NIDA T32 Neuroimaging Sciences Training in Substance Abuse Program (1T32DA022975-14). SA is partially supported by funds provided by The Regents of the University of California, Tobacco-Related Diseases Research Program, Grant Number No. T32IR5355. The funding organizations, including the National Institutes of Health (NIH), had no role in the preparation, review, or approval of the manuscript. This manuscript reflects the views of the authors and do not reflect the opinions or views of the NIH.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. The authors confirm that generative AI was used to refine the grammar and language of this manuscript, but all ideas, analyses, and conclusions are our own. We take full responsibility for the final content.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. (2018) 3:223–30. doi: 10.1016/j.bpsc.2017.11.007
- Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. (2018) 14:91–118. doi: 10.1146/annurev-clinpsy-032816-045037
- Glocker B, Jones C, Bernhardt M, Winzeck S. Algorithmic encoding of protected characteristics in chest x-ray disease detection models. *EBioMedicine*. (2023) 89:104467. doi: 10.1016/j.ebiom.2023.104467
- Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med*. (2021) 27:2176–82. doi: 10.1038/s41591-021-01595-0
- Greene AS, Shen X, Noble S, Horien C, Hahn CA, Arora J, et al. Brain-phenotype models fail for individuals who defy sample stereotypes. *Nature*. (2022) 609:109–18. doi: 10.1038/s41586-022-05118-w
- Li J, Bzdok D, Chen J, Tam A, Ooi LQR, Holmes AJ, et al. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. *Sci Adv*. (2022) 8:eabj1812. doi: 10.1126/sciadv.abj1812
- Benjamin R. Assessing risk, automating racism. *Science*. (2019) 366:421–2. doi: 10.1126/science.aaz3873
- Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. (2019) 366:447–53. doi: 10.1126/science.aax2342
- Petersen E, Ferrante E, Ganz M, Feragen A. Are demographically invariant models and representations in medical imaging fair? *arXiv [Preprint]*. (2024) *arXiv:2305.01397*. Available at: 10.48550/arXiv.2305.01397 (Accessed July 15, 2024).
- Danks D, London AJ. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (2017). p. 4691–4697.
- Lechner T, Ben-David S, Agarwal S, Ananthakrishnan N. Impossibility results for fair representations. *arXiv [Preprint]*. (2021) *arXiv:2107.03483*. Available at: 10.48550/arXiv.2107.03483 (Accessed February 08, 2024).
- Anderson E. *Black in White Space: The Enduring Impact of Color in Everyday Life*. Chicago, IL: The University of Chicago Press (2022).
- Gee GC, Hing A, Mohammed S, Tabor DC, Williams DR. Racism and the life course: taking time seriously. *Am J Public Health*. (2019) 109:S43–7. doi: 10.2105/AJPH.2018.304766
- Williams DR. Stress and the mental health of populations of color: advancing our understanding of race-related stressors. *J Health Soc Behav*. (2018) 59:466–85. doi: 10.1177/0022146518814251
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. (2010) 167:748–51. doi: 10.1176/appi.ajp.2010.09091379
- Wakefield JC. The concept of mental disorder: on the boundary between biological facts and social values. *Am Psychol*. (1992) 47:373–88. doi: 10.1037/0003-066x.47.3.373
- Kapur S. Reducing racial bias in AI models for clinical use requires a top-down intervention. *Nat Mach Intell*. (2021) 3:460. doi: 10.1038/s42256-021-00362-7
- Desai M. *Travel and Movement in Clinical Psychology: The World Outside the Clinic*. London: Palgrave Macmillan (2018).
- Metzl J. *The Protest Psychosis: How Schizophrenia Became a Black Disease*. Boston, MA: Beacon Press (2009).
- Black Parker C, McCall WV, Spearman-McCarthy EV, Rosenquist P, Cortese N. Clinicians' racial bias contributing to disparities in electroconvulsive therapy for patients from racial-ethnic minority groups. *Psychiatr Serv*. (2021) 72:684–90. doi: 10.1176/appi.ps.202000142
- Fields C, Rosenblatt M, Aina J, Thind JK, Harper A, Bellamy C, et al. Racialized heterospecificity in connectome-based predictions of brain-behavior relationships. *PsyArXiv [Preprint]*. (2024). Available at: 10.31234/osf.io/yqehw (Accessed October 05, 2024).
