Statistical Discrimination

Sonja Starr*

The Supreme Court has emphatically and repeatedly rejected efforts to justify otherwise-illegal discrimination against individuals by resort to statistical generalizations about groups. But practices that violate this principle are pervasive and largely ignored or even embraced by courts, lawyers, and law scholars. For example, many health care fields, in seeming contravention of antidiscrimination statutes, make use of explicitly racialized diagnostic algorithms that make it harder for Black patients to access care than non-Black patients with identical symptoms. Moreover, the justice system itself has embraced numerous similar practices, including demographic and "sociocultural" adjustments of intellectual-capacity assessments for defendants facing the death penalty, explicit class-based discrimination in criminal justice risk assessments, and the use of race- and sex-specific actuarial data to calculate tort damages. This Article examines these practices, the law governing them, and the reasons for these disconnects between law and practice.

TABLE OF CONTENTS

INTRODUCTION

The National Football League ("NFL") began its 2020 season with a rebranding befitting the times: as a racial justice ally. After years of criticism for failing to support Black players' political protests, the League offered a new look: "End Racism" printed on end zones, "Say Their Stories" videos, and Black Lives Matter stickers on helmets.[1] But as the season commenced, a new race-related controversy emerged, this one intertwined with the NFL's other recent public relations crisis. In 2015, the League had settled a class action related to concussions, covering over 20,000 former players.[2] The settlement required claimants to undergo cognitive testing to determine the extent of impairment. As a lawsuit two players filed in August 2020 revealed, these test results were quietly subjected to an adjustment called "race-norming": each player was evaluated relative to the normal range on those tests *for his racial group*. Because the Black average scores were lower, scores that would qualify a white player as impaired were deemed normal for Black players. Black players needed more serious impairments to qualify for the settlement's six- and seven-figure payouts.

The race-norming scandal brought deserved blowback to the NFL. Under pressure, in June 2021, the NFL announced an intention to end the practice; in October, the parties to the concussion settlement agreed on a new framework.[3] The NFL has publicly blamed norms in neuropsychology, and implied that the League now deserves credit for driving change in that

---

[1] *See* Patti Williams, *Rebranding the NFL: How the League Shifted Its Message on Racial Justice*, WHARTON BUS. DAILY (Sep. 11, 2020), https://knowledge.wharton.upenn.edu/podcast/knowledge-at-wharton-podcast/shropshire-athlete-activism/ [https://perma.cc/6SFD-4E9B]; Rob Maaddi, *NFL Returns Social Justice Helmet Decals, End Zone Stencils*, ASSOC. PRESS (Sep. 2, 2021), https://apnews.com/article/sports-nfl-c10867c92924cc9a59c99e220c8ee84e [https://perma.cc/9WEJ-NHJK].

[2] *See* Jason M. Breslow, *NFL Concussion Settlement Wins Final Approval from Judge*, PBS FRONTLINE (Apr. 22, 2015), pbs.org/wgbh/frontline/article/nfl-concussion-settlement-wins-final-approval-from-judge/ [https://perma.cc/6CVG-8CLE]; *Read the Proposed N.F.L. Concussion Settlement*, N.Y. TIMES (Oct. 21, 2021), https://www.nytimes.com/interactive/2021/10/21/sports/football/proposed-nfl-concussion-settlement.html [https://perma.cc/YMT8-WHQP].

[3] *See Proposed NFL Concussion Settlement, supra* note 2; Ken Belson, *Plan Filed to Scrap Race as Factor in N.F.L. Concussion. Settlement*, N.Y. TIMES (Oct. 20, 2021), https://

discipline.[4] This is breathtakingly cynical: although some doctors apparently submitted race-normed scores on their own, others reported being pushed by settlement administrators to do so, and the NFL appealed many claims in which they didn't.

Yet there's a grain of truth in the NFL's argument: the practice of race-norming was hardly invented by the NFL, and indeed, race-specific protocols have long pervaded many areas of health care. And indeed, similar practices pervade the legal system itself. This Article will explore a range of facially discriminatory practices in health care and law that are baked into algorithms, "norming" of test scores, and the like. All of these are practices that explicitly purport to justify discrimination based on statistics. As I'll show, they all run afoul of clearly established legal doctrine—and yet somehow, haven't been struck down by courts. Often, they haven't even been challenged, and law scholars have had nearly nothing to say about them. I explore this strange disconnect between law and practice in this Article. I examine why we so often tolerate discrimination that has a veneer of scientific objectivity, highlight a set of very troubling and almost surely illegal practices, and lay out a roadmap for those who might wish to challenge or change them.

Let's begin with the context the NFL cited to justify its use of race: the widespread reliance on racialized clinical practice guidelines in health care. Many diagnostic or other algorithms explicitly treat race as a variable—and not in the direction that one might expect, i.e., to recognize and seek to reduce racial inequities in health-affecting conditions. Rather, these algorithms commonly make it harder for Black and other minority patients to get treatment, by treating as "normal for the patient's race" symptoms that actually represent health risks. Echoing an ugly history of medical racism, they often characterize group disparities in the prevalence of such symptoms as essential racial differences, rather than what they overwhelmingly are: the byproduct of race gaps in health-influencing socioeconomic factors. And they lump together racial groups that contain vast individual variation, effectively mischaracterizing the health of large swaths of those groups.

For these reasons, many health disciplines are reconsidering these guidelines' scientific and ethical merit—although this movement is in its early days, and has a long way to go. Meanwhile, though, there's another question that nobody seems to be asking: How is all this legal? As I detail in this Article, racial discrimination in health care is prohibited by many overlapping statutes; in public settings (including one quarter of U.S. hospitals), it is also subject to constitutional restrictions. But there haven't been direct legal challenges to these algorithms' use in medicine. No legal scholarship

---

www.nytimes.com/2021/10/20/sports/football/nfl-concussion-settlement-race.html [https://perma.cc/5JVZ-SC9G].

[4] *See* Will Hobson, How 'race-norming' was built into the NFL concussion settlement, WASH. POST (Aug. 2, 2021), https://www.washingtonpost.com/sports/2021/08/02/race-norming-nfl-concussion-settlement/ [https://perma.cc/EPH5-GPYR].

investigates their lawfulness in any detail. In commentaries within the health disciplines, despite thoughtful ethical critiques, the possibility that the *law* might constrain these algorithms' use goes unmentioned. Meanwhile, patients are surely overwhelmingly unaware that these race-based distinctions exist.

Racialized medical algorithms have occasionally made their way into other litigation contexts, as the NFL example illustrates, but even then, courts have rarely second-guessed them. For example, racial adjustments to kidney function measurements—now rejected in the most recent nephrology standards on medical grounds—have been invoked by the federal Bureau of Prisons when kidney-impaired Black inmates seek medical release.[5] One prisoner sued to challenge this Bureau policy in April 2022, seeking class status—a case that could provide a novel opportunity for a court to address the discrimination issue.[6] To date, however, courts have not questioned these race adjustments. One potentially consequential legal development is that the Department of Health and Human Services has recently issued a proposed rule that, among many other provisions, would explicitly bar racial discrimination in clinical algorithms; its commentary makes critical references to the race-norming controversy, although it does not squarely reject all uses of race. [7] If the rule is adopted, much will turn on the Department's willingness to enforce it, especially against practices that are entrenched medically.

It's surprising how little role law has played in the race-norming controversy so far, given that litigation over denial of health care is common and fear of litigation heavily influences the medical system. Yet perhaps we shouldn't be surprised, as several quite similar practices are widespread in our legal system itself. Consider a few examples, each of which this Article will explore:

- In civil suits, damage awards are routinely calculated based on race- and sex-specific actuarial predictions. For example, a Black girl's life will typically be valued as being worth much less than that of an otherwise-identical white boy, because actuarial tables show that Black women have on average earned less than white men. In a wrongful death lawsuit, her family will receive a much smaller payout, explicitly based on race and sex.
- Prosecutors have repeatedly introduced, and courts have permitted, race-based adjustments to intellectual-capacity scores used to deter-

---

[5] *See* Joseph Goldstein, *How a Race-Based Medical Formula is Keeping Some Black Men in Prison*, N.Y. Times (Apr. 22, 2022), https://www.nytimes.com/2022/04/22/nyregion/prison-kidney-federal-courts-race.html [https://perma.cc/Z4KN-CRRL].

[6] *See* Complaint, Robinson v. Fed. Bureau Prisons, No. 1:22-cv-01098 (Apr. 20, 2022); *see also* Goldstein, *supra* note 5.

[7] *See* Nondiscrimination in Health Programs and Activities: Use of Clinical Algorithms in Decision-Making, 87 Fed. Reg. 47824 § 92.210 (proposed Aug. 24, 2022) [hereinafter "DHHS Notice of Rulemaking"].

mine whether, under *Atkins v. Virginia*,[8] an individual may be subject to the death penalty. Black and Hispanic defendants have been sentenced to death, and some executed, even though, if they were white, their test performance would likely have barred execution. "Sociocultural" norming also makes poor and otherwise disadvantaged defendants more likely to be executed.

- Beyond capital cases, many criminal justice decisions now incorporate risk assessments that extrapolate individual crime risk from past patterns among people with similar characteristics. These algorithms do not typically use race as a predictor, but they do often make outcomes turn explicitly on measures of socioeconomic disadvantage.

These examples from medicine and law all constitute "statistical discrimination": disparate treatment of individuals based on statistical generalizations about the groups to which they belong. Statistical discrimination can also be more subtle or covert. But these examples aren't; they involve explicit use of discriminatory factors in formal, quantitative algorithms or tests. Their persistence in legally regulated contexts and in the justice system itself is somewhat surprising, because each violates longstanding doctrine that cannot be meaningfully distinguished. That is, these practices aren't just troubling; they're illegal, in ways that (once brought to light) could readily be proven in court.

As Part II of this Article details, the Supreme Court has repeatedly held that otherwise-illegal discrimination cannot be justified based on statistical generalizations about groups, *even if those generalizations are empirically supported.* As a shorthand, I call this principle the "prohibition of statistical discrimination," although it is not an *absolute* prohibition; its most important limit is that it generally applies only to certain classifications that receive heightened scrutiny. It has been applied to race and sex discrimination in many contexts, constitutional and statutory. And under a special constitutional doctrine specific to criminal cases, the Supreme Court has also applied the same principle to bar treating poor and unemployed criminal defendants adversely based on the statistical generalization that they pose higher crime risks.

These binding precedents are squarely applicable to the above-described practices, both in the law and in medicine. Yet with almost no exceptions, the courts have not weighed in, and the practices persist. Their legality has not been subject to serious judicial analysis, and the lawyers that one might expect to raise these issues have largely failed to do so. So we live with a strange disconnect in the law governing statistical discrimination: while forbidden in some contexts, in some other contexts where it appears equally illegal, it is tolerated with little objection. This Article describes and critiques this disconnect. It seeks to provide the first sustained examination,

---

[8] 536 U.S. 304 (2002).

cutting across substantive areas, of the law's incongruous treatment of statistical discrimination.

While the Supreme Court's hostility to statistical generalizations will be familiar to some readers, legal scholars have given relatively little attention to this principle. No existing scholarship examines in depth what exactly the doctrine prohibits, what the limitations of this prohibition are, and what its implications are for existing practices; this area of doctrine is typically not given more than passing mention in constitutional law instruction.[9] Many of the specific examples I focus on have also been nearly or completely ignored by legal literature. There *is* a recently burgeoning literature (cutting across disciplines including law, data sciences, and philosophy) on "algorithmic fairness," but that work has not focused on the express use of prohibited categories like race, and it does not explore doctrine or theory related to statistical discrimination.[10] Rather, it has focused on subtler algorithmic sources of disparity, such as facially neutral algorithms with racially disparate impacts or disparate predictive parity, and/or on the development of algorithmic methods to *counter* disparities. From a constitutional law perspective, those examples differ sharply from those I consider; existing doctrine is poorly equipped to handle them, whereas it is well equipped to address the disparate-treatment-type examples that I focus on here. And yet, as important as those harder algorithmic-fairness problems are, it's important not to forget about the purportedly "easier" problems of statistical discrimination—which, after all, have not been solved, and have largely been ignored or tolerated. In 2022, somehow, many people of color are being denied medical care, receiving lesser remedies for injuries, or even being put

---

[9] Frederick Schauer's book PROFILES, PROBABILITIES, AND STEREOTYPES (2006) is the most detailed examination of the role of generalizations in justifying discrimination, but is framed as a qualified *defense* of profiling, and in any event does not closely examine much doctrine or inconsistencies in its application, focusing more abstractly on the moral permissibility of relying on statistical evidence.

[10] Some of this literature will be discussed as relevant to specific topics below, especially in Part VI, which focuses on pro-equity algorithm design. For a sampling of this broad literature, see generally SOLON BAROCAS ET AL., FAIRNESS AND MACHINE LEARNING (forthcoming MIT Press 2023, digital version available at https://fairmlbook.org/ [https://perma.cc/UC7J-EY3X]); MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM (2019); VIRGINIA EUBANKS, AUTOMATING INEQUALITY (2019); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION (2016); Alessandro Fabris et al., *Algorithmic Fairness Datasets: The Story So Far*, 36 DATA MINING AND KNOWLEDGE DISCOVERY 2074 (2022); Jie Xu et al., *Algorithmic Fairness in Computational Medicine*, 84 EBIOMEDICINE 1 (2022); Xiaomeng Wang et al., *A Brief Review on Algorithmic Fairness*, 1 MGMT. SYS. ENG'G 7 (2022); Dana Pessach & Erez Shmueli, *Algorithmic Fairness*, ARXIV (Jan. 21, 2020); Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Legal and Statistical Framework*, 119 MICH. L. REV. 291 (2020); Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811 (2020); Sahil Verma & Julia Rubin, *Fairness Definitions Explained*, PROC. OF THE INT'L WORKSHOP ON SOFTWARE FAIRNESS (2018); Sam Corbett-Davies & Sharad Goel, *The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning*, ARXIV (2018); Jon Kleinberg et al., *Discrimination in the Age of Algorithms* (Nat'l Bureau of Econ. Rsch., Working Paper No. 25548), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3332296 [https://perma.cc/4TW9-NKPH]; Pauline Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189 (2017).

to death explicitly because of their race. These are situations our legal system can address, hasn't, but should.

The lacuna in scholarship may be one reason that courts and lawyers appear routinely to ignore or misunderstand the doctrine surrounding statistical discrimination. It is a somewhat surprising gap, given that, as I'll show, the prohibition on statistical discrimination is both a key animating principle of existing constitutional and antidiscrimination law, and normatively central from a wide variety of perspectives on what vision of equality the law should seek to promote.

The Article's first objectives are thus doctrinal and practical. After Part I discusses the NFL scandal (which serves as a useful entry point to these issues), I turn in Part II to the governing legal principles on statistical discrimination. Next, I examine and critique divergences from those principles in the legal system (Part III, which includes case studies on civil damage awards, capital punishment, and criminal justice risk assessments) and in medical care (Part IV, which offers the first detailed legal analysis of the use of race in clinical algorithms). In these Parts, I hope to provide practitioners with useful legal arguments against racist and otherwise discriminatory practices, and to make the moral and policy case that these practices must not be condoned. The NFL debacle, along with related current debates within medicine, has created a moment in which many people are alert to these issues for the first time. It shouldn't be wasted.

After exploring these doctrinal problems, the Article will then turn in Part V to a sociological puzzle: why do we continue to tolerate these examples of apparently illegal statistical discrimination? I'll consider research on how non-technical audiences defer to scientific expertise; how, conversely, technical fields (including economics, where statistical-discrimination theory originated) fail to grapple with the normative assumptions underlying their descriptive models; and how "system justification" tendencies and professional courtesy mean nobody calls offensive practices out. This is a story about how science sanitizes inequities that we would not tolerate if they were framed in less technical terms. And it's also a story of systemic racism and injustice—of how easy it is to turn a blind eye to discriminatory practices that are long embedded, or that themselves embed and ratify the products of past discrimination. Scientific justifications for discrimination have a long pedigree, and so does systemic racism. But old stories are constantly being told in new ways, and sometimes they can be hard to recognize.

Finally, in Part VI, I address a possible counterpoint: Should we worry about embracing legal principles that might also impede the use of statistical techniques to *promote* equity? I argue that opposing practices that heighten racial and other disparities certainly does not mean that one must oppose the use of analogous methods to mitigate disparities. Like many scholars, I believe that a policy's justness and legal permissibility ought not to turn on the classifications used but on whether it amplifies inequality and subordination. But this answer merits an important caveat. The Supreme Court does not

embrace this substantive approach, and its increasing tendency toward "colorblind" ideology can't be ignored. I'll discuss how race-conscious efforts to promote equity can be designed to be more likely to withstand legal challenges.

I. RACE-NORMING IN THE NFL's CONCUSSION SETTLEMENT

The NFL's race-norming debacle arose within the context of another scandal that has been unfolding since the 1990s: the evidence that head impacts during play can seriously injure the brain.[11] In 2011, several players filed a lawsuit against the NFL over its handling of concussion-related injuries, seeking class status.[12] Many other suits followed; a complaint consolidating suits on behalf of approximately 2000 players was filed in the Eastern District of Pennsylvania in 2012.[13] The case was assigned to Senior Judge Anita Brody, who ordered the parties into mediation. The parties negotiated a proposed settlement in 2013,[14] and in 2015 Judge Brody certified a class encompassing over 20,000 former players[15] and approved the settlement, which was approved by the Third Circuit in April 2016 and took effect in January 2017.[16] The settlement framework required a neuropsychological evaluation and established a standard payout based on age and degree of impairment. Payouts ranged well into the millions,[17] with the total settlement estimated to cost the NFL over a billion dollars.[18]

---

[11] *See* Daniel Rapaport, *Timeline: Six Studies of Head Trauma in Football That Helped Establish Link to CTE*, SPORTS ILLUSTRATED (July 26, 2017), https://www.si.com/nfl/2017/07/26/nfl-concussion-head-trauma-studies-football-timeline [https://perma.cc/7REQ-J27P]; Lauren Ezell, *Timeline: The NFL's Concussion Crisis*, PBS FRONTLINE (Oct. 8, 2013), https://www.pbs.org/wgbh/pages/frontline/sports/league-of-denial/timeline-the-nfls-concussion-crisis/ [https://perma.cc/A8RZ-DL9B].

[12] *See id.*

[13] *See* Mallory Simon, *2,000 Players Unite in Suing NFL Over Head Injuries,* CNN (June 7, 2012), https://www.cnn.com/2012/06/07/sport/football/nfl-concussion-lawsuit/index.html [https://perma.cc/SAM3-UVJJ].

[14] *See NFL, Ex-Players Agree to $765M Settlement in Concussions Suit*, NFL (Aug. 29, 2013), https://www.nfl.com/news/nfl-ex-players-agree-to-765m-settlement-in-concussions-suit-0ap1000000235494 [https://perma.cc/33S4-A77H].

[15] *See Proposed NFL Concussion Settlement*, *supra* note 2.

[16] *See In re: Nat'l Football League Players Concussion Injury Litigation*, 821 F.3d 410 (3d Cir. 2016), *as amended* (May 2, 2016); Lawrence Hurley, *Supreme Court Ends Fight Over $1 Billion NFL Concussion Deal*, REUTERS (Dec. 12, 2016), https://www.reuters.com/article/us-usa-court-nfl-supreme-court-ends-fight-over-1-billion-nfl-concussion-deal-id USKBN1411O3 [https://perma.cc/T664-MJ87].

[17] *See Monetary Award Grid by Age at Time of Qualifying Diagnosis*, *NFL Concussion Settlement* (2018), https://www.nflconcussionsettlement.com/Docs/Monetary%20Award %20Grid%20-%20Option%202.pdf [https://perma.cc/XW7R-JA5Y]; Maryclaire Dale, *NFL pledges to halt 'race-norming,' review Black claims*, ASSOC. PRESS (June 2, 2021), https://apnews.com/article/pa-state-wire-race-and-ethnicity-health-nfl-sports-205b304c0c3724532d74 fc54e58b4d1d [https://perma.cc/N62X-YL5F].

[18] *See Proposed NFL Concussion Settlement*, *supra* note 2; *NFL concussion claims hit $500 million in less than 2 years*, ASSOC. PRESS (July 30, 2018), https://apnews.com/article/north-america-pa-state-wire-ap-top-news-nfl-football-football-cd3ddb8ed41641a8a52ed9608 102624f [https://perma.cc/W9CT-HQU7].

Controversially, the settlement did *not* include a standard medical assessment for dementia, including brain imaging, but instead required a battery of cognitive skills tests.[19] "Race-norming" refers to a practice involved in scoring those tests. Players' raw test scores are translated into adjusted scores, and the translation differs based on race. Players' cognitive performance is effectively compared to people with the same demographics in a past sample used for test development. These demographic adjustments were not mentioned in the public settlement terms—only in the "confidential manual given to doctors."[20]

When the scandal emerged later on, the NFL—backed by the players' association's lead lawyer—claimed that race-norming was within the professional discretion of assessing doctors.[21] And it was probably true that doctors sometimes adopted this practice on their own—as I'll discuss further in Part IV, it is a common practice in neuropsychology. But it's not universal, and when some doctors resisted race-norming, the League pushed back. For example, in appealing an award to Najeh Davenport, the NFL objected to the failure to use "full demographic norms," which it characterized as an "industry standard."[22] This appeal was reviewed by court-appointed special masters, who concluded in August 2020 that the settlement did not strictly mandate race-norming scores, but ordered Davenport's assessing doctor to justify his decision not to do so.[23] Davenport and Kevin Henry, whose claim had been rejected based on race-normed scores,[24] then sued the NFL, alleging racial discrimination in violation of 42 U.S.C. 1981.[25] A media firestorm

---

[19] *See Black NFL Players Call for End of Algorithm That Assumes Black Men Have Lower Cognitive Abilities*, Mkt. Watch (May 14, 2021), https://www.marketwatch.com/story/retired-black-nfl-players-and-their-families-call-for-race-norming-practice-to-end-01621018741#:~:text=under%20the%20settlement%2C%20however%2C%C20the,decline%20to%20win%20an%20award [https://perma.cc/VCP6-2XLX].

[20] Ken Belson, *Black Former NFL Players Say Racial Bias Skews Concussion Payouts*, N.Y. Times (Aug. 25, 2020), https://www.nytimes.com/2020/08/25/sports/football/nfl-concussion-racial-bias.html [https://perma.cc/65AR-CWZE].

[21] *See NFL pledges to halt 'race-norming', supra* note 17; Dave Zirin, *So What the Hell Is Race Norming?*, The Nation (Mar. 12, 2021), https://www.thenation.com/article/society/race-norming-nfl-concussions/ [https://perma.cc/S4R2-FUFG] (citing Commissioner Roger Goodell).

[22] Belson, *supra* note 20; *see also* Pete Madden et al., *Clinicians Fear NFL's Concussion Settlement Program Protocols Discriminate Against Black Players*, ABC News (Feb. 3, 2021), https://abcnews.go.com/Sports/clinicians-fear-nfls-concussion-settlement-program-protocols-discriminate/story?id=75646704 [https://perma.cc/5YWE-PSYN] (citing neuropsychologists' statements that they did not feel free not to apply race norms); *NFL families seek to end 'race-norming' in $1B settlement of brain injury claims*, ABC News (May 14, 2021), https://abcnews.go.com/Sports/nfl-families-seek-end-race-norming-1b-settlement/story?id=77695019 [https://perma.cc/D9KJ-K8ZS].

[23] *See* Belson, *supra* note 20.

[24] *See id.*

[25] *See* Henry v. Nat'l Football League, No. 20-4165 (E.D. Pa. Aug. 25, 2020); Belson, *supra* note 20.

grew, aided by activism from players' families, including a petition to the court with 50,000 signatures.[26]

In March 2021, Judge Brody dismissed Davenport and Henry's lawsuit on procedural grounds, finding it an impermissible collateral attack on the settlement.[27] But she also took the unusual step of ordering the parties to the underlying settlement into mediation on the issue.[28] On June 2, 2021, the League announced that it would end race-norming,[29] and on October 20, 2021, the parties (including Davenport and Henry as intervenors) submitted a revised testing plan to the court. In addition to adopting a new, race-neutral process, it offered Black players whose claims were denied the opportunity for reevaluation,[30] a process that may lead to "hundreds" of new awards.[31] The NFL denied wrongdoing[32] and "said it hoped the new testing formula, developed with input from a panel of experts, would be widely adopted in medicine."[33] In August 2022, the settlement administrator announced that an initial group of sixty-one players who had had their claims denied would now receive payouts.[34] That number is likely to go up, as hundreds of retested players are now having mild dementia symptoms monitored, while thousands more players are still eligible for non-race-normed testing.[35]

The NFL has offered limited commentary on race-norming and has largely sought to distance itself publicly from it. But the League had legally defended it and sought to enforce it. In an August 2020 statement defending the practice, the League's position was that the settlement

> always contemplated the use of recognized statistical techniques to account for demographic differences such as age, education and race. The point of such adjustments . . . [is] to ensure that individuals are treated fairly and compared against comparable groups.[36]

---

[26] *See* Ken Belson, *NFL Asked to Address Race-Based Evaluations in Concussion Settlement*, N.Y. Times (Mar. 9, 2021), https://www.nytimes.com/2021/03/09/sports/football/nfl-concussions-settlement-race.html [https://perma.cc/GU9H-MASL]; *Black NFL Players Call for End of Algorithm That Assumes Black Men Have Lower Cognitive Abilities*, *supra* note 19.

[27] *See Henry v. Nat'l Football League*, No. 20-4165 (Mar. 8, 2021).

[28] *See Black NFL Players Call for End of Algorithm That Assumes Black Men Have Lower Cognitive Abilities*, *supra* note 19.

[29] *See NFL pledges to halt 'race-norming'*, *supra* note 17.

[30] *See NFL Agrees to End Race-Based Brain Testing in $1B Settlement on Concussions*, NPR (Oct. 20, 2021), https://www.npr.org/2021/10/20/1047793751/nfl-concussion-settlement-race-norming-cte [https://perma.cc/E3EC-8GAU].

[31] *See* Jodi Balsam, *NFL Concussion Settlement Five Years Later*, Brook. Sports & Ent. L. Blog (June 7, 2021) https://sports-entertainment.brooklaw.edu/sports/nfl-concussion-settlement-five-years-later/ [https://perma.cc/89VF-KJZD].

[32] *See Proposed NFL Concussion Settlement*, *supra* note 2.

[33] *See NFL Agrees to End Race-Based Brain Testing in $1B Settlement on Concussions*, *supra* note 30.

[34] *See* Ken Belson, *More Black Former NFL Players Eligible for Concussion Payouts*, N.Y. Times (Aug. 12, 2022), https://www.nytimes.com/2022/08/12/sports/football/nfl-concussion-settlement-race.html [https://perma.cc/QKN4-WAX4].

[35] *Id.*

[36] Belson, *supra* note 20.

At least at first, even Christopher Seeger, the players' lead lawyer in the concussion litigation, agreed with this characterization. Seeger, in comments he later apologized for,[37] attributed the testing criteria to the guidance of leading experts and stated that he had "not seen any evidence of racial bias."[38]

One could have imagined the NFL defending race-norming as an adjustment for biases in testing. But this would have been transparently disingenuous, and the League didn't try.[39] Instead, its filings presented race-norming (or "racial demographic adjustments") as a way to more accurately "estimate an individual's premorbid intellectual functioning"[40]—i.e., to account for the presumed lower pre-injury starting point of Black players. The media widely characterized the League as assuming Black players were less intelligent,[41] and the NFL did not dispel this characterization. It thus squarely, if quietly, endorsed one of the most toxic racial generalizations that exists. This stereotype has a long history and some contemporary adherents,[42] but one might expect a public-facing twenty-first-century business to stay far from it, especially when trying to remake its image on race.

But there are reasons the NFL's position is not so surprising. Its racial politics had long been blundering at best; that's why the public-image campaign was needed.[43] The settlement was expensive; race-norming saved the League money. It did have some support in common practices in neuropsychology. The League avoided pushback on race-norming from the players' lawyers, and because the practice wasn't publicly documented, it had reason

---

[37] *See* Matt Stieb, *NFL Will Stop 'Race-Norming' Policy That Diminished Black Brain Injuries*, INTELLIGENCER (June 2, 2021), https://nymag.com/intelligencer/2021/06/nfl-to-end-race-norming-that-diminished-black-injuries.html [https://perma.cc/S83R-KLTY].

[38] Belson, *supra* note 20.

