

Clinical algorithms, racism, and “fairness” in healthcare: A case of bounded justice

Sarah El-Azab¹ and Paige Nong²

Big Data & Society
July–December: 1–13
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517231213820
journals.sagepub.com/home/bds



Abstract

To date, attempts to address racially discriminatory clinical algorithms have largely focused on fairness and the development of models that “do no harm.” While the push for fairness is rooted in a desire to avoid or ameliorate health disparities, it generally neglects the role of racism in shaping health outcomes and does little to repair harm to patients. These limitations necessitate reconceptualizing how clinical algorithms should be designed and employed in pursuit of racial justice and health equity. A useful lens for this work is bounded justice, a concept and research analytic proposed by Melissa Creary to guide multidisciplinary health equity interventions. We describe how bounded justice offers a lens for (1) articulating the deep injustices embedded in the datasets, methodologies, and sociotechnical infrastructure underlying design and implementation of clinical algorithms and (2) envisioning how these algorithms can be redesigned to contribute to larger efforts that not only address current inequities, but to redress the historical mistreatment of communities of color by biomedical institutions. Thus, the aim of this article is two-fold. First, we apply the bounded justice analytic to fairness and clinical algorithms by describing structural constraints on health equity efforts such as medical device regulatory frameworks, race-based medicine, and racism in data. We then reimagine how clinical algorithms could function as a reparative technology to support justice and empower patients in the healthcare system.

Keywords

Bounded justice, racism, healthcare, clinical algorithms, algorithmic reparation, health equity

This article is a part of special theme on Reparative Approaches to Algorithmic Justice. To see a full list of all articles in this special theme, please click here: <https://journals.sagepub.com/page/bds/collections/reparativeapproachestoalgorithmicjustice>

Introduction

Clinical algorithms are structured, data-driven tools used to inform and facilitate decision making in healthcare (Dennstädt et al., 2021). Their widespread adoption in healthcare is driven by the desire to improve clinician decision making, reduce medical errors, optimize clinical workflows, and improve patient outcomes (Grant et al., 2018; Kishore et al., 2022; Luo et al., 2020; Rubashkin, 2022; Topol, 2019). Common applications of clinical algorithms include the estimation of the risk of a particular health outcome, assessment of therapeutic need, and allocation of critical resources (Steyerberg and Vergouwe, 2014; Van Calster et al., 2019). For example, throughout the COVID-19 pandemic, algorithms were developed to triage symptoms and forecast likelihood of outcomes such as hospitalization and death (Jehi et al., 2020; Rozenbaum

et al., 2021; Wynants et al., 2020). Clinical algorithms have traditionally been derived from clinical guidelines and developed using statistical methods such as regression. However, with the advent of precision medicine and the concurrent growth of both data volume and computational power, artificial intelligence (AI) and machine learning (ML) methods, such as support vector machines (SVM),

¹Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, USA

²Division of Health Policy and Management, University of Minnesota, Minneapolis, MN, USA

Corresponding author:

Sarah El-Azab, Department of Health Management and Policy, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA.
Email: saelazab@umich.edu



random forest, and deep learning, are increasingly used to develop clinical algorithms (Luo et al., 2020). Clinical algorithms have grown more technologically advanced over time, evolving from paper-based charts and risk scores to web-based calculators, proprietary software, and clinical decision support systems embedded within electronic health records (EHRs) (Cresswell et al., 2020; Steyerberg and Vergouwe, 2014).

Despite the perceived potential of clinical algorithms to positively transform healthcare, a growing body of evidence has established that they can be racially discriminatory, encoding biases that exacerbate pre-existing racial and ethnic health inequities and harming patients of color. Clinical algorithms often unfairly assign risk to Black, Hispanic, and other patients of color and subsequently direct them away from necessary clinical resources (Owens and Walker, 2020). In some cases this bias is an intentional design choice, as in the case of clinical algorithms that are race “corrected” to differentially assign risk of a particular diagnostic or prognostic outcome along racial and ethnic lines (Roberts, 2021). In these cases, the risk calculation is explicitly adjusted by either increasing or decreasing risk of a certain outcome based solely on the perceived racial or ethnic identity of the patient. For example, the American Heart Association’s heart failure calculator is used to help cardiologists determine when they should initiate medical therapy for patients with acute heart failure (Vyas et al., 2020). The algorithm predicts risk of death and arbitrarily assigns lower risk to Black patients, thereby making it more difficult for Black patients to receive lifesaving treatment than non-Black patients.

In other cases, such as the development of AI and ML-based prediction models, the data sources and design-choices reflect the mechanisms through which racism and intersecting systems of oppression shape health inequities, both at the point of care where the algorithm is implemented and outside of the healthcare system (Huang et al., 2022; Owens and Walker, 2020). This is exemplified by Medical Information Mart for Intensive Care (MIMIC), the largest publicly available clinical database and one of the most common data sources used to develop AI and ML-based clinical algorithms (Johnson et al., 2023). A study assessing the performance of deep learning models on this dataset found that not only are Black and Hispanic patients less likely than White patients to receive intensive care, but that the duration of their treatment is also shorter (Meng et al., 2022). Thus, the models built using this data reflect racial inequities in access to intensive care.

In response to the growing evidence of how these tools encode racism, algorithmic fairness has emerged as a set of frameworks for quantifying racial bias in clinical algorithms and optimizing model performance to ensure equality in outcomes. Algorithmic fairness has become

increasingly popular as a response to racist algorithmic output. While the push for algorithmic fairness in healthcare is motivated by the ideals of health equity and justice, debiasing a racially discriminatory algorithm does little to disrupt the mechanisms that connect racism to differences in health outcomes. To truly advance racial health equity, we must employ reparative approaches that remedy harms stemming both from clinical algorithms and longstanding disenfranchisement of communities of color by biomedical institutions (Davis et al., 2021). Whereas algorithmic fairness seeks to eliminate bias to ensure that certain groups are not favored over others by the algorithm itself, algorithmic reparation “assumes and leverages bias” to guide reallocation and redistribution of resources to the margins (Davis et al., 2021). Pursuing algorithmic reparation in healthcare therefore requires understanding how entrenched injustices constrain the desired impact of algorithms. To that end, we apply bounded justice, a conceptual framework and analytical tool proposed by Melissa Creary to guide the design of multidisciplinary health equity interventions (Creary, 2021). Bounded justice identifies how historical embodiments and consequences of racialized marginalization persist despite well-intentioned health equity efforts precisely because of how foundational racism and other intersecting structures of oppression are to society.

In this article, we begin by arguing that algorithmic fairness frameworks as applied to clinical algorithms are a case of bounded justice because they do not adequately engage with how structural racism shapes health outcomes. We then apply bounded justice as a conceptual framework and examine how race-based medicine, the values embedded in the healthcare system, data work practices, and insufficient regulatory structures bound efforts to use clinical algorithms and fairness frameworks in pursuit of health equity. Finally, we apply bounded justice as an analytical tool to guide a practice of algorithmic reparation to move beyond the limitations of fairness toward justice.

Algorithmic fairness and bounded justice

Algorithmic fairness in healthcare is a nascent field of study focused on the creation of nondiscriminatory algorithmic systems that are optimized to “ensure equality in patient outcomes, performance, and resource allocation” (Rajkomar et al., 2018). These efforts are either computational or procedural in nature. Computational approaches work toward fairness by mitigating bias throughout the model development pipeline in three stages: (1) preprocessing to remove biases originating from data, (2) in-processing to explicitly train models to be nondiscriminatory by implementing fairness constraints, and (3) postprocessing to audit and correct unfair model outputs (Chen et al., 2021; Xu et al., 2022).