- Forde AT, Crookes DM, Suglia SF, Demmer RT. The weathering hypothesis as an explanation for racial disparities in health: a systematic review. *Ann Epidemiol*. (2019) 33:1–18.e3. doi: 10.1016/j.annepidem.2019.02.011
- Brody GH, Lei M, Chae DH, Yu T, Kogan SM, Beach SRH. Perceived discrimination among African American adolescents and allostatic load: a longitudinal analysis with buffering effects. *Child Dev*. (2014) 85:989–1002. doi: 10.1111/cdev.12213
- Lugo-Candelas C, Polanco-Roman L, Duarte CS. Intergenerational effects of racism: can psychiatry and psychology make a difference for future generations? *JAMA Psychiatry*. (2021) 78:1065–6. doi: 10.1001/jamapsychiatry.2021.1852
- Crenshaw KW. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics (1989). Available at: https://scholarship.law.columbia.edu/faculty_scholarship/3007 (Accessed March 01, 2024).
- Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. *Lancet*. (2017) 389:1453–63. doi: 10.1016/S0140-6736(17)30569-X
- 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
- Cogburn CD, Roberts SK, Ransome Y, Addy N, Hansen H, Jordan A. The impact of racism on black American mental health. *Lancet Psychiatry*. (2024) 11:56–64. doi: 10.1016/S2215-0366(23)00361-9
- Maness SB, Merrell L, Thompson EL, Griner SB, Kline N, Wheldon C. Social determinants of health and health disparities: COVID-19 exposures and mortality among African American people in the United States. *Public Health Rep*. (2021) 136:18–22. doi: 10.1177/0033354920969169
- Basu A. Use of race in clinical algorithms. *Sci Adv*. (2023) 9:eadd2704. doi: 10.1126/sciadv.add2704
- Sacks DA, Incerpi MH. Of aspirin, preeclampsia, and racism. *N Engl J Med*. (2024) 390:968–9. doi: 10.1056/NEJMp2311019
- Hanna A, Denton E, Smart A, Smith-Loud J. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (2020). p. 501–12
- Vyas DA, Eisenstein LG, Jones DS. Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *N Engl J Med*. (2020) 383:874–82. doi: 10.1056/NEJMs2004740
- Black C, Temple S, Acquaye A, Fields C, Konopasky A. Sociohistorical justice: a corrective framework to mend the modern harms of medical history. *Lancet Reg Health Am*. (2024) 38:100874. doi: 10.1016/j.lana.2024.100874
- Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, 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) 13:e007988. doi: 10.1161/CIRCEP.119.007988
- Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. (2018) 2:158–64. doi: 10.1038/s41551-018-0195-0
- Breland-Noble A, Streets FJ, Jordan A. Community-based participatory research with black people and black scientists: the power and the promise. *Lancet Psychiatry*. (2024) 11:75–80. doi: 10.1016/S2215-0366(23)00338-3
- Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. (2019) 25:1337–40. doi: 10.1038/s41591-019-0548-6
- Kleinberg G, Diaz MJ, Batchu S, Lucke-Wold B. Racial underrepresentation in dermatological datasets leads to biased machine learning models and inequitable healthcare. *J Biomed Res*. (2022) 3:42–47. doi: 10.46439/biomedres.3.025
- Burlew AK, Weekes JC, Montgomery L, Feaster DJ, Robbins MS, Rosa CL, et al. Conducting research with racial/ethnic minorities: methodological lessons from the NIDA clinical trials network. *Am J Drug Alcohol Abuse*. (2011) 37:324–32. doi: 10.3109/00952990.2011.596973
- Ricard JA, Parker TC, Dhamala E, Kwasa J, Allsop A, Holmes AJ. Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data. *Nat Neurosci*. (2023) 26:4–11. doi: 10.1038/s41593-022-01218-y
- Alegria M, Chatterji P, Wells K, Cao Z, Chen C-n, Takeuchi D, et al. Disparity in depression treatment among racial and ethnic minority populations in the United States. *Psychiatr Serv*. (2008) 59:1264–72. doi: 10.1176/ps.2008.59.11.1264

43. Murray P. Mental Health Disparities: African Americans (2017). Available at: <https://www.psychiatry.org/psychiatrists/diversity/education/mental-health-facts> (Accessed March 06, 2024).