[39] Race-norming was being used squarely against Black players' interests, plus the NFL itself was the one requiring the (non-standard) testing protocol, which experts criticized. *See* Suzanne Leigh, *'Race Norming' Blamed for Denying Payouts to Ex-NFL Players With Dementia*, UNIV. OF CAL. SAN FRANCISCO (Dec. 21, 2020), https://www.ucsf.edu/news/2020/12/419426/race-norming-blamed-denying-payouts-ex-nfl-players-dementia [https://perma.cc/6WAY-CM98]; Rich McHugh, *Doctors Call NFL Concussion Settlement 'Junk Science'*, NEWS NATION NOW, (July 27, 2021) https://www.newsnationnow.com/investigation/doctors-call-nfl-concussion-settlement-junk-science/ [https://perma.cc/H7VC-7YKY]. Also, the NFL gives all draft entrants a cognitive test that is *not* race-normed, despite longstanding critiques that it is racially biased and unrelated to job performance. *See* Joseph Stromberg, *Reminder: The NFL's Wonderlic Aptitude Test is Totally Worthless*, VOX, (May 8, 2014) https://www.vox.com/2014/5/8/5694518/why-the-nfls-wonderlic-aptitude-test-is-totally-worthless [https://perma.cc/CZ6B-CBAG].

[40] *Proposed NFL Concussion Settlement*, *supra* note 2 (Article I, defining terms).

[41] *See, e.g.*, *NFL pledges to halt 'race-norming'*, *supra* note 17.

[42] *See, e.g.*, Matthew Yglesias, *The Bell Curve is About Policy. And it's Wrong*, VOX (Apr. 10, 2018), https://www.vox.com/2018/4/10/17182692/bell-curve-charles-murray-policy-wrong [https://perma.cc/NGS4-BD3E] discussing The BELL CURVE controversy and the continued influence of author Charles Murray).

[43] *See e.g.*, Rodger Sherman, *The NFL's Colin Kaepernick Excuses Look Even More Damning in Retrospect,* THE RINGER (Jun. 15, 2020), https://www.theringer.com/nfl/2020/6/15/21289577/colin-kaepernick-protest-black-lives-matter-nfl-team-excuses [https://perma.cc/ZXY2-CLFX].

to hope it could avoid pushback from anyone else. In fact, until Davenport and Henry sued, three years into the settlement's operation, it did avoid that pushback.

Perhaps the NFL wasn't crazy, then, to bet on escaping accountability. As I explore below, when discrimination takes the form of statistical adjustments, it's often easier for it to escape scrutiny, especially when it appears sanctioned by experts. And race-norming does have a medical pedigree—even though it's quite clearly legally problematic.

II.   RACE-NORMING, STATISTICAL DISCRIMINATION, AND LEGAL DOCTRINE

In the NFL case, the court never reached the merits of Davenport and Henry's racial discrimination claim. Here, I'll show that that claim was well supported by existing doctrine—for reasons carrying much broader implications that the rest of the Article will explore. Section A reviews the Supreme Court's doctrine rejecting statistical justifications for discrimination; Section B explains why this principle matters normatively; and Section C applies it to the NFL controversy. For some readers, this Part's detailed analysis may seem like overkill, given that (as I hope all readers will come away convinced) the doctrine is, indeed, quite clear. But it's worth emphasizing that (as Parts III and IV will show in detail) courts and lawyers routinely misunderstand or ignore this doctrine, or wrongly believe it to be more narrowly confined than it is. Moreover, even for those who are generally on board with the legal principles laid out here, it's worth taking some time to think carefully about their contours and limitations, so as to understand whether any of these practices and statistical justifications can be distinguished from those the Court has rejected. Looking at this body of doctrine as a whole—across legal and factual contexts—can help us to do so, and it's a project that, surprisingly, legal scholars have not engaged in.

A.   The Prohibition of Statistical Discrimination

In both constitutional and statutory contexts, the Supreme Court has repeatedly endorsed what I call the "prohibition of statistical discrimination": with very limited exceptions, otherwise-illegal discrimination on the basis of certain types of classifications cannot be justified by statistical generalizations about groups, even if the generalizations are empirically supported. This principle isn't laid out in precisely those terms by the Court, but it's an animating theme of its equality jurisprudence, which consistently rejects statistical defenses of race and sex discrimination,[44] as well as discrimination against indigent criminal defendants. Here, I'll lay out this case law and its limits.

---

[44] *See, e.g.*, Frontiero v. Richardson, 411 U.S. 677, 690–91 (1973).

Consider *Craig v. Boren*,[45] one of the Court's early sex discrimination cases. *Craig* involved a challenge to a law that imposed a higher drinking age on men than on women.[46] To defend it, the state submitted statistical evidence that young men caused drunk driving accidents at more than ten times the rate of young women.[47] The Court deemed these statistics irrelevant, deeming it unfair to many young men who *don't* drive drunk to lump them in with those who do:

> [P]rior cases have consistently rejected the use of sex as a decisionmaking factor even though the statutes in question certainly rested on far more predictive empirical relationships . . . [P]roving broad sociological propositions by statistics is a dubious business, and one that inevitably is in tension with the normative philosophy that underlies the Equal Protection Clause.[48]

The "prior cases" in question were a series of seminal challenges to sex classifications that rested on empirical assumptions about sex differences. For example, both *Frontiero v. Richardson*,[49] in 1973, and *Weinberger v. Wiesenfeld*,[50] involved administrative schemes that applied a rebuttable presumption that married women depended financially on their husbands.[51] This presumption had indisputable statistical support, yet the Court called it an "overbroad generalization[ ] that could not be tolerated under the Constitution."[52] Why? Because it was unfair to families that it didn't accurately describe: "gender-based generalization cannot suffice to justify the denigration of the efforts of women who do work and whose earnings contribute significantly to their families' support."[53] These cases emphasize that equal protection law protects *individuals*' right to be treated as such, not subjected to group-based generalizations.[54] This individualistic approach has some justly criticized disadvantages, as I'll discuss below. But it's inescapably core to existing doctrine—what the *Craig* Court meant by "the normative philosophy that underlies the Equal Protection Clause."[55]

The prohibition of statistical discrimination has been reiterated in many other cases. In *United States v. Virginia*,[56] the Court struck down the Virginia

---

[45] 429 U.S. 190 (1976).

[46] *See id.* at 192.

[47] *See id.* at 200–01.

[48] *Id.* at 202–04.

[49] 411 U.S. 677 (1973).

[50] 420 U.S. 636 (1975).

[51] *See Frontiero*, 411 U.S. at 681; *Wiesenfeld*, 420 U.S. at 644.

[52] *Id.* at 643 (quoting Schlesinger v. Ballard, 419 U.S. 498, 507 (1975)).

[53] *Id.* at 643.

[54] *See* Sonja B. Starr, *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*, 66 STAN. L. REV. 803, 827 (2014); Richard A. Primus, Equal Protection and Disparate Impact: Round Three, 117 HARV. L. REV. 493, 553 (2003).

[55] *Craig*, 429 U.S. at 204.

[56] 518 U.S. 515 (1996).

Military Institute's exclusion of women.[57] Virginia had offered, and the district court credited, expert testimony that VMI's "adversarial" method was typically ill-suited to women.[58] The Court deemed this evidence constitutionally insufficient *even if accurate*.[59] The Court wrote that permissible justifications for gender discrimination

> must not rely on overbroad generalizations about the different talents, capacities, or preferences of males and females . . . It may be assumed, for purposes of this decision, that most women would not choose VMI's adversarial method. . . . [T]he question is whether the Commonwealth can constitutionally [exclude] women who have the will and capacity . . .[60]

Similar cases abound.[61] Moreover, the Court's language in *Virginia* suggests that this restriction is *additional to* the other requirements of intermediate scrutiny. It doesn't become okay to rely on statistical generalizations when an important state interest is at stake; if it were, surely the prevention of drunk-driving deaths in *Craig* would have qualified. Rather, the prohibition constrains the types of arguments that can be put forth to show a substantial relationship to such an interest.

Much of the key case law involves sex, not race, but if anything, the prohibition on statistical discrimination seems stronger in the race context, where strict(er) scrutiny applies. Defendants accused of race discrimination typically deny it, rather than defend it, so statistical defenses do not often arise.[62] But in *Virginia*, the Court observed that "[s]upposed 'inherent differences' are no longer accepted as a ground for race or national origin classifications," whereas "inherent differences" *could* support sex classifications in some limited contexts.[63] Indeed, the only statistical general-

---

[57] *See id.* at 519.

[58] *See id.* at 524.

[59] *See id.* at 550.

[60] *Id.* at 533, 542.

[61] *See, e.g.*, J.E.B. v. Alabama ex rel. T.*B.*, 511 U.S. 127, 139 n.11. (1994) (rejecting gender-based peremptory strikes notwithstanding empirical claims about gender predicting voting); *see also id.* at 148–49 (O'Connor, J., concurring) (agreeing with this conclusion despite a "plethora of studies"); Nev. Dep't of Hum. Res. v. Hibbs, 538 U.S. 721, 738 (2003) (finding pattern of unconstitutionality in practices relying on gendered assumptions about caregiving, observing that the "faultline between work and family [is] precisely where sex-based overgeneralization has been and remains strongest.").

[62] *But see* Batson v. Kentucky, 476 U.S. 79, 98 (1986) (holding unconstitutional peremptory juror challenges grounded in race-based assumptions about voting tendencies); Palmore v. Sidoti, 466 U.S. 429, 433 (1984) (holding that stepparent's race could not be considered in custody proceeding, notwithstanding the documented existence of widespread prejudice against mixed-race families).

[63] *Virginia*, 518 U.S. at 533. The Court's language offering "celebration" of the differences between the "two sexes" has not aged terribly well in an era with a richer sense of sex and gender possibilities, but its core points opposing the use of generalizations that inflict subordination remain good law.

izations the Court has sometimes accepted in heightened-scrutiny cases have involved physical differences related to childbearing.[64]

In *Buck v. Davis*,[65] the Supreme Court resoundingly rejected race-based statistical discrimination.[66] *Buck* was a capital habeas case in which a psychiatric expert (astoundingly, called by the defense) had testified that defendant Buck's Black race put him at higher statistical risk of future dangerousness, citing racial disparities in arrest and incarceration.[67] The Supreme Court held that introducing this testimony was an egregious and prejudicial mistake by counsel, notwithstanding the expert's ultimate conclusion of non-dangerousness.[68] The Court first observed that if the *prosecution* introduced similar evidence, the case would be even more straightforward: "It would be patently unconstitutional for a state to argue that a defendant is liable to be a future danger because of his race."[69] It then fleshed out the reason such predictive claims are so harmful:

> Here was hard statistical evidence—from an expert—to guide an otherwise speculative inquiry. And it was potent evidence . . . [that] appealed to a powerful racial stereotype—that of black men as "violence prone." . . . [This] opinion coincided precisely with a particularly noxious strain of racial prejudice. . . . For these reasons, we cannot accept the District Court's conclusion that "the introduction of any mention of race" during the penalty phase was "de minimis." . . . Some toxins can be deadly in small doses.[70]

The Court has similarly rejected statistical justifications for discrimination against indigent criminal defendants. Socioeconomic discrimination is in most cases subject only to rational basis review. But in the criminal context, decisions drawing on equal protection and due process principles have applied a distinct, demanding standard of scrutiny to discrimination against indigent defendants.[71] In *Bearden v. Georgia*, petitioner had his probation revoked when he lost his job; the Court unanimously deemed this unconstitutional wealth discrimination.[72] While the case is most remembered for its discussion of ability to pay restitution, it has a crucial passage on statistical

---

[64] *See, e.g.*, Nguyen v. Immigr. & Naturalization Serv. 533 U.S. 53, 73 (2001); Michael M. v. Superior. Ct of Sonoma Cnty., 450 U.S. 464, 467 (1981). In the Title VII context, the Court has rejected similar distinctions. *See, e.g.*, Int'l Union, United Auto., Aerospace & Agric. Implement Works of America UAW vs. Johnson Controls, Inc., 499 U.S. 187, 211 (1991) (holding that sex is not a bona fide occupational disqualification from jobs with fetal-endangering lead exposure).

[65] 137 S. Ct. 759 (2017).

[66] *See id.* at 777. *Buck* was a 6–2 decision; the dissent focused on procedural questions and did not dispute the impermissibility of the race testimony.

[67] *See id.* at 768.

[68] *See id.* at 777.

[69] *Id.* at 775 (citing Zant v. Stephens, 462 U.S. 862, 885 (1983) for the proposition that race arguments are "constitutionally impermissible or totally irrelevant to the sentencing process").

[70] *Id.* at 776–77.

[71] *See, e.g.*, Griffin v. Illinois, 351 U.S. 12, 17 (1956) (plurality opinion); Starr, *supra* note 54, at 830–34 (discussing this line of cases).

[72] *See* Bearden v. Georgia, 461 U.S. 660, 663, 672–73 (1983).

discrimination as well. The state had argued that the petitioner's job loss and resulting poverty put him at higher risk of recidivism, supporting this claim with "several empirical studies."[73] The Court did not question these studies' validity, but squarely rejected this argument nonetheless:

> This is no more than a naked assertion that a probationer's poverty by itself indicates he may commit crimes in the future. . . . [T]he State cannot justify incarcerating a probationer who has demonstrated sufficient bona fide efforts to repay his debt to society, solely by lumping him together with other poor persons and thereby classifying him as dangerous. This would be little more than punishing a person for his poverty.[74]

This resistance to "lumping" is the prohibition of statistical discrimination in action.

These are all constitutional cases, but the Court has similarly rejected statistical justifications for statutorily prohibited discrimination. For example, in *City of Los Angeles Department of Water & Power v. Manhart*,[75] a Title VII case, the Court held that an employer could not rely on women's higher life expectancy to require them to pay higher pension-plan premiums.[76] The Court observed that the life-expectancy generalization was "unquestionably true: Women, as a class, do live longer than men."[77] However, because not *all* women live longer than *all* men, the Court found that this did not provide a sound basis for disparate treatment; Title VII "precludes treatment of individuals as simply components of a racial, religious, sexual, or national class. . . . [even based on] a true generalization."[78] Discussing *Manhart* recently in *Bostock v. Clayton County*,[79] the Court emphasized that sex discrimination doesn't cease to be sex discrimination when it's labeled a "life expectancy adjustment," and also rejected the idea that fairness *requires* sex-specific standards in order to ensure that predictions for women and men are on average equally accurate.[80]

Beyond the applicable Supreme Court doctrine, in its 1991 amendments to Title VII, Congress specifically banned race-norming of employment-related tests. This ban was spurred by practices designed for equal opportunity purposes.[81] At the center of the controversy was the General Aptitude Test Battery (GATB), which was promoted by the Department of Labor. Expecting employers to be reluctant to use a test that could expose them to dispa-

---

[73] *Id.* at 671 n.11.
[74] *Id.* at 671.
[75] 435 U.S. 702 (1978).
[76] *See id.* at 722.
[77] *Id.* at 707.
[78] *Id.* at 707–09.
[79] 140 S. Ct. 1731 (2020).
[80] *Id.* at 1740.
[81] *See* Linda S. Gottfredson, *The Science and Politics of Race Norming*, 49 Am. Psych. 955, 955 (1994).

rate impact liability, the Department's solution was race-norming of the reported percentile ranks of each test-taker.[82] This was done quietly, but when the Department of Justice learned of the practice, it threatened to sue the responsible office at Labor.[83] In the 1991 Civil Rights Act, acting with "strong public support and virtually no opposition," Congress banned race norming of employment tests.[84] Note that before this legislation, this use of race-norming was legally permissible because, in general, federal employment discrimination law permits affirmative action.[85] Even before Congress acted, it would have been illegal for employers to adopt statistical adjustments that *disadvantaged* minority applicants.[86]

The scope of the prohibition of statistical discrimination has limits; it is not truly an absolute "prohibition," although I use that term as a shorthand.[87] It doesn't sweep *more* broadly than prohibitions of other forms of intentional discrimination. It simply rejects the invocation of statistical generalizations to defend discrimination along certain vectors, especially race and sex discrimination, as well as socioeconomic discrimination in the context of criminal justice. It is less absolute with respect to sex discrimination (particularly regarding physical sex differences) than with respect to race (at least outside the affirmative action context). This too tracks the strictness of the underlying substantive norm: sex discrimination receives only intermediate constitutional scrutiny, is not prohibited at all by some antidiscrimination statutes that reach only race, and is less strictly prohibited by others than race is.[88]

---

[82] *See id.* at 956.

[83] *See id.* at 956–58.

[84] *Id.* at 955.

[85] *See* 29 C.F.R. § 1608 (2022).

[86] Another paradigmatic example of private-sector statistical discrimination is race-based risk rating by insurance companies, which was ubiquitous before being gradually abandoned in the mid-twentieth century. It is now widely viewed as illegal, although a recent review found a surprising absence of specific statutes in many jurisdictions. Ronen Avraham, Kyle D. Logue & Daniel Schwarcz, *Understanding Insurance Antidiscrimination Laws*, 87 S. Cal. L. Rev. 195, 240–44 (2014). It's plausible that race-based risk rating violates 42 U.S.C. § 1981, which covers private contracting, but there's virtually no case law. *See id.* at 242–43. *Compare* Guidry v. Pellerin Life Ins. Co., 364 F. Supp. 2d 592, 599 (W.D. La. 2005) (finding, in my view wrongly, no race discrimination if race is merely one factor shaping a risk rating) *with* Thompson v. Metropolitan Life Ins., 149 F. Supp. 2d 38, 40 (S.D.N.Y. 2001) (denying motion to dismiss a Section 1981 challenge to an insurer's past racially discriminatory policies). The *Thompson* defendant soon settled, as did other insurers faced with similar suits. Mary L. Heen, *Ending Jim Crow Life Insurance Rates*, 4 Nw. J. L. Soc. Pol'y 360, 360–61 (2009). Health insurers *are* prohibited by the Affordable Care Act from sex-based rate-setting, and insurers in several other fields widely decline to do so. *See* Avraham et al., *supra*, at 213–14.

[87] Aziz Huq argues that the federal courts' treatment of statistical discrimination is "hesitant and equivocal." Aziz Z. Huq, *What Is Discriminatory Intent?*, 103 Cornell L. Rev. 1211, 1248 (2018). But the prohibition is quite clear at the Supreme Court level, although I certainly agree (as this Article argues) that courts don't consistently follow it.

[88] Title VII of the Civil Rights Act allows sex, but not race, to be a bona fide occupational qualification under limited circumstances. Meanwhile, two of the other major statutes discussed in this Article (Title VI of the Civil Rights Act and 42 U.S.C. § 1981) do not cover sex discrimination.

But it's not *generally* illegal for the government or private actors to rely on statistical generalizations, nor are such generalizations always irrelevant to a practice's legal permissibility.[89] For example, universities rely on grades for admissions, explicitly or implicitly assuming that they predict college performance. Countless other examples pervade public and private decision-making every day. Generalizations like these are routinely imperfect, but nobody thinks that imperfection poses a constitutional problem. And indeed, statistical generalizations may often provide a rationale that establishes that such classifications aren't completely arbitrary or dispels suspicions of underlying animus, and thus helps them to survive rational basis review.

For some classifications, the permissibility of statistical generalizations may be more complicated. Consider age discrimination. Legal age classifications, which are pervasive, typically explicitly or implicitly turn on broad-brush empirical generalizations—for example, about maturity (e.g., driving and drinking ages and the age of majority), or about how aging will influence health, work, and retirement choices (e.g., Medicare and Social Security eligibility). While much policy debate surrounds some of these cutoffs, few contend that the government should never generalize based on age or may not legally do so, and these rules trigger only rational basis review. Yet none of them would be easy to defend even under that deferential standard *absent* some kind of empirically plausible generalization. (A law barring only people in their fifties from drinking would probably be struck down as arbitrary, lacking a plausible reason.)

In constitutional law, statistical generalizations *do* make the difference in justifying age discrimination. Yet this is not so in employment law, thanks to the Age Discrimination in Employment Act ("ADEA"), which prohibits most statistical age discrimination. The ADEA endorses the principle that individuals should be judged "based on their ability rather than age."[90] It offers a bona fide occupational qualification (BFOQ) defense, which effectively permits some statistical discrimination, but this defense is "extremely narrow."[91] Age discrimination illustrates a broader point: whether statistical generalizations are permitted as justifications typically tracks (and helps to define) the scope and stringency of the underlying restriction on discrimination.

Finally, one might wonder whether the doctrinal picture I have painted is likely to change. The Supreme Court has become more conservative, in-

---

[89] SCHAUER, *supra* note 9, offers a strong defense of statistical profiling (and generally, reliance on probabilistic statistical evidence to make decisions) outside the context of race, gender, and sexual orientation, and I have no objection to many of the examples he provides. Broadly speaking, however, I don't share his worry, which motivates the book, that society is too *intolerant* of group generalizations. I'm much more concerned that, as this Article shows, we often tolerate them inappropriately even in the context of race and other vectors of subordination.

[90] 29 U.S.C. § 621(b) (listing congressional findings).

[91] *See, e.g.*, W. Air Lines v. Criswell, 472 U.S. 400, 412–15 (1985) (upholding jury's rejection of age BFOQ for flight engineers).

cluding on matters of racial equality, and virtually everyone expects more shoes to drop soon.[92] But these changes in the Court don't provide any reason to expect it to be more open to statistical discrimination. If anything, the Court has become more allergic to race-conscious governmental decision-making, and more committed to its individualistic Equal Protection Clause vision. This trend may be most visible in cases where race-consciousness cuts in the direction of reducing racial disparities, a point I consider further below. But *Buck v. Davis*, for example, is a recent case that strongly rejected an anti-Black use of a racial classification. Even beyond race, as *Bostock* also suggests, I think we can expect the Supreme Court to continue to hold a fairly stringent line against—if nothing else—the use of explicitly discriminatory classifications.[93]

B.   Why This Principle Matters

An aversion to statistical justifications for discrimination is a feature of our law—and should be. Adherents of a wide range of perspectives on equality law should be able to agree on this point, which has enjoyed the support of Supreme Court justices across the ideological spectrum.[94] Without this principle, indeed, there would be little left of even the most basic protections of equality law.

First, let's consider the principle from the perspective that dominates U.S. constitutional law: the anticlassification approach. This should require little elaboration, as the above-discussed doctrine is already grounded in this perspective. The anticlassification approach is fundamentally individualistic; it protects individuals treated adversely because of their race, sex, or other protected status. From this perspective the prohibition of statistical discrimination is a *defining* feature of legal equality. If the harm of classification is the lumping in of the individual with the group, it obviously cannot be defended via a statistical generalization about the group.

But the anticlassification approach has many critics. Scholars writing from an "antisubordination" perspective have argued that the focus on classifications misses the point; what we should worry about is enforcement of

---

[92] *See, e.g.*, Nina Totenberg, *Can Race Play a Role in College Admissions? The Supreme Court Hears the Arguments*, NPR (Oct. 31, 2022), https://www.npr.org/2022/10/31/1131789230/supreme-court-affirmative-action-harvard-unc [https://perma.cc/9L5W-2MQR].

[93] In *Bostock*, which extended Title VII to cover sexual orientation and gender identity, the defendants didn't make a statistical argument. But the Court endorsed *Manhart*'s reasoning and if anything rejected actuarial fairness reasoning even more explicitly. *See Bostock*, 140 S. Ct. at 1748.

[94] Consider the authors of some of the opinions cited above: *Hibbs* was written by Chief Justice Rehnquist, *Bostock* by Justice Gorsuch, *Virginia* by Justice Ginsburg, *Craig*, *Weinberger*, and *Frontier* by Justice Brennan, and *Manhart* by Justice Stevens. Some, like *Bearden* (written by Justice O'Connor) and *Palmore* (written by Chief Justice Burger), were unanimous. *See Hibbs*, 538 U.S. at 721; *Bostock*, 140 S. Ct. at 1731; *Virginia*, 518 U.S. at 515; *Craig*, 429 U.S. at 190; *Wiesenfeld*, 420 U.S. at 636; *Frontiero*, 411 U.S. at 677; *Manhart*, 435 U.S. at 702; *Bearden*, 461 U.S. at 660; *Palmore*, 466 U.S. at 429.

racial and other hierarchies.[95] Contemporary scholars of systemic racism and critical race theory have likewise emphasized the counterproductiveness of "colorblindness" as an approach to redressing racism.[96] They argue that the anticlassification approach, on the one hand, fails to recognize the way seemingly "neutral" actions can exacerbate deeply embedded social inequalities. And on the other hand, its focus on classifications themselves, rather than what they are used for, means that it interferes with conscious efforts to redress those inequalities (for example, affirmative action).

A reader who sympathizes with these critiques might understandably wonder whether the prohibition of statistical discrimination should be rejected instead of celebrated. This, however, would be the wrong conclusion. The prohibition does crucial work in advancing substantive equality objectives, and antisubordinationists need not reject it just because conservatives embrace it; this should be common ground. I'll return in Part VI to the question of equality-promoting uses of race-conscious statistical tools. Let's set that aside for now, because those tools' permissibility doesn't turn on what we think about statistical justifications for discrimination per se. The anti-statistical-discrimination principle, as I've defined it, is that otherwise-impermissible discrimination typically can't be justified by group generalizations. The debate about affirmative action, other race-conscious policymaking, and colorblindness is about what should count as otherwise-impermissible discrimination.

When classifications are deployed in a way that amplifies disparities, they are problematic from an antisubordination perspective, not just from an anticlassification one. Statistical justifications don't get around those impacts, and indeed, can even make those them worse. For example, generalizations about racial groups are often expressively noxious; this is illustrated by the NFL's claims about race and intelligence, and we'll see other examples below. Gender generalizations, too, tend to reinforce traditional gender roles that have historically been used to subordinate.

Indeed, if courts *did* accept statistical justifications for discrimination, constitutional and statutory protections against disparate treatment would unravel to nearly nothing. The examples of statistical discrimination that this Article highlights are unusually explicit in form—they involve quantified metrics, plainly identifiable as "statistical." But the concept of statistical discrimination is much broader than that (as many of the above-discussed cases illustrate). Economists use the phrase to refer generally to discrimination motivated by beliefs about group differences (especially beliefs that have some empirical support), as opposed to sheer animus. As I discuss in

---

[95] See, *e.g.*, Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 Phil. & Pub. Aff. 107, 157–58 (1976); Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 Univ. of Miami L. Rev. 9, 9 (2003).

[96] For one example of this large literature, see generally Eduardo Bonilla-Silva, Racism Without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America (5th ed. 2022).

Part V, many economists argue that *most* discrimination in labor markets and other contexts is statistical. And even when sheer animus *is* at work, it would often be difficult for a plaintiff to prove it.

In a society deeply shaped by centuries of racial oppression and other inequalities, countless persistent disparities pervade nearly every aspect of life. These disparities could in turn produce potential statistical justifications for discrimination. For example, suppose a hiring manager defended his reluctance to hire young women because they are statistically more likely than young men to take costly maternity leave. Or suppose a police officer admitted considering race when deciding who to search, citing statistics showing racial disparities in criminal justice involvement. Or suppose a loan officer could point to data showing racial disparities in default rates—very plausible given the intertwining of race and poverty—and used that to justify race discrimination in lending.

The implication of the case law discussed in Section A is that none of these arguments could excuse discrimination even if the discriminator could cite empirical support.[97] This is a crucial principle, and not just from an anticlassification perspective. Without it, existing disparities could be used to rationalize continued discrimination that, in a vicious cycle, amplifies them. In addition, it would be easy for statistical claims to be invoked as pretexts to justify discrimination that *is* based on animus. Proof problems already bedevil disparate-treatment cases, even when the only demand is to show *that* the defendant discriminated, not *why* they did so.

Some readers may protest that *accurate* statistical discrimination can lead to better predictions about individuals, and that better predictions can serve a variety of social purposes (for example, better medical care or more effective policing). As I'll discuss further in Part V, many economists have advanced this view. In my view, this counterargument cannot outweigh the above-discussed costs of legally tolerating statistical discrimination against disadvantaged racial groups, women, and the poor: expressive harms, exacerbation of inequality, and the unraveling of legal protections against disparate treatment even when the statistical justification is pretextual. I'll develop this position below in the context of this Article's principal examples, each of which involve putative defenses along these lines.

But even if one *does* prioritize predictive accuracy over equality objectives, there's good reason to worry about tolerating statistical discrimination. As the examples discussed in Parts III and IV will demonstrate, it is *very* easy for purportedly accurate statistical discrimination to drift into *inaccurate* discrimination—or into the misuse of statistical evidence of group differences for purposes that don't serve the interests they're supposedly meant

---

[97] The Supreme Court has never squarely addressed whether the Fourteenth Amendment prohibits police racial profiling, which is deeply unfortunate, especially because its Fourth Amendment case law has made it easier for police to get away with it; still, the principles discussed in Section A plainly prohibit it. *See* Sonja B. Starr, *Testing Racial Profiling: Empirical Assessment of Disparate Treatment by Police*, 2016 U. CHI. L. REV. 484, 488–93 (2016).

to serve. We'll see this again and again—for example, lost-earnings calculations based on badly outdated data, and IQ adjustments in the *Atkins* context based on sloppy, pseudoscientific practices. We'll see it even in the medical context, in which racial adjustments are made not for strategic litigation purposes, but rather to try to improve treatment. You might thus expect the evidence supporting them to be strong, but we'll see many examples in which concerning health disparities are inappropriately essentialized as "normal" group differences, with results that field experts are now beginning to recognize as medically counterproductive.