Even so, there is debate over how to operationalize algorithmic fairness. In a scoping review of approaches to evaluating and mitigating racial bias in clinical algorithms,

Huang et al. (2022) found that the application of computational fairness metrics varied, with studies implementing some measure of group fairness such as equal opportunity difference, accuracy, or disparate impact. For example, the goal of one study was to minimize bias in inpatient mortality predictions between white and non-White patient groups. To do so, the authors only included patients with complete race and ethnicity data, preprocessed their training data to account for systemic inequities reflected in the data, and used equal opportunity difference as their fairness metric (Allen et al., 2020). Importantly, while Huang et al. found that most studies were successful at increasing the fairness of the clinical algorithm, they were only able to do so with regard to a single fairness metric. This highlights a key limitation of computational algorithmic fairness known as the “impossibility of fairness”—there are many definitions and mathematical formulations of fairness criteria and there is no way to satisfy them all at once (Green, 2022; Pfohl et al., 2021). As such, the reach of the intended justice is contingent on selecting the appropriate fairness metric and balancing trade-offs between model performance and equal outcomes.

On the other hand, procedural approaches to fairness are concerned with policies and frameworks for developing and implementing algorithms. They are used to reshape clinical workflows, implement governance structures and reporting guidelines, and create normative standards for the development and deployment of algorithms (Reddy et al., 2020; Sikstrom et al., 2022). This approach moves beyond algorithm development to promote transparency and ameliorate bias. Examples include checklists and frameworks for algorithmic transparency (interpretability, explainability, accountability), inclusion, and impartiality (Sikstrom et al., 2022). Although it expands beyond computational fairness, procedural fairness remains limited in focus to the algorithm within its organizational context rather than the structural racism and inequities that shape its impact on patients. At minimum, computational and procedural fairness approaches seek to ensure that clinical algorithms are neutral and “do no harm” (McCradden et al., 2020). However, fair clinical algorithms are increasingly interpreted as tools for achieving health equity through distributive justice, or the equal allocation of resources across protected groups, such as race and ethnicity (Pfohl et al., 2021; Rajkomar et al., 2018).

Even if a clinical algorithm is optimized to provide unbiased recommendations to clinicians, it is unlikely that it will perform as intended or have the desired effect. This is because algorithmic fairness is an example of bounded justice. Bounded justice is a concept, framework, and analytical tool created by Creary (2021) to highlight the fundamental limitations of well-intentioned health equity efforts. Per Creary, “bounded justice suggests that it is impossible to attend to fairness, entitlement, and equity when the basic social and physical infrastructures underlying them

have been eroded by racism and other historically entrenched -isms.” Bounded justice therefore calls attention to how applications of distributive justice in pursuit of health equity are constrained and undermined by the cumulative effects of historical and contemporary racialized marginalization. Racism and intersecting forms of oppression operate as fundamental causes of health outcomes and racial disparities, structuring access to resources and shaping complex causal mechanisms that tie these resources to health outcomes (Phelan and Link, 2015). Put simply, “health outcomes are the products and expressions of unjust social, economic, and political institutions” (Creary, 2021). However, instead of acknowledging or engaging with these upstream systems and structures, health equity interventions frequently target downstream health outcomes in narrowly defined ways that fundamentally limit progress toward equity.

Algorithmic fairness frameworks exemplify this bounded downstream focus by overestimating the impact of the clinical algorithm and point-of-care decisions within which the algorithm operates (Mitchell et al., 2021). They do not substantively engage with key determinants of health that exist outside of that narrow decision space and how they constrain the benefit of fair algorithmic recommendations for patients of color (Cerdeña et al., 2020; McCradden et al., 2020). Put another way, fairness frameworks are bounded precisely because they do not grapple directly with the larger structural forces harming patients of color and preventing them from benefiting from equity-focused efforts. Regrettably, the designation of algorithms as “fair” or “neutral” obfuscates this reality and lends credence to their outputs, resulting in the “erroneous conflation of model performance with accrual of benefit” (Pfohl et al., 2021). In this way, algorithmic fairness in healthcare is a form of political and algorithmic idealism that does little to advance health equity (Creary, 2021; Davis et al., 2021; Hoffmann, 2019). Bounded justice posits that in order to achieve racial health equity through algorithmically mediated interventions, we must foreground how racism socially patterns the material conditions that differentially shape patient health outcomes well before they even reach the decision space within the healthcare system (Braveman et al., 2011; Phelan et al., 2010). This necessitates a shift from technodeterministic fairness frameworks to a practice of algorithmic reparation that decenters algorithms and reimagines their role as vehicles for structural redress (Davis et al., 2021).

Implementing algorithmic reparation in pursuit of health equity entails reforming biomedical institutions and the sociotechnical infrastructures embedded within them to promote both the retroactive redress and the proactive repair of the compounding impacts of systemic oppression (Creary, 2021; Davis et al., 2021). While bounded justice tells us that this work will inevitably be bounded to some extent, it also allows us to appreciate enduring

sociohistorical constraints and account for them in reparative algorithmic design processes. In the following sections we first apply bounded justice as a conceptual framework to understand how efforts at health equity are bounded within the healthcare system before clinical algorithms are even created. We then illustrate how bounded justice can be used as an analytical tool and praxis by examining current approaches to algorithmic reparation in healthcare.

Bounded justice as a conceptual framework: Algorithmic (un)fairness

Per Davis et al. (2021), the pursuit of algorithmic reparation and reform “requires unvarnished realism about the conditions under which any sociotechnical intervention will go into effect.” As a conceptual framework, bounded justice is woven together with concepts such as structural violence, embodiment, intersectionality, and social exclusion to expose these conditions. When applied to the healthcare system and the biomedical establishment, a bounded justice conceptual framework reveals how the intersecting power structures, values, and priorities of the healthcare system reflect those of the larger society. Through inequitable treatment and systematically unequal access to quality care, the healthcare system has a legacy of perpetuating racism and other inequities that bounds both algorithmic fairness and reparative efforts (Delaney et al., 2021; Dimick et al., 2013; Hoffman et al., 2016; Ryn et al., 2011).

Race-based medicine

Efforts to use clinical algorithms as a form of distributive justice are bounded by race-based medicine. Race-based medicine is the misuse of race within clinical practice guidelines, biomedical knowledge production, and decision making wherein individuals in different racial and ethnic groups are characterized as being innately distinct from one another (Braun, 2021; Braun et al., 2007; Duster, 2005). This is a deep-seated practice originating from a white supremacist agenda to reinforce a racial hierarchy that positions white people as biologically superior to Black people (Cerdeña et al., 2020; Duster, 2003; Yearby, 2021). Race-based medicine has persisted with recurrent reification of racial essentialism (Cerdeña et al., 2020; Lett et al., 2022; Roberts, 2011). For example, in the early 1990s and 2000s, overlapping government and biomedical priorities motivated the use of genomic technologies to address racial health disparities (Duster, 2003; Kahn, 2004: 200, 2006). These technologies promoted and legitimized the conflation of race with genetics (Duster, 2003: 114–115; Duster, 2005). To be clear, race and ethnicity are not biological or genetic attributes—they are multidimensional sociopolitical constructs that “precisely captures the social classification of people in a race-

conscious society” and reflect outcomes and mechanisms of racialization (Jones, 2000; Roth, 2016). However, while there is increased recognition that race and ethnicity are socially constructed, in practice they continue to be conflated with physiological and even behavioral, cultural, or other social differences in health. For example, in a content analysis of UpToDate, a point-of-care software that aggregates clinical practice guidelines, Cerdeña et al. (2022) found that race was consistently biologized or treated as a risk factor for behavioral difference across racial and ethnic groups.