44. Denckla CA, Cicchetti D, Kubzansky LD, Seedat S, Teicher MH, Williams DR, et al. Psychological resilience: an update on definitions, a critical appraisal, and research recommendations. *Eur J Psychotraumatol*. (2020) 11:1822064. doi: 10.1080/20008198.2020.1822064
45. Baek EJ, Jung H-U, Chung JY, Jung HI, Kwon SY, Lim JE, et al. The effect of heteroscedasticity on the prediction efficiency of genome-wide polygenic score for body mass index. *Front Genet*. (2022) 13:1025568. doi: 10.3389/fgene.2022.1025568
46. Gamage J, Weerahandi S. Size performance of some tests in one-way anova. *Commun Stat Simul Comput*. (1998) 27:625–40. doi: 10.1080/03610919808813500
47. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health*. (2002) 23:151–69. doi: 10.1146/annurev.publhealth.23.100901.140546
48. Yang K, Tu J, Chen T. Homoscedasticity: an overlooked critical assumption for linear regression. *Gen Psychiatr*. (2019) 32:e100148. doi: 10.1136/gpsych-2019-100148
49. Zimmerman DW. A note on preliminary tests of equality of variances. *Br J Math Stat Psychol*. (2004) 57:173–81. doi: 10.1348/000711004849222
50. Chekroud AM, Hawrilenko M, Loh H, Bondar J, Gueorguieva R, Hasan A, et al. Illusory generalizability of clinical prediction models. *Science*. (2024) 383:164–7. doi: 10.1126/science.adg8538
51. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med*. (2018) 178:1544–7. doi: 10.1001/jamainternmed.2018.3763
52. Breslau J, Kendler KS, Su M, Gaxiola-Aguilar S, Kessler RC. Lifetime risk and persistence of psychiatric disorders across ethnic groups in the United States. *Psychol Med*. (2005) 35:317–27. doi: 10.1017/s0033291704003514
53. Breslau J, Aguilar-gaxiola S, Kendler KS, Su M, Williams D, Kessler RC. Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. *Psychol Med*. (2006) 36:57–68. doi: 10.1017/S0033291705006161
54. Jackson JS, Torres M, Caldwell CH, Neighbors HW, Nesse RM, Taylor RJ, et al. The national survey of American life: a study of racial, ethnic and cultural influences on mental disorders and mental health. *Int J Methods Psychiatr Res*. (2004) 13:196–207. doi: 10.1002/mpr.177
55. Williams DR, González HM, Neighbors H, Nesse R, Abelson JM, Sweetman J, et al. Prevalence and distribution of major depressive disorder in African Americans, Caribbean blacks, and non-hispanic whites: results from the national survey of American life. *Arch Gen Psychiatry*. (2007) 64:305–15. doi: 10.1001/archpsyc.64.3.305
56. Barnes DM, Bates LM. Do racial patterns in psychological distress shed light on the Black-White depression paradox? A systematic review. *Soc Psychiatry Psychiatr Epidemiol*. (2017) 52:913–28. doi: 10.1007/s00127-017-1394-9
57. Jegarl AM, Jegede O, Isom J, Ciarleglio N, Black C. Psychotic misdiagnosis of racially minoritized patients: a case-based ethics, equity, and educational exploration. *Harv Rev Psychiatry*. (2023) 31:28–36. doi: 10.1097/HRP.0000000000000353
58. Black C, Calhoun A. How biased and carceral responses to persons with mental illness in acute medical care settings constitute iatrogenic harms. *AMA J Ethics*. (2022) 24:E781–787. doi: 10.1001/amajethics.2022.781
59. Chae DH, Nuru-Jeter AM, Adler NE, Brody GH, Lin J, Blackburn EH, et al. Discrimination, racial bias, and telomere length in African-American men. *Am J Prev Med*. (2014) 46:103–11. doi: 10.1016/j.amepre.2013.10.020