These aren't, I think, readily dismissible as unfortunate cases of "statistical discrimination done incorrectly." Rather, the proliferation of empirically mistaken practices is a predictable hazard of tolerating statistical discrimination, especially based on categories regarding which rank stereotypes have so long pervaded our society. In a world awash in stereotypes (and also awash in badly conducted studies, strategically manipulated empirical claims, and innumeracy), discriminating agents often cannot or will not distinguish between statistically accurate and inaccurate bases for discrimination. Economists have begun to write about this issue, too, offering models of how stereotypes and inaccurate statistical discrimination can take hold even among purportedly rational actors.[98]

The high likelihood of inaccuracy links to one of the main stated reasons courts apply heightened scrutiny to certain types of classifications in the first place. As the Supreme Court has put it, classifications triggering strict scrutiny are very "seldom relevant to the achievement of any legitimate state interest."[99] This isn't because those employing them never put forth any claim to empirical justification, however; it's because the claims regularly fail to hold up. Claims about essential racial differences (a crucial historical foundation for white supremacy) have time and again been exposed as false and are even more presumptively dubious.[100] Claims of sex difference, meanwhile "very likely reflect outmoded notions of the relative capabilities of men and women," and thus courts approach them with skepticism.[101]

For all these reasons, it's fortunate that the line between not liking a group and holding negative beliefs about it is not, generally, one that the law recognizes. Our governing constitutional doctrine recognizes only a narrow vision of equality. But it is, at least, not as narrow as that.

---

[98] *See infra* note 328.

[99] *City of Cleburne*, 473 U.S. at 440.

[100] The literature critiquing racial essentialism and documenting its impact is vast. For one recent example in the medical context, see Jennifer Tsai, *How Should Educators and Publishers Eliminate Racial Essentialism?*, AMA J. ETHICS (Mar. 1, 2022 https://edhub.ama-assn.org/ama-journal-of-ethics/module/2789184 [https://perma.cc/SEU2-CM59].

[101] *City of Cleburne*, 473 U.S. at 441.

C. The NFL Scandal and the Law

With this background in mind, consider again Davenport and Henry's suit against the NFL under 42 U.S.C. § 1981. Let's set aside the procedural basis for its dismissal, and focus on the substantive question: was what the NFL did unlawful racial discrimination? The above-discussed case law implies a clear yes. Section 1981 applies to racial discrimination in contracting, including the enjoyment of benefits under private contracts, which would encompass settlement administration. Race must be a but-for cause of the defendant's injury,[102] but it appears that Davenport, Henry, and many other players could readily show that their settlement payouts were affected by their race, and indeed, as noted above, many claims previously denied have now been granted.[103]

Under this statute, racial discrimination is subject to strict scrutiny, because the Supreme Court has described Section 1981 as "coextensive with the Equal Protection Clause."[104] These cases involved government defendants, and the Court has provided no guidance on applying strict scrutiny to private contracting (for example, defining the equivalent of a "compelling state interest"). But it's hard to imagine the NFL's race-norming satisfying any variant of strict scrutiny. It's not clear what interest could plausibly be "compelling." Even if diagnostic accuracy could meet that standard (a stretch here), race-norming wasn't narrowly tailored to satisfy that interest. It was possible to use non-race-normed standards (as some doctors sought to do and as the new settlement framework requires) or other approaches discussed in medical literature, which I'll discuss in Part IV.

Most glaringly, any attempt to get the NFL's race-norming past strict scrutiny would run afoul of the prohibition of statistical discrimination. The NFL's defense of race-norming depended on the statistical generalization that race predicts cognitive ability. Indeed, the generalization that Black people are less intelligent is at least as repugnant as any that the Court has considered and rejected.

In short, the race-norming of players' test results wasn't just wrong and shocking; it was illegal. And as we'll see, the reasons this is so are also true of other practices that are still, somehow, pervasive.

---

[102] *See* Comcast Corp. v. Nat'l Ass'n of Afr. Am.-Owned Media, 140 S. Ct. 1009, 1021 (2020).

[103] For a useful overview of Section 1981 case law, see U.S.C.A. for the Third Circuit, Instructions For Race Discrimination Claims Under 42 U.S.C. § 1981 (Oct. 2014), https://www.ca3.uscourts.gov/sites/ca3/files/6_Chap_6_2014_fall.pdf [https://perma.cc/G2ZX-6QDX].

[104] Grutter v. Bollinger, 539 U.S. 306, 343 (2003); *see also* Gen. Bldg. Contractors Ass'n v. Pa., 458 U.S. 375, 389–90 (1982).

III.   EXPLICIT STATISTICAL DISCRIMINATION IN LEGAL SETTINGS

Notwithstanding the clarity with which the Supreme Court has often condemned statistical discrimination, explicit and high-stakes examples of it persist, mostly without legal challenge. Here I discuss three examples within the justice system itself: use of race- and sex-based earnings and life-expectancy predictions for damages calculations; use of race-adjusted intelligence tests to determine death-penalty eligibility; and use of socioeconomic factors in criminal justice risk assessments. I argue that each runs afoul of the doctrine described above, and is also indefensible from a policy perspective.

A.   Calculation of Compensatory Damages and Restitution

Civil plaintiffs are often awarded damages for lost future earnings, calculated based on experts' counterfactual estimates of what they would have earned absent their injury. Those experts (usually forensic economists) generally rely on actuarial tables for key terms in these calculations: expected future wages, life expectancy, and work-life expectancy (remaining years of work).[105] Such tables also shape other components of compensatory damages—for example, medical-expense awards often account for life expectancy.[106] Routinely, these experts use race- and sex-specific actuarial tables to make these calculations. In a few states, this practice is required by statutes or included in pattern jury instructions.[107] The same practices often shape criminal restitution awards.[108] These practices produce, in otherwise-identical cases, lower awards for nonwhite plaintiffs and for women, for no reason other than race and sex.

Much lower, in fact. For example, in *United States v. Bedonie* (a 2004 decision on two homicide cases involving restitution), an expert's calculation of the expected earnings of the Native American male victim was fifty-eight percent of what a race-neutral calculation would have been.[109] For the other victim, a Native American baby girl, the expert produced estimates ranging from forty-four to fifty-five percent (depending on assumptions about her education) of the corresponding estimates unadjusted by race and sex.[110]

---

[105] *See generally* Thomas R. Ireland, *The Role of a Forensic Economist in a Damage Assessment for Personal Injuries*, in MEASURING LOSS IN CATASTROPHIC INJURY CASES 15 (Kevin S. Marshall & Thomas R. Ireland eds., 2006).

[106] See, *e.g.*, Smith v. United States Dep't of Veterans Affs., 865 F.Supp. 433, 441 (1994).

[107] *See, e.g.*, Colo. Rev. Stat. § 13-25-102 (2016); Ga. Code Ann. § 24-14-44 (2013); 9 R.I. Gen. Laws § 9-19-38 (2012); Kan. Civil Pattern Jury Instructions § 171.45 (2016); Ky. Wrongful Death Actions § 13:3 (2012-13 ed.) (Life Expectancy Tables); N.D. Civil Pattern Jury Instructions § 70.47 (2012) (Personal Injury); 8 Tenn. Civil Pattern Jury Instructions, App. C (2012).

[108] *See, e.g.*, United States v. Bedonie, 317 F. Supp. 2d 1285 (D. Utah 2004), *rev'd sub nom.* United States v. Serawop, 410 F.3d 656 (10th Cir. 2005) (reversal was based on the criminal conviction and not the life expectancy calculation).

[109] *See Bedonie*, 317 F.2d. at 1313.

[110] *See id.* at 1314.

This method bakes into every award the cumulative effect of centuries of racism and sexism. Sharp disparities in earnings and life expectancy result from discrimination and structural inequality in the job market, health care, and other dimensions of life. The prevailing method of damage calculations makes these disparities define what an individual's life is worth. Moreover, the calculations turn back the clock on progress in redressing disparities, because disparities observed among previous generations generate projections for future ones.[111]

This practice has persisted for many decades yet has received vanishingly little judicial scrutiny. A review was published recently by Professors Ronen Avraham and Kimberley Yuracko.[112] Neither their review nor my own search produced any examples of appellate decisions reversing a trial court's decision to admit race- or sex-specific calculations, or otherwise identifying constitutional problems with this practice.[113] A handful of decisions (including *Bedonie*) involve trial courts excluding race-specific tables from evidence and/or permitting neutral tables; although clearly motivated by discomfort with the equity implications of using race-specific tables, these decisions have not weighed in on that method's constitutionality.[114] For example, in *Wheeler Tarpeh-Doe v. United States*,[115] the district court held that it would be "inappropriate" to incorporate "discrimination" into a damage calculation, and observed that the practice's application to the biracial plaintiff presented additional problems.[116] In *Bedonie*, the court noted the constitutional concern but, citing the principle of constitutional avoidance, invoked its discretion to choose an alternate approach.[117] In *Reilly v. United States*,[118] the district court rejected sex-specific tables because the old data that generated them undermined their probative value, and the First Circuit affirmed,[119] but no constitutional arguments were raised.[120]

---

[111] *See infra note* 101 and accompanying text.

[112] *See* Kimberly A. Yuracko & Ronen Avraham, *Valuing Black Lives: A Constitutional Challenge to the Use of Race-Based Tables in Calculating Tort Damages*, 106 Calif. L. Rev. 325, 327–29 (2018).

[113] *Id.* at 329 (stating that "use of race-based tables in the calculation of tort damages remains both standard and largely unnoticed"). I cannot definitively prove a negative, but I searched for a range of terms likely to appear in such cases, and for subsequent cases citing each of the critical cases discussed here (none of which themselves cite such an opinion).

[114] *See, e.g.*, *Bedonie*, 317 F. Supp. 2d at 1319, *rev'd and remanded sub nom. Serawop*, 410 F.3d 656; Reilly v. United States, 665 F. Supp. 976, 997 (D.R.I. 1987); *see also* Martha Chamallas & Jennifer B. Wriggins, The Measure of Injury: Race, Gender, and Tort Law 156 (2010) (discussing Sept. 11 Victim Compensation Fund decision).

[115] 771 F. Supp. 427 (D.D.C. 1991).

[116] *Id.* at 455–56.

[117] *See Bedonie*, 317 F. Supp. 2d at 1319; *see also Serawop*, 505 F.3d at 1126–27 (affirming this exercise of discretion).

[118] 665 F. Supp. 976 (D.R.I. 1987).

[119] 863 F.2d 149, 167 (1st Cir. 1988) (noting the "antiquated premise" that "women will absent themselves from the work force for prolonged intervals").

[120] *See Reilly*, 665 F. Supp at 997.

The two most notable decisions that *did* explicitly hold race-specific tables unconstitutional both were written by the late Judge Jack Weinstein of the Southern District of New York, one of the federal judiciary's most vigorous civil rights supporters. In *McMillan v. City of New York*,[121] he refused to allow race-specific life-expectancy estimates to shape medical damages.[122] The opinion included a sweeping discussion of the social construction of race and the history of racial essentialism.[123] The equal protection discussion itself was more concise, listing Supreme Court decisions overturning racial classifications and finding that relying on a race-specific calculation would constitute such a classification. [124]

In their 2010 book, Professors Martha Chamallas and Jennifer Wiggins argued that if this "landmark" case were "followed by other courts, [it] could significantly alter the valuation of tort claims."[125] But in the years since, that hasn't yet happened, at least not often.[126] In 2015, in *G.M.M. ex rel. Hernandez-Adams v. Kimpson*,[127] Judge Weinstein himself revisited the issue, as applied to a lifetime-earnings prediction involving a Hispanic child who suffered lead poisoning. Judge Weinstein interrupted the defense expert's testimony as to the expected educational attainment of Hispanics and ordered the expert to exclude this ethnic consideration from his calculations. His written opinion held: "Propelling race and ethnicity to the forefront of predictions about an individual's future achievement ignores the myriad factors affecting an individual's capacity to fulfill his or her potential."[128]

There also aren't many cases explicitly *upholding* the use of race- or sex-specific tables against constitutional challenges—perhaps because it appears that lawyers haven't usually objected to them, even when they disadvantage their clients.[129] In *Bedonie*, the criminal restitution case, the prosecution argued *for* race- and sex-specific calculations, even though these methods produced lesser restitution awards.[130] We only know how race and sex affected those calculations because the court, *sua sponte*, asked the expert to submit alternative calculations that did not account for them. As

---

[121] 253 F.R.D. 247 (E.D.N.Y. 2008).

[122] *See id.* at 256.

[123] *See id.* at 249–53.

[124] *See id*. at 255. The decision also makes a due-process argument, *see id.* at 255–56, which I do not focus on here.

[125] CHAMALLAS & WRIGGINS, *supra* note 114, at 156. The authors note that even Judge Weinstein relied on *male* life expectancy tables. *See id.* at 166.

[126] Magistrate Judge Robert Levy has done so twice with brief statements. See Sung-Ho Hwang v. Grace Rd. Church, 2018 WL 4921638, *7 fn.9 (E.D.N.Y. 2018) ("I have not considered plaintiff's race or ethnicity, as "'[r]ace-based statistics and other race-centric data cannot be relied upon' in calculating life expectancy"); Cedeno v. Broan-Nutone, LLC, 2019 WL 4751913, *10 fn.15 (E.D.N.Y. 2019) (same).

[127] 116 F. Supp.3d 126 (E.D.N.Y. 2015).

[128] *Id.* at 152.

[129] Even plaintiffs' proposed estimates are often discounted to their detriment. See, e.g., Bulala v. Boyd, 239 Va. 218, 232 (1990) (mentioning plaintiff's estimate being discounted by "age, race, and sex").

[130] *See Bedonie*, 317 F.2d. at 1313.

Chamallas and Wriggins observe, this was a rare move: "although the expert had performed thousands of lost-income analyses, he testified that he had never before been asked to provide race- and sex-neutral calculations in a wrongful death case."[131]

Legal scholarship and teaching materials in torts and remedies have also given little attention to the issue for decades, although there has been a small uptick recently; meanwhile, criminal sentencing literature completely ignores it. Until a few years ago, the only law scholars who had substantially focused on the question were Chamallas and Wriggins.[132] Avraham and Yuracko published two papers more recently, and Catherine Sharkey has offered an alternative suggestion: use of race- and sex-neutral calculations drawn from administrative cost-benefit analyses.[133] In their 2018 piece, Avraham and Yuracko surveyed sixteen torts and remedies casebooks and found that "only seven torts and two remedies casebooks even mention the role of race and gender in damage calculations, and among those nine, most give the issue only a few sentences worth of attention."[134] A handful of student notes have addressed the issue, mostly in the past few years.[135] The most detailed constitutional arguments are found in Chamallas's 1994 piece and Avraham and Yuracko's 2018 paper. Interestingly, Avraham and

---

[131] CHAMALLAS & WRIGGINS, *supra* note 114 at 160.

[132] *See generally* CHAMALLAS & WRIGGINS, *supra* note 114; Jennifer Wriggins, *Damages in Tort Litigation: Thoughts on Race and Remedies, 1865–2007*, 27 REV. LITIG. 37 (2007); Jennifer Wriggins, *Torts, Race, and the Value of Injury, 1900–1949*, 49 HOW. L.J. 99 (2005); Jennifer Wriggins, *Constitution Day Lecture, Constitutional Law and Tort Law: Injury, Race, Gender, and Equal Protection*, 63 ME. L. REV. 263 (2010); Martha Chamallas, *Civil Rights in Ordinary Tort Cases: Race, Gender and the Calculation of Economic Loss*, 38 LOY. L.A. L. REV. 1435 (2005); Martha Chamallas, *The September 11th Victim Compensation Fund: Rethinking the Damages Element in Injury Law*, 71 TENN. L. REV. 51 (2003); Martha Chamallas, *The Architecture of Bias: Deep Structures in Tort Law*, 146 U. PA. L. REV. 463 (1998); Martha Chamallas, *Questioning the Use of Race-Specific and Gender-Specific Economic Data in Tort Litigation: A Constitutional Argument*, 63 FORDHAM L. REV. 73 (1994) [hereinafter *Questioning the Use*].

[133] *See* Yuracko & Avraham, *supra* note 112 at 325 (2018); Ronen Avraham & Kimberly A. Yuracko, *Torts and Discrimination*, 78 OHIO ST. L.J. 661 (2017); Catherine M. Sharkey, *Valuing Black and Female Lives: A Proposal for Incorporating Agency VSL into Tort Damages*, 96 NOTRE DAME L. REV. 1479 (2021).

[134] Yuracko & Avraham, *supra* note 112, at 329 n.15 (providing details on all sixteen casebooks). The casebook the authors credit with the most detailed discussion is DOMINICK VETRI ET AL., TORT LAW AND PRACTICE (5th ed. 2016).

[135] *See generally* Anne M. Anderson, Note, *How Much Are You Worth?*, 73 WASH. & LEE. L. REV. ONLINE 206 (2016); Goran Dominioni, Note, *Biased Damages Awards: Gender and Race Discrimination in Tort Trials*, 1 INT'L COMP. POL'Y & ETHICS L. REV. 269 (2018); Agustin Paneque, Note, *Civil Rights and Tort Calculation: Challenging the Reliability and Constitutionality of Race-Based and Gender-Based Life Expectancy and Future Wage Earning Calculations*, 19 RUTGERS RACE & L. REV. 133 (2018); Dhruti J. Patel, Note, *Policing Corporate Conduct Toward Minority Communities: An Insurance Law Perspective on the Use of Race in Calculating Tort Damages*, 53 U. MICH. J.L. Reform 227 (2019); Loren D. Goodman, Note, *For What It's Worth: The Role of Race- and Gender-Based Data in Civil Damages Awards*, 70 VAND. L. REV. 1353 (2017); Sherri R. Lamb, Note, *Toward Gender-Neutral Data for Adjudicating Lost Future Earning Damages: An Evidentiary Perspective*, 72 CHI.-KENT L. REV. 299 (1996).

Yuracko do not rely on the constitutional doctrine surrounding statistical discrimination.[136] Taking that doctrine into account strengthens the case further, because without the ability to rely on statistical generalizations, any coherent defense of these tables falls apart.

In the existing literature, considerable attention is devoted to a threshold question: Is there state action, such that the Equal Protection Clause applies? Or is the discrimination here that of the private expert who performs the calculations? The two leading constitutional papers present this question as a significant potential hurdle to challenges, although both ultimately argue that there is state action.[137] Chamallas (1994) centers her discussion on the judge's decision to admit the race- and sex-specific tables, arguing that this constitutes an implicit legal endorsement.[138] Avraham and Yuracko observe that there are limits to the principle, associated with *Shelley v. Kramer,* that the state may not enforce private prejudices.[139] Exploring the post-*Shelley* case law, however, they find three principles that imply that *Shelley*'s holding should apply to the jury's reliance on expert calculations. First, state adoption of those calculations burdens "important social and economic interests" of plaintiffs; second, the state's adoption of these reports constitutes "symbolic encouragement" of discrimination because it incentivizes tortfeasors to commit harms in minority communities; and third, the reasons for the disparities reflected in the tables include prior state discriminatory action.[140]

In my view, the state action question has a more straightforward answer. It is the jury (or judge, in a bench trial), a state actor, that decides the amount of damages awarded.[141] That decision, if based on a race- or sex-

---

[136] Avraham and Yuracko do cite *Manhart*'s rejection of actuarial reasoning, but don't rely on it heavily because it is a Title VII case. Yuracko & Avraham, *supra* note 112, at 367 n.218. Chamallas's original constitutional argument does cite the Supreme Court's older statistical discrimination precedents. *See* Chamallas, *supra* note 132, at 120. *See also* CHAMALLAS & WIGGINS, *supra* note 114, at 158 (arguing that the "use of gender- and race-based tables saddles nonconforming individuals with generalizations about their group").

[137] *See* Anderson, *supra* note 135, at 229–50 (arguing that there is no state action).

[138] *See* Chamallas, *supra* note 132, at 106.

[139] *See* Avraham & Yuracko, *supra* note 133, at 350 (discussing Shelley v. Kraemer, 334 U.S. 1, 20 (1948)).

[140] *See id.* at 352–57. The authors, somewhat confusingly, use this third point to argue that the state's use of the report would "materially facilitate" private discrimination.

[141] That the jury itself is a state actor seems clear, although there is surprisingly little case law directly on point. To be sure, it's composed of private citizens and is often described as a bulwark *against* state power. But juries can authorize or bar the exercise of the most profound of state powers, implicating the most profound of other citizens' rights; consider their role in criminal cases, including capital cases. Many cases imply that juries may violate the Constitution. For example, in *McCleskey v. Kemp,* 481 U.S. 279 (1987), the Court held that statistical evidence alone could not establish that the jury in a particular case engaged in "constitutionally unacceptable" racial discrimination violating the Eighth Amendment, *id.* at 309, or intended to discriminate in violation of the Equal Protection Clause, *id.* at 292–97, but never suggested that the jury was not a state actor, and appeared to take for granted that if the jury *did* racially discriminate it would be unconstitutional. Likewise, in *Pena-Rodriguez v. Colorado,* 137 S. Ct. 855 (2017), the Court held that a state bar on post-verdict juror testimony could not be used to exclude testimony on juror racial bias; although the case was a Sixth

specific calculation, discriminates based on race or sex. The expert's report itself is neither a state action nor an act of discrimination. It's not even, like in *Shelley*, a discriminatory private legal instrument that a private party is asking the state to enforce. The report and accompanying testimony are just an *argument for state discrimination*, much like testimony before a legislature urging passage of a discriminatory law (which nobody would suggest makes the ensuing law any less "state action"). The expert has only influence; the jury has power. One need not parse the case law on *Shelley*'s limits, because the state's role here is primary in a way that it wasn't even in *Shelley* itself. The fact that the jury is following an expert suggestion when it issues a discriminatory verdict does not mean it's not the jury's own verdict.[142]

Perhaps one could respond: suppose the jury isn't told that the tables are race- and sex-specific? Or suppose the judge instructs the jury to rely on the expert's calculation? Then arguably the jury did not have the discriminatory purpose and/or the freedom of action that it would take to say that it discriminated unconstitutionally. But in either of these circumstances, something else has gone seriously wrong. If counsel has failed to object to discriminatory tables that disadvantage their client, that's a bad lawyering mistake. If the problem is that the judge is keeping key information from the jury or constraining its ability to choose a nondiscriminatory alternative (e.g., by excluding neutral tables), then the judge is effectuating discrimination.[143] Surely the application of the Equal Protection Clause cannot be defeated by one state actor forcing another to discriminate.[144]

Amendment challenge, the Court emphasized the important equal protection interests at stake in avoiding juror racial bias and referred to the use of voir dire to ensure that jurors "are free of racial bias" as an example of a necessary step to avoid "state-sponsored racial discrimination in the jury system." *Id.* at 867. In general, private citizens may engage in "state action" when their actions are sufficiently intertwined with the state; notably, private litigants' use of peremptory strikes are state action because "without the overt, significant participation of the government, the peremptory challenge system, as well as the jury trial system of which it is a part, simply could not exist." Edmondson v. Leesville Concrete Co., 500 U.S. 614, 622 (1991). This logic applies inescapably to the jury itself.

[142] By analogy, in *Palmore v. Sidoti*, 466 U.S. 429 (1984), the state court judge issued a custody decision that took race into account; it made no difference that one of the (private) parties had argued for this decision or that it cited the (private) prejudice of others in society. Unlike with *Shelley*, there is no controversy over whether *Palmore* involved state action. *See* Jody D. Armour, *Race Ipsa Loquitur: Of Reasonable Racists, Intelligent Bayesians, and Involuntary Negrophobes*, 46 STAN. L. REV. 781 (1994).

[143] This is somewhat similar to Chamallas's argument but does not depend on the claim that merely admitting a discriminatory argument into evidence constitutes a discriminatory state action.

[144] What about settlements based on race- and sex-based calculations? If it involved an individual plaintiff, the Equal Protection Clause probably wouldn't be implicated. Most courts have resisted extending *Shelley* to all judicial enforcement of private contracts. Avraham & Yuracko, *supra* note 133, at 351–52. However, discriminatory *class action* settlements are different, because the settlement's impact on class members depends entirely on the court's actions in certifying the class and approving the settlement. *See* Tulsa Prof'l Collection Servs., Inc. v. Pope, 485 U.S. 478, 486 (1988) ("[W]hen private parties make use of state procedures with the overt, significant assistance of state officials, state action may be found."). In any event, if courts declined to allow race- or sex-based calculations, presumably parties disadvantaged by them would be less likely to agree to settlements based on them.

Given that the Equal Protection Clause applies, the case that race- and sex-based damage calculations violate it is inescapable. If otherwise-unconstitutional race- and sex-based distinctions cannot be justified by statistical generalizations about race and sex, then there is no case for these calculations at all. It would obviously be unconstitutional to award lower damages to people of color and women *without* the basis the actuarial tables provide—but the actuarial tables don't save them, either. Indeed, the generalization that men earn more than women is exactly the one at issue in *Frontiero* and *Weinberger*. Reliance on actuarial tables is what the Court rejected in *Manhart* (in the Title VII context). As the Court reiterated in *Bostock* (discussing *Manhart*), sex discrimination can't be defined away simply by relabeling it as discrimination based on an actuarial prediction.[145] And it's implausible that the Court would accept equivalent generalizations based on race.[146]

Moreover, the state's interest in relying on these generalizations is far too weak to survive strict or even intermediate scrutiny, even if the prohibition on doing so were less absolute than the Court has implied. The only state interest that one could argue is served by race- and gender-specific calculations is a general interest in more "accurate" compensatory damage awards.[147] But the claim that these tables improve "accuracy" can only be true *at best* in the aggregate, averaged across many cases; in many individual cases, they will reduce accuracy. Group averages—even when very accurately estimated—are typically of very limited value in producing forecasts for particular individuals for variables that vary widely *within* a group.[148] This is true for wages, life expectancy, and work life; notwithstanding group-level disparities, individual variation is so wide that knowing somebody's race or gender tells you little about their expected life or earnings.[149]

---

[145] *See Bostock*, 140 S.Ct. at 1744.

[146] Avraham and Yuracko persuasively reject several arguments for why race-based damage calculations could be characterized as *not racial classifications*, and not subject to strict scrutiny. One is that race is being used as a "biomarker," analogous to its use in criminal suspect descriptions. *See* Avraham & Yuracko, *supra* note 133, at 338-40. This argument is implausible here, but I consider its application to the medical-diagnosis context in Part IV. The second is that race is only one of several factors considered; as Avraham and Yuracko explain, this doesn't matter legally, so long as an award would have been different but for race. *Id.* at 340–46. The third is that the approach is "race neutral" because all plaintiffs' damages (including white plaintiffs') are calculated based on their respective races. This sort of argument was squarely rejected in *Brown v. Board of Education*, 347 U.S. 483 (1954) and *Loving v. Virginia*, 388 U.S. 1 (1967). Unsurprisingly, Avraham and Yuracko decisively dispose of it. Avraham & Yuracko, *supra* note 133, at 347.

[147] Anthony Sebok has laid out this accuracy argument (sympathetically but ambivalently) in a blog post. Anthony Sebok, *Can Tort Damages Discriminate?*, New Private Law (Aug. 17, 2015), https://blogs.harvard.edu/nplblog/2015/08/17/can-tort-damages-discriminate-anthony-sebok/ [https://perma.cc/8RFH-VUT7].

[148] *See* Starr, *supra* note 54, at 842–45 (providing examples that illustrate this point).

[149] The standard deviation for life expectancy at birth in the U.S. is about fifteen years. *See* Ryan D. Edwards, *The Cost of Uncertain Lifespan*, 26 J. Population Econ. 1485 (2013). In a normal distribution, about sixty-eight percent of individuals fall within one standard deviation of the mean—here, a thirty-year range; the lifespan distribution is not quite normal, but this is

And this statistical point amplifies the reasons that the Court has rejected relying on such generalizations—there are *many* individuals, not just a few outliers, that they don't describe well.