Using bounded justice as a conceptual framework, we see that the enmeshment of race-based medicine into “fair” clinical algorithms parallels racialization of genetics. These algorithms incorporate race and ethnicity as fixed attributes into clinical decision making, decoupled from racism as an underlying fundamental cause of health inequities (Phelan and Link, 2015; Ukoha et al., 2022; Yearby, 2021). This perpetuates white supremacy and inequity within the healthcare system as opposed to ameliorating it. Liao and Carbonell refer to this as materialized oppression, wherein clinical algorithms “reflect past oppression, do the work of oppression in the present day, and carry oppression into the future” (Liao and Carbonell, 2022). For example, the modification of diet in renal disease (MDRD) and chronic kidney disease epidemiology collaboration (CKD-EPI) equations are race-corrected algorithms used to calculate estimated glomerular filtration rate (eGFR), a measure of kidney function used to determine kidney transplant eligibility. Both equations adjust eGFR by race to assign higher eGFR and perceived kidney function to Black patients, which in turn increases the threshold they must meet to qualify for a kidney transplant (Braun, 2021; Vyas et al., 2020). While this practice originated from debunked claims that there are race-based genetic differences in kidney function or that Black people have greater muscle mass and higher levels of creatinine (and therefore kidney function), it continues to deprive Black patients of critical, lifesaving treatment and decreases their life expectancy (Braun, 2021; Roberts, 2021). This is a materialization of antiblack racism and injustice via algorithms that becomes embodied by Black patients.

Despite recognition that incorporating race-based medicine into clinical tools and technologies is harmful, there is still reluctance to abolish this practice on the part of the biomedical establishment. For example, in response to a congressional inquiry into the widespread use of race corrections, the Society for Thoracic Surgeons (STS) stated that “elimination of race from all STS risk models at this time would be scientifically inaccurate and result in an intentional, unethical misrepresentation of facts to certain patient populations, most notably Black patients” (House Committee on Ways and Means, 2021). This statement is emblematic of bounded justice as it highlights the tension between the entrenchment of ideas of racialized

risk and the larger goal of health equity (Gordon, 2021; Madhusoodanan, 2021). In this way, race-based medicine further contributes to a circular reification of race and not racism as the risk factor for poor health, thereby upholding a white racial framing, or “the overarching worldview that encompasses important racial ideas, terms, images, emotion and interpretation, and lens by which white supremacy is perpetuated” (Feagin and Bennefield, 2014; Hardeman and Karbeah, 2020).

Values and decision making

The clinical problems chosen for algorithmic intervention, the data collected, outcome definitions, algorithm design, and postimplementation evaluations are shaped by social values and biases embedded within the healthcare system (Chen et al., 2021). In the United States, the dominant paradigm of care is value-based healthcare (VBHC), wherein payers and providers are financially incentivized to improve patient outcomes while controlling costs (i.e. maximizing “value”) (Abduljawad and Al-Assaf, 2011; Brown et al., 2003; Lin et al., 2021). Profit motive is therefore a salient value driving healthcare decision making and thus the design of clinical algorithms, often at the expense of patients (Johnson and Kane, 2010). For example, given that compensation is tied to performance on measures of quality of care, algorithms are often designed to target the quality measures themselves (e.g. reducing 30-day readmissions) as opposed to the underlying goal of improving patient outcomes (Cox et al., 2022; Li and Evans, 2022). Furthermore, the pursuit of profit in the U.S. healthcare system has contributed to the entrenchment of systemic barriers restricting patient access to care such as underinsurance, predatory debt collection, and high administrative costs for billing (Bai et al., 2021; Yearby et al., 2022). Clinical algorithms can perpetuate such inequities when their design and intended use center the financial priorities of health systems and insurers instead of serving historically marginalized communities and working toward justice.

Another constraining value of the U.S. healthcare system is the widely held belief that health is the responsibility of the individual as opposed to the collective (Ferryman and Pitcan, 2018; Johnson and Kane, 2010). This value is embedded into population health management, precision medicine, and patient-centered care paradigms that focus on downstream patient-level risk factors, as opposed to the upstream systems of oppression that generate and reinforce racial health inequities (Epstein et al., 2010; Ferryman and Pitcan, 2018; Lantz, 2019; Manzer and Bell, 2022). Individualization of risk in this manner places the locus of control and therefore the target of intervention on the patient, with implications for data collection and algorithmic design (Boyd et al., 2020; Ferryman and Pitcan, 2018; Lantz, 2019). Even efforts that seemingly focus upstream on social determinants of health are

frequently transformed into individualized strategies and interventions (Freij et al., 2019; Kasthurirathne et al., 2018; Lantz, 2019; Lantz et al., 2023). Instead of implementing policy change to address structures of racism, unaffordable housing, low wages, inaccessible nutrition, wealth inequality, and environmental risks, health systems collect data on patient-level social risk factors and social needs that are in turn integrated into and scope the design of individual level medical treatments (Lantz, 2019; Lantz et al., 2023). Social needs themselves are also targeted for intervention and health systems may attempt to facilitate patient referrals to a social worker or community organization to address immediate social needs, such as emergency access to food (Lans et al., 2022; Zhao et al., 2021). Alternatives where hospitals engage at the structural level have been conceptualized (Dave et al., 2021), but are not widely implemented.

While profit motive and individualization are just two examples of values shaping clinical decision making, they both exemplify how attempts to include patients “at the table” are fundamentally bounded because the healthcare system prioritizes the values associated with power rather than antiracism (Creary, 2021). While VBHC attempts to improve patient outcomes, the financial interests of hospitals and insurance companies are prioritized in practice over patient needs in the conceptualization, operationalization, and pursuit of quality. Similarly, the targeting of individual social needs reflects what Lantz describes as “the fallacy that societal problems having to do with health primarily need health care solutions” as opposed to policy change to address social determinants at a community-level (Lantz, 2019). In both cases, the power of biomedical institutions and private interests to guide decision making is maintained or amplified. As described by boyd (2023), “technology is consistently leveraged to codify the values that its makers or users wish to make rigid.” Even if a clinical algorithm is “fair,” it is bounded in its ability to ameliorate inequities because it embodies the values that uphold them (Ghassemi and Mohamed, 2022; Johnson and Kane, 2010). As Ford and Airhihenbuwa highlight in the Public Health Critical Race praxis, critical engagement with the values that underlie assumptions and knowledge is a key step in redesigning systems toward equity (Ford and Airhihenbuwa, 2010).

Data work

Clinical algorithms are developed using patient-level data aggregated from a variety of sources, including electronic health records, insurance claims, government health surveys, and social media (Chen et al., 2021; Mhasawade et al., 2021). Importantly, these datasets are products of human decision making that reflect efforts at aggregation and measurement within social contexts structured by historical medical racism and inequitable health policy

(Benjamin, 2019; Cruz, 2022). As such, the data underlying clinical algorithms are artifacts of biomedical institutions and sociotechnical infrastructures that systematically harm patients of color (Benjamin, 2019; Bonham et al., 2018). Algorithmic fairness approaches take care to remove this bias throughout the model development pipeline. However, these efforts are bounded because enactments of structural violence are obscured behind the layers of methodological decisions intended to neutralize them.

Muller and Strohmayr (2022) refer to this as the forgettance stack wherein each step in the pipeline of algorithmic design and development pushes “previous actions into the infrastructure, where the action itself and its consequences are easily forgotten.” The bottom layers of this stack encompass data work practices such as data capture, data cleaning, and data curation and are the practices that are most likely to be forgotten as fairness frameworks are applied (Bossen et al., 2019). By way of illustration, race and ethnicity data are critical attributes in algorithmic fairness methodologies that address racial bias, with great emphasis placed on ensuring that clinical data sets are representative of marginalized racial and ethnic groups (Hanna et al., 2020). However, there is extensive data work that occurs before race and ethnicity are used to ensure a “fair” algorithm. In the United States, standards used for collection of race and ethnicity data are embedded into technical infrastructure within health systems, such as EHRs (Polubriaginof et al., 2019). Race and ethnicity data are collected from patients by healthcare workers, using these standard categories as a guide (Cruz and Smith, 2021). Once stored within the EHR, this data may be exchanged and aggregated with data from other health systems (Cook et al., 2022). When this data is extracted for analysis, missing data points may be imputed, new categories may be created, and others may be dropped entirely. The dataset may even be resampled to artificially create a “representative” sample of data.