60. Mays VM, Cochran SD, Barnes NW. Race, race-based discrimination, and health outcomes among African Americans. *Annu Rev Psychol*. (2007) 58:201–25. doi: 10.1146/annurev.psych.57.102904.190212
61. Paradies Y, Ben J, Denson N, Elias A, Priest N, Pieterse A, et al. Racism as a determinant of health: a systematic review and meta-analysis. *PLoS One*. (2015) 10:e0138511. doi: 10.1371/journal.pone.0138511
62. Brown TT, Partanen J, Chuong L, Villaverde V, Chantal Griffin A, Mendelson A. Discrimination hurts: the effect of discrimination on the development of chronic pain. *Soc Sci Med*. (2018) 204:1–8. doi: 10.1016/j.socscimed.2018.03.015
63. Lewis TT, Everson-Rose SA, Powell LH, Matthews KA, Brown C, Karavolos K, et al. Chronic exposure to everyday discrimination and coronary artery calcification in African-American women: the SWAN heart study. *Psychosom Med*. (2006) 68:362–8. doi: 10.1097/01.psy.0000221360.94700.16
64. Slopen N, Lewis TT, Williams DR. Discrimination and sleep: a systematic review. *Sleep Med*. (2016) 18:88–95. doi: 10.1016/j.sleep.2015.01.012
65. Anderson E. *A Place on the Corner*. Chicago, IL: University of Chicago Press (1978).
66. Anderson E. *Streetwise: Race, Class, and Change in an Urban Community*. Chicago, IL: University of Chicago Press (1990).
67. Anderson E. *Code of the Street: Decency, Violence, and the Moral Life of the Inner City*. 1st edn. New York, NY: W.W. Norton (1999).
68. Anderson E. *The Cosmopolitan Canopy: Race and Civility in Everyday Life*. 1st edn. New York, NY: W.W. Norton & Co (2011).
69. Geronimus AT. The weathering hypothesis and the health of African-American women and infants: evidence and speculations. *Ethn Dis*. (1992) 2:207–21.
70. Manton K, Stallard E. National research council (US) committee on population vol. 3. In: Soldo BJ, Martin LG, editors. *Racial and Ethnic Differences in the Health of Older Americans*. Washington, DC: National Academies Press (1997). p. 1–300.
71. Caraballo C, Herrin J, Mahajan S, Massey D, Lu Y, Ndumele CD, et al. Temporal trends in racial and ethnic disparities in multimorbidity prevalence in the United States, 1999–2018. *Am J Med*. (2022) 135:1083–1092.e14. doi: 10.1016/j.amjmed.2022.04.010
72. Chae DH, Epel ES, Nuru-Jeter AM, Lincoln KD, Taylor RJ, Lin J, et al. Discrimination, mental health, and leukocyte telomere length among African American men. *Psychoneuroendocrinology*. (2016) 63:10–6. doi: 10.1016/j.psychenue.2015.09.001
73. Caraballo C, Massey DS, Ndumele CD, Haywood T, Kaleem S, King T, et al. Excess mortality and years of potential life lost among the black population in the US, 1999–2020. *JAMA*. (2023) 329:1662–70. doi: 10.1001/jama.2023.7022
74. Simons RL, Lei M-K, Beach SRH, Barr AB, Simons LG, Gibbons FX, et al. Discrimination, segregation, and chronic inflammation: testing the weathering explanation for the poor health of black Americans. *Dev Psychol*. (2018) 54:1993–2006. doi: 10.1037/dev0000511
75. Brody GH, Yu T, Chen E, Miller GE, Kogan SM, Beach SRH. Is resilience only skin deep?: rural African Americans' socioeconomic status-related risk and competence in preadolescence and psychological adjustment and allostatic load at age 19. *Psychol Sci*. (2013) 24:1285–93. doi: 10.1177/0956797612471954
76. Assari S. Health disparities due to diminished return among black Americans: public policy solutions. *Soc Issues Policy Rev*. (2018) 12:112–45. doi: 10.1111/sipr.12042