Moreover, even *aggregate* accuracy of predictions across many cases may not be helped and could potentially be *harmed* by using these tables, even relative to using race- and sex-neutral tables (much less compared to more individualized prediction methods). This may seem incongruous, given that race and sex have some predictive value. But it's possible because of the way they use disparities from the past to make predictions about the future. Sex and race disparities in earnings have gotten smaller over time, and one can hope that they will shrink further. If they shrink enough, then relying on race- and sex-based estimates from the past may be counterproductive even if aggregate accuracy across all cases is the only goal. This is rendered more likely given that statistical tables are especially relied on where the injured party is a child, with no actual earnings history to consult. A child's counterfactual earnings are typically estimated based on data from their grandparents' and great-grandparents' generations.[150]

Each of these problems means that the fit between the state's interest in accurately estimating damages and the use of race or sex to do so falls short of a "substantial relationship" per the intermediate-scrutiny standard (a demanding standard),[151] much less the narrow tailoring required by strict scrutiny. Moreover, race-neutral alternatives, which must be ruled out under the narrow tailoring test, could likely achieve better predictions—for example, using richer information about the individual's life and family circumstances.

Even more fundamentally, it's not obvious that the state's interest in maximizing accuracy of damages calculations is important, much less compelling. Accuracy makes the plaintiff more exactly whole, but it's not clear how crucial that objective is. Torts theorists disagree on how central the objective of accuracy is, and there are many ways tort law departs from it.[152]

---

a good enough approximation to make the point. The black-white race gap in life expectancy is six years and the sex gap is five years—large, but dwarfed by individual variation. Farida Ahmad, Elizabeth Arias, Kenneth D. Kochanek, & Betzaida Tejada-Vera, *Provisional Life Expectancy Estimates for January through June, 2020*, CDC VITAL STAT. RAPID RELEASE (Feb. 2021). Race and gender gaps in average earnings are larger than for life expectancy; for example, the average Black man in 2019 earned fifty-six percent of what the average white man did. *See* Eduardo Porter, *Black Workers Stopped Making Progress on Pay. Is It Racism?*, N.Y. TIMES (June 28, 2021), https://www.nytimes.com/2021/06/28/business/economy/black-workers-racial-pay-gap.html [https://perma.cc/JEE2-8LQ8]. Even so, variability *within* groups is even larger. *See* Rakesh Kochhar & Anthony Cilluffo, *Income Inequality in the U.S. Is Rising Most Rapidly Among Asians*, PEW RES. CENTER (July 12, 2018) https://www.pewresearch.org/social-trends/2018/07/12/income-inequality-in-the-u-s-is-rising-most-rapidly-among-asians/ [https://perma.cc/9KPT-5MTP (showing graphs of income distributions by race).

[150] *See* Robert W. Johnson, *The Impact of Race and Gender on Earnings Capacity and Damages: An Economist's Perspective*, 1 Ann.2005 ATLA-CLE 931 (2005) (table 3) (observing that as of 2004, the then-current tables included data for workers whose careers started in the early 1950s).

[151] *See Virginia*, 518 U.S. at 531 (requiring an "exceedingly persuasive justification").

[152] *See* John C. P. Goldberg, *Two Conceptions of Tort Damages: Fair v. Full Compensation*, 55 DEPAUL L. REV. 435, 436–38 (2006). *See generally* Jules Coleman, Scott Hershovitz,

It's notoriously difficult for plaintiffs to recover for non-economic damages, even though it's obvious that for many, those are the most important harms suffered.[153] The very fact that wrongful death damages are predominantly based on lost earnings is itself in deep tension with the make-whole principle; nobody thinks of the value of their loved ones' lives as being substantially defined this way. Attorneys' fees can't be recovered in most cases,[154] even though this typically means plaintiffs are never made whole. Meanwhile, we sometimes allow punitive damages, which serve deterrence and expressive aims, not corrective justice. Beyond damages, other doctrines also compromise corrective justice to serve other aims—for example, evidence of settlement offers is excluded for public policy reasons (even though it may in fact be probative of liability).[155] Liability based on purely statistical evidence is disfavored, even though this rule in the aggregate may reduce outcome accuracy.[156] And the importance of accurate calculations is even less obvious with respect to criminal restitution; criminal sentencing usually does not focus on corrective objectives at all.

If we look beyond corrective justice toward other purposes of tort law (or toward the principal purposes of criminal sentencing), the state's interest in using race- and gender-based calculations falls apart further. If we consider incentives, one would have to believe it's *good* that these calculations encourage potential tortfeasors to concentrate the risk of injury on people of color and women—for example, sitting hazardous facilities in communities of color.[157] But surely the broader social disparities produced by a classification cannot be treated, in an equal protection analysis, as a justification for it. Likewise, from an expressive perspective, the message sent by race- and sex-based damage awards—that the lives of people of color and women are worth less than white men—is poisonous. It ratifies the effects of centuries of discrimination.

Finally, there's no reasonable argument that failing to consider race or sex would be unfair to defendants whose victims are women and/or people of color, or to white male plaintiffs. A common basis given for race- and sex-based estimation is that the defendant has a right to take their victim as they find them, and if "as they find them" is defined by a legacy of societal discrimination, that's not the defendant's duty to correct via a more-than-make-whole damage payment.[158] But this argument is facile. The defendant

---

& Gabriel Mendlow, *Theories of the Common Law of Torts,* The Stanf. Encyc. of Phil. (June 2, 2022), https://plato.stanford.edu/archives/win2021/entries/tort-theories/ [https://perma.cc/2N2H-J3FM] (briefly reviewing competing theories of tort law, of which corrective justice is one, and citing sources).

[153] Chamallas & Wiggins, *supra* note 114, at 170–82.

[154] *See, e.g.*, Hardt v. Reliance Standard Life Ins. Co., 560 U.S. 242, 253 (2010).

[155] *See, e.g.*, Fed. R. Evid. 408.

[156] *See* Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. Personality & Soc. Psych. 739, 739 (1992).

[157] *See* Avraham & Yuracko, *supra* note 133, at 335; Sharkey, *supra* note 133, at 1489–90.

[158] *See* Anthony Sebok, *Judge Jack Weinstein's Ruling Barring the Use of Race in Calculating the Expected Lifespan of a Man Seeking Tort Damages*, FindLaw (Oct. 22, 2008),

is not responsible for societal discrimination, but neither are they entitled to benefit from it. Fairness does not entitle a defendant to special credit for having (strategically or luckily) picked a victim from a disadvantaged social group. Similarly, fairness does not entitle a white male plaintiff to have the advantages of his race and sex translated into larger damage payments. If a particular white male plaintiff is undercompensated for his actual harm, his complaint should be that particular evidence in his case contravenes the statistical prediction—not that the statistical prediction should have credited him for being a white male. The latter kind of "actuarial fairness" reasoning is deeply distasteful and irreconcilable with the logic of the cases discussed in Part II.[159]

If race- and sex-specific damage calculations were forbidden, what would the likely alternative be? When courts *have* declined to rely on these tables, the typical approach isn't to abandon actuarial prediction entirely. Rather, it's to use race- and sex-neutral tables. As the example from *Bedonie* illustrates, this shift can make a big difference, especially in awards to women and girls of color, who otherwise suffer intersectional race and sex discrimination.[160]

The blended-tables alternative would be a substantial step toward closing the race and sex gaps in damage awards, and of countering the toxic expressive messages entailed in those classifications. It would not fully *eliminate* racial disparities, because race-correlated socioeconomic variables (especially education) often also shape calculations.[161] In addition, when damages do not turn on actuarial predictions but on individualized information about the plaintiff's earnings, we can expect them to differ in the aggregate across racial and gender lines because disparities in plaintiffs' actual earnings presumably reflect those in society at large. To eliminate this problem would require a deeper rethinking of how we value individuals' lives and health. This question is beyond this paper's scope, but it's worth considering Professor Sharkey's recent proposal to shift to agencies' "value of a statistical life" approach, which does not vary based on race, gender, or income.[162] A system in which a human life is priced based on how much

---

https://supreme.findlaw.com/legal-commentary/judge-jack-weinsteins-ruling-barring-the-use-of-race-in-calculating-the-expected-lifespan-of-a-man-seeking-tort-damages-an-isolated-decision-or-the-beginning-of-a-legal-revolution.html [https://perma.cc/42H4-SJPG]; CHAMALLAS & WRIGGINS, *supra* note 114, at 166–67 (describing this as the "most familiar" defense); *see also* Sebok, *supra* note 147 ("Awarding damages in excess of our best guess as to actual expected losses . . . shades into a strong form of instrumentalism that, for various reasons, is unattractive.").

[159] *See Bostock*, 140 S.Ct. at 1744 (approvingly discussing *Manhart*'s rejection of such reasoning).

[160] For white women or men of color, race and gender have opposing effects, which might explain some failures to challenge the tables. *See* Chamallas, *Questioning the Use*, *supra* note 132, at 122 (observing, in 1994, that at the time the gender disparities were larger than the race disparities).

[161] *See* Avraham & Yuracko, *supra* note 133, at 369–71.

[162] Sharkey, *supra* note 133, at 1480–83. *See generally* W. KIP. VISCUSI, PRICING LIVES: GUIDEPOSTS FOR A SAFER SOCIETY (2018) (outlining the VSL method). Current case law on

money the person could have made is always going to be inequitable. To make that prediction based on race and sex is even worse, and absolutely a practice worth fighting, but nobody should be fully satisfied with its elimination.

B.   Intellectual Disability Determinations in Capital Cases

In 2002, in *Atkins v. Virginia*, [163] the Supreme Court barred the execution of a defendant with an intellectual disability on Eighth Amendment grounds. [164] In *Atkins* proceedings in at least eight states, prosecution experts have argued that intelligence tests and/or adaptive behavior assessment scores must be adjusted based on race, ethnicity, or national origin (also a suspect classification).[165] In some cases, the government has introduced scores that have either been normed based on race, national origin, "sociocultural group," and/or socioeconomic status, or have had "corrections" made for such factors.[166] Courts have also permitted experts to testify that test scores that aren't normed along these lines should be adjusted upward.[167] The U.S. government has argued that foreign defendants must take tests normed to their home country's population.[168] Prosecution arguments along

---

VSL calculations, and "hedonic damages" generally, is varied; in federal courts, for example, hedonic damages may be available but courts have barred experts from quantifying them. *See* Hart v. Corr. Corp. of Am., No. 211CV00267MCAWPL, 2014 WL 12670796 (D.N.M. May 6, 2014) (observing that the "non-economic value of human life is the same for every person, regardless of race"). Treating this feature as a bug, one district court in a police-shooting case rejected a VSL calculation precisely because it failed to adjust for race, sex, and age, and thus was too ill fitting to "value the life of a specific individual." Ayers v. Robinson, 887 F. Supp. 1049 (N.D. Ill. 1995). When life expectancy is used to calculate hedonic damages, courts *have* incorporated race and gender in calculations. *See* Smith v. United States Dep't of Veterans Affairs, 865 F.Supp. 433, 441 (1994).

[163] 536 U.S. 304 (2002).

[164] *See id.* at 304 (2002); *accord* Hall v. Florida, 572 U.S. 701, 723 (2014); Moore v. Texas, 137 S.Ct. 1039, 1039 (2017).

[165] *See* Robert M. Sanger, *IQ, Intelligence Tests, "Ethnic Adjustments" and Atkins*, 65 Am. U. L. Rev. 87 (2015) (reviewing cases); Brief for Public Law Scholars as Amici Curiae Supporting Petitioner, Hernandez v. Stephens, 134 S. Ct. 1760 (2014) (No. 13-8004), 2014 WL 333536.

[166] *See, e.g.*, Maldonado v. Thaler, 625 F.3d 229, 238–39 (5th Cir. 2010) (describing expert's upward adjustments of scores for "cultural" reasons); *In re* Champion, 322 P.3d 50, 67 (Cal. 2014) (prosecution's expert, who a postconviction-stage referee found convincing, "explained that because Blacks ordinarily perform more poorly than Whites on those tests, it is preferable to use ethnically corrected norms").

[167] See, e.g., Hodges v. State, 55 So. 3d 515, 525 (Fla. 2010) (quoting expert dismissing defendant's low IQ scores because such tests "tend to underestimate particularly the intelligence of African-Americans"); Brown v. State, 982 So. 2d 565, 604 (Ala. Ct. Crim. App. 2006) (similar); Black v. State, 2005 WL 2662577, at *8-*9 (Tenn. Ct. Crim. App. 2005) (similar); State v. Were, 890 N.E.2d 263, 292–93 (Ohio 2008) (finding no error in the trial court's conclusion that IQ tests understated a Black defendant's intelligence because they were "culturally biased").

[168] *See* United States v. Salad, 959 F. Supp. 2d 865, 876 (E.D. Va. 2013).

these various lines have "been regularly accepted without challenge" from defense counsel (though we'll see some exceptions below).[169]

Courts have rarely offered much analysis of these score adjustments. Some have rejected *Atkins* claims while citing the government's adjusted scores (or criticisms of unadjusted scores). Some trial courts have refused to rely on these adjustments, but mainly based on scientific skepticism; there is no case law considering whether demographic adjustments might, regardless of the empirical merits*,* violate the Equal Protection Clause. Moreover, while many articles touch on the issue, these practices and their constitutionality have been subject to little sustained examination in the legal academy.[170] In September 2020, however, California's governor signed a bill concerning *Atkins* assessments, which among other things prohibited adjustments "based on race, ethnicity, national origin, or socioeconomic status"—a significant development that occasioned virtually no commentary.[171]

It's hard to pinpoint just how widespread these practices and expert arguments are in *Atkins* proceedings, since adjustments may take place without being mentioned in courts' decisions. It's even harder to pinpoint their effect on outcomes, since *Atkins* decisions typically cite a wide range of facts. But it's clear that race-norming and other such adjustments can substantially shift scores, enough to potentially sway outcomes. "[P]rosecution experts regularly assert that the standard testing instruments based on general population norms underestimate the IQ of African-Americans by approximately 10-15 points."[172]

Consider some examples. Ramiro Hernandez-Llanas was executed in 2012 even though he had scored between 52 and 57 on several IQ cases, well below the range around 70 that *Atkins* and its progeny suggest is on the legal borderline.[173] However, on another IQ test "scaled to Mexican norms" (Hernandez-Llanas was from Mexico), he scored a 70, a point on which the

---

[169] *See* Michael L. Perlin, '*Your Corrupt Ways had Finally Made you Blind': Prosecutorial Misconduct and the Use of 'Ethnic Adjustments' in Death Penalty Cases of Defendants with Intellectual Disabilities*, 65 Am. U. L. Rev. 1437, 1452 (2016) (internal quotations omitted); David L. Shapiro et al., *Ethnic Adjustment Abuses in Forensic Assessment of Intellectual Abilities*, Practice Innovations 2 (Oct. 2019).

[170] The leading law review article is by practitioner Robert Sanger, *supra* note 165, and offers a useful description of the practices at issue. Sanger's constitutional argument is somewhat confusing, however, relying on a misreading of Washington v. Davis, 426 U.S. 229 (1976) (which he interprets to prohibit race-norming of employment tests, an issue that case did not raise). *See* Sanger, *supra* note 165, at 140–41. A 2014 law professors' amicus brief in support of an unsuccessful petition for certiorari offered a more straightforward, concise constitutional argument. Public Law Scholars' Amicus Brief, *supra* note 165. *See also* Perlin, *supra* note 169, at 1441 (arguing that these practices are prosecutorial misconduct).

[171] A.B. 2512, 2020 Leg. (Cal. 2020).

[172] Nancy Haydt, *Intellectual Disability: A Digest of Complex Concepts in* Atkins *Proceedings*, Champion 44 (Jan./Feb. 2014). *See also* Shapiro et al., *supra* note 169, at 7 (observing that this testimony relies selectively on outlier studies that show especially large score gaps).

[173] *See* Hernandez v. Stephens, 537 F. App'x 531, 536–43 (5th Cir. 2013) (per curiam).

Fifth Circuit relied in affirming the denial of habeas relief.[174] Beside IQ, the state postconviction proceedings also included considerable evidence concerning Hernandez-Llanas's "adaptive functioning deficits," which included the inability to perform a litany of basic life tasks, such as bathing himself.[175] On this point also, the prosecutor's expert testified that these deficits were "normal for [his] 'cultural group,'" and the lower court agreed, even though the expert apparently neither had met Hernandez-Llanas nor had any knowledge of his "cultural group" beyond the testimony in the proceedings.[176] In the federal habeas proceedings, the Fifth Circuit didn't consider the permissibility of the IQ score adjustment (mentioning it only in passing) or of the other claims concerning his cultural group, and affirmed the denial of relief. The Supreme Court denied certiorari.[177]

In Texas, as of 2010, a single psychologist named George Denkowski had testified for the state in "29 cases—nearly two-thirds of all *Atkins* appeals in that state."[178] Denkowski was a leading proponent of "sociocultural" adjustments of IQ and adaptive-behavior assessments, including, in *Atkins* cases, assessing defendants relative to a "criminal socioculture."[179] His case for this relied explicitly on generalizations about poor families—for example, an argument that "criminal offenders from poor families would not have learned such skills as counting money,"[180] or even "maintaining hygiene,"[181] such that absence of such skills should not be interpreted as an adaptive deficit. For example, in the case of Daniel Plata, a Mexican man with developmental disabilities stemming from a birth injury, Denkowski "used his assumptions about Plata's upbringing to re-score his responses on a test measuring basic skills like how to count money, groom oneself, or use

---

[174] *See id.* at 536. Mexican norms for IQ tests have been "widely criticized for overstating IQ." Johnson, *supra* note 150, at 291 (citing Hoi K. Suen & Stephen Greenspan, *Linguistic Sensitivity Does not Require one to use Grossly Deficient Norms: Why U.S. Norms Should be Used with the Mexican WAIS-III in Capital Cases*, 34 Psych. Intell. & Developmental Disabilities 2 (2008); Hoi K. Suen & Stephen Greenspan, *Serious Problems with the Mexican Norms for the WAIS-III when Assessing Mental Retardation in Capital Cases*, 16 Applied Neuropsychology 214, 215 (2009).

[175] Hernandez v. Thaler, 2011 WL 4437091 at *4 –5, *22–24 (W.D. Tex. Sept. 23, 2011).

[176] Sheri Lynn Johnson, *A Legal Obituary for Ramiro*, 50 U. Mich. J.L. Reform 291, 303–04 (2016).

[177] *See* Hernandez v. Stephens, 572 U.S. 1036 (2014).

[178] Karen Franklin, *Atkins Claim: Did Texas Psychologist Skew Data for Death?*, In the News: Forensic Psych. Blog (Jan. 10, 2010), https://forensicpsychologist.blogspot.com/2010/01/atkins-claims-did-texas-psychologist.html?m=1 [https://perma.cc/T5LS-F5CZ].

[179] George C. Denkowski & K.M. Denkowski, *Adaptive Behavior Assessment of Criminal Defendants with a Mental Retardation Claim*, 26 Am. J. Forensic Psych. 43 (2008).

[180] Renée Feltz, *Cracked: Despite a U.S. Supreme Court ban, Texas has Continued to Send Mentally Retarded Criminals to Death row. Will a Mexican Immigrant's Case Correct this Injustice?*, Tex. Observer (Jan. 8, 2010), https://www.texasobserver.org/cracked/ [https://perma.cc/5LTS-M3A6].

[181] Brandi Grissom, *Psychologist Who Cleared Death Row Inmates is Reprimanded*, N.Y. Times (Apr. 14, 2011), https://www.nytimes.com/2011/04/15/us/15ttpsychologist.html [https://perma.cc/YZY2-L5JA].

a microwave."[182] His approach did not rely on "norming" in the sense of percentile ranking within a subgroup; rather, he simply added points to various components of scores. These adjustments were large; Plata's IQ score went from 70 to 77, for example, and his adaptive-behavior score went from 61 to 71.[183] In Plata's case, though, the defense successfully fought these adjustments; the trial court blasted Denkowski's approach as devoid of scientific foundation, accepted other tests in which Plata scored lower, and commuted Plata's sentence.[184]

Critics described Denkowski's method as "voodoo psychometrics" and "junk science," and Denkowski himself as "Dr. Death."[185] The approach was also critiqued as racially biased. Forensic psychologist Karen Franklin wrote: "This subtly racist argument of cultural deficit seems to be becoming increasingly popular as a way to explain away the deficits of low-functioning Mexican immigrants in particular."[186] In 2009, after an ethics complaint, the state psychology board voted to revoke Denkowski's license. He appealed and ultimately settled, agreeing never to testify in another *Atkins* case.[187] Two defendants that he testified against had already been executed. The Texas courts reconsidered many other cases, with mixed results; in some, requests for resentencing were denied with little explanation and over dissents.[188] Stopping Denkowski from testifying didn't eliminate his influence; his "strategy has been widely adopted by prosecution witnesses."[189]

At least when they turn on suspect classifications (race, ethnicity, or national origin), the score-adjustment procedures described in these cases are irreconcilable with the case law reviewed in Part II. In many of these cases there's no basis for the distinctions being drawn other than a statistical generalization about a group, or sometimes a mere stereotype. It's true that in

---

[182] Feltz, *supra* note 180, at 10.

[183] *See id.* at 10.

[184] *Ex parte* Plata, No. 693143-B (351st Dist. Ct., Harris County, Tex., Sept. 28, 2007). Notably, even this opinion, the most searing denunciation of demographic adjustments in the *Atkins* case law, does not really consider the permissibility of this kind of group generalization. Rather, the court's objection was to the lack of statistical support for Denkowski's adjustments; in a sense, the discrimination here was not "statistical" enough for the court. *Id.* at *49.

[185] Kevin McGrew, *Excellent Article on Role of "Dr. Death" Psychologist and Junk Science in Texas Atkins MR Death Penalty Cases*, INTEL. COMPETENCE & DEATH PENALTY BLOG (Jan. 8, 2010), https://iqmrdeathpenalty.blogspot.com/2010/01/excellent-article-on-role-of-death.html [https://perma.cc/EU6P-UJ9R]; Feltz, *supra* note 180, at 9.

[186] Franklin, *supra* note 178.

[187] Grissom, *supra* note 181.

[188] *See, e.g.*, *Ex parte* Matamoros, No. WR-50,791-02, 2012 WL 4713563 (Tex. Crim. App. Oct. 3, 2012) (denying relief); *Ex parte* Gallo, No. WR-77,940-01, 2013 WL 105277 (Tex. Crim. App. Jan. 9, 2013); *Ex parte* Hunter, No. WR-69,291-01, 2015 WL 2159808 (Tex. Crim. App. Apr. 22, 2015); *Ex parte* Butler, 416 S.W.3d 863 (Tex. Crim. App. 2012); *Ex parte* Wesbrook, No. WR-52,120-03, 2016 WL 930747 (Tex. Crim. App. Mar. 9, 2016); *Ex parte* Escobedo, No. WR-56,818-01, 2012 WL 982907 (Tex. Crim. App. Mar. 21, 2012); *Ex parte* Maldonado, No. WR-51,612-02, 2012 WL 1439056 (Tex. Crim. App. Apr. 25, 2012); *Ex parte* Davis, No. WR-40,339-09, 2020 WL 1557291 (Tex. Crim. App. Apr. 1, 2020).

[189] Haydt, *supra* note 172, at 51; *see also* Shapiro et al., *supra* note 169, at 16 (observing that outside Denkowski's case, professional boards have not weighed in).

appeals or postconviction proceedings, it may be unclear what weight was given the prosecution's improper argument. But even in such cases, the remedy should be guided by the Court's warnings in *Buck v. Davis* about the likelihood of prejudice based on experts' race-related claims: "some toxins can be deadly in small doses."[190]

Norming or other score adjustments based on class is also likely forbidden, given the line of cases protecting criminal defendants from socioeconomic discrimination. If the state could not, to justify probation revocation in *Bearden*, rely on the empirically supported generalization that unemployment predicts crime risk, then surely it cannot rely in a capital proceeding on assertions that poor families do not teach their children basic life skills.[191] Even if the adjustments were carried out more rigorously than we see in some *Atkins* cases (e.g., formal socioeconomic norming), they would still entail "lumping" the defendant "together with other poor persons" and "punishing a person for his poverty."[192]

What about "cultural" or "sociocultural" adjustments? These labels may seem to skirt these legal restrictions, since they don't quite track any suspect classification. And in some instances, especially with foreign defendants, cultural differences or linguistic barriers really might interfere in identifiable ways with comprehension of test questions.[193] If assessments are designed in a culturally appropriate way and administered in a defendant's language by a culturally competent person, however, back-end score corrections should rarely be needed.[194] In any case, in practice sociocultural adjustments are often not used in a nuanced, individualized way. Rather, "culture" is treated as synonymous with suspect classifications (e.g., Mexican origin or Black race) and/or with poverty, and the resulting adjustments are crude.

One's views of the state's interest in executing any given defendant might differ based on one's broader death penalty views. But it would be difficult to argue that it has a compelling or even cognizable interest in executing people more readily based on race, national origin, or poverty, or that allowing effectively different IQ cutoffs on these bases is a narrowly tailored response to its general penological interests. The Supreme Court's recent decision in *Buck v. Davis* was a ringing rejection of race-based statistical claims in the capital context, and would be hard to distinguish in *Atkins* cases involving race-based score adjustments.[195]

---

[190] 137 S. Ct. at 776–77.

[191] *See Bearden*, 461 U.S. at 671–72.

[192] *Id.* at 671.

[193] For instance, in *Salad*, 959 F.Supp.2d at 876, the government plausibly argued this was so for the defendant, a Somali nomad arrested in a piracy-related case in Somalia (although its demand for Somali-normed tests instead was, per the defense, impossible to meet, since no such tests exist).

[194] *See* Shapiro et al., *supra* note 169, at 11.

[195] *See Buck*, 137 S. Ct. at 759.

Moreover, although I focus on the equal protection question, these adjustments also contravene the core Eighth Amendment principles reflected by *Atkins* and its progeny. Nothing in these cases' logic about retributive and deterrent objectives of punishing disabled individuals turns on how the individual ranks in intellectual ability versus other people of his race, ethnicity, and class.

To defend these score adjustments, one can imagine the government relying on the Court's exhortations in *Hall v. Florida*[196] and *Moore v. Texas*[197] to follow medical-community standards. Race-norming is common in many medical contexts, and the intellectual disability diagnostic guidelines found in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) advise that "instruments must be normed for sociocultural background."[198] The manual does not specify what this norming should entail, and does not mention race specifically. Another note states: "Cultural sensitivity and knowledge are needed during assessment, and the individual's ethnic, cultural, and linguistic background, available experiences, and adaptive functioning within his or her community and cultural setting must be taken into account."[199] It's plausible for the government to argue that adjustments of both IQ and adaptive-functioning scores are consistent with the DSM-5.

But this argument is ultimately unconvincing. First, it would be over-reading the Court's emphasis on medical practice norms—which, it held, must inform but "do not dictate a court's intellectual-disability determination."[200] In *Hall*, the Court emphasized that it is "the Court's duty to interpret the Constitution," and that the "legal determination of intellectual disability is distinct from a medical diagnosis."[201] It quoted its 2012 decision in *Kansas v. Crane*,[202] which stated:

The science of psychiatry, which informs but does not control ultimate legal determinations, is an ever-advancing science, whose distinctions do not seek precisely to mirror those of the law . . . Consequently, we have sought to provide constitutional guidance in this area by proceeding deliberately and contextually, elaborating generally stated constitutional standards and objectives as specific circumstances require.[203]

The Court in *Hall* and *Moore* was not asked to decide whether states should follow medical norms even if they entail discrimination based on

---

[196] 572 U.S. 701 (2014).

[197] 137 S. Ct. 1039 (2017).

[198] Am. Psychiartric Ass'n, Diagnostic and Stat. Manual of Mental Disorders (5th ed. 2013); *see also Moore*, 581 U.S. at 7 (citing the DSM-5 as an authority).

[199] *Id; see also* K.F. Widaman & G.F. Siperstein, *Assessing Adaptive Behavior of Criminal Defendants in Capital Cases: A Reconsideration*, 27 Am. J. Forensic Psych. 5, 12 (2009) (observing that "diagnostic manuals . . . provide essentially no guidance" on sociocultural adjustment).

[200] *Moore*, 581 U.S. at 13 (citing *Hall*, 572 U.S. at 721).

[201] *Id.* at 720.

[202] 534 U.S. 407 (2002).

[203] *Id.* at 413.

suspect classifications. It would be hard to imagine a stronger case for departing from such norms in service of an overriding constitutional value—especially given the Court's holding in *Buck* that race is *never* a permissible sentencing consideration. Moreover, in *Hall* and *Moore*, the Court's concern was with states making it *easier* to execute defendants by ratcheting up the bar for an intellectual-disability diagnosis, threatening to render a "nullity" the "protection of human dignity" that *Atkins* offered.[204] Here, it is *following* the purported medical norm that threatens that protection.