Despite how extensive this data work is, it is often forgotten. For example, studies examining the reporting and representativeness of demographic data used to develop and validate clinical algorithms have found that the most underreported data points are race and ethnicity (Bozkurt et al., 2020; Crowley et al., 2020). As such, there is no recourse for recovering the provenance of the data. It becomes impossible to examine the extent to which race and ethnicity data became divorced from the context of its production with every preprocessing decision made. While the resultant clinical algorithm may appear to perform fairly across racial and ethnic groups, the fact that these groups are abstractions of abstractions is lost. Essentially, while a claim can be made that the algorithm addressed an inequity, the underlying structural violence remains buried within the forgettance stack. This represents a form of structural gaslighting, a term coined by Nora Berenstein to describe how institutional entities “invoke

oppressive ideologies, disappear or obscure the actual causes and mechanisms of oppression, and conceptually sever acts of oppression from the structures that produce them” (Berenstein, 2020). By obscuring the sources of injustice and harm, algorithmic fairness is bounded.

Regulation

Biomedical institutions and the structures that regulate them reflect the historical and contemporary violence of racism (Yearby et al., 2022). Current regulation of algorithms in healthcare is no exception. The regulatory frameworks that do exist, while nascent, are severely limited in terms of their engagement with inequity (Ferryman, 2020). While multiple federal bodies have released reports and statements about valuing equity and fairness, these largely represent “the political idealism of equity-based policies” (Creary, 2021) whereby stated values themselves are bounded in their scope and policymakers fail to engage with the breadth and depth of the impacts of racism. As an illustration, a recently proposed rule (Section 1557 of the Affordable Care Act) from the Department of Health and Human Services (HHS) describes potential protection against algorithmic bias:

Proposed § 92.210 states that a covered entity must not discriminate against any individual on the basis of race, color, national origin, sex, age, or disability through the use of clinical algorithms in its decision-making. This is a new provision, and this topic has not been addressed in previous Section 1557 rulemaking. The Department believes it is critical to address this issue explicitly in this rulemaking given recent research demonstrating the prevalence of clinical algorithms that may result in discrimination. (HHS et al., 2022)

Notably, this proposed rule does not engage with the ways that racism and discrimination fundamentally shape all algorithms through the data used to create them, placing the onus on individual clinicians to evaluate complex algorithmic tools (Shachar and Gerke, 2023). It does not consider how algorithms could be designed specifically to ameliorate algorithmic racism that has already occurred. The proposed rule states that it “would put covered entities on notice that they cannot use discriminatory clinical algorithms and may need to make reasonable modifications in their use of the algorithms, unless doing so would cause a fundamental alteration to their health program or activity” (Goodman et al., 2023; HHS et al., 2022; Khazanchi et al., 2022). The limitations inherent in this proposed rule forestall organizational transformation and explicitly prevent meaningful change that could work toward justice.

Additional regulatory guidance generally falls under a fairness paradigm, whereby algorithms are evaluated by subpopulation performance or a variety of fairness

metrics. Some recommendations for using regulation to ensure fairness include reporting the representation of racial and ethnic minority patients in training datasets and analyzing algorithmic output by subpopulation (Ferryman, 2020). However, these data-centric regulatory frameworks fail to account for or consider how data reflects the erosion of basic infrastructures caused by racism and do not sufficiently ensure that inequitable implications are identified. As an example, the Food and Drug Administration (FDA) currently regulates some algorithms as medical devices, requiring performance studies with development and validation details (FDA, 2020). However, in an assessment of the FDA's database of approved algorithms and associated performance studies, researchers found that 97% of approved devices were only evaluated by retrospective studies instead of prospective studies that could fully characterize potential impacts on patients of color (Wu et al., 2021). They additionally found that 72% of device studies did not report whether the device was tested at multiple sites, an important method of assessing bias, and only 17% of device studies reported the performance of the device on demographic subgroups including racial and ethnic populations.

The inconsistent reporting of development and validation details for approved devices underscores the fact that, even if computational fairness approaches are implemented by developers, the FDA does not yet have the regulatory capacity to effectively evaluate clinical algorithms for racial bias or the potential for group harms (El-Sayed, 2021; Ferryman, 2020). Developers are not beholden to the FDA when they decide how to implement fairness considerations, such as selection of fairness metrics, data preprocessing, or operationalization of the attributes they use to assess bias (e.g. race and ethnicity). When regulatory bodies cannot engage with baseline evaluation of bias or fairness, this precludes deeper analysis of how racism becomes embedded within and perpetuated by clinical algorithms. This is especially concerning given the FDA's history of reifying race-based medicine in regulatory decisions, as in the case of the race-based heart medication BiDil (Kahn, 2008). In this way, existing regulatory structures are bounded and constrain the effectiveness of algorithmic fairness approaches as a means of advancing health equity.

Bounded justice as an analytical tool: Algorithmic reparation

There are myriad examples of racist harms perpetrated in the healthcare system, both algorithmically and otherwise, that urgently require reparation. Algorithmic reparations could potentially be designed to make progress on acknowledging these harms, working to repair them, and improving health for marginalized patients. However, a defining

characteristic of a reparations approach informed by bounded justice is the acknowledgement that technology itself is not a complete solution. Rather, it can be designed as an important aspect of larger interventions and programs explicitly designed to counteract and repair the harms of racism. Clinical algorithms designed for these goals with an awareness of technology's limitations are an important departure from the bounded fairness approaches described above that individualize structural harms and fundamentally limit transformative efforts. To support the design of these kinds of algorithms, bounded justice can operate as an analytical tool to (1) identify and name how racism operates in algorithms, (2) contextualize the depth and breadth of the effects of racism across all phases of algorithm design and use, (3) anticipate the ways racism will necessarily shape any technology even when it is designed for equity, and (4) quantify the success of interventions.

While bounded justice can inform critical design, deployment, and evaluation of clinical algorithms, it can also expand beyond algorithms to more comprehensive approaches to reparation in healthcare. Bounded justice as an analytical tool guides acknowledgment of the many ways that justice is forestalled by structural aspects of the healthcare system including its values, data sources, and continued reliance on race-based medicine as described above. Bounded justice does this in concert with principles of the public health critical race praxis in calling for an unlearning of disciplinary norms and critical engagement with the assumptions and values underlying the systems that perpetuate structural violence (Creary, 2021; Ford and Airhihenbuwa, 2010). The issues of individualism and social exclusion described above can be repaired or reimagined through bounded justice. For example, algorithmic decision making that accounts for structural barriers and social determinants of health at the community and population levels can minimize the negative effects of individualization. Structural measures like area deprivation indices rather than individual experiences of housing insecurity, for example, can allow for population-level analysis of barriers and relationships between these factors and clinical outcomes. This can push back against individualization and inadequate interventions, bringing attention to the structural and designing interventions that remove barriers to health equity.

Similarly, the values driving clinical algorithm design in healthcare can be reprioritized. Rather than uncritically pursuing predictive capacity in an algorithm for cost control, healthcare systems can consider and center patient values. Systematic data collection and analysis of patient values is a critical step in designing responsive approaches to algorithmic decision making. This necessarily involves a reconceptualization of policy as a mechanism for justice that is rooted in community. It also requires the dedication of limited resources to antiracism. For example, directing IT resources to the design and implementation of an algorithm

that produces racist outcomes (e.g. eGFR) is the norm. However, using bounded justice as an analytic, health systems can direct those resources to algorithm redesign that specifically corrects for historical racist inequities in concert with additional community-level interventions (Ahmed et al., 2021). This entails critical consideration of how algorithms have contributed to these inequities and whether they are an appropriate target for intervention. We highlight examples of these types of efforts below.