77. Assari S. Combined racial and gender differences in the long-term predictive role of education on depressive symptoms and chronic medical conditions. *J Racial Ethn Health Disparities*. (2017) 4:385–96. doi: 10.1007/s40615-016-0239-7
78. Assari S. Blacks' diminished return of education attainment on subjective health: mediating effect of income. *Brain Sci*. (2018) 8:176. doi: 10.3390/brainsci8090176
79. Agyemang C. Negro, black, black African, African Caribbean, African American or what? Labelling African origin populations in the health arena in the 21st century. *J Epidemiol Community Health*. (2005) 59:1014–8. doi: 10.1136/jech.2005.035964
80. Sternthal MJ, Slopen N, Williams DR. Racial disparities in health: how much does stress really matter? *Du Bois Rev*. (2011) 8:95–113. doi: 10.1017/S1742058X11000087
81. Pascoe EA, Smart Richman L. Perceived discrimination and health: a meta-analytic review. *Psychol Bull*. (2009) 135:531–54. doi: 10.1037/a0016059
82. Sellers RM, Caldwell CH, Schmeelk-Cone KH, Zimmerman MA. Racial identity, racial discrimination, perceived stress, and psychological distress among African American young adults. *J Health Soc Behav*. (2003) 44:302–17. doi: 10.2307/1519781
83. Sellers RM, Rowley SAJ, Chavous TM, Shelton JN, Smith MA. Multidimensional inventory of black identity: a preliminary investigation of reliability and construct validity. *J Pers Soc Psychol*. (1997) 73:805–15. doi: 10.1037/0022-3514.73.4.805
84. Joseph-Salisbury R. Institutionalised whiteness, racial microaggressions and black bodies out of place in higher education. *Whiteness Educ*. (2019) 4:1–17. doi: 10.1080/23793406.2019.1620629
85. Mosher CJ, Mahon-Haft T. Race, crime and criminal justice in Canada. In: Kalunta-Crumpton A, editor. *Race, Crime and Criminal Justice* London: Palgrave Macmillan (2010). p. 242–69. doi: 10.1057/9780230283954_11
86. Henry F, James CE, Tator C. *Racism in the Canadian University*. Toronto, ON: University of Toronto Press (2009). p. 128–59.
87. Mouzon DM, McLean JS. Internalized racism and mental health among African-Americans, US-born Caribbean blacks, and foreign-born Caribbean blacks. *Ethn Health*. (2017) 22:36–48. doi: 10.1080/13557858.2016.1196652
88. Farmer PE, Nizeye B, Stulac S, Keshavjee S. Structural violence and clinical medicine. *PLoS Med*. (2006) 3:e449. doi: 10.1371/journal.pmed.0030449
89. Aryal S, Blankenship JM, Bachman SL, Hwang S, Zhai Y, Richards JC, et al. Patient-centricity in digital measure development: co-evolution of best practice and regulatory guidance. *NPJ Digit Med*. (2024) 7:128. doi: 10.1038/s41746-024-01110-y
90. Israel BA. *Methods for Community-Based Participatory Research for Health*. 2nd edn. Hoboken, NJ: Jossey-Bass (2013).
91. Wallerstein N, Duran B. Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. *Am J Public Health*. (2010) 100(Suppl 1):S40–46. doi: 10.2105/AJPH.2009.184036
92. Umaña-Taylor AJ, Quintana SM, Lee RM, Cross WE, Rivas-Drake D, Schwartz SJ, et al. Ethnic and racial identity during adolescence and into young adulthood: an integrated conceptualization. *Child Dev*. (2014) 85:21–39. doi: 10.1111/cdev.12196

93. Almeida J, Molnar BE, Kawachi I, Subramanian SV. Ethnicity and nativity status as determinants of perceived social support: testing the concept of familism. *Soc Sci Med.* (2009) 68:1852–8. doi: 10.1016/j.socscimed.2009.02.029
94. Markides KS, Coreil J. The health of Hispanics in the southwestern United States: an epidemiologic paradox. *Public Health Rep.* (1986) 101:253–65.