In addition, it's not at all a medical consensus position that intellectual disability tests *should* be scored differently depending on group characteristics. This point is dramatically illustrated by the saga of Dr. Denkowski, who nearly lost his license, and was also condemned by the American Association on Intellectual and Developmental Disabilities (AAIDD).[205] The DSM-5 is light on details, potentially allowing for a Denkowski-type approach to "sociocultural" norming, but capable of other interpretations. J. Gregory Olley wrote that the necessary accounting for "sociocultural factors" is already done when a single U.S. standard is used, because "the tests' norm groups match the demographics of the most recent U.S. census, which includes people of varied sociocultural backgrounds."[206] Olley cited a caution from the AAIDD's diagnostic guide: "Do not allow cultural or linguistic diversity to overshadow or minimize actual disability."[207] AAIDD itself filed an amicus brief supporting the unsuccessful certiorari petition in Hernandez-Llanas's case, arguing that the requirement that assessments account for cultural context means something quite far from the "superficial, reductionist exercise" that the prosecution engaged in; an assessor must bring both "actual knowledge of a *precise* culture" and a detailed, individualized assessment of the way cultural and linguistic issues affected the particular individual.[208]

Psychologists Stephen Greenspan and George Woods have similarly argued that the DSM only means to encourage cultural sensitivity, not norming by race, ethnicity, or SES, a practice that "has been discredited by intelligence scholars and test developers."[209] Greenspan has also observed

---

[204] *Moore*, 581 U.S. at 18–22 (citing *Hall*, 572 U.S. at 714–22).

[205] Grissom, *supra* note 181.

[206] J. Gregory Olley, *Definition of Intellectual Disability in Criminal Court Cases*, 51 INTELL. & DEVEL. DISABILITIES 117, 119 (2013).

[207] *Id.* (citing ROBERT L. SCHALOCK, WIL BUNTINX, SHARON BORTHWICK-DUFFY, RUTH LUCKASSON, MARTI SNELL, MARC J. TASSÉ, & MICHAEL WEHMEYER, AAIDD 2002 SYSTEM IMPLEMENTATION COMMITTEE, USER'S GUIDE WORK GROUP, USER'S GUIDE: MENTAL RETARDATION: DEFINITION, CLASSIFICATION, AND SYSTEMS OF SUPPORTS: APPLICATIONS FOR CLINICIANS, EDUCATORS, DISABILITY PROGRAM MANAGERS, AND POLICY MAKERS (10th ed. 2007).

[208] Brief of American Association on Intellectual and Developmental Disabilities and the Arc of the United States as Amici Curiae in Support of Petitioner at 11, *Hernandez v. Stephens*, 537 F. App'x 531 (5th Cir. 2013) (per curiam).

[209] Stephen Greenspan & George W. Woods, *Intellectual Disability as a Disorder of Reasoning and Judgement: The Gradual Move Away From Intelligence Quotient-Ceilings*, 27 CURRENT OPINION 110, 113–14 (2014); *see also An Introduction to Assessment*, *in* MASSACHUSETTS GENERAL HOSPITAL GUIDE TO LEARNING DISABILITIES 3, 15 (H. Kent Wilson & Ellen B. Braaten eds., 2019) (describing group-normed tests as "less widely accepted").

that forensic experts in *Atkins* cases appear to be *more* likely to conduct such norming than those carrying out assessments for other purposes—even though many in the field consider the *Atkins* context to be the *least* defensible application of such norms. The authors of two leading articles often invoked to support race-norming have "disavowed" that practice in the *Atkins* context.[210]   Another group of psychologists have described "ethnic adjustments" in *Atkins* cases as a "flagrant misuse of psychological testing" and "pseudoscience.[211]

Relatedly, another justification often offered for score adjustments is that IQ tests are otherwise biased against Black and/or Hispanic defendants.[212] This explanation has a superficial appeal, implying that the corrections are an *antiracist* move. But it is ultimately unpersuasive, at least in the *Atkins* context.

First, even if test bias *does* shape score disparities at the group level, this explanation for low scores is a generalization that doesn't hold true in all individual cases. In many cases, the experts opining that a test is biased don't point to any specific, individualized reasons that a test got it wrong for a particular defendant, nor for why a particular adjustment is individually appropriate.[213]   Obviously, not *every* defendant of color who scores in the disabled range got there because of testing bias. Some individuals in every racial group have legitimate intellectual disabilities. Moreover, many factors associated with poverty (toxic exposures, for example) make such disabilities more likely, and disproportionately affect communities of color.[214] These are problems of systemic racism, not biological difference, but the resulting disabilities are real, not artifacts of test bias.

This is not to say that there are no biases in intelligence tests, and rough group-level adjustments to offset estimated biases might be defensible in some contexts where the purpose is benign.[215] But they cannot justify deeming an individual eligible for the death penalty. In general, the Supreme Court has emphasized the importance of *individualized* determination of the facts permitting execution,[216] and its rejection of bright-line IQ cutoffs is

---

[210] *See* Haydt, *supra* note 172, at 48 & nn.72–73.

[211] Shapiro et al., *supra* note 169, at 2, 9.

[212] *See* Brief for Public Law Scholars as Amici Curiae Supporting Petitioner, *supra* note 165, at 14–15 (collecting cases).

[213] For example, experts refer to a general "cultural bias" or tendency of tests to understate the intelligence of Black test-takers. *See id.*

[214] Hernandez, for example, grew up in extreme poverty with extensive neurotoxin exposure. Johnson, *supra* note 176, at 292. Lead exposure sharply increases risk of intellectual disability, *see* Christine F. Delgado et al., *Lead Exposure and Developmental Disabilities in Preschool-Aged Children*, 24 J. Pub. Health Mgmt. & Prac. 10 (2018), and disproportionately impacts people of color and the poor, *see* Spencer Banzhaf, Lala Ma & Christopher Timmins, *Environmental Justice: The Economics of Race, Place, and Pollution*, 33 J. Econ. Persp. 185, 193 (2019) (discussing lead and other pollutants).

[215] The broader topic of equity-promoting uses of racial adjustments is taken up in Part IV, *infra*.

[216] *See, e.g.*, Lockett v. Ohio, 438 U.S. 586, 602–05 (1978).

consistent with this emphasis.[217] This principle complements the above-discussed equal protection doctrine similarly emphasizing the right of individuals not to be lumped in with groups.

Finally, various aptitude tests shape access to all kinds of opportunities, and many have been critiqued for biases. But we typically don't correct these biases with rough group-based score changes. Recall the story of the General Aptitude Test Battery: when this *was* done, and threatened white interests, a backlash and Congressional ban swiftly followed.[218] It is deeply perverse that our society and courts would then embrace "bias correction" in the one context in which the correction's impact cuts the most dramatically in the opposite direction: allowing the supposed victim of bias not access to an opportunity, but rather to be put to death.[219]

C.   Punishing Poverty: Risk Assessment in Criminal Justice

Criminal justice is undergoing a prediction revolution. Over the past fifteen years or so, data-driven algorithms have begun to transform the criminal process. Today, at least some courts in at least 47 states are using actuarial risk assessments (focused on risk of crime and/or failure to appear) to guide bail decisions.[220] In at least 20 states, courts use crime-prediction instruments at sentencing.[221] Parole boards in nearly every state with discretionary parole use them too, a practice with precursors dating back many decades.[222] Risk assessments, along with related "needs assessments," are also used for other corrections purposes, such as tailoring the terms of supervision or the provision of services.[223]

How do these predictive tools work? They typically do not involve machine learning or "big data." Many are simply scored checklists—a list of perhaps eight or ten questions about various risk factors, with points assigned to each answer.[224] Very simple instruments are especially popular at

---

[217] *See Hall*, 572 U.S. at 724.

[218] *See supra* notes 81–85 and accompanying text.

[219] Inconsistency across contexts may also undermine a state's justification in Eighth Amendment analysis. *See Moore*, 137 S.Ct. at 1052 ("Texas cannot satisfactorily explain why it applies current [diagnostic standards] in other contexts, yet clings to superseded standards when an individual's life is at stake.").

[220] *See Where Are Risk Assessments Being Used?*, MAPPING PRETRIAL INJUSTICE, https://pretrialrisk.com/national-landscape/where-are-prai-being-used/ [https://perma.cc/NA5V-TH9A] (citing "11 entire states" and 178 counties in other states).

[221] *See* Michael Brenner, Jeannie Suk Gersen, Michael Haley, Matthew Lin, Amil Merchant, Richard Jagdishwar Millett, Suproteem K. Sarkar & Drew Wegner, *Constitutional Dimensions of Predictive Algorithms in Criminal Justice*, 55 HARV. C.R.-C.L. L. REV. 267, 268 (2020).

[222] *See* BERNARD E. HARCOURT, AGAINST PREDICTION: PROFILING, POLICING, AND PUNISHING IN AN ACTUARIAL AGE 78–80 (2007) (reviewing state parole practices).

[223] *See, e.g., Practitioner's Guide to COMPAS Core*, EQUIVANT (April 4, 2019), http://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf. 1-2 and 32. [https://perma.cc/A23Q-P59Q].

[224] *See* Starr, *supra* note 54, at 813–14 (describing example of Missouri's checklist).

the bail stage, because the time available before bail proceedings generally does not permit extensive interviews.[225] But other risk tools used at other stages are much more detailed. Either way, the weights associated with each risk factor are typically based loosely on the regression coefficients estimated in studies of past offenders.[226]

The risk factors that determine the score vary, and for some instruments, it is not possible to identify them with certainty, because the score formulas are corporate trade secrets not made available even to defendants.[227] But some information is public. Consider, for example, the national market leader, the Level of Service Inventory-Revised (LSI-R); scoring guides for test administrators can be found, albeit not the weights given to each response.[228] Its questionnaire has 54 items. The first ten relate to criminal history.[229] The next ten relate to employment and educational history.[230] Then come financial stability (including past and present difficulty paying bills), reliance on social assistance, and family factors, including whether any family members have criminal records.[231] Other questions address social networks, substance abuse, mental health, and attitudes toward society and the law.[232] The LSI-R does not include demographic variables like age or gender; many other instruments do. No actively used risk assessment instrument relies on race, although most of the variables described above are race-correlated.

In the early years of this predictive revolution (i.e., until the early 2010s), these instruments were adopted with virtually no pushback.[233] They were widely celebrated for adopting evidence-based practices in lieu of relying upon decision-makers' intuition. Risk assessment was embraced as a

---

[225] For example, the popular COMPAS tool made by Equivant is based on a questionnaire with over 100 questions, but Equivant offers a separate product for pretrial risk assessment that contains just eight factors. *Practitioner's Guide to COMPAS Core*, *supra* note 223, at 31. The earliest sentencing-stage risk instruments were likewise very basic checklists—for example, Missouri's checklist contained only eleven questions. *See* Michael A. Wolff, *Missouri's Information-Based Discretionary Sentencing System*, 4 Ohio St. J. Crim. L. 95, 113–14 (2006). A popular bail tool developed by Arnold Ventures requires no interview at all. *See Public Safety Assessment FAQs*, Arnold Ventures (Mar. 18, 2019), https://craftmediabucket.s3.amazonaws.com/uploads/Public-Safety-Assessment-101_190319_140124.pdf [https://perma.cc/Q9KV-F2RE].

[226] For further description, see Starr, *supra* note 54, at 809–20.

[227] *See* Kirsten Martin, *Ethical Implications and Accountability of Algorithms*, 160 J. Bus. Ethics, 835–50 (2019); *Vision for Justice 2020 and Beyond: A New Paradigm for Public Safety*, Leadership Conf. on Civ. and Hum. Rts. & Civ. Rts. Corps (Sept. 2019) http://civilrightsdocs.info/pdf/reports/Vision-For-Justice-2020-SHORT.pdf [https://perma.cc/HCF3-SSBW].

[228] *See Idaho LSI-R Scoring Guide 3–27* (Apr. 2015), https://archive.epic.org/EPIC-19-11-21-ID-FOIA20191206-ID-lsi-scoring-guide-v-3.pdf [https://perma.cc/RC9U-LCLM].

[229] *See id.* at 3–6.

[230] *See id.* at 6–10.

[231] *See id.* at 11–14.

[232] *See id.* at 16–20, 26–27.

[233] *See* Starr, *supra* note 54, at 814–17 (surveying the then-current discourse surrounding sentencing instruments).

public safety measure, helping to identify the most dangerous individuals.[234] It was also embraced by many criminal justice reform advocates as a way to reduce incarceration by identifying defendants who could be safely released; politically, including risk assessment in reforms also allowed them to be sold as "smart on crime."[235]

In recent years, the discourse surrounding risk assessment has become much richer and more contested—even while the growth in its use continued. Critics have focused on numerous issues, including transparency and accuracy.[236] A growing empirical literature assesses these instruments' real-world effects on outcomes including overall detention rates and disparities.[237] Meanwhile, the development of machine-learning-based crime prediction tools has become a burgeoning, if still relatively small, subfield of the data-science world.[238]

There is now a fairly substantial body of work grappling with equity concerns, focusing mostly on racial disparities. This issue received widespread public attention after a 2016 investigative report by ProPublica, which focused on another popular commercial product called COMPAS; it found that the instrument produced more "false positives" for Black defendants (predicting high crime risk when the individual did not commit a crime) while producing more "false negatives" for white defendants (predicting low crime risk when the individual did commit a crime).[239] In law reviews, philosophy journals, and statistics and data science publications, scholars have begun to debate the question of how we should think about equality in

---

[234] *See, e.g.*, Nat'l Ctr. For State Cts, Evidence-Based Sentencing To Improve Public Safety & Reduce Recidivism: A Model Curriculum For Judges (2009); Roger K. Warren, Pew Ctr. On The States, Arming The Courts With Research: 10 Evidence-Based Sentencing Initiatives To Control Crime And Reduce Costs 3 (May 2009), https://www.ncsc.org/__data/assets/pdf_file/0018/25551/pew_armingthecourtwithresearch.pdf [https://perma.cc/B47N-7TK4].

[235] *See, e.g.*, Pamela M. Casey, Roger K. Warren & Jennifer K. Elek, Nat'l Ctr. For State Cts., Using Offender Risk And Needs Assessment Information At Sentencing: Guidance For Courts From A National Working Group 3 (2011), https://www.ncsc.org/__data/assets/pdf_file/0019/25174/rna-guide-final.pdf [https://perma.cc/3WF5-3PFF]; Matthew Kleiman, *Using Evidence-Based Practices in Sentencing Criminal Offenders,* 44 The Book of the States 299.

[236] *See, e.g.*, Hannah Bloch-Wehba, *Access to Algorithms*, 88 Fordham L. Rev. 1265, 1265 (2020); Dan L. Burk, *Algorithmic Legal Metrics*, 96 Notre Dame L. Rev 1147, 1183 (2021).

[237] *See, e.g.*, Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans* 1–2 (IZA Inst. of Labor Econ. Discussion Paper No. 12853, Dec. 2019); Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions* 6 (Nat'l Bureau of Econ. Rsch., Working Paper No. 23180).

[238] *See, e.g.*, Neil Shah et al., *Crime Forecasting: A Machine Learning and Computer Vision Approach to Crime Prediction and Prevention*, 4 Visual Computing for Indus., Biomed., and Art 2 (2021); Steven Walczak, *Predicting Crime and Other Uses of Neural Networks in Police Decision Making*, 12 Frontiers in Psych. 8 (2021).

[239] *See* Julia Angwin et al., *Machine Bias*, 3 ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/E7LW-58AK].

the context of algorithmic crime prediction—is it about equal predictive accuracy? If so, should we aim to equalize rates of false positives or false negatives, or the ratio between them, or some other measure?[240] Or are we worried about inappropriate decision inputs (disparate treatment), or disparate impacts across racial groups? And how do any of these equality concerns weigh against the objectives of reducing crime and/or incarceration? Some of the data-science research has focused on incorporating equity or fairness concerns into the design of algorithms.[241] And the possibility of this type of adaptation has given rise to a more optimistic literature: scholars who see in algorithmic prediction a potential *solution* to, or at least mitigation of, entrenched disparities and discrimination.[242]

Today's risk-assessment literature accordingly runs the gamut from what might be described as Luddite (those for whom the entire project smacks of a *Minority Report*-like dystopia, and/or seems inescapably racist)[243] to techno-utopians, with many points in between. This criminal-justice debate is a microcosm of—and frequent case study within—the broader literature on algorithmic fairness throughout society.[244] As someone who began publishing, very critically, about risk assessment and inequality back in 2014—when little was written on this—I appreciate the many thoughtful voices in this increasingly rich debate.[245] And in Part VI, I will return to the topic of designing risk assessments with equity in mind. Notably, though, these sophisticated design approaches are not yet having much impact on the actual risk-assessment market; the tools most jurisdictions are using are far simpler and contain many features that are deeply troubling. As Jens Ludwig and Sendhil Mullainathan, prominent economists who have been mostly optimistic voices regarding risk assessment's potential, wrote recently:

---

[240] *See* Deborah Hellman, *Measuring Algorithmic Fairness*, 106 Va. L. Rev. 811, 823 (2020).

[241] *See, e.g.*, Sahil Verma & Julia Rubin, *Fairness Definitions Explained*, Int'l Workshop on Software Fairness (2018), https://fairware.cs.umass.edu/papers/Verma.pdf [https://perma.cc/HV2C-9QS6].

[242] *See, e.g.*, Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination* in *the Age of Algorithms* (Nat'l Bureau of Econ. Rsch., Working Paper No. 25548), https://www.nber.org/papers/w25548 [https://perma.cc/HR79-2AY6]; Aziz Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 Duke L. J. 1043, 1133 (2019); Hellman, *supra* note 240, at 846–860.

[243] *See, e.g.*, Cathy O'Neil, Weapons of Math Destruction (2016); Katia Schwerzmann, *Abolish! Against the Use of Risk Assessment Algorithms at Sentencing in the US Criminal Justice System*, 34 Phil. & Tech 1883–1904 (2021).

[244] *See, e.g.*, Dana Pessach & Erez Shmueli, *Algorithmic Fairness*, arXiv 1, 31 (Jan. 21, 2020), https://arxiv.org/pdf/2001.09784.pdf [https://perma.cc/TBT7-D4U3]; O'Neil, *supra* note 243.

[245] *See generally* Starr, *supra* note 54; Sonja B. Starr, *The Risk Assessment Era: An Overdue Debate* 27 Fed. Sentencing Rep. 205–06 (2015); Sonja B. Starr, *The New Profiling: Why Punishing Based on Poverty and Identity is Unconstitutional and Wrong* 27 Fed. Sentencing Rep. 229–36 (2015); Sonja B. Starr, *The Odds of Justice: Actuarial Risk Prediction and the Criminal Justice System* 29 Chance 49–51 (2016); Sonja B. Starr, *Sentencing by the Numbers*, N.Y. Times (Aug. 11, 2014), https://www.nytimes.com/2014/08/11/opinion/sentencing-by-the-numbers.html [https://perma.cc/CFZ4-T4FV].

[T]he optimism . . . did not last long. *In practice*, algorithms often proved less helpful than anticipated. In many cases, they were even actively harmful. Some algorithms proved to be no more accurate than the judges whose prediction errors they were purported to correct. Reports emerged of algorithms that were themselves discriminatory, producing racially disparate outcomes at a high enough rate that the phrase "algorithmic bias" has entered the lexicon . . . Machine learning algorithms in criminal justice are not doomed to fail, but algorithms are fragile: if crucial design choices are made poorly, the end result can be (and is often) disastrous.[246]

To advance this Article's examination of statistical discrimination, I will home in on just one problem with the market-dominating algorithms: the use of socioeconomic status (SES) variables to predict risk.[247] Although it has not been the focus of most of the literature, this is one of the most troubling features of most criminal justice risk assessment tools: they punish poverty. The LSI-R is not an outlier on this. COMPAS, for example, includes dozens of SES-related questions in its questionnaire, such as:
- "Do you have a job?"[248]
- "How hard is it for you to find a job ABOVE minimum wage compared to others?"[249]
- "How often do you have trouble paying bills?"[250]
- "How often do you worry about financial survival?"[251]

---

[246] Jens Ludwig & Sendhil Mullainathan, *Fragile Algorithms and Fallible Decision-Makers: Lessons from the Justice System*, 35 J. Econ. Persps. 71, 72 (2021).

[247] Note that the poverty indicators are not the *only* troubling input variables. For example, I think it is repugnant to treat people adversely because their *parents* were once incarcerated, or in general because relatives have criminal records, as LSI-R and COMPAS both do. It is also indefensible to punish people for having been crime *victims* in the past, *see* Starr, *supra* note 54, at 812–13, or for living in high-crime neighborhoods, *see id.* at 806. If a court could be persuaded to treat discrimination along those vectors as legally problematic to begin with, the arguments against statistical generalizations about these risk factors would apply in those contexts, too. But there are currently no special legal protections for people from justice-involved families, crime victims, or residents of high-crime neighborhoods, so the doctrine in Part II would be harder to apply directly. In Starr, *supra* note 54, at 821–22, I argue that equal protection scrutiny *should* apply to parental incarceration and similar variables, drawing on an analogy to illegitimacy. I also show there that statistical discrimination doctrine *would* support male defendants challenging the use of gender in some of the instruments; that argument would track the SES argument closely, and defense counsel absolutely should not ignore it when it could help their clients. *Id.* at 823–29. But I focus in this Article on the SES issue, because I find it more morally urgent.

[248] COMPAS-CORE Sample Risk Assessment 5, https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE#document/p1/a296558 [https://perma.cc/Y8TC-5B5V] (question 80).

[249] *Id.* at 6 (question 89).

[250] *Id.* (question 92).

[251] *Id.* (question 94).

Other questions vary on these themes, and also examine housing and educa-tion.[252] The more parsimonious risk tools go into less detail, but do often include one or two SES variables, such as employment status and years of education.[253] This is because these variables are, unquestionably, statistically associated with crime risk. In the examples discussed in Sections A and B, one could plausibly attack the *accuracy* of the statistical discrimination at issue. The SES-in-risk-assessment debate is stronger ground for defenders of statistical discrimination. The statistical relationship is real, and the public safety objectives are significant.

Nonetheless, I will argue that the use of these variables to shape risk scores and, thereby, criminal justice outcomes is both wrong and unconstitu-tional, a violation of the intertwined equal protection and due process princi-ples applied in *Bearden* and other cases to discrimination against indigent defendants. To advance this argument (which I detailed in a 2014 article),[254] one need not attack the entire risk assessment project, which does not de-pend on those variables' inclusion. But this specific feature of the risk tools is particularly at odds with the statistical-discrimination case law. Unlike the various racial-disparity concerns that the algorithmic fairness literature has overwhelmingly focused on, the SES issue is not one of disparate impact; it involves explicit classifications. These SES variables are highly race-corre-lated, so their inclusion presumably has a racially disparate impact, and chal-lenging them legally would likely have a racial-equity payoff. But the most effective legal strategy wouldn't turn on that racial impact; it would simply challenge them directly for discriminating against the poor.

It may seem almost quaint, in the context of the algorithmic-fairness literature, to focus on the permissibility of inputs,[255] and not, for example, predictive parity. But as discussed in Part II.B, while disparate treatment of disadvantaged groups certainly isn't the *only* form of discrimination that raises moral concerns, its prohibition is still a bedrock of equality law that is crucial for substantive reasons, not just formalistic ones. I suspect that if the risk assessments explicitly increased Black and Hispanic defendants' risk scores based on race, they would not be the focus of an elaborate scholarly debate about the nuances of racial parity measures. Virtually everyone in-volved in that debate would dismiss such an approach as racist (and it would be squarely barred by *Buck v. Davis*). Few would defend this practice even if the resulting system produced acceptable results on any of the proposed pre-dictive-parity metrics.[256] And I will argue here that even if discrimination

---

[252] *See id.*

[253] *See* Starr, *supra* note 54, at 805.

[254] *See id.* at 821–62.

[255] *See* Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. Chi. L. Rev. 459, 467–73 (2019) (critiquing input-focused measures of fairness).

[256] The wrongfulness and illegality of treating race as a risk factor is largely taken for granted in the literature. *See, e.g.*, Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Legal and Statistical Framework*, 119 Mich. L. Rev. 291, 302 (2020). For

against poor defendants is not *quite* as vile as that would be (given the unique social role of race), it's still really bad, and it's still illegal.

The principle barring discrimination against indigent criminal defendants is often referred to as the "*Griffin* equality principle," after *Griffin v. Illinois*,[257] a case about indigent would-be appellants' access to expensive trial transcripts. Justice Black wrote for a plurality:

> Providing equal justice for poor and rich, weak and powerful alike, is an age-old problem. . . . In this tradition, our own constitutional guaranties of due process and equal protection both call for procedures in criminal trials which allow no invidious discriminations between persons and different groups of persons. Both equal protection and due process emphasize the *central aim of our entire judicial system*—all people charged with crime must, so far as the law is concerned, "stand on an equality before the bar of justice in every American court." . . . *In criminal trials, a State can no more discriminate on account of poverty than on account of religion, race, or color.* . . . There can be no equal justice where the kind of trial a man gets depends on the amount of money he has.[258]

This language is strong; "religion, race, [and] color" are classifications triggering strict scrutiny. The plurality further explained that the Illinois system, by effectively excluding many poor people from the right to appeal, was unconstitutional even though the right to appeal itself was not protected by the Constitution.[259] That is, poverty-based discrimination is a problem even if the same result could lawfully have been reached absent that discrimination.

Griffin has mostly been followed in other cases involving financial barriers to participation in the criminal process.[260] Because it recognizes that formally equal requirements impose disparate burdens on the indigent, and requires affirmative steps to achieve equal access, this line of cases represents a more ambitious vision of equality than a mere anti-classification principle (and this point is what divided the Court in *Griffin* itself).[261] One might object that the principle can't be taken seriously, given that our system falls short in so many ways of realizing that ideal of "equal justice for poor and rich"; for example, problems with the effectiveness of indigent defense

---

a contrary argument from a federal judge, see Richard G. Kopf, *Federal Supervised Release and Actuarial Data (including Age, Race, and Gender),* 27 Fed. Sent'g Rep., 207, 214 (2015).

[257] 351 U.S. 12 (1956).

[258] *Id.* at 16–19 (emphasis added).

[259] *See id.* at 13.

[260] *See, e.g.*, Lewis v. Casey, 518 U.S. 343, 372 (1996); Douglas v. California, 372 U.S. 353, 355 (1963).

[261] *See Griffin*, 351 U.S. at 24 (Frankfurter, J., concurring) (describing the Illinois system as a "squalid discrimination"). The dissenters emphasized the lack of a formal classification but accepted the premise that the Constitution prohibited discrimination against indigent defendants in the anticlassification sense. *See id.* at 28–29 (Burton, J., dissenting); *id.* at 34 (Harlan, J., dissenting).

are rife and overwhelming. Nonetheless, the type of discrimination represented by SES variables in risk assessment is different from this kind of failure (and, on the bright side, ought to be easier to eliminate). It's facial discrimination against the poor in the disparate-treatment sense, which is unusual in the criminal justice system. Outside the context of risk assessment, it would be quite extraordinary to hear a prosecutor argue that a defendant's poverty ought to be held against them. Indeed, traditionally, when evidence of poverty is introduced, it's by the defense: in mitigation in sentencing, or to argue against excessive bail amounts in bail proceedings.[262] But in risk assessment, SES variables cut the other way, because poverty and other markers of disadvantage are statistically associated with increased crime risk.

The Court did confront a very similar use of SES once, in *Bearden*—and emphatically and unanimously rejected it. As explained in Part II, the state defended Bearden's probation revocation by pointing to the statistical increase in recidivism risk associated with his job loss and resulting poverty. The Court's rejection of this argument should be understood as decisive for the risk assessment instruments. Just as the state's statistical argument in *Bearden* did, punishment based on risk assessment "lump[s]" indigent defendants "together with other poor persons," and amounts to "little more than punishing a person for his poverty."[263] Moreover, the *Griffin-Bearden* line of cases does not draw distinctions between poverty itself and other SES factors; most of the cases deal with inability to pay fees, while *Bearden* dealt with unemployment and inability to pay restitution. The Court treated these as synonymous with poverty. All of the above-mentioned SES variables in risk assessments likewise run afoul of *Bearden*'s holding.

Bearden also provides useful guidance as to what kind of scrutiny this "intertwined" equal protection/due process analysis entails; it isn't identical to any one tier of traditional equal protection analysis. The Court held:

> Whether analyzed in terms of equal protection or due process, the issue cannot be resolved by resort to easy slogans or pigeonhole analysis, but rather requires a careful inquiry into such factors as "the nature of the individual interest affected, the extent to which it is affected, the rationality of the connection between legislative means and purpose, [and] the existence of alternative means for effectuating the purpose.[264]

In bail, sentencing, and parole, at least, criminal defendants have a strong interest at stake: their liberty. [265] The "extent to which" that interest

---

[262] Starr, *supra* note 54, at 818.

[263] *Bearden*, 461 U.S. at 671.

[264] *Id.* at 666–67 (citing Williams v. Illinois, 399 U.S. 235, 260 (1970) (Harlan, J., concurring)).