Use of algorithms for repair

The concept of reparations is not new in healthcare. Driven by the understanding that racial health disparities are an outcome of systematic discrimination and exclusion, these approaches focus on increasing access to resources for marginalized communities, often in the form of federal financial restitution (Nelson, 2016; Williams et al., 2019). For example, Bassett and Galea (2020) proposed reparations for slavery as a means of closing Black–White health disparities. They argue that reparations would mitigate health disparities by (1) providing Black Americans with the financial means “to obtain health-producing resources such as better neighborhoods, better schools, and access to cleaner air,” (2) alleviate stress to improve health, and (3) proactively address the intergenerational effects of marginalization over time. While this type of macro-level restitution is necessary, bounded justice as an analytical tool encourages us to additionally consider more localized, meso-level manifestations of these resource-based injustices as we design interventions.

An example of this is medical restitution, a reparative approach piloted by Wispelwey and Morse (2021) that is focused on addressing the harms of institutional racism in healthcare and medicine. Morse and Wispelwey applied Darity’s framework of *acknowledgement*, *redress*, and *closure* in the context of heart failure management to design a program they call Healing ARC (Darity and Mullen, 2022; Wispelwey et al., 2022). They first identified a racial disparity in referrals to specialty care for heart failure, whereby Black and Latinx or Hispanic patients were not receiving referrals at the same rate as their white counterparts. Using a multistep approach in partnership with affected patients and the community, the team designed a clinical decision support tool in the form of an EHR alert to prompt providers to consider referrals for patients who historically were denied this path to specialty care. Importantly, the team also worked with community members to ensure that their work was responsive to community needs.

In this way, Healing ARC provides an example of a reimagined value system. Rather than prioritizing the status quo vis a vis power and profit, Wispelwey and Morse centered patient values and concerns. They designed a program that placed patient needs and redress of harm at

the center. Importantly, Healing ARC provides a specific example of a reparative approach that uses clinical algorithms for repair but does not conceptualize technology as a complete solution. The algorithmic tool was incorporated as one piece of a larger effort. This is a particularly crucial insight. By avoiding the techno-optimism that portrays algorithms as saviors, focusing on the sociotechnical system holistically and explicitly centering the affected communities, the program avoided many pitfalls in its pursuit of more equitable treatment.

Redress of algorithmic Harms

As described previously, clinical algorithms can materialize oppression, propagating race-based medicine and reinforcing societal injustice. Bounded justice pushes us to both recognize and redress these algorithmic harms. An example of this is type of work is the New York City Coalition to End Racism in Clinical Algorithms (CERCA), a collaborative effort on the part of the New York City health department and a coalition of local health systems to end the misuse of race modifiers in clinical algorithms and to remedy harm to affected patients. To pursue this goal CERCA applies antiracist principles to clinician education, policy work to change clinical practice guidelines, patient advocacy and support, and the use of reparative paradigms for racial justice (CERCA, 2020). Drawing an important distinction between race-based and race-conscious algorithms, the effort to redress health inequities necessarily engages with the realities of race and racism (Khazanchi et al., 2022). Specific efforts to end the use of algorithmic race-based medicine continue to expand under this umbrella, such as research that exposes the many instances biologization of race in clinical practice (Burns et al., 2023).

CERCA and similar efforts emphasize the importance of engaging with and dismantling digital race-based medicine. They also represent the type of interrogation and departure from disciplinary norms that bounded justice promotes. Instead of looking to the next technological intervention as a solution, these approaches engage explicitly with the historical and contemporary realities of embodied and embedded racism. They also critically engage with how data reflects these realities and create alternative reparative processes. By working to transform medical education and practice, collaborating across health system boundaries, building patient power, and explicitly working to repair the harms of racism, these efforts serve as examples of the bounded justice analytic in practice.

Beyond the algorithm

As both Healing ARC and CERCA demonstrate, an approach to algorithmic reparation informed by bounded justice must engage with the social world beyond the

algorithm itself (Davis et al., 2021). As analyzed in other domains like predictive policing and facial recognition software, algorithms function in social contexts as methods of categorizing and sorting people, routinely perpetuating structural racism (boyd, 2023; Le Bui and Noble, 2020; Richardson et al., 2019; Shelby and Henne, 2022). Algorithms and the data underlying them reinforce inequitable access to resources and, as Shelby and Henne (2022) describe, “co-produce racialized social phenomena,” legitimizing the racial categories along which resources are stratified (boyd, 2023; Duster, 2003, 2005; Liao and Carbonell, 2022). Reparative approaches to health equity and racial justice through the reallocation of resources engage with and use these very same categories (Davis et al., 2021). Bounded justice can serve as a praxis to build alternatives, facilitating historical analysis of these racial categories and how they have been utilized to maintain power and structure access to critical resources. Given that racialized groups do not experience racism in the same way, bounded justice also encourages a nuanced approach to designing reparative solutions through the examination and leveraging of localized mechanisms tying resources to health outcomes at a community level.

Importantly, these localized mechanisms operate largely outside of the healthcare system. Structural racism operates through mutually reinforcing systems and structures, including criminal justice, education, built environment, and housing. These are all domains that shape people’s abilities to lead healthy lives. They also impact the healthcare people receive. Reparative approaches to algorithmic systems like insurance, credit, housing, and policing will therefore have the potential to impact healthcare and the health of individuals and communities. Bounded justice encourages holistic evaluations of distributive approaches to health equity, which necessitates awareness of the various institutions and contexts through which racism operates and impacts health outcomes, in the present and in the past. Given that the healthcare system does not operate in isolation from other institutions or social forces, there is both opportunity and need for interdisciplinary collaboration on algorithmic reparation that appreciates the deep connections between health and society.

Importantly, Creary also calls for the use of bounded justice as a tool to “quantify justice” and assess the impact of interventions implemented in pursuit of equity (Creary, 2021). It can be tempting to apply an algorithmic lens to any work that entails quantification or measurement. However, the communities for whom reparative interventions are implemented might not define success such that it can be quantified in that way. For example, in *The Social Life of DNA*, Alondra Nelson explores how DNA testing was used as a tool to pursue racial reconciliation and reparations for slavery by the African-American community. Nelson describes how reparations had not only material dimensions, but also symbolic and psychological

components tied to the historical amnesia surrounding the cumulative disadvantage and social exclusion faced by Black people in the United States (Nelson, 2016). In this case, reparations entails what Ta-Nehisi Coates describes as the “the full acceptance of our collective biography and its consequences” (The Case for Reparations by Ta-Nehisi Coates - The Atlantic, 2014). With this in mind, we can both design and benchmark the success of reparative approaches to algorithmic justice based explicitly on what communities want repaired and redressed.

This pushes us to reconceptualize our approach to algorithmic systems and their use for reparation. For example, what would it look like if algorithmic systems and data infrastructure were built to acknowledge and foreground historical harms, as opposed to burying them within the forgettance stack? How would our approaches to quantifying the impact of our pursuit of health equity and restitution change if the standard for success was set by the communities who have borne the brunt of injustice? In the case of clinical algorithms, prioritizing communities in design and assessment of reparative interventions entails recognition of the risks of perpetuating the privilege granted to algorithms both within and outside of the healthcare system. For example, assumptions that algorithms outperform human capabilities are problematic because they fuel techno-optimism and foreclose contestation. They also run the risk of promoting paternalistic implementations of resource allocation that do not account for what communities need or desire. Algorithmic reparations risk falling into this pattern if they are not explicit about their limitations and the ways they must be paired with nonalgorithmic work. Bounded justice serves as a “biopolitical lever” to draw attention to these types of unforeseen consequences. It also encourages the pursuit of structural interventions that address cumulative disadvantage, such as community-level interventions or public policy change, outside of the healthcare system and outside of algorithmic systems.