95. Ogbu JU, Simons HD. Voluntary and involuntary minorities: a cultural-ecological theory of school performance with some implications for education. *Anthropol Educ Q.* (1998) 29:155–88. doi: 10.1525/aeq.1998.29.2.155
96. Jordan A, Costa M, Nich C, Swarbrick M, Babuscio T, Wyatt J, et al. Breaking through social determinants of health: results from a feasibility study of Imani breakthrough, a community developed substance use intervention for Black and Latinx people. *J Subst Use Addict Treat.* (2023) 153:209057. doi: 10.1016/j.jsat.2023.209057
97. Stahler GJ, Mennis J, DuCette JP. Residential and outpatient treatment completion for substance use disorders in the U.S.: moderation analysis by demographics and drug of choice. *Addict Behav.* (2016) 58:129–35. doi: 10.1016/j.addbeh.2016.02.030
98. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep.* (2014) 129(Suppl 2):19–31. doi: 10.1177/00333549141291S206
99. Fiscella K, Sanders MR. Racial and ethnic disparities in the quality of health care. *Annu Rev Public Health.* (2016) 37:375–94. doi: 10.1146/annurev-publhealth-032315-021439
100. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the public health disparities geocoding project. *Am J Public Health.* (2005) 95:312–23. doi: 10.2105/AJPH.2003.032482
101. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* (2002) 16:321–57. doi: 10.1613/jair.953
102. Friedler SA, Scheidegger C, Venkatasubramanian S, Choudhary S, Hamilton EP, Roth D. *Proceedings of the Conference on Fairness, Accountability, and Transparency* (2019). p. 329–38
103. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med.* (2018) 378:981–3. doi: 10.1056/NEJMp1714229
104. Tobin MJ. Fiftieth anniversary of uncovering the tuskegee syphilis study: the story and timeless lessons. *Am J Respir Crit Care Med.* (2022) 205:1145–58. doi: 10.1164/rccm.202201-0136SO
105. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* (2019) 1:206–15. doi: 10.1038/s42256-019-0048-x
106. Parthasarathy S, Katzman J. Bringing communities in, achieving AI for all. *Issues Sci Technol.* (2024) 40:41–4. doi: 10.58875/SLRG2529
107. Benjamin R. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity (2020).
108. Delgado C, Baweja M, Burrows NR, Crews DC, Eneanya ND, Gadegebeku CA, et al. Reassessing the inclusion of race in diagnosing kidney diseases: an interim report from the NKF-ASN task force. *Am J Kidney Dis.* (2021) 78:103–15. doi: 10.1053/j.ajkd.2021.03.008
109. Diao JA, Powe NR, Manrai AK. Race-free equations for eGFR: comparing effects on CKD classification. *J Am Soc Nephrol.* (2021) 32:1868–70. doi: 10.1681/ASN.2021020224
110. Beaudoin JR, Curran J, Alexander GC. Impact of race on classification of atherosclerotic risk using a national cardiovascular risk prediction tool. *AJPM Focus.* (2024) 3:100200. doi: 10.1016/j.focus.2024.100200
111. Muntner P, Colantonio LD, Cushman M, Goff DC, Howard G, Howard VJ, et al. Validation of the atherosclerotic cardiovascular disease pooled cohort risk equations. *JAMA.* (2014) 311:1406–15. doi: 10.1001/jama.2014.2630
112. Admon LK, Winkelmann TNA, Zivin K, Terplan M, Mhyre JM, Dalton VK. Racial and ethnic disparities in the incidence of severe maternal morbidity in the United States, 2012–2015. *Obstet Gynecol.* (2018) 132:1158–66. doi: 10.1097/AOG.0000000000002937
113. Kawakita T, Mokhtari N, Huang JC, Landy HJ. Evaluation of risk-assessment tools for severe postpartum hemorrhage in women undergoing cesarean delivery. *Obstet Gynecol.* (2019) 134:1308–16. doi: 10.1097/AOG.0000000000003574
114. McCarthy AM, Guan Z, Welch M, Griffin ME, Sippo DA, Deng Z, et al. Performance of breast cancer risk-assessment models in a large mammography cohort. *J Natl Cancer Inst.* (2020) 112:489–97. doi: 10.1093/jnci/djz177
115. Lawrence JA, Kawachi I, White K, Bassett MT, Priest N, Masunga JG, et al. A systematic review and meta-analysis of the everyday discrimination scale and biomarker outcomes. *Psychoneuroendocrinology.* (2022) 142:105772. doi: 10.1016/j.psyneuen.2022.105772
116. Krieger N. *Small Books, Big Ideas in Population Health 4 1 Online Resource*. New York, NY: Oxford University Press (2021).