[265] Their interests may be less acute when risks-needs assessments are used to assign services, for example.

"is affected" by the inclusion of SES variables may be hard to prove in any given case, because of the non-transparency of some algorithms and the fact that judges and parole boards don't always explain the risk scores' impact on their decisions. But it's obvious that any risk factor that counts toward risk scores will affect outcomes in *some* cases.[266] Even absent clear proof that consideration of discriminatory risk factors shifted the outcome in their case, a defendant could argue that they are entitled to a decision-making process free from discriminatory factors. In *Buck*, analogously, the Court was not swayed by the fact that the impact of the statistical evidence linking race to crime risk was unknown; Buck was entitled to a remedy because race *might* have shaped the outcome.[267] If we take seriously *Griffin*'s statement equating discrimination against indigent defendants with race discrimination, the same logic should apply here.

What about the state's purpose, and its alternate means for effectuating that purpose? Here it bears emphasis, again, that the *Bearden* Court squarely rejected the state's attempt to cite recidivism risk reduction to justify poverty-based discrimination. This rejection wasn't because the state has no interest in preventing crime; other parts of the decision give careful consideration to this interest.[268] Rather, what offended the Court was the use of a statistical generalization in order to establish the connection between what the state wanted to do (revoke probation) and its penological purposes.[269] The Court's reasoning here is fatal to the use of SES variables in risk assessment, which has no justification other than the statistical generalization that disadvantaged people commit more crime. Indeed, risk assessment *is* statistical generalization.

As to alternatives, a defendant challenging the use of SES need not contest the state's asserted interest in risk assessments *per se*. An obvious alternative is to drop just the challenged variables, and indeed, some commonly used risk tools don't include SES factors and still perform well pre-

---

[266] *See* Starr, *supra* note 54, at 862–70 (showing survey-experiment evidence that providing a risk assessment can greatly change preferred outcomes).

[267] *See* Buck v. Davis, 137 S.Ct. 759, 775 (2017).

[268] *See Bearden*, 461 U.S. at 669–70.

[269] The Court gave much more weight to another asserted state interest that did *not* depend on such a generalization: that Bearden's job loss stopped him from paying restitution, which, the state argued, meant that he wouldn't be adequately punished for his crime absent incarceration. *See id.* at 670–71. The Court left open the possibility that this argument *could* justify revocation, but not automatically; less burdensome alternatives must be considered first. Relatedly, the Court in *Bearden* stated that a sentencing court could consider "employment history and financial resources" as "part of evaluating the entire background of the defendant." *Id.* at 671. Defenders of the SES variables could potentially try to rely on this language to limit *Bearden*'s impact, but this would misread the Court, which in this passage was referring back to its earlier discussion of the legitimate interest in ensuring that a defendant's inability to pay restitution does not "immunize[ ] him from punishment." *Id.* at 669. Nothing in this discussion suggests that it would be acceptable for a sentencing court to shape the initial sentence based on the statistical generalization that the poor or unemployed are more dangerous (which the Court considered offensive).

dictively.[270] In this part of the inquiry, states using non-transparent corporate instruments may actually be at a disadvantage. Without the ability to share the proprietary information involved in the algorithm and its design process, it will be harder for states to establish that its designers carefully considered the alternative of eliminating the SES variables, or that doing so would have come at unacceptable cost to accuracy.

Moreover, even if these variables *are* important to the instruments' overall group-level accuracy, their value in improving forecast accuracy for each *individual* defendant is likely quite minimal. After all, as discussed in Section A, even when statistical tools make very accurate group-level predictions, they often tell you little about what to expect for a particular individual.[271] This is especially true when the tool is predicting the rate at which a probabilistic outcome (e.g., pretrial crime) occurs in a group, in which the actual outcome for any given individual is all-or-nothing (*no* individual matches the group average).[272] For example, suppose a pretrial risk tool predicts a likelihood of pretrial crime (if released) of forty-five percent, which would typically be labeled "high risk."[273] But fifty-five percent of those assigned that probability prediction will *not* commit a crime if released, even if the risk tool is perfectly accurate at the group level. Given the huge uncertainties already associated with individual prediction, it's hard to see a slight reduction of predictive accuracy as a major detriment to the state's penological interests—even if statistical generalizations *were* permitted to establish those interests.[274]

Bearden remains good law. Most criminal defendants are indigent, and many are likely harmed by inclusion of SES variables. Yet defense attorneys don't seem to be invoking *Bearden* to challenge risk assessments, or at least, there's no discussion of it in published decisions. Defendants are often unrepresented at bail and parole proceedings. But even at sentencing, which presents a better chance for counsel to advance well-developed constitutional arguments, it's very likely that many defense counsel don't understand how the risk assessments work or how SES factors play in (they *can't* fully

---

[270] For example, the Arnold Foundation's popular pretrial Public Safety Assessment considers only age and criminal history, *see About the Public Safety Assessment*: *What Is the PSA?*, Advancing Pretrial Pol'y & Rsch., https://advancingpretrial.org/psa/about/ [https://perma.cc/WD6J-EFJH], and has performed well in numerous validation studies, *About the Public Safety Assessment*: *Research*, Pretrial Pol'y & Rsch., https://advancingpretrial.org/psa/research/ [https://perma.cc/62PU-NMAA] (collecting studies).

[271] *See* Starr, *supra* note 54, at 842–50.

[272] *See id.* at 845.

[273] *See, e.g.*, *Colorado Pretrial Assessment Tool (CPAT) Administration, Scoring, and Reporting Manual Version 2*, Colorado Ass'n of Pretrial Servs. 9 (2015) (stating that the highest risk category has on average a forty-two percent probability of recidivism).

[274] In Starr, *supra* note 54, at 856, I give another reason the link between the state's interest and the classification is weak: the risk tools do not assess the impact of the decision being considered on crime risk, but rather estimate risk conditional on release. Those questions are not identical; the characteristics that predict especially an especially negative effect of detention on crime are not the same as those that predict higher risk.

understand this when instruments are non-transparent).[275] Institutionally, public defenders' offices may be reluctant to take strong public positions against risk assessment, which benefits some of their clients, although few presumably are advantaged by the SES variables specifically.[276] Nor has the inclusion of SES variables received much public attention.

This pattern should change. Arguments against this practice stand on strong ground doctrinally, and the doctrine stands on strong ground morally. The generalization that the poor are criminally dangerous is noxious, as *Bearden* recognized. And the *Griffin* equality principle resonates, many decades later, because it's still true that providing equal justice to the poor is a central challenge for our criminal process—and that our failure to meet this challenge is to our collective shame. Resource-related problems like access to effective counsel are stubbornly difficult to solve—at least, absent sufficient political will. But it wouldn't be difficult to stop making incarceration decisions turn on poverty in the disparate-treatment sense. Algorithm developers could drop these variables; legislatures or court systems could refuse to use algorithms that include them; and courts could very easily strike them down.

IV.   RACE AND HEALTH CARE ALGORITHMS

As the NFL correctly (albeit hypocritically) observed during its recent controversy, race-norming practices like those it used have support in neuropsychology—and indeed, pervade health sciences generally. In many contexts, race is routinely incorporated into the algorithms that shape diagnosis and treatment, sometimes called Clinical Practice Guidelines (CPGs). Recently, some doctors have critiqued those algorithms, arguing that they have weak empirical support and contribute to racial disparities in care.[277] But racialized CPGs have escaped legal challenge to date, legal scholarship has largely ignored them, and participants in the health-care debates have not framed the issue in legal terms—even though, as I'll argue, the above-discussed doctrine on statistical discrimination provides a potent basis for

---

[275] *See* Jason Tashea, *Calculating Justice*, 103 ABA. J. 54, 58 (2017) (arguing that non-transparency impedes counsel from bringing constitutional challenges to risk assessments).

[276] *See* Mathew DeMichele et al., *What Do Criminal Justice Professionals Think About Risk Assessment at Pretrial?*, 83 FED. PROB. J. 31, 37 (2019) (finding that ninety-two percent of sampled defense attorneys saw at least some strengths of Arnold's PSA tool). *But see* Julian Adler et al., *Arguing the Algorithm: Pretrial Risk Assessment and the Zealous Defender*, 21 CARDOZO J. OF CONFLICT RESOL. 581, 586 (2020) (advocating for defense attorneys' obligations to challenge pretrial risk algorithms).

[277] For broad discussions of the problem, see Darshali A. Vyas et al., *Hidden in Plain Sight: Reconsidering the Use of Race Correction in Clinical Algorithms*, 383 NEW ENG. J. MED. 874, 874 (2020); Ashley C. Rondini, *Meso-Level Racism in Medicine*, 20 CONTEXTS 56, 58 (2021); *see also* Tracie Canada & Chelsey R. Carter, *The NFL's Racist 'Race-Norming' Is an Afterlife of Slavery*, SCI. AM. (July 8, 2021), (criticizing both the NFL and several fields of medicine).

legal challenge. I'll first illustrate the issue with examples (Section A) and then turn to legal analysis and potential remedies (Section B).

A. Examples of Racialized Clinical Practice Guidelines

Here, I give four examples of racialized CPGs that have recently received attention: diagnosis of cognitive impairment, lung function testing, kidney disease diagnosis and treatment, and recommendations regarding cesarean sections. These are illustrative, not exhaustive—racialized approaches have been described as pervasive in medical practice,[278] and medical education has been critiqued for reinforcing inaccurate, essentialist conceptions of race.[279] These examples illustrate some common themes in this broader debate.

1. Dementia and Cognitive Injury

The use of race-normed cognitive tests to diagnose dementia and other impairments is common, albeit not universal.[280] The way race adjustments work in this context has already been discussed in Part I. Arguments *for* race-norming typically emphasize the dangers of overdiagnosis, including stigma.[281] But while this concern may be significant in some neuropsycho-

---

[278] The professional debate over norms used in intellectual-disability diagnosis is discussed in Section III.B, *infra*. Another high-stakes discussion concerns an American Heart Association CPG that treats "nonblack" race as a risk factor for heart-failure death; a recent review noted that "[t]he AHA does not provide a rationale for this adjustment," which "direct[s] care away from black patients," who must show more serious symptoms to qualify for hospital admissions and other care referrals. Vyas et al., *supra* note 277, at 874; *see also* Aditi Nayak et al., *Understanding the Complexity of Heart Failure Risk and Treatment in Black Patients*, 13 Circulation: Heart Failure 301, 301 (2020) (observing that Black Americans actually have particularly *high* age-adjusted heart failure risks, "as well as the worst clinical outcomes"). Other examples abound. *See, e.g.*, Rachel H. Kowalsky et al., *The Case for Removing Race from the American Academy of Pediatrics Clinical Practice Guideline for Urinary Tract Infection in Infants and Young Children with Fever*, 174 JAMA Pediatrics 229, 229 (2020) (critiquing a CPG that makes it harder for nonwhite children to get care for urinary tract infections).

[279] *See, e.g.*, Jessica Cerdeña et al., *From Race-Based to Race-Conscious Medicine*, 396 The Lancet 1125, 1125 (2021), https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)32076-6/fulltext [https://perma.cc/2KSN-Y4CW] (describing pervasive racial pseudoscience in medical training, including "racialized belief in diminished pain sensitivity of Black patients," which "translates to consistently inadequate pain management"); Tsai, *supra* note 100.

[280] *See, e.g.*, Philip G. Gasquoine, *Race-Norming of Neuropsychological Tests*, 19 Neuropsychol. Rev. 250, 252 (2009).

[281] *See, e.g.*, Desiree A. Byrd & Monica G. Rivera-Mindt, *Neuropsychology's Race Problem Does Not Begin or End with Demographically Adjusted Norms*, 18 Nature Reviews: Neurology 125, 126 (2022). Byrd and Rivera-Mindt also argue that citations to group differences in cognitive scores have long been a "pillar of scientific racism," *id.*, which is inarguable. But they suggest that this problem stems from a *failure* to apply demographically tailored norms (which, they suggest, is why the score differences exist). Expressive meaning is in the eye of the beholder, I suppose, but (like many other observers, as illustrated by the public reaction to the NFL scandal) I see the adjustment of scores by race as itself implying a claim

logical contexts, critics have argued that it is far outweighed by the risks of underdiagnosis in the context of dementia, a condition usually diagnosed when patients come to providers seeking help for cognitive changes that they and their families have observed. For example, Katherine Possin and colleagues observe that while race-norming "was developed to reduce" the "overpathologizing and overtreating" of people of color, "in many clinical situations, false negatives cause even greater harm, such as when needed services are deemed unnecessary."[282] The practice echoes, they argue, a "damaging, century-long history of assuming that [IQ] differences are primarily inherited and then using this false assumption to legitimize unequal distribution of resources by social class."[283] In truth, they observe, there is vanishingly little human genetic variation that is identifiably associated with distinct racial groups; the best explanations for racial differences in test scores are grounded in socioeconomic disadvantage.[284] Philip Gasquoine has similarly observed that race-norming is intended to reduce misdiagnosis, but has argued that it's a very crude tool for doing so and serves to reinforce damaging social stereotypes.[285] Underdiagnosis also risks denying access to financial support needed by those suffering cognitive decline—not just legal claimants like the NFL players, but more commonly, seekers of disability benefits, including Social Security.[286]

In contexts in which overdiagnosis of cognitive decline (and resulting fear and stigma) is a serious patient-welfare concern, is there an alternative solution to it? What Possin et al. propose is a "precision medicine approach to normative standard adjustments," directly estimating the impact of the SES factors that underlie racial differences:

Black individuals and other marginalized race/ethnic groups in the US have experienced social and economic disparities that have well-documented associations with poor cognitive outcomes. . . . A more scientifically sound alternative to race norms would be regression-based normative approaches that explicitly measure and adjust for social determinants of brain health.[287]

---

about group difference. *See generally* Michelle Fernandes et al., *Addressing Racial Inequities in Neuropsychological Assessment Requires International Prescriptive Standards, Not Demographically Adjusted Norms*, 18 Nature Reviews: Neurology 377 (2022) (agreeing with Byrd and Rivera-Mindt about some of the problems they identify with testing, but arguing that addressing them via "continued use of race-based adjustments risks propagation of entrenched scientific racism by justifying differential thresholds for disparate, ill-defined racial populations."). *Id*. In any case, the practice plainly results in some patients of color losing access to dementia care.

[282] Katherine L. Possin, Elena Tsoy, & Charles C. Windon, *Perils of Race-Based Norms in Cognitive Testing: The Case of Former NFL Players*, 78 JAMA Neurol. 377, 377–78 (2021).

[283] *Id.*

[284] *Id*.

[285] *See* Gasquoine, *supra* note 280, at 254–55.

[286] For a broad overview of how the Social Security Administration uses psychological tests, see *Program Operations Manual System*, Soc. Sec. Admin. (May 22, 2023), https://secure.ssa.gov/poms.nsf/lnx/0424583050 [https://perma.cc/KQ9C-8BSX].

[287] Possin et al., *supra* note 282, at 377.

A similar position was formally adopted by the American Academy of Clinical Neuropsychology ("AACN") in November 2021, extending not just to dementia diagnosis but to neuropsychological testing more broadly.[288] The AACN's statement is quite hedged, discussing the need for clinicians to engage in a context-specific assessment of whether overdiagnosis or underdiagnosis is the greater potential harm to the patient, and to allow that assessment to shape the choice of scoring approach. Ultimately, the AACN did state that it "supports the elimination of race as a variable in demographically-based normative test interpretation," but it warned that this must be understood as a long-term objective, because alternative norming strategies had to be developed and validated.[289] Like Possin et al., it invoked the idea of "precision medicine" and the importance of SES factors in describing its vision of the ideal approach.[290]

It's worth noting what would be different about this kind of alternative—and what wouldn't be. This "precision medicine"-oriented approach isn't an abandonment of group-specific norming, but rather a formally race-blind variant on it, norming on SES-related measures instead (along with other predictors). This race-blindness has some advantages versus the race-norming that occurs now. It would make the procedure easier to defend legally. It would avoid the especially pernicious expressive messages associated with race-norming. It does not rely on race as a biological category, or traffic in the long cultural history of scientific racism, or label any racial group less intelligent. And SES factors themselves are far less likely to be interpreted as fixed or essential.

But even a race-blind instrument of this sort would effectively discriminate based on poverty, such that a poor person with the same unadjusted scores would be less likely to obtain a diagnosis, and thus harder to access care. Moreover, it would still make it, on average, harder for Black and Hispanic individuals than white individuals to obtain a diagnosis of cognitive impairment, precisely because of the strong correlation between race and SES that the authors point to. If racial gaps in average cognitive test scores are largely explained by SES factors, then an instrument that norms test results based on a rich set of socioeconomic variables will likely result in adjustments that are, on average, pretty similar to the race adjustments that currently take place (although producing different results for some subgroups, such as affluent Black people).

Perhaps a more appealing alternative is the one suggested by Gasquoine, who has argued that subgroup-based norming and statistical generalizations can be avoided entirely. Gasquoine argues that clinicians can readily

---

[288] *See Position Statement on Use of Race as a Factor in Neuropsychological Test Norming and Performance Prediction*, Am. Acad. of Clinical Neuropsychology (Nov. 2021), https://theaacn.org/wp-content/uploads/2021/11/AACN-Position-Statement-on-Race-Norms.pdf [https://perma.cc/5WP5-Y8J5].

[289] *See id.* at 5–6.

[290] *See id.* at 5.

gather enough individualized information about a person's baseline to apply an "individual comparison standard" instead.[291] This can often be done even without access to pre-decline or pre-injury sources of information, by extrapolating from current performance on "hold measures," which are tests of particular dimensions of cognitive performance that are "relatively resistant to the particular type of brain injury suspected."[292] In dementia diagnosis, which involves detailed assessments, time isn't of the essence as it might be for, say, emergency-room triage. The individualized information that physicians gather might inevitably be incomplete, but it is hard to imagine it being *less* informative than a crude statistical generalization about a racial group.

2. *Lung function and pulmonary care*

In pulmonary diagnosis and care, a key tool is the spirometer, which measures the volume of air inhaled and exhaled. As sociologist Lundy Braun has documented in detail, race-norming of spirometry results has been ubiquitous for so long that doctors no longer need to do it by hand. With modern electronic spirometers, the "entire process is so fully automated that users are often unaware that in selecting a patient's race, they are activating a 'correction process.'" [293] The adjustment assumes that "normal" lung function for Black individuals is ten to fifteen percent lower than for white individuals; individuals of Asian descent are often subject to a smaller downward correction.[294] As Adam Gaffney and his coauthors observe, this adjustment can dramatically affect medical referrals:

> Using one commonly used prediction equation, an FEV1 [exhalation volume] of 3.5 liters would be in the normal range (91% of predicted) for a 40-year-old, 6-foot-tall Black man, but the exact same value would be considered abnormally low (78% of predicted) for a 40-year-old, 6-foot-tall white man. . . . The upshot is that two otherwise identical patients may be treated differently based on the color of their skin. . . . If the aforementioned Black man . . . actually had a better prognosis than the white man . . . the use of these race-specific equations might be appropriate . . . But

---

[291] Gasquoine, *supra* note 280, at 257–59.

[292] Philip G. Gasquoine, *Performance-Based Alternatives to Race-Norms in Neuropsychological Assessment*, 148 CORTEX 231, 234. (2022).

[293] LINDSAY LUNDY BRAUN, BREATHING RACE INTO THE MACHINE: THE SURPRISING CAREER OF THE SPIROMETER FROM PLANTATION TO GENETICS xvii (Univ. of Minn. Press 2014); *see also* Lucia Trimbur & Lundy Braun, *Race-ing the NFL: How Anti-Black Standards Are Built into the 2015 Concussion Settlement*, TROPICS OF META (Nov. 1, 2020), https://tropicsofmeta.com/2020/11/01/race-ing-the-nfl-how-anti-black-standards-are-built-into-the-2015-concussion-settlement [https://perma.cc/G7MG-N8QA] (observing providers in "nephrology, pulmonology, cardiology, and neuropsychology" are "often not at all aware that they are adjusting by 'race'").

[294] *See* BRAUN, *supra* note 293 at xv.

there is no evidence this is the case, and the lone study of the issue suggests the opposite.[295]

The study referred to in the last sentence concludes that exhalation volume, unadjusted by race, "has the same prognostic significance for 'normal' African-American and white participants."[296]

Moreover, this is no longer the lone study. A wave of recent research has confirmed both how significant the race adjustment's diagnostic impact is, and how seriously it undermines the accuracy of diagnosis. For example, Magnus Ekström and David Mannino used data from a large population health study that included spirometry.[297] They found that when race-specific norms were used, about the same share of white and Black study participants appeared to have some level of lung impairment (8.5 percent and 9.3 percent, respectively), but when white reference values were used for Black participants, their impairment rate jumped to 37 percent.[298] Similarly, reported levels of moderate to severe impairment were 0.8 percent for both white and Black participants when race-specific norms were used, but jumped to 1.7 percent for Black participants when the white norms were used for them.[299] Critically, though, the actual symptoms (breathlessness) reported by participants as well as their subsequent mortality were predicted equally well for both groups by the white-normed numbers; using the Black reference norms led to substantial underdiagnosis, relative to symptoms and outcomes.[300] In other words, the race-norming practice effectively forces lung impairment to be identified by spirometers at similar rates across groups, even though actual lung impairment prevalence sharply differs. The authors note that their findings suggest that the range of mild impairment that is effectively defined away by the race norms affects 28 percent of the U.S. Black population, and more dangerously, that over half of moderate-to-severe impairment cases are being missed.[301] Another 2022 study by Arielle Elmaleh-Sachs et al. had more qualified findings, but likewise concluded that there was no diagnostic advantage to using race-specific norms, and that they should be abandoned.[302]

---

[295] Adam W. Gaffney, Steffie Woolhandler, & David U. Himmelstein, *Are Lung Function Algorithms Perpetuating Health Disparities Experienced by Black People?*, STAT. (Sept. 15, 2020), https://www.statnews.com/2020/09/15/lung-function-algorithms-health-disparities-black-people [https://perma.cc/EFG4-3XJJ].

[296] P.G.J. Burney & R.L. Hooper, *The Use of Ethnically Specific Norms for Ventilatory Function in African American and White Populations*, 41 INT'L J. EPIDEMIOLOGY 782, 782 (2012).

[297] *See generally* Magnus Ekström & David Mannino, *Research Race-Specific Reference Values and Lung Function Impairment, Breathlessness, and Prognosis: Analysis of NHANES 2007–2012*, 23 RESPIRATORY RSCH. 271 (2022).

[298] *See id.* at 274 (table 2).

[299] *See id.*

[300] *See id.* at 274–76.

[301] *See id.* at 276.

[302] *See* Arielle Elmaleh-Sachs et al., *Race/Ethnicity, Spirometry Reference Equations, and Prediction of Incident Clinical Events*, 205 AM. J. RESPIRATORY & CRITICAL CARE MED. 700

If the unadjusted numbers mean the same thing medically across races, then the solution appears straightforward: take race out of the algorithm. Perhaps the recent research will help to push the field in that direction.[303] Yet at least as of this writing, race norms for spirometry are still embraced by international standards, in addition to being programmed into the machines.[304]

So how did this apparently unfounded practice come to be? As Braun explains, the premise of intrinsic racial differences in lung capacity has deep historical roots in nineteenth- and twentieth-century arguments for Black inferiority and white supremacy. [305] The medical foundation for it, however, was always lacking. Braun shows that studies on which race corrections are and were grounded (from nineteenth-century America to twentieth-century apartheid South Africa through work published in the 21st century) attributed differences to underlying racial biology.[306] But this reasoning overlooked many other reasons Black subjects might on average show reduced lung capacity, such as environmental exposures during work or at home, "inferior medical care [and] higher rates of infectious diseases,"[307] nutritional differences, and the like.[308] The spirometer takes a racial disparity in a health indicator (driven by other underlying racial disparities) and allows it to redefine what is "normal" for Black patients.

This redefinition makes it harder for any Black patient to get treatment for lung impairments, even though (indeed, precisely because) at a group level, Black people disproportionately suffer from and die of such impairments, from asthma to pneumonia to COVID-19.[309] As Ekström and Mannino argue:

> [U]sing lung function references that are specific for each race/ ethnicity may contribute to under diagnosis of lung function impairment and disability, failure to identify impaired lung function

---

(2022). This study had a somewhat smaller sample and was predicting relatively rare clinical events, and compared the overall prognostic value across all groups of race-specific and race-neutral equations; many comparisons produced no statistically significant difference, and none favored the race-specific approach.

[303] *See* Neil W. Schluger, *The Vanishing Rationale for the Race Adjustment in Pulmonary Function Test Interpretation*, 205 Am. J. Respiratory Critical Care Med. 612–14 (2022) (calling for abandonment of race norms based on the recent research); Nicole B. Ramsey et al., *Deconstructing the Way We Use Pulmonary Function Test Race-Based Adjustments*, 10 J. Allergy Clinical Immunology Prac. 972–78 (2022) (raising a range of criticisms of race norms and calling for change).

[304] *See, e.g.*, Sanja Stanojevic et al., *ERS/ATS Technical Standard on Interpretive Strategies for Routine Lung Function Tests*, 60 Eur. Resp. J. 25 (2022) ("When interpreting PFT results, a clinician must interpret a particular result as within or outside the normal range for an individual of that age, sex, height and ethnic background based on reference equations . . ."). *See also* Ekström & Mannino, *supra* note 297, at *2 (citing standards).

[305] *See* Braun, *supra* note 293, at 27–54.

[306] *See id.* at 128.

[307] *Id.* at 37.

[308] *See id.* at 190–93, 202–06.

[309] *See* Gaffney et. al, *supra* note 292.

as a contributing cause in evaluation of breathlessness, misclassify the association between lung function and outcomes, and potentially lead to insufficient or delayed treatment and compounded race-related health inequities in the community.[310]

Race correction also makes it harder for Black claimants to obtain compensation for workplace exposures to substances like cotton dust, coal, and asbestos, because race defines the expectation of their baseline lung function; OSHA's framework for assessing the harms of cotton dust exposure builds in a fifteen percent race correction factor for Black lung capacity, for example.[311] This is so even though (despite the large aggregate racial disparities in pulmonary health) within-group variation in lung capacity far exceeds differences across groups, such that these group generalizations about pre-workplace-injury baselines will often be quite inaccurate for individuals.[312]

3. Kidney disease diagnosis

A long-used algorithm that uses blood serum creatinine levels to estimate kidney function is race-normed, apparently because of an underlying assumption that Black people have more muscle mass and normally produce more creatinine. This makes the same test results less likely to translate into a kidney disease diagnosis for Black people, even though the evidence for this assumption is dubious, and even though Black people have higher rates of end-stage kidney disease and resulting death.[313] This underdiagnosis can lead to denial of care.[314] Since 2019, critiques of this adjustment within nephrology have emerged, as well as medical research on alternative methods.[315] In 2020 a joint task force of the American Society of Nephrology and

---

[310] Ekström & Mannino, *supra* note 294, at *6.

[311] *See* BRAUN, *supra* note 278, at 200; *see also id.* at xiii-xv (discussing manufacturer Owens Corning's attempt to use race-normed standards to deny asbestos-related disability claims); *id.* at 167–94 (discussing race correction in work-related silicosis diagnosis).

[312] *See* Gaffney et. al, *supra* note 295, at xiv.

[313] *See* Vyas et al., *supra* note 277, at 875; *see also* Emily Henderson, *National Task Force Advocates Removal of Race from the Kidney Function Algorithm*, NEWS MED. LIFE SCI. (Sep. 21, 2021), https://www.news-medical.net/news/20210924/National-task-force-advocates-removal-of-race-from-thekidney-function-algorithm.aspx [https://perma.cc/3FWQ-7962]; *see also* Salman Ahmed, Cameron T. Nutt, Nwamaka D. Eneanya, Peter P. Reese, Karthik Sivashanker, Michelle Morse, Thomas Sequist, & Mallika L. Mendu, *Examining the Potential Impact of Race Multiplier Utilization in Estimated Glomerular Filtration Rate Calculation on African-American Care Outcomes*, 36 J. GEN. INTERNAL MED. 464, 466 (2020) (finding that one-third of Black patients with chronic kidney disease would be moved to more severe disease categorization if race correction were removed).

[314] *See* Bessie A. Young, *Removal of Race from Estimation of Kidney Function*, 18 NATURE REV. NEPHROLOGY 201, 202 (2022).