Conclusion

In our discussion of clinical algorithms and health equity, we highlight how bounded justice offers a praxis to guide the pursuit of reparative approaches to algorithmic justice. As a conceptual framework, it offers the opportunity for deep engagement with the ways in which manifestations of racism and other interlocking systems of oppression constrain well-intentioned interventions, such as algorithmic fairness. It also serves as an analytical tool to direct the design and evaluation of sociotechnical and structural interventions for health equity that contend with the enduring effects of historical injustice and elevate community voices. Perhaps the most important characteristic of a reparations approach informed by bounded justice is the acknowledgement that technology itself will not be the

solution. It cannot be. However, these tools can be designed with that understanding in mind. Algorithmic systems with explicitly named and defined limitations are better equipped to work toward justice than those built for fairness that fail to engage with their inadequacies. A reparative approach to clinical algorithms in healthcare must thus be comprehensive, informed by multidisciplinary expertise and deep engagement with the fundamental causes of racism and other inequities (Davis et al., 2021). It should explicitly name and identify the ways reparative efforts *will* represent examples of bounded justice. By designing a system of algorithmic reparations in healthcare to specifically address the embodied structural inequities patients experience, the boundedness of justice can be mitigated.

While the primary focus of this work has been on healthcare, there are implications for other contexts. First, the work of algorithmic reparation in domains such as financial institutions or systems of policing and surveillance have the potential to shape health outcomes and should be considered part of the work of pursuing health equity. Furthermore, bounded justice can be applied to the algorithmic functions of other domains in the ways we have applied it to healthcare. This includes identifying how the system or institution has perpetuated racism and explicit recognition of what this means for the data currently being used to design algorithms. It also implies (1) critical engagement with the values shaping algorithmic systems, (2) identification and reimagining of the power dynamics therein, and (3) appreciation of the sociohistorical constraints that undermine even the most well-intentioned fairness efforts. Critically, algorithmic reparations in any context must be inextricably connected to nontechnological equity work.

We end by noting that there is no checklist for applying bounded justice in pursuit of health equity or algorithmic justice. This is because there are no quick solutions for addressing systemic racism, the cumulative impacts on disenfranchised communities, or the entrenchment of injustice in our algorithmic systems. Furthermore, applying the same template to the design of interventions is limited given that the mechanisms tying racism to health outcomes are continuously shifting, both temporally and geographically. The challenge of applying bounded justice as a praxis is that it takes time, effort, and creativity to comprehend the extent to which even our most well-intentioned health equity efforts are constrained and do the work to create alternatives. However, this is also the strength of bounded justice. By recognizing the extent to which the pursuit of justice is bounded, we begin with the knowledge that the work of repair and redress will necessarily be continuous and iterative, balancing urgent short-term needs with the long-term work of systems transformation. As Creary describes, “[if] we are compelled to create a more just world, we must theorize, study, evaluate, redesign, re-evaluate, act and then return to the cycle with reflection

embedded throughout—to determine the most effective ways to distribute justice” (Creary, 2021).

Acknowledgements

We would like to thank Melissa Creary for her feedback, generosity, and guidance in applying bounded justice to this work. We are also grateful to the Algorithmic Reparation workshop organizers and participants.


Declaration of conflicting interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sarah El-Azab  <https://orcid.org/0000-0003-4435-4700>

Paige Nong  <https://orcid.org/0000-0002-2849-9005>

References

- Abduljawad A and Al-Assaf AF (2011) Incentives for better performance in health care. *Sultan Qaboos University Medical Journal* 11(2): 201–206.
- Ahmed S, Nutt CT, Eneanya ND, et al. (2021) Examining the potential impact of race multiplier utilization in estimated glomerular filtration rate calculation on African-American care outcomes. *Journal of General Internal Medicine* 36(2): 464–471.
- Allen A, Mataraso S, Siefkas A, et al. (2020) A racially unbiased, machine learning approach to prediction of mortality: Algorithm development study. *JMIR Public Health and Surveillance* 6(4): e22400.
- Bai G, Zare H, Eisenberg MD, et al. (2021) Analysis suggests government and nonprofit hospitals’ charity care is not aligned with their favorable tax treatment. *Health Affairs* 40(4): 629–636.
- Bassett MT and Galea S (2020) Reparations as a public health priority—a strategy for ending black–white health disparities. *New England Journal of Medicine* 383(22): 2101–2103. Massachusetts Medical Society.
- Benjamin R (2019) Assessing risk, automating racism. *Science* 366(6464): 421–422.
- Berenstein N (2020) White feminist gaslighting. *Hypatia* 35(4): 733–758. Cambridge University Press.
- Bonham VL, Green ED and Pérez-Stable EJ (2018) Examining how race, ethnicity, and ancestry data are used in biomedical research. *JAMA* 320(15): 1533.
- Bossen C, Pine KH, Cabitza F, et al. (2019) Data work in healthcare: An Introduction. *Health Informatics Journal* 25(3): 465–474. SAGE Publications Ltd.
- Boyd D (2023) The structuring work of algorithms. *Daedalus* 152(1): 236–240.
- Boyd R, Lindo E, Weeks L, et al (2020) On Racism: A New Standard For Publishing On Racial Health Inequities. In: *Health Affairs Forefront*. Available at: <https://>