117. Trent M, Dooley DG, Dougé J, Cavanaugh RM, Lacroix AE, Fanburg J, et al. The impact of racism on child and adolescent health. *Pediatrics.* (2019) 144(2):e20191765. doi: 10.1542/peds.2019-1765
118. Jones CP. Levels of racism: a theoretic framework and a gardener's tale. *Am J Public Health.* (2000) 90:1212–5. doi: 10.2105/ajph.90.8.1212
119. Reeves A, Elliott MR, Lewis TT, Karvonen-Gutierrez CA, Herman WH, Harlow SD. Study selection bias and racial or ethnic disparities in estimated age at onset of cardiometabolic disease among midlife women in the US. *JAMA Netw Open.* (2022) 5:e2240665. doi: 10.1001/jamanetworkopen.2022.40665
120. Javed Z, Haisum Maqsood M, Yahya T, Amin Z, Acquah I, Valero-Elizondo J, et al. Race, racism, and cardiovascular health: applying a social determinants of health framework to racial/ethnic disparities in cardiovascular disease. *Circ Cardiovasc Qual Outcomes.* (2022) 15:e007917. doi: 10.1161/CIRCOUTCOMES.121.007917
121. Chae DH, Drenkard CM, Lewis TT, Lim SS. Discrimination and cumulative disease damage among African American women with systemic lupus erythematosus. *Am J Public Health.* (2015) 105:2099–107. doi: 10.2105/AJPH.2015.302727
122. Assari S, Mincy R. Racism may interrupt age-related brain growth of African American children in the United States. *J Pediatr Child Health Care.* (2021) 6:1047. doi: 10.26420/jpediatrchildhealthcare.2021.1047
123. Assari S, Bazargan M. Minorities' diminished returns of educational attainment on hospitalization risk: national health interview survey (NHIS). *Hosp Pract Res.* (2019) 4:86–91. doi: 10.15171/HPR.2019.17
124. Assari S. Parental education better helps white than black families escape poverty: national survey of children's health. *Economies.* (2018) 6:30. doi: 10.3390/economies6020030
125. Bradby H. Institutional racism in mental health services: the consequences of compromised conceptualisation. *Sociol Res Online.* (2010) 15:86–96. doi: 10.5153/sro.2197
126. Ghezze FT, Zungu Z, John A, Abdinasir K, Bhui K, Adebisi A, et al. The association between experiences of racism and mental health on children and young people in the UK: rapid scoping review. *BJPsych Open.* (2025) 11:e26. doi: 10.1192/bjo.2024.836
127. Medeiros S, Coelho R, Millett C, Saraceni V, Coeli CM, Trajman A, et al. Racial inequalities in mental healthcare use and mortality: a cross-sectional analysis of 1.2 million low-income individuals in Rio de Janeiro, Brazil 2010–2016. *BMJ Glob Health.* (2023) 8:e013327. doi: 10.1136/bmjgh-2023-013327