[315] *See* Chi-yuan Hsu, Wei Yang, Rishi V. Parikh, Amanda H. Anderson, Teresa K. Chen, Debbie L. Cohen, Jiang He, Madhumita J. Mohanty, James P. Lash, Katherine T. Mills, Anthony N. Muiru, & Afshin Parsa, *Race, Genetic Ancestry, and Estimating Kidney Function in CKD*, 385 NEW ENG. J. MED. 1750, 1755 (2021) (presenting evidence that alternative marker of kidney function, other than serum creatinine, is equally accurate and does not vary by race).

the National Kidney Foundation was established to consider the issue.[316] In September 2021, the task force published a recommendation for new, race-neutral guidelines.[317] However, because the old algorithm is used in countless individual labs throughout the country, the shift away from it will probably not be quick, and will require "sustained advocacy."[318] In addition, debates about best practices persist—some experts have endorsed different race-neutral formulas relying on a different blood marker, for example, arguing that that marker is more accurate across racial groups.[319]

One potentially legally consequential illustration of the problem of inertia in medical practice is the fact that the Bureau of Prisons (BOP) is, as of this writing, apparently still using the old algorithm when considering the requests of kidney-impaired prisoners for compassionate release (either to seek treatment or because kidney impairment places them at heightened risk of COVID). In March 2022, a federal judge in New Jersey rejected a Black prisoner's COVID-related request for release, in part because his kidney impairment was insufficient to justify it.[320] But the data cited by the judge make clear that had a racial adjustment not been applied, the prisoner could potentially have qualified as having "chronic kidney disease," which, the opinion notes, is considered a COVID risk factor.[321] Another prisoner, Jonte Robinson, in April 2022 filed a challenge to the BOP policy in a Virginia federal court.[322] The complaint, which seeks class certification, alleges violations of

---

[316] *See* Nwamaka D. Eneanya, Wei Yang, & Peter Philip Reese, *Reconsidering the Consequences of Using Race to Estimate Kidney Function*, 322 JAMA 113 (2019); *see also* Henderson, *supra* note 313, at 2 (describing these events).

[317] *See* Henderson, *supra* note 313, at 1–2; *see also* Cynthia Delgado, Mukta Baweja, Nilka Rios Burrows, Deidra C. Crews, Nwamaka D. Eneanya, Crystal A. Gadegbeku, Lesley A. Inker, Mallika L. Mendu, Greg W. Miller, Marva M. Moxey-Mims, Glenda V. Roberts, Wendy L. St. Peter, Curtis Warfield, & Neil R. Powe, *Reassessing the Inclusion of Race in Diagnosing Kidney Diseases: An Interim Report from the NKF-ASN Task Force*, 32 J. AM. SOC'Y NEPHROLOGY 1305 (2021) (presenting interim report).

[318] Katie Palmer, *Changing the Equation: Researchers Remove Race from a Calculator for Childbirth*, STAT. (June 3, 2021), https://www.statnews.com/2021/06/03/vbac-calculator-birth-cesarean/ [https://perma.cc/T2JS-AN9M].

[319] *See* Hsu et al., *supra* note 315; Inker et al., *supra* note 317; *see also* Chi-yuan Hsu & Alan S. Go, *The Race Coefficient in Glomerular Filtration Rate-Estimating Equations and its Removal*, 31 CURRENT OPINION IN NEPHROLOGY AND HYPERTENSION 527, 531 (2022) (observing that the cystatin C approach is more accurate with no racial bias, but is more expensive); Holly J. Kramer et al., *An Endorsement of the Removal of Race From GFR Estimation Equations: A Position Statement From the National Kidney Foundation Kidney Disease Outcomes Quality Initiative*, 80 AM. J. KIDNEY DISEASES 691 (forthcoming) (arguing that the race-neutral serum creatinine-based approach of the 2021 recommendation is sufficiently accurate for most purposes, but that a formula incorporating cystatin C levels is preferable where high accuracy is essential); Young, *supra* note 314, at 202 (similarly endorsing the cystatin C approach based on recent research).

[320] *See* United States v. McPhatter, No. 18-578, 2022 WL 874457, at *6–7 (D.N.J. March 23, 2022).

[321] *See id.* at *7 n.4 and accompanying text (citing these figures in passing and without questioning racial adjustment); Goldstein, *supra* note 5.

[322] *See* Complaint at 21, Robinson v. Fed. Bur. Prisons, No. 1:22-cv-01098 (D.D.C. Apr. 20, 2022).

the Equal Protection and Cruel and Unusual Punishments clauses, as well as the Administrative Procedure Act.

In November 2021, a dozen scholars published a commentary describing the kidney-function race correction as a classic example of "how structural racism [shapes] adverse health outcomes."[323] The authors argued that the notion that Black creatinine levels are "naturally" higher had weak scientific support. Some studies used samples in which Black participants were more likely to have kidney disease and/or socioeconomic risk factors that might impact creatinine levels but also could adversely impact health outcomes.[324] As with the lung capacity example, then, racial disparities in health and health-related risk factors may have effectively been locked into what the formulas treated as "normal" for Black people.[325]

4.   *Vaginal Birth After Caesarean Section (VBAC).*

Beyond diagnostic decisions, race also figures into other algorithms that guide treatment decisions. For example, in obstetrics, patients who have previously had a Caesarean section often face a difficult decision as to whether to attempt a vaginal delivery (VBAC) or schedule another C-section. Until recently, an online calculator widely used to guide this decision used Black race as a risk factor predicting VBAC failure—a prediction that guided some Black patients to forgo the attempt and some doctors to refuse to offer it to them.[326] C-sections are major surgeries, carrying higher risks than vaginal births; recovery takes six weeks absent complications. After criticism within the field, the calculator was rendered race-neutral in June 2021—medicine's "first example of race correction being abandoned sys-

---

[323] Nwamaka D. Eneanya, L. Ebony Boulware, Jennifer Tsai, Marino A. Bruce, Chandra L. Ford, Christina Harris, Leo S. Morales, Michael J. Ryan, Peter P. Reese, Roland J. Thorpe Jr., Michelle Morse, Valencia Walker, Fatiu A. Arogundade, Antonio A. Lopes, & Keith C. Norris, *Health Inequities and the Inappropriate Use of Race in Nephrology*, 18 NATURE REV. NEPHROLOGY 84, 85 (2021).

[324] *See generally id.* Hsu & Go, *supra* note 319, at 532, raise the concern that the newly recommended race-neutral formula, by labeling more Black Americans with kidney disease or with more severe kidney disease, will mask racial disparities in outcomes for kidney disease at any given severity level. But if the new formula predicts adverse health outcomes more similarly well across races (as Hsu and Go observe, citing another recent study, Guofen Yan et al., *Estimation of Black-White Disparities in CKD Outcomes*, 80 AM. J. KIDNEY DIS. 423 (2022)), that seems potentially like an asset for clinical purposes, as it means people facing the same risks are more likely to get the same diagnoses. The change does, of course, need to be taken into account when comparing outcome-disparity studies that use different diagnostic standards, as Yan et al. observe.

[325] Another example from nephrology is the algorithm used to predict kidney transplant success, used by the national Kidney Allocation System. This algorithm predicts that kidneys of Black *donors* are less likely to be successfully transplanted; it disproportionately affects Black recipients because transplants are mostly intraracial. *See* Vyas et al., *supra* note 277, at 875. The result is longer wait times for Black recipients. *Id.*

[326] *See* Palmer, *supra* note 318; *see also* Shakeela Faulkner et al., *The Effects of Removing Race from the VBAC Calculator: Implications for Counseling*, 224 AM. J. OBSTETRICS & GYNECOLOGY S467 (Feb. 2021); Nicholas Rubashkin, *Why Equitable Access to Vaginal Birth Requires Abolition of Race-Based Medicine*, 24 AMA J. ETHICS E233 (2022).

tematically in a tool in response to these equity concerns."[327] Still, it's worth considering the example, because it illustrates a somewhat different way that racialized CPGs can lock in and amplify existing disparities.

If VBAC attempts fail more often for Black patients, why is this so? Darshali Vyas and coauthors observe that the biological explanations that some have proffered ("ethnic variation in pelvic architecture") have "historically racist antecedents" and are "not supported by biological plausibility."[328] Other possible factors include race-correlated socioeconomic and socioeconomically mediated health risk factors (e.g., insurance status; hypertension), as well as disparities in care across providers.[329] But other risk factors could be accounted for directly without relying on a race proxy, and if disparities across providers is the issue, it's hard to explain why race-contingent success rates should be relevant to the advice a given care team gives their own patients, which is what the calculator is used for. Black patients may have less access systematically to care that will help VBAC attempts to succeed—a social problem demanding policy solutions. But presumably a given care team expects to provide equally good care to its Black patients as to other patients. If it fails to do so, then directly addressing that problem should be the first-order concern, rather than steering Black patients into surgeries.[330] To the extent that there are differences *across* providers that have an aggregate racial effect (which would not be accounted for in a race-neutral algorithm), these could perhaps be addressed by having providers or facilities report their own local success and attempt rates, which would provide more specific information to their patients than a national race-specific average would. Researchers have found that hospital practices dramatically affect VBAC access, and have called for dissemination of information to patients, as well as other interventions to reduce those disparities.[331] In any event, one recent study found that removing race from the VBAC calculator had no effect on its predictive accuracy,[332] and another found that the race-

---

[327] Palmer, *supra* note 318 (quoting Darshali Vyas).

[328] Darshali Vyas et al., *Challenging the Use of Race in the Vaginal Birth After Cesarean Section Calculator*, 29 WOMEN'S HEALTH ISSUES 201, 202 (2019).

[329] *See* Laura B. Attanasio & Mary T. Paterno, *Racial/Ethnic Differences in Socioeconomic Status and Medical Correlates of Trial of Labor After Cesarean and Vaginal Birth After Cesarean*, 30 J. WOMEN'S HEALTH 1788, 1792 (2021) (discussing possible mediating factors).

[330] A substantial literature exists on causes and responses to racial disparities and discrimination in obstetrics generally. *See, e.g.*, Rebecca F. Hamm et al., *Addressing Disparities in Care on Labor and Delivery*, 11 CURRENT OBSTETRICS & GYN. REPS. 143 (2022).

[331] *See* Jourdan E. Triebwasser et al., *Hospital Contribution to Variation in Rates of Vaginal Birth After Cesarean*, 39 J. PERINATOLOGY 904 (2019) (finding "ten-fold difference in risk-adjusted rates of VBAC" across hospitals in a region, and concluding that there is "an urgent need to identify and disseminate institutional practices associated with increased access to TOLAC and VBAC in order to reverse the trend of increasing repeat cesarean deliveries and concomitant maternal morbidity."). *Id.* at *6–7.

[332] *See* Ayisha Buckley et al., *Racial and Ethnic Disparities Among Women Undergoing a Trial of Labor After Cesarean Delivery: Performance of the VBAC Calculator With and Without Patients' Ethnicity*, 29 REPROD. SCI. 2030 (2022).

specific calculator had systematically underpredicted Black and Hispanic VBAC success rates.[333]

Moreover, the VBAC example is a useful illustration of how race-based medical claims can amount to self-fulfilling prophecies. In many instances, the choice of how soon to abandon a labor attempt in favor of a C-section is discretionary; barring an emergency, doctors and patients have choices. And predicted success probabilities can shape those choices:

> [I]f the cesarean health disparity is driven by implicit bias and systemic racism, then the VBAC calculator may contribute to this in an ongoing fashion. . . . First, because the calculator will provide lower estimated success rates, it may dissuade Black women from even attempting a [VBAC]. But, it also may influence the outcomes among those who attempt a trial of labor as well. It has been shown that having lower expectations for success actually leads to lower success rates.[334]

Speaking of the recent change in the calculator, top New York City health officer Michelle Morse stated: "It's a really critical acknowledgement by the medical community that we got it wrong [and that] we have not always been transparent . . . about how racism shapes the clinical questions we're asking."[335]

B.   The Law and Racialized CPGs

Many of the health literature's recently published commentaries on racialized CPGs offer very thoughtful, detailed reflections on what medical science, ethics, and/or racial equity demand.[336] But one thing is strikingly missing from them: any discussion of the law. Law is absent even from the reports of expert commissions appointed to reconsider particular guidelines. Meanwhile, legal scholarship has paid very limited attention to racialized CPGs. The legal academy's most prominent voice on the issue has for years been Professor Dorothy Roberts, who has given talks calling for abolition of race correction.[337] But Professor Roberts's commentary, much like the criticism within the medical field itself, has focused on the scientific failings and

[333] *See* Faulkner et al., *supra* note 326.

[334] Editorial Comment, *Hidden in Plain Sight: Reconsidering the Use of Race Corrections in Clinical Algorithms*, 76 OBSTETRICS & GYNECOLOGY SURVEY 5, 6 (2021).

[335] Palmer, *supra* note 318.

[336] *See, e.g.,* Kowalsky et al., *supra* note 278; Vyas et al., *supra* note 277; Canada & Carter, *supra* note 277; Trimbur & Braun, *supra* note 293.

[337] *See* Dorothy Roberts, *TEDTalk*: *The Problem with Race-Based Medicine* YOUTUBE (March 4, 2016), https://www.youtube.com/watch?v=KXLMjn4WPBY [https://perma.cc/45TP-XX6L]; *see also* CPM Oxford, *Session 4: Dorothy Roberts: The Past and Future of Race, Health and Justice*, YOUTUBE (June 29, 2021), https://www.youtube.com/watch?v=MUYdOFS8JbU [https://perma.cc/QKY6-6ED2]; Pete Madden, *Neuropsychologists Call For Elimination Of Race-Norming In Clinical Tests Following NFL Concussion Controversy*, ABC NEWS (Dec. 2, 2021) https://abcnews.go.com/US/neuropsychologists-call-

racist history of these practices. What I seek to add here is a more traditional legal inquiry: are these practices permitted by current statutes, regulations, and doctrine? Can patients denied care sue? Neither medical nor legal literature to date appears to have focused on these questions.[338]

This should surprise us. Medicine is a highly regulated profession, and its modern practice is also heavily shaped by the specter of litigation.[339] Medical practitioners, hospitals and other employers, health insurers, liability insurers and the like, not to mention lawyers and law professors, think about the way law governs medicine *all the time*. But not here, apparently— even though it should be obvious that it might be illegal to base care on explicitly racialized algorithms. As noted above, the kidney algorithm is now finding its way into court because of its collateral effect on prison release decisions, and this may provide an opportunity for a federal court to pass on it soon (at least in that specific context). But one might expect that even without a connection to the justice system like this, patients denied care could file suit. It doesn't seem to have happened yet, but it could.

Numerous legal restrictions on racial discrimination govern U.S. health care providers. Title VI of the 1964 Civil Rights Act prohibits racial discrimination by entities receiving federal funds for any aspect of their operations, which encompasses the vast majority of U.S. health facilities, and Section 1557 of the Patient Protection and Affordable Care Act reinforces that protection and extends it to protect participants in ACA health plans.[340] Health care is also governed by state antidiscrimination statutes, which vary, and various specialized federal statutes.[341] Providers in public facilities are state actors who are additionally subject to federal and state constitutional restrictions, and can be sued for breaching patients' constitutional rights under 42 U.S.C. § 1983.[342]

---

elimination-race-norming-clinical-tests-nfl/story?id=81493363   [https://perma.cc/K7A8-TGW4].

[338] Aside from Professor Roberts's talks, the most detailed treatment by legal scholars that I have located is in one recent article on algorithmic discrimination in health care, which devotes a few pages to the topic of racialized CPGs. *See* Sharona Hoffman & Andy Podgurski, *Artificial Intelligence and Discrimination in Health Care*, 19 Yale J. Health, Pol., L. & Ethics, 1, 19–23 (2020). But this discussion focuses on disparate impact claims for other sorts of algorithmic bias, stating briefly that disparate treatment claims might be available in "extreme" cases involving "malevolent" providers or "deliberate indifference"; it does not address the basic question whether the routine usage of racialized CPGs is illegal disparate treatment. *See also* Priya Desai, Note, *The Use of Race in Medical Artificial Intelligence*, 212 U. Pitt. J. Tech L. & Pol'y, 149, 156–66 (2021) (arguing for FDA regulation of racialized algorithms, but not considering whether they violate existing disparate-treatment bans).

[339] *See generally* Robert I. Field, *Why is Health Care Regulation So Complex?*, 33 Pharmacy and Therapeutics 607 (2008).

[340] *See* 42 U.S.C. § 18116.

[341] The Department of Health and Human Services lists applicable federal nondiscrimination statutes. *See Laws and Regulations Enforced by OCR*, U.S. Dep't of Health and Human Servs., https://www.hhs.gov/civil-rights/for-providers/laws-regulations-guidance/laws/index.html [https://perma.cc/XR49-722T] .

[342] *See* McCabe v. Nassau Cnty. Med. Ctr., 453 F.2d 698, 703 (2nd Cir. 1971).

I'll focus the analysis that follows on Title VI, since its applicability is perhaps the most sweeping, but the substantive argument is largely similar with respect to other legal tools, including Section 1557. Title VI had a massive effect on healthcare when first passed, effectively ending the segregation of thousands of facilities. Since then, it has been referred to as a "sleeping giant" because, despite its scope, Title VI litigation is infrequent.[343] This infrequency stems largely from widely-criticized limitations, including courts' refusal to allow private lawsuits for disparate impact discrimination.[344] Meanwhile, as with other civil rights laws, claims of intentional discrimination typically founder on issues of proof.[345] But these obstacles would not impede challenges to racialized CPGs, which draw explicit racial classifications on their face. Title VI provides a powerful potential remedy against hospitals or other facilities (albeit not against individual physicians) for individuals denied care because of race.

Although the statistical discrimination precedents discussed in Part II do not come from the Title VI context, there's every reason to believe that the Court would apply them in that context. The Supreme Court has held that Title VI's prohibition of intentional racial discrimination is substantively coterminous with the Equal Protection Clause,[346] meaning that Title VI effectively extends the strict-scrutiny standard that applies to governmental race discrimination to "any recipient of Title VI funds."[347]

Strict scrutiny is a demanding standard, and each of the CPGs discussed above seems at least fairly likely to fail it. To be sure, in medical practice (much more readily than in settlement claims administration, as in the NFL case), one can at least *imagine* interests that might clear this bar. Hypothetically, there could be situations where reliance on race was so valuable in predicting a particular kind of medical need or outcome that failing to do so would exact an unacceptable cost in human life and well-being. In such a situation, the identification of a compelling interest would not exhaust the strict scrutiny analysis; rather, the courts would have to consider whether the CPG was narrowly tailored—i.e., that no race-neutral approach could accomplish the interest in question.

As noted in Part II, it may be that, legally, statistical generalizations *can't be invoked* to show that strict scrutiny is satisfied, in which case any

---

[343] Wendy Dunne DiChristina, *"So Sue Me": Medical Professionals Should Support Strengthening Title VI Anti-Discrimination Laws*, 7 Voices in Bioethics (2021).

[344] *See* Alexander v. Sandoval, 532 U.S. 275 (2001); Amitabh Chandra, Michael Frakes, & Anup Malani, *Challenges To Reducing Discrimination And Health Inequity Through Existing Civil Rights Laws*, 36 Health Affs. 1041 (2017) (criticizing this limitation and others); Jamille Fields Allsbrook & Katie Keith, *ACA Section 1557 As A Tool For Anti-Racist Health Care*, Health Aff. Blog (Dec. 8, 2021), https://www.healthaffairs.org/content/forefront/aca-section-1557-tool-anti-racist-health-care [https://perma.cc/9KMX-38RM].

[345] *See* Chandra et al., *supra* note 344.

[346] *See Grutter,* 539 U.S. at 343 (citing Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 287 (1978)).

[347] Title VI Legal Manual, Civil Rights Division U.S. Dep't of Just. (2021) at § VI.C.2.

defense of the CPG falls apart. But even if the courts were less absolutist in excluding such defenses, it seems unlikely that the evidence supporting many racialized CPGs could survive strict scrutiny. CPGs generally have *some* empirical grounding, but the examples in Part A suggest that many are effectively misusing empirical facts about racial disparities. A recurring pattern is this: nonwhite populations are found to disproportionately bear some characteristic that predicts adverse outcomes. But rather than treat this as evidence of a worrisome health disparity, diagnostic CPGs use the disparity to redefine each group's "normal" baseline, which in turn tends to magnify disparities via denials of care.[348] Tellingly, the data that underlies these CPGs doesn't tend to speak to care outcomes, and the lung-capacity example illustrates that differences in group averages in the prevalence of some health indicator don't imply any racial difference, conditional on that indicator, in the need for care. And while treatment-success algorithms like the VBAC calculator operate slightly differently, they again treat troubling disparities as though they are intrinsic, with the effect of magnifying them.

Finally, the studies on which the CPGs rely don't tend to compare them to race-neutral alternatives, and without that comparison, it would be difficult to satisfy strict scrutiny. In several of the contexts above, critics have proposed race-neutral approaches that appear viable (sometimes involving another approach entirely, sometimes simply removing race). And some fields that have been using racialized CPGs for decades are beginning to embrace these alternatives—although the use of race in medicine is so ubiquitous that it's far too soon to declare that this is a problem likely to disappear without legal intervention.

In identifying these problems, I do not rely just on my own assessment, but on the critiques of experts within the fields in question. When specialists are calling attention to the weak scientific basis for racialized CPGs, their inequitable consequences, and their connection to historic prejudices and scientific racism, lawyers and courts should take note. We ought not to throw up our hands and assume the dominant practice must have a strong scientific foundation or that moving away from it would have unacceptable costs in health outcomes. The demand of modern equality law is that practices that discriminate based on race cannot be blindly accepted; they must be subject to rigorous scrutiny.

One potential argument defendants might raise in court is that the use of race in CPGs *isn't a racial classification*—or at least not a "suspect" one that would trigger strict scrutiny. Although this may seem curious on its face, it's not a frivolous argument doctrinally, although it is ultimately wrong. Defendants would likely draw an analogy to police use of racially

---

[348] As the dementia discussion illustrates, defenders of race corrections have suggested that CPGs that make it harder to get care actually *benefit* Black patients by avoiding overdiagnosis. But it's unclear how to value this vague "benefit" versus costs of undertreatment, especially given that CPGs are typically used in contexts in which patients are seeking care.

specific witness descriptions to identify criminal suspects, to which courts have not applied strict scrutiny.[349] Instead, courts have treated race in this context as just a physical characteristic, like hair and eye color. In *Brown v. City of Oneonta*,[350] the Second Circuit expressly differentiated this use from police racial profiling and/or stereotyping, finding that the police policy was "race-neutral": they relied on all the physical characteristics provided by witnesses, and would do the same regardless of the race of the suspect.[351] One can imagine an argument that CPGs are the same: they use race information from all patients, treating them simply as another biological fact (like age, sex, weight, and symptoms) that helps the provider come to an accurate diagnosis or prognosis. Perhaps using race to tailor medical care is just accommodating a physical difference, no different from a cosmetologist tailoring makeup to skin tone, which nobody would deem discriminatory.

But this argument doesn't hold up, even if one accepts the contestable premise of the police-suspect-description cases.[352] The above-discussed CPGs are much more like racial profiling than suspect descriptions or the hypothetical cosmetologist. The CPGs depend on statistical generalizations about groups by lumping people together. They are not using race to describe a particular person's appearance; rather, they make probabilistic predictions based on group averages and norms. The fact that these are predictions about the body (or mind) instead of behavior or other tendencies doesn't change that. In *United States v. Virginia*,[353] the Court made clear that while "irreducible physical differences" may exist in the sex context (a necessary distinction in this context—nobody argues that doctors must ignore sex differences),[354] its doctrine squarely rejected the possibility of inherent racial differences.[355] And this is an important point to insist on, because the United States' history of racial oppression is rife with claims of biological differences between races. As scholars Tracie Canada and Chelsey Carter have recently written:

> [D]espite its contemporary uses, race norming can be traced back to plantation slavery, eugenics efforts globally and a long history of racial science used to justify the belief in inferior racial groups.

---

[349] *See* R. Richard Banks, *Race-Based Suspect Selection and Colorblind Equal Protection Doctrine and Discourse*, 48 UCLA L. Rev. 1075, 1077–78 (2001).

[350] 221 F.3d 329 (2d Cir. 2000).

[351] *See id.* at 337–38.

[352] R. Richard Banks offers a compelling critique of these cases. *See generally* Banks, *supra* note 349.

[353] 518 U.S. 515 (1996).

[354] Sex-norming of diagnostic standards does raise some issues, including the accommodation of individuals whose bodies do not map neatly onto a male/female binary. But recognizing the complexity of sex categorization clearly doesn't require that all genetic and physical differences associated with sex be ignored by doctors. Because the case law distinguishes between sex and race when physical differences are concerned, however, there is no reason to worry that a legal bar on race-norming would need to be extended to sex-norming.

[355] *See Virginia*, 516 U.S. at 533.

> These misguided scientific endeavors are rooted in an idea that Black people's bodies are inherently different from white people's bodies.[356]

It cannot be the case that a race-based group generalization may escape strict scrutiny merely by framing it as a claim of an essential biological difference; if anything, given this history, courts should be *more* suspicious of such claims.

Finally, another legal authority that could potentially soon change the legal landscape governing racialized CPGs is a recent proposed rule from the Department of Health and Human Services (specifically, the Center for Medicaid and Medicaid Service and the Office for Civil Rights), interpreting Section 1557 of the Affordable Care Act, under which the Department has the statutory authority to pass regulations.[357] The Rule, which broadly addresses discrimination issues related to health care, was open for notice and comment from August to October 2022. Section 92.210 of the proposed Rule states: "[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."[358]

If it passes, the effect of this provision will depend on how aggressively the Department interprets and enforces it. In its Notice of Proposed Rulemaking, in the stated reasons for this provision, the Department directly—but somewhat noncommittally—addresses the race-norming issue, acknowledging the NFL controversy, describing the problems with and change in the kidney function guidelines, and listing related controversies in other medical fields (including all those discussed in this paper).[359] It concludes that reliance on the older kidney-function algorithm "may lead to discrimination against patients based on race and ethnicity . . . if a covered entity takes action based on the algorithmic output that results in less favorable treatment of a Black patient as compared to white patients with similar or healthier kidneys because an algorithm determined that a Black patient's kidney function is better than it actually is."[360] Regarding CPGs in other fields, the Notice states:

> Covered entities must be mindful when using tools that rely on racial or ethnic variables to ensure their reliance on such tools does not result in discriminatory clinical decisions. We encourage covered entities to use updated tools that have removed or do not have known biases, such as the updated eGFR discussed above. The Department notes that the use of algorithms that rely upon race and ethnicity-conscious variables may be appropriate and justified

---

[356] Canada & Carter, *supra* note 277.
[357] *See* DHHS Notice of Rulemaking, *supra* note 7.
[358] *Id.*
[359] *Id.*
[360] *Id.*

under certain circumstances, such as when used as a means to identify, evaluate, and address health disparities.[361]

"Must be mindful" and "encourage" are fairly soft, hortatory phrases, and it is possible that the Department will be deferential in its application of this provision, even if it passes unchanged.[362] Still, this document is a pretty strong sign that race-norming is at least on the Department's radar as a potential civil rights violation. The specific emphasis on the kidney example makes it very plausible that OCR will be friendly to Section 1557 complaints when covered entities fail to adopt the new race-neutral standard for estimating glomerular filteration—a possibility that could help to hasten what has otherwise been expected to be a slow shift.

It's probably no accident that the example the Department focuses on is one in which leading authorities in the field have already embraced change, and it might similarly be more willing to dive into other race-norming controversies (like the VBAC example) where the field's norms have already shifted. Still, even where this isn't the case, there are other examples in which race-norming leads to exactly the sort of discrimination they illustrate with the kidney example, and where there is no plausible justification grounded in redressing disparity. The example of spirometry is a good one: the practice clearly adversely affects Black patients and appears completely lacking in a medical justification, and a patient denied care on that basis should have a strong basis for a Section 1557 complaint.