- www.healthaffairs.org/doi/10.1377/forefront.20200630.939347/ (accessed 22 February 2022).
- Bozkurt S, Cahan EM, Seneviratne MG, et al. (2020) Reporting of demographic data and representativeness in machine learning models using electronic health records. *Journal of the American Medical Informatics Association* 27(12): 1878–1884.
- Braun L (2021) Race correction and spirometry. *Chest* 159(4): 1670–1675.
- Braun L, Fausto-Sterling A, Fullwiley D, et al. (2007) Racial categories in medical practice: How useful are they? *PLOS Medicine* 4(9): e271. Public Library of Science.
- Braveman PA, Kumanyika S, Fielding J, et al. (2011) Health disparities and health equity: The issue is justice. *American Journal of Public Health* 101(S1): S149–S155.
- Brown GC, Brown MM and Sharma S (2003) Value-based medicine: Evidence-based medicine and beyond. *Ocular Immunology and Inflammation* 11(3): 157–170.
- Burns NR, Kolarova T, Katz R, et al. (2023) Reconsidering race adjustment in prenatal alpha-fetoprotein screening. *Obstetrics & Gynecology* 141(3): 438.
- CERCA (2020) *New York City Coalition to End Racism in Clinical Algorithms (CERCA) Inaugural Report*. New York City Department of Health and Mental Hygiene. Available at: <https://www.nyc.gov/assets/doh/downloads/pdf/cmo/cerca-report.pdf>.
- Cerdeña JP, Asabor EN, Plaisime MV, et al. (2022) Race-based medicine in the point-of-care clinical resource UpToDate: A systematic content analysis. *eClinicalMedicine* 52: 101581.
- Cerdeña JP, Plaisime MV and Tsai J (2020) From race-based to race-conscious medicine: How anti-racist uprisings call us to act. *The Lancet* 396(10257): 1125–1128.
- Chen IY, Pierson E, Rose S, et al. (2021) Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science* 4(1): 123–144.
- Coates T-N (2014) The Case for Reparations. *The Atlantic*. Available at: <https://www.theatlantic.com/magazine/archive/2014/06/the-case-for-reparations/361631/> (accessed 7 July 2023).
- Cook L, Espinoza J, Weiskopf NG, et al. (2022) Issues with variability in electronic health record data about race and ethnicity: Descriptive analysis of the national COVID cohort collaborative data enclave. *JMIR Medical Informatics* 10(9): e39235.
- Cox M, Panagides JC, Tabari A, et al. (2022) Risk stratification with explainable machine learning for 30-day procedure-related mortality and 30-day unplanned readmission in patients with peripheral arterial disease. *PLoS ONE* 17(11): 1–18.
- Creary MS (2021) Bounded justice and the limits of health equity. *Journal of Law, Medicine & Ethics* 49(2): 241–256.
- Cresswell K, Callaghan M, Khan S, et al. (2020) Investigating the use of data-driven artificial intelligence in computerised decision support systems for health and social care: A systematic review. *Health Informatics Journal* 26(3): 2138–2147. SAGE Publications Ltd.
- Crowley RJ, Tan YJ and Ioannidis JPA (2020) Empirical assessment of bias in machine learning diagnostic test accuracy studies. *Journal of the American Medical Informatics Association* 27(7): 1092–1101.
- Cruz TM (2022) The social life of biomedical data: Capturing, obscuring, and envisioning care in the digital safety-net. *Social Science & Medicine* 294: 114670.
- Cruz TM and Smith SA (2021) Health equity beyond data: Health care worker perceptions of race, ethnicity, and language data collection in electronic health records. *Medical Care* 59(5): 379–385.
- Darity WA Jr and Mullen AK (2022) *From Here to Equality: Reparations for Black Americans in the Twenty-First Century*. Chapel Hill: UNC Press Books.
- Dave G, Wolfe MK and Corbie-Smith G (2021) Role of hospitals in addressing social determinants of health: A groundwater approach. *Preventive Medicine Reports* 21: 101315.
- Davis JL, Williams A and Yang MW (2021) Algorithmic reparation. *Big Data & Society* 8(2): 1–12.
- Delaney SW, Essien UR and Navathe A (2021) Disparate impact: How colorblind policies exacerbate black–white health inequities. *Annals of Internal Medicine* 174(10): 1450–1451. American College of Physicians.
- Dennstädt F, Treffers T, Iseli T, et al. (2021) Creation of clinical algorithms for decision-making in oncology: An example with dose prescription in radiation oncology. *BMC Medical Informatics and Decision Making* 21(1): 212.
- Dimick JB, Ruhter J, Sarrazin MV, et al. (2013) Black patients are more likely to undergo surgery at low quality hospitals in segregated regions. *Health Affairs (Project Hope)* 32(6): 1046–1053.
- Duster T (2003) *Backdoor to Eugenics*, 2nd edn New York: Routledge.
- Duster T (2005) Race and reification in science. *Science* 307(5712): 1050–1051. American Association for the Advancement of Science.
- El-Sayed SS (2021) Medical algorithms need better regulation. *Scientific American*. Available at: <https://www.scientificamerican.com/article/the-fda-should-better-regulate-medical-algorithms/> (accessed 7 October 2021).
- Epstein RM, Fiscella K, Lesser CS, et al. (2010) Why the nation needs a policy push on patient-centered health care. *Health Affairs* 29: 8. Health Affairs: 1489–1495.
- FDA (2020) Software as a Medical Device (SaMD). FDA. Available at: <https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd> (accessed 23 February 2023).
- Feagin J and Bennefield Z (2014) Systemic racism and U.S. Health care. *Social Science & Medicine* 103: 7–14.
- Ferryman K (2020) Addressing health disparities in the Food and Drug Administration's artificial intelligence and machine learning regulatory framework. *Journal of the American Medical Informatics Association* 27(12): 2016–2019.
- Ferryman K and Pitcan M (2018) *Fairness in Precision Medicine*. February. Data & Society.
- Ford CL and Airhihenbuwa CO (2010) The public health critical race methodology: Praxis for antiracism research. *Social Science & Medicine* 71(8): 1390–1398.
- Freij M, Dullabh P, Lewis S, et al. (2019) Incorporating social determinants of health in electronic health records: Qualitative study of current practices among top vendors. *JMIR Medical Informatics* 7(2): e13849.
- Ghassemi M and Mohamed S (2022) Machine learning and health need better values. *NPJ Digital Medicine* 5(1): 1–4. 1. Nature Publishing Group.
- Goodman KE, Morgan DJ and Hoffmann DE (2023) Clinical algorithms, antidiscrimination laws, and medical device regulation. *JAMA* 329(4): 285–286.

- Gordon M and Di Bartolo IM (2021) Using race with caution in the ASCVD calculator. *American Family Physician* 104(2): 292–294.
- Grant SW, Collins GS and Nashef SAM (2018) Statistical primer: Developing and validating a risk prediction model†. *European Journal of Cardio-Thoracic Surgery* 54(2): 203–208.
- Green B (2022) Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. *Philosophy & Technology* 35(4): 90.
- Hanna A, Denton E and Smart A (2020) Towards a Critical Race Methodology in Algorithmic Fairness. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*: 501–512.
- Hardeman RR and Karbeah J (2020) Examining racism in health services research: A disciplinary self-critique. *Health Services Research* 55(S2): 777–780.
- HHS, CMS and OCR (2022) Nondiscrimination in Health Programs and Activities. Available at: <https://www.federalregister.gov/documents/2022/08/04/2022-16217/nondiscrimination-in-health-programs-and-activities> (accessed 2 July 2023).
- Hoffman KM, Trawalter S, Axt JR, et al. (2016) Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America* 113(16): 4296–4301.
- Hoffmann AL (2019) Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22(7): 900–915.
- House Committee on Ways and Means (2021) *Fact Versus Fiction: Clinical Decision Support Tools and the (Mis)Use of Race*. Majority Staff Report.
- Huang J, Galal G, Etemadi M, et al. (2022) Evaluation and mitigation of racial bias in clinical machine learning models: Scoping review. *JMIR Medical Informatics* 10(5): e36388.
- Jehi L, Ji X, Milinovich A, et al. (2020) Development and validation of a model for individualized prediction of hospitalization risk in 4,536 patients with COVID-19. *PLoS ONE* 15(8). Public Library of Science: e0237419.
- Johnson AEW, Bulgarelli L, Shen L, et al. (2023) MIMIC-IV, a freely accessible electronic health record dataset. *Scientific Data* 10(1): 1. Nature Publishing Group: 1.
- Johnson DW and Kane NM (2010) The U.S. Health Care System: A Product of American History and Values. In: Elhauge E (ed) *The Fragmentation of U.S. Health Care: Causes and Solutions*. Oxford: Oxford University Press, 323–342.
- Jones CP (2000) Levels of racism: A theoretic framework and a gardener's tale. *American Journal of Public Health* 90(8): 1212–1215.
- Kahn J (2004) How a drug becomes 'ethnic': Law, commerce, and the production of racial categories in medicine. *Yale Journal of Health Policy, Law, and Ethics* 4(1): 1–46.
- Kahn J (2006) Genes, race, and population: Avoiding a collision of categories. *American Journal of Public Health* 96(11): 1965–1970.
- Kahn J (2008) Exploiting race in drug development: BiDil's interim model of pharmacogenomics. *Social Studies of Science* 38(5): 737–758.
- Kasthurirathne SN, Vest JR, Menachemi N, et al. (2018) Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. *Journal of the American Medical Informatics Association* 25(1): 47–53.
- Khazanchi R, Tsai J, Eneanya ND, et al. (2022) Leveraging Affordable Care Act Section 1557 To Address Racism In Clinical Algorithms. In: *Health Affairs Forefront*. Available at: <https://www.healthaffairs.org/doi/10.1377/forefront.20220930.182927/full/> (accessed 8 January 2023).
- Kishore AK, Hossain MJ, Cameron A, et al. (2022) Use of risk scores for predicting new atrial fibrillation after ischemic stroke or transient ischemic attack—a systematic review. *International Journal of Stroke* 17(6): 608–617. SAGE Publications.
- Lans A, Kanbier LN, Bernstein DN, et al. (2022) Social determinants of health in prognostic machine learning models for orthopaedic outcomes: A systematic review. *Journal of Evaluation in Clinical Practice* 29: 292–299.
- Lantz PM (2019) The medicalization of population health: Who will stay upstream? *The Milbank Quarterly* 97(1): 36–39.
- Lantz PM, Goldberg DS and Gollust SE (2023) The perils of medicalization for population health and health equity. *The Milbank Quarterly* 101(S1): 61–82.
- Le Bui M and Noble SU (2020) We're missing a moral framework of justice in artificial intelligence: On the limits, failings, and ethics of fairness. In: Dubber MD, Pasquale F and Das S (eds) *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press, 161–179.
- Lett E, Asabor E, Beltrán S, et al. (2022) Conceptualizing, contextualizing, and operationalizing race in quantitative health sciences research. *The Annals of Family Medicine* 20(2): 157–163.
- Li X and Evans JM (2022) Incentivizing performance in health care: A rapid review, typology and qualitative study of unintended consequences. *BMC Health Services Research* 22(1): 690.
- Liao S and Carbonell V (2022) Materialized oppression in medical tools and technologies. *The American Journal of Bioethics* 23(4): 9–23.
- Lin E, Sage WM, Bozic KJ, et al. (2021) Value-based healthcare: The politics of value-based care and its impact on orthopaedic surgery. *Clinical Orthopaedics and Related Research* 479(4): 674–678.
- Luo J-C, Zhao Q-Y and Tu G-W (2020) Clinical prediction models in the precision medicine era: Old and new algorithms. *Annals of Translational Medicine* 8(6): 274.
- Madhusoodanan J (2021) A troubled calculus. *Science* 373(6553): 380–383.
- Manzer JL and Bell AV (2022) The limitations of patient-centered care: The case of early long-acting reversible contraception (LARC) removal. *Social Science & Medicine* 292: 114632.
- McCadden MD, Joshi S, Mazwi M, et al. (2020) Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health* 2(5): e221–e223.
- Meng C, Trinh L, Xu N, et al. (2022) Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Scientific Reports* 12: 7166.
- Mhasawade V, Zhao Y and Chunara R (2021) Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence* 3(8): 659–666.
- Mitchell S, Potash E, Barocas S, et al. (2021) Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and its Application* 8(1): 141–163.

- Muller M and Strohmayer A (2022) Forgetting practices in the data sciences. In: CHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 29 April 2022, 1–19. ACM.
- Nelson A (2016) *The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome*. Boston, MA: Beacon Press.
- Owens K and Walker A (2020) Those designing healthcare algorithms must become actively anti-racist. *Nature Medicine* 26(9): 1327–1328.
- Pfohl SR, Foryciarz A and Shah NH (2021) An empirical characterization of fair machine learning for clinical risk prediction. *Journal of Biomedical Informatics* 113: 103621.
- Phelan JC and Link BG (2015) Is racism a fundamental cause of inequalities in health? *Annual Review of Sociology* 41(1): 311–330.
- Phelan JC, Link BG and Tehranifar P (2010) Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. *Journal of Health and Social Behavior* 51(1_suppl): S28–S40.
- Polubriagino FCG, Ryan P, Salmasian H, et al. (2019) Challenges with quality of race and ethnicity data in observational databases. *Journal of the American Medical Informatics Association* 26(8–9): 730–736.
- Rajkomar A, Hardt M, Howell MD, et al. (2018) Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine* 169(12): 866–872. American College of Physicians.
- Reddy S, Allan S, Coghlan S, et al. (2020) A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association* 27(3): 491–497.
- Richardson R, Schultz JM and Crawford K (2019) Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online* 94: 41.
- Roberts DE (2011) What's wrong with race-based medicine? Genes, drugs, and health disparities. *Minnesota Journal of Law, Science, & Technology* 12(1): 1–21.
- Roberts DE (2021) Abolish race correction. *The Lancet* 397(10268): 17–18.
- Roth WD (2016) The multiple dimensions of race. *Ethnic and Racial Studies* 39(8): 1310–1338.
- Rozenbaum D, Shreve J, Radakovich N, et al. (2021) Personalized prediction of hospital mortality in COVID-19-positive patients. *Mayo Clinic Proceedings. Innovations, Quality & Outcomes* 5(4): 795–801.
- Rubashkin N (2022) “You don’t really know until you try”: VBAC prediction from the patient perspective. *American Journal of Obstetrics and Gynecology* 226(1): S531.
- Ryn M van, Burgess DJ, Dovidio JF, et al. (2011) The impact of racism on clinician cognition, behavior, and clinical decision making. *Du Bois Review* 8(1): 199–218.
- Shachar C and Gerke S (2023) Prevention of bias and discrimination in clinical practice algorithms. *JAMA* 329(4): 283–284.
- Shelby R and Henne K (2022) Situating questions of data, power, and racial formation. *Big Data & Society* 9(1): 1–4.
- Sikstrom L, Maslej MM, Hui K, et al. (2022) Conceptualising fairness: Three pillars for medical algorithms and health equity. *BMJ Health & Care Informatics* 29(1): e100459.
- Steyerberg EW and Vergouwe Y (2014) Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal* 35(29): 1925–1931.
- Topol EJ (2019) High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine* 25(1): 44–56. 1. Nature Publishing Group.
- Ukoha EP, Snaveley ME, Hahn MU, et al. (2022) Toward the elimination of race-based medicine: Replace race with racism as preeclampsia risk factor. *American Journal of Obstetrics and Gynecology* 227(4): 593–596.
- Van Calster B, Wynants L, Timmerman D, et al. (2019) Predictive analytics in health care: How can we know it works? *Journal of the American Medical Informatics Association: JAMIA* 26(12): 1651–1654.
- Vyas DA, Eisenstein LG and Jones DS (2020) Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *The New England Journal of Medicine* 383(9): 874–882. Massachusetts Medical Society.
- Williams DR, Lawrence JA and Davis BA (2019) Racism and health: Evidence and needed research. *Annual Review of Public Health* 40(1): 105–125.
- Wispelwey B, Marsh RWilson, et al. (2022) Leveraging Clinical Decision Support for Racial Equity: A Sociotechnical Innovation. *NEJM Catalyst Innovations in Care Delivery*. Massachusetts Medical Society. Available at: <https://catalyst.nejm.org/doi/full/10.1056/CAT.22.0076> (accessed 11 October 2022).
- Wispelwey B and Morse M (2021) An antiracist agenda for medicine. *Boston Review*, 17 March. Available at: <https://www.bostonreview.net/articles/michelle-morsebram-wispelwey-what-we-owe-patients-case-medical-reparations/> (accessed 8 January 2023).
- Wu E, Wu K, Daneshjou R, et al. (2021) How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. *Nature Medicine* 27(4): 582–584. 4. Nature Publishing Group.
- Wynants L, Calster BV, Collins GS, et al. (2020) Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *BMJ* 369: m1328. British Medical Journal Publishing Group.
- Xu J, Xiao Y, Wang WH, et al. (2022) Algorithmic fairness in computational medicine. *eBioMedicine* 84: 1–10.
- Yearby R (2021) Race based medicine, colorblind disease: How racism in medicine harms us all. *The American Journal of Bioethics* 21(2): 19–27. Taylor & Francis.
- Yearby R, Clark B and Figueroa JF (2022) Structural racism in historical and modern us health care policy. *Health Affairs* 41(2): 187–194. Health Affairs.
- Zhao Y, Wood EP, Mirin N, et al. (2021) Social determinants in machine learning cardiovascular disease prediction models: A systematic review. *American Journal of Preventive Medicine* 61(4): 596–605.