A successful complaint would create a strong incentives for the funded entity to comply voluntarily and change their algorithms, since enforcement could potentially entail loss of federal funds.[363] Moreover, even if the facility fights an adverse finding in court or even if OCR does not act at all, its regulation is potentially entitled to *Chevron* deference as an interpretation of Section 1557, and thus should be helpful to civil litigants challenging denials of care (although the Supreme Court appears to be becoming less *Chevron-*

---

[361] *Id.*

[362] It can expect some pushback. The American Hospital Association submitted comments in which the algorithms section was one of just two provisions in the lengthy proposed regulation that it contested. *See AHA Comments to HHS on Nondiscrimination in Health Programs, Activities in Sec. 1557 of the Affordable Care Act*, Aᴍ. Hᴏsᴘ. Assoᴄ. (Sept. 28, 2022), https://www.aha.org/lettercomment/2022-09-28-aha-comments-hhs-nondiscrimination-health-pro-grams-and-activities [https://perma.cc/QZP5-P9TE]. It argued that race-conscious algorithms are sometimes indeed necessary to redress racial disparities, and that it was "vital that non-discriminatory and beneficial uses of such algorithms not be over-deterred." *Id.* Ignoring all of the examples of discriminatory race-norming listed in the Notice, it gave one counterexample, stating that Black men have higher rates of death from prostate cancer due to less early access to antigen testing. *See id.* The comments do not explain why race must be incorporated into any treatment algorithm to compensate for this disparity. *See id.*

[363] The Department's website lays out its civil rights enforcement methods, including negotiation of voluntary compliance. *See Section 1557 of the Patient Protection and Affordable Care Act*, U.S. Dᴇᴘ'ᴛ ᴏғ Hᴇᴀʟᴛʜ ᴀɴᴅ Hᴜᴍᴀɴ Sᴇʀᴠs., https://www.hhs.gov/civil-rights/for-individuals/section-1557/index.html#:~:text=the%20Office%20for%20Civil%20Rights,in%20covered%20health%20programs%20or [https://perma.cc/KSJ7-4AW4].

friendly).[364] The proposed rule also explains that Section 1557 provides a private right of action by incorporating by reference the enforcement mechanisms of several other civil rights statutes.[365]

In short, healthcare providers who exclude patients from care on the basis of race are typically acting illegally, and there's little reason to believe we should think otherwise when this exclusion is based on a racialized CPG. Presumably, the reason doctors and hospitals don't seem worried about this is because thus far, nobody has been bringing lawsuits. But this might change now that racial CPGs have gotten more public attention, not to mention the attention of the federal government—and it *should* change. I imagine that field experts might protest that the evidence for some racialized CPGs is stronger than I've suggested. It's possible they could be right. But if so, those relying on racialized CPGs should be prepared to present that evidence to defend them in court, rather than expecting that race-based decisions will evade legal scrutiny entirely. Our civil rights laws provide a potential remedy for individuals denied medical care on the basis of their race, and both those individuals and those who deny that care should know it.

V.   Why Is Statistical Discrimination (Sometimes) Tolerated?

Our law seems to powerfully reject statistical justifications for otherwise-unlawful discrimination—but in practice, a number of practices relying on such justifications have escaped legal scrutiny or been promoted by courts themselves. Why does this disconnect exist? Part of the answer must simply be that racist, sexist, and classist attitudes are deeply entrenched; many forms of discrimination that the law formally prohibits are persistent. But this isn't the whole answer. In each of the examples discussed above, it's hard to imagine the same arguments for race, sex, or class discrimination being accepted today if they were *not* embedded in statistical adjustments. Explicit classifications like these are the sort of discrimination that our law is designed to address effectively. Why hasn't it? I don't have a definitive answer, but I'll briefly suggest a few possibilities.

---

[364] Commentators have noted the trend of "ignoring *Chevron*" in recent cases involving DHHS interpretations of the Affordable Care Act specifically. Richard J. Pierce, Jr., *Is Chevron Deference Still Alive?*, Reg. Rev. (July 14, 2022), https://www.theregreview.org/2022/07/14/pierce-chevron-deference/ [https://perma.cc/AM78-RKWN]. *But see* Christopher J. Walker, *What American Hospital Association v. Becerra Means for the Future of Chevron Deference: Probably Not Much*, Yale J. Reg. Notice & Comment (June 15, 2002), https://www.yalejreg.com/nc/what-american-hospital-association-v-becerra-means-for-the-future-of-chevron-deference-probably-not-much/ [https://perma.cc/73HH-NVVU] (arguing that the Court's decision not to apply *Chevron* was consistent with *Chevron* itself).

[365] DHHS Notice of Rulemaking, *supra* note 7 (§ 92.301).

A.   Deference to Experts and Technical Knowledge

A large literature spanning numerous fields explores the human tendency to defer to "expert" knowledge.[366] In particular, Americans are not very quantitatively literate and are easily daunt by even quite basic statistical and scientific claims, which can insulate such claims from serious scrutiny. Elaine Sutherland writes:

Somewhat paradoxically, it is this very ignorance of science that often results in non-scientists being mesmerized by it. Science is perceived as solid, knowable, measurable: in short, science offers certainty. These factors combine to place the person who does understand science, the expert, in an incredibly powerful position.[367]

The tendency to defer to experts is not intrinsically problematic. It would be hard to live without it, given that we each personally possess only tiny fragments of humanity's vast store of collective knowledge.[368] But excessive deference to experts carries dangers. Obviously, experts can be wrong about facts. Experts hired for litigation may have incentives to spin the facts strategically.[369] In addition, many purportedly "objective" claims to empirical knowledge actually embed highly contestable normative judgments within them.[370] For example, the question whether basing earnings predictions on race and gender makes them more accurate is an empirical one; the question whether it's *right* to do so is a moral and legal one. Forensic economists have no particular expertise as to the latter question. And yet the fact that demographically specific estimates are "standard practice" in forensic economics has helped to largely immunize them from scrutiny.

Susan Stefan has similarly critiqued courts' deference, in due process cases, to the "professional judgment" of administrators and professionals.[371] She argues that courts have a romanticized view of certain types of professionals, especially those in health professions, as carrying *moral* authority in addition to technical expertise—and are thus loath to recognize situations where they commit moral errors.[372] This point might help to explain courts' acceptance of troubling testimony from neuropsychologists, for example— and also highlights the value of professional organizations weighing in

---

[366] *See, e.g.*, Mark Button & Kevin Mattson, *Deliberative Democracy in Practice: Challenges and Prospects for Civic Deliberation*, 31 POLITY 609, 620 (1999); *see* Starr, *supra* note 54, at 866 n.255.

[367] Elaine E. Sutherland, *Undue Deference to Experts Syndrome?*, 16 IND. INT'L & COMP. L. REV. 375, 381–82 (2006).

[368] Indeed, lately some have worried that an ongoing breakdown of trust in experts poses an epistemic crisis, imperiling collective action against threats such as climate change and COVID-19. *See generally* NAOMI ORESKES, WHY TRUST SCIENCE (2021).

[369] *See generally* Sutherland, *supra* note 367.

[370] *See* Sonja B. Starr & M. Marit Rehavi, *Mandatory Sentencing and Racial Disparity: Assessing the Role of Prosecutors and the Effects of Booker*, 123 YALE L.J. 2, 17–18 (2013).

[371] *See* Susan Stefan, *Leaving Civil Rights to the "Experts": From Deference to Abdication Under the Professional Judgment Standard,* 102 YALE L.J. 639, 643 (1992).

[372] *See id.* at 646–54.

against discriminatory practices, as happened in the Denkowski affair. Courts' deference to doctors could potentially pose a challenge for those challenging medical CPGs, especially because juries, too, tend to trust medical experts.[373]

Beyond trust in experts, a related phenomenon is trust in, or even fetishization of, technical tools and technical progress. Professor Jessica Eaglin has written forcefully about the "technological advancement narrative" in the criminal justice risk assessment movement.[374] She argues: "The introduction of sentencing technologies facilitated interpreting. . .inequities as natural. As such, sentencing technologies reified structural racism under the auspice of scientific objectivity."[375] That is, the technological narrative encourages us to see risk factors in instruments as objective predictors, rather than as products of social inequities that racial justice requires an answer to. Something similar could be said of the use of race and sex to predict earnings, and of the various diagnostic algorithms that take racial disparities in health indicators and use them to redefine the expected norms for different racial groups.

B.   Norms Within Economics and Other Empirical Disciplines

The flip side of the deference-to-experts problem is that scientific experts may view discrimination as lawyers' and policymakers' job to worry about. Many scholars in empirical disciplines conceive of their roles as being purely descriptive or predictive ("I'm just reporting what the data say"); they may deliberately avoid making openly normative claims. But as noted above, the design of empirical models often involves implicit normative judgments, especially when the model is used to shape real-world decisions. For example, there's nothing wrong with academic research assessing race and sex as predictors of earnings; such research is essential to understand disparities. But when such predictions shape damage awards, they don't merely *describe* disparities; they amplify them.

Parts I, III and IV have already explored the role of disciplinary norms (in, for example, neuropsychology, criminology, forensic economics, and medicine) in shaping many of the practices I have focused on here, and this discussion need not be repeated. But I want to point to one additional influence not yet discussed: the role in economics of the distinction between statistical and taste-based discrimination. This distinction's impact goes far beyond forensic economists.

Serious economic research on discrimination began in the 1950s, with Professor Gary Becker's seminal work.[376] Becker sought to understand why

---

[373] *See* Sutherland, *supra* note 367, at 382.

[374] Jessica M. Eaglin, *Technologically Distorted Conceptions of Punishment*, 97 Wash. Univ. L. Rev. 483, 486 (2019).

[375] *Id.* at 487.

[376] *See generally* Gary S. Becker, The Economics of Discrimination (1957).

discrimination occurs in labor markets and its expected effects on wages and employment.[377] His early model simply assumed that employers have a "taste for discrimination."[378] Other economists built on and critiqued this work; many argued that it could not explain discrimination's persistence in competitive markets. The theory of statistical discrimination emerged in the 1970s (initially in work by Kenneth Arrow and Edmund Phelps) as a way of explaining why discrimination does persist.[379] It argued that employers use race rationally as a low-cost proxy for characteristics they seek in workers but lack individualized information on. This "taste-based" versus "statistical" distinction has, ever since, been a dominant theme of both empirical and theoretical economic work on discrimination in many decision-making contexts.[380] Some more recent work has also pointed to categories between these two—stereotyping or "inaccurate statistical discrimination"—both of which involve group-based generalizations that are empirically ill-founded.[381]

In this literature, there is some ambiguity as to whether the statistical/taste-based distinction has *normative* content. Statistical discrimination is described as "rational" in the sense of serving the employer's interests; does that mean it is *good*, or at least less bad? Some economists resist that characterization,[382] and some research simply seeks to better understand statistical discrimination, which might help efforts to reduce it. For example, one implication of the theory is that restricting decision-makers' access to individualized information about other characteristics they care about might inadvertently encourage statistical discrimination.[383] And some empirical economic literature—including "auditing" studies that test discrimination

---

[377] *See id.* at 10–11.

[378] *Id.*

[379] *See* Kenneth Arrow, The Theory of Discrimination, *in* Discrimination and Labor Mkts., 3 (Orley Ashenfelter & Albert Rees eds., 2015); Edmund S. Phelps, *The Statistical Theory of Racism and Sexism*, 62 Am. Econ. Rev. 659 (1972); *see also* Dennis J Aigner & Glen G. Cain, *Statistical Theories of Discrimination in Labor Markets*, 30 Indus. & Lab. Rel. Rev. 175, 175–87 (1977).

[380] *See* András Tilcsik, *Statistical Discrimination and the Rationalization of Stereotypes*, 86 Am. Soc. Review 93, 94 (2021); Hanming Fang & Andrea Moro, *Theories of Statistical Discrimination and Affirmative Action: A Survey*, *in* 1A Handbook of Social Economics, 133, 134 (Jess Benhabib, Alberto Bisin & Matthew O. Jackson eds., 2011).

[381] J. Aislinn Bohren, Kareem Haggag, Alex Imas, & Devin G. Pope, *Inaccurate Statistical Discrimination* (Becker Friedman Institute, Working Paper No. 2019-86), https://bfi.uchicago.edu/wp-content/uploads/BFI_WP_201986-3.pdf [https://perma.cc/8AUM-7DGM]; Petro Bordalo, Katherine Coffman, Nicola Gennaioli, & Andrei Shleifer, *Stereotypes*, 131 Q. J. Econ. 1753 (2016).

[382] *See, e.g.*, Evan K. Rose, *A Constructivist Perspective on Empirical Discrimination Research* (Univ. Chi. Working Paper, 2022), https://ekrose.github.io/files/constructivism.pdf [https://perma.cc/G6BF-3JB3]).

[383] My own empirical research has documented this unintended impact of restrictions on criminal-record information. *See* Amanda Y. Agan & Sonja B. Starr, *Ban the Box, Criminal Records, and Racial Discrimination*, 133 Q.J. Econ. 191 (2018).

through field experiments—uses methods that do not differentiate between statistical and taste-based discrimination.[384]

But other economic literature does suggest that statistical discrimination is less normatively problematic than taste-based discrimination—that there's a morally relevant distinction between *disliking* and *believing negative things about* a group. Economists Marianne Bertrand and Esther Duflo have written:

> While taste-based discrimination is clearly inefficient . . . statistical discrimination is theoretically efficient and hence more easily defendable in ethical terms under the utilitarian argument. Moreover, statistical discrimination can also be argued to be "fair" in that it treats identical people with the same expected productivity. . .and is not motivated by animus. In fact, many economists would most likely support allowing statistical discrimination as a good policy, even where it is now illegal.[385]

Sociologist Andras Tilcsik cited this and many other examples in a recent study of the role of statistical discrimination theory in economics and its impact on people's normative assessments of discrimination.[386] Reviewing ten leading introductory economics textbooks, he found that all ten presented statistical discrimination as "rational," only four "include any critical commentary," and only one mentioned the possibility of inaccurate statistical generalizations.[387] He observed that economic research also portrays statistical discrimination as "pervasive, inescapable, and normal," and argues that the expectation that nearly everyone will "rely on group-level generalizations" effectively provides "a license and justification for doing so."[388] Tilcsik presented evidence from a survey experiment bearing this out; participants exposed to statistical-discrimination theory (especially without critical commentary) were more likely to endorse gender stereotypes and less likely, in a hypothetical situation, to "hire" women.[389]

An important example of how statistical discrimination has been normalized within economics—and law-and-economics specifically—is the prevalence of "outcome-test" (or "hit-rate") approaches to measuring discrimination. These methods dominate the empirical economic literature on discrimination in criminal justice, and are also prevalent in other substantive areas. In a seminal paper, John Knowles, Nicola Persico, and Petra Todd

---

[384] *See, e.g.*, Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 Am. Econ. Rev. 991 (2004).

[385] Marianne Bertrand & Esther Duflo, *Field Experiments on Discrimination*, (Nat'l Bureau of Econ. Rsch., Working Paper No. 22014), http://www.nber.org/papers/w22014 [https://perma.cc/ZTM2-C4SM].

[386] *See* Tilcsik, *supra* note 380, at 102.

[387] *Id.* at 102.

[388] *Id.* at 101.

[389] *Id.* at 107–13.

presented an equilibrium model of traffic stops which assumes that "unbiased" police *would* consider race when deciding who to search—specifically, that they would take into account race-specific "hit rates" of searches in uncovering contraband and adjust their race-specific stop criteria until (at equilibrium) those rates equalize.[390] The authors present this as a rational way to maximize hit rates.[391] The only discrimination estimated by this model is taste-based, which is detected if hit rates are *not* equal across groups.[392] If hit rates *are* equal across groups, the conclusion is that there is *no* discrimination—as Knowles, Persico, and Todd found in their study of Maryland traffic stops—even though, given the assumptions of the model, that finding actually means that the police *are* taking race into account.[393]

It's a model that utterly does not track how the law defines discrimination—and yet it's highly influential in the law-and-economics world. Similar papers have been published concerning bail and parole, for example.[394] Economists have sometimes explicitly proposed that courts *should* use outcome tests of discrimination.[395] And these tests have made their way into litigation, even though they do not answer any question the law cares about. In the *Floyd v. City of New York* stop-and-frisk litigation, for example, N.Y.P.D. attempted to defend itself by introducing evidence that its stops produced equal hit rates across races (a self-damning finding, although N.Y.P.D. didn't seem to realize it).[396] Outcome-test models are particularly technical and challenging for lay readers, and it's very likely that lawyers would not pick up on the fact that the models are defining discrimination much more narrowly than the law does. Meanwhile, if economists largely do not understand the legal irrelevance of the taste-based versus statistical distinction, it is not surprising that forensic economists, for example, would so routinely urge juries to issue statistically discriminatory damage awards.

C. *Professional Courtesy and Not Rocking the Boat*

In addition to the respect paid by lawyers to members of other professions, lawyers and courts also may be driven by norms of courtesy *within* the legal profession. Many critics have pointed to downsides of this courtesy—for example, the fact that prosecutors are virtually never sanctioned for mis-

---

[390] *See* John Knowles & Nicola Persico, *Racial Bias in Motor Vehicle Searches: Theory and Evidence*, 109 J. OF. POL. ECON. 203, 203–29 (2001).

[391] *See id.* at 210–11.

[392] *See id.* at 205.

[393] *See id.*

[394] *See* David Arnold, Will Dobbie, & Crystal S. Yang, *Racial Bias in Bail Decisions*, 133 Q. J. OF ECON. 1885, 1885–1932 (2018); Stéphane Mechoulan & Nicolas Sahuguet, *Assessing Racial Disparities in Parole Release*, 44 J. LEGAL STUD. 39, 39–74 (2015).

[395] Nicola Persico & David A. Castleman, *Detecting Bias: Using Statistical Evidence to Establish Intentional Discrimination in Racial Profiling Cases,* 2005 U. CHI. LEGAL F. 217, 233 (2005).

[396] Floyd v. City of New York, 813 F.Supp.2d 457, 462–63 (S.D.N.Y. 2011).

conduct.[397] Even where a court reverses a conviction based on prosecutorial misconduct, the court rarely names the prosecutor, apparently because doing so would be embarrassing.[398] Bar discipline is even rarer; nobody brings complaints.[399]

These norms may help to explain why, when a practice is longstanding within the justice system (race- and sex-based damage calculations), endorsed by prosecutors (ethnically adjusted *Atkins* assessments), or broadly embraced in a courthouse community (risk assessment), lawyers and courts may be reluctant to call out racial and other discrimination within it. It may explain why, even when courts *are* evidently uncomfortable with the equity implications of a practice, they so frequently find ways to reject it (e.g., invoking their own discretion) that do not involve calling it unconstitutional discrimination. This approach is more polite; it calls nobody out. It may achieve the same result in the case and is consistent with constitutional avoidance principles. But it's much less likely to highlight a practice's problems and discourage its future use.

So, for example, almost no judge (aside from Judge Weinstein) has held it unconstitutional to base damages on race and gender, even though the case for this holding is well supported by Supreme Court doctrine. Why not? Perhaps nobody wants to rock the boat. Nobody wants to say that something courts have done for decades is wrong, much less racist or sexist—an especially fraught accusation. Calculating damages this way is just *what we do*. Of course, history gives many examples of courts eventually becoming willing to strike down entrenched practices, from Jim Crow segregation to racially restrictive covenants to the exclusion of same-sex couples from marriage. But perhaps this was easier when it involved civil society or nonjudicial entities. It may take more courage and introspection for courts to identify similar problems in their own house.

These features of the legal profession aren't unique; communities of all sorts are characterized by a reluctance to resist widely or long-accepted practices. Social psychologists have a variety of interrelated explanations for this reluctance. One is "system justification": the powerful human tendency "to defend and justify the status quo and to bolster the legitimacy of the existing social order."[400] When the practice involves a group to which one belongs

---

[397] *See, e.g.*, Angela J. Davis, *The Legal Profession's Failure to Discipline Unethical Prosecutors*, 36 HOFSTRA L. REV. 275, 275–310 (2007); Sonja B. Starr, *Sentence Reduction as a Remedy for Prosecutorial Misconduct*, 97 GEO. L.J. 1509, 1513–18 (2009) (reviewing this literature).

[398] *See* Adam M. Gershowitz, *Prosecutorial Shaming: Naming Attorneys to Reduce Prosecutorial Misconduct*, 43 U.C. DAVIS L. REV. 1059, 1060 (2009).

[399] *See, e.g.*, Tracey L. Meares, *Rewards for Good Behavior: Influencing Prosecutorial Discretion and Conduct with Financial Incentives*, 64 FORDHAM L. REV. 852, 893–99 (1995); Peter A. Joy, *The Relationship Between Prosecutorial Misconduct and Wrongful Convictions: Shaping Remedies for a Broken System*, 2006 WIS. L. REV. 399, 399–401.

[400] John T. Jost, Mahzarin R. Banaji, & Brian A. Nosek, *A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo*, 25 POL. PSYCH. 881, 887 (2004).

(e.g., lawyers), or one's *own* choices, this is bolstered by "group justification" and/or "ego justification."[401] A related concept is the "just world theory": a human need to believe that people mostly get what they deserve—which, decades of research has shown, leads people to find ways to justify injustices and blame their victims.[402]

Humans, in short, are excellent rationalizers. That tendency may have some advantages for our day-to-day happiness, but it has also surely enabled all kinds of horrors, including hundreds of years of slavery and Jim Crow. Against that background, this paper's examples are small ones that shouldn't surprise anyone. It's worth noting that these psychological phenomena might help explain not just courts' and lawyers' tolerance of the practices discussed here, but also the nature of the mistake made in some of those practices: the normalization and reification of inequality by treating socially created disparities as though they are grounded in intrinsic group difference.

VI. Possible Equity-Promoting Uses of Statistical Generalizations

This Article so far has focused largely on practices that I consider to be easy cases for the prohibition on statistical discrimination. In these cases, antisubordination and anticlassification arguments cut in the same direction, because negative statistical generalizations about disadvantaged groups are being used to justify adverse treatment of them. But harder cases exist, in which classifications and statistical adjustments that are mechanically similar to those discussed above are proposed to counter the disparate impacts of facially neutral approaches. We've already discussed the General Aptitude Test Battery example, in which race-norming was (until Congress banned it) used to adjust employment test scores. Consider a few other examples:

- In the medical context, some current critics of race-norming propose as an alternative not race neutrality, but an approach that actively accounts for and seeks to mitigate racial disparities that affect health.[403]
- Since the 1990s, government agencies have routinely incorporated into environmental assessments (for example, those required by the National Environmental Policy Act) consideration of the impacts of their decisions on communities of color.[404] These race-conscious quantitative analyses are motivated by the idea that race predicts vul-

---

[401] *Id.*

[402] *See generally* Melvin J. Lerner & Leo Montada, Responses to Victimizations and Beliefs in a Just World (1998).

[403] *See, e.g.*, Cerdeña et al., *supra* note 279, at 1125–28; Joyce Frieden, *Can Medicine Be 'Race-Conscious' Without Being Racist?*, MedPage Today (Sep. 28, 2021), https://www.medpagetoday.com/practicemanagement/informationtechnology/94754 [https://perma.cc/HE5P-VD89].

[404] *See* Exec. Order No. 12898, 3 C.F.R. § 651.17 (1994); *Promising Practices for EJ Methodologies in NEPA Reviews*, Fed. Interagency Working Grp. on Env't Just. & NEPA

nerability to accumulated environmental hazards, and that "neutral" metrics of health risk understate the dangers that exposures can pose to more vulnerable groups.

• In research on criminal justice risk assessments, an oft-suggested remedy for racially disparate impacts or disparate misprediction rates is race-conscious design of the algorithms to mitigate those flaws.[405] Specific proposals vary, but the usual idea is to use race data not in the final algorithm but in development, to test which variables and design choices exacerbate or minimize racial disparities.

The details and merits of these various approaches are beyond this Article's scope. But let's stipulate that there are surely contexts in which inclusion of race (or other heightened-scrutiny classifications) in a statistical tool could reduce problematic disparities. If so, should we worry that a strong doctrinal prohibition on statistical discrimination could interfere with equity-promoting applications? The modern movement against systemic racism critiques color-blindness and demands a conscious focus on antiracism and substantive justice.[406] Should those sympathetic with those demands view race-norming and other statistical techniques as tools to promote those ends?

I'm going to start with how I think such applications *should* be treated doctrinally and then explain how courts *would* likely treat them. As discussed in Part II.B, although I've grounded my principal arguments in current doctrine for practical reasons, my preference would be to conceptualize constitutional equality in substantive terms, focusing on the effects of practices on vectors of social stratification. If that perspective were adopted, none of the examples above are analogous to statistical discrimination against disadvantaged groups. I agree with many, including some Supreme Court justices, who have argued that it's inappropriate to apply strict scrutiny to interventions meant to close racial gaps in society.[407] We don't apply strict scrutiny to every classification, and the reasons reliance on race is "suspect" relate to the shameful legacy of white supremacy, not to efforts to dismantle it. Even if strict scrutiny applies, in my view, mitigation of disparities grounded in that legacy should be considered a compelling governmental interest, and well-tailored race-conscious efforts should be recognized as sometimes necessary to achieve that interest.

But U.S. courts have not embraced this view. Courts have consistently applied strict scrutiny to so-called "reverse discrimination" cases. And the Supreme Court has rejected the idea of remediating societal discrimination

---

Comm., (Mar. 2016), https://www.epa.gov/sites/default/files/2016-08/documents/nepa_promising_practices_document_2016.pdf [https://perma.cc/355M-EBRE].

[405] *See supra* section III.C.

[406] *See, e.g.*, Ashley ("Woody") Doane, *Beyond Color-blindness: (Re)Theorizing Racial Ideology*, 60 Socio. Persp. 975, 975–91 (2017); Neil Gotanda, *A Critique of "Our Constitution is Color-Blind,"* 44 Stan. L. Rev. 1, 1–68 (1991).

[407] *See* Adarand Constructors, Inc. v. Peña, 515 U.S. 200, 243 (1995) (Stevens, J., dissenting) ("There is no moral or constitutional equivalence between a policy that is designed to perpetuate a caste system and one that seeks to eradicate racial subordination.").

as a compelling state interest, allowing only more focused remedies for iden-tifiable state discrimination. As of this writing, it *has* treated educational diversity as a compelling interest and has allowed use of race as a "plus factor" to achieve that interest[408]—but it's expected to change direction on this very soon, in a case pending as of this writing.[409] Beyond education, the Court has increasingly resisted government efforts to respond to racial ine-quality; its voting rights jurisprudence is a high-stakes example.[410]

Still, this situation doesn't mean there is *no* room for race-conscious statistical approaches. Such a bar is not implied by the case law on statistical discrimination,[411] and more generally, courts have not yet insisted on a wholly "color-blind" America, even though the Supreme Court is likely to move somewhat in that direction. Affirmative action and other race-con-scious policies are, as of this writing, common in employment and other contexts, not just in education.[412] The Department of Labor requires such steps (at least "training programs, outreach efforts," and the like) for federal contractors,[413] and its regulations enforcing Title VII encourage affirmative action.[414] Race-conscious environmental assessments have been routine for decades.[415] There are countless such examples, and most don't occasion law-suits. The legal landscape will likely soon shift, beginning with this sum-mer's expected bar on affirmative action in higher education, and some policies are already changing in anticipation of it; for example, the Biden Administration has shifted to race-blind criteria for environmental cleanup

---

[408] *See Grutter*, 539 U.S. at 328, 334.

[409] *See, e.g.*, Joan Biskupic, *Supreme Court Conservatives May Have Their Chance to End Affirmative Action at Universities*, CNN (Dec. 9, 2021), https://www.cnn.com/2021/12/09/politics/affirmative-action-supreme-court-conservatives-harvard/index.html [https://perma.cc/C9SS-XLAF].

[410] *See* Carrie Johnson, *U.S. Judges are Narrowing Voting Protections. Some Fear Lasting Damage*, NPR (Feb. 25, 2022), https://www.npr.org/2022/02/25/1082626791/u-s-judges-are-narrowing-voting-protections-some-fear-lasting-damage [https://perma.cc/PCK7-R8MZ]; *see also* Shelby Cnty. v. Holder, 570 U.S. 529 (2013) (finding unconstitutional Section 4(b) of the Voting Rights Act).

[411] For example, in *Virginia,* the Court approvingly cited several cases that approved "sex classifications" that "promote equal employment opportunity," differentiating these from clas-sifications that "perpetuate the legal, social, and economic inferiority of women." 518 U.S. at 533–34 (citing Califano v. Webster, 430 U.S. 313, 320 (1977) (per curiam); Cal. Fed. Sav. & Loan Ass'n v. Guerra, 479 U.S. 272, 289 (1987).

[412] The NFL, notably, requires teams to interview minority head-coaching candidates and uses draft picks to incentivize minority staff development. *See The Rooney Rule*, NFL, https://operations.nfl.com/inside-football-ops/diversity-inclusion/the-rooney-rule/ [https://perma.cc/NVE2-39XT].

[413] *See* UNITED STATES DEP'T OF LABOR, *Affirmative Action*, https://www.dol.gov/general/topic/hiring/affirmativeact [https://perma.cc/9FW9-Y8CG].

[414] *See* 29 C.F.R. § 1608 (2012).

[415] *See Environmental Equity: Reducing Risk for All Communities*, 2 U.S. ENV'T PROT. AGENCY (June 1992), https://www.epa.gov/sites/default/files/2015-02/documents/reduc-ing_risk_com_vol2.pdf [https://perma.cc/GYG7-ZV59].

fund administration in order to insulate against legal challenge.[416] Still, we shouldn't expect governmental or private use of race-conscious equity measures to disappear overnight, even assuming the use of affirmative action in higher education is struck down. And given the prevalence of higher-profile race-conscious policies, algorithmic approaches to racial equity may not be the likeliest immediate targets for legal attack.

Second, when classifications not subject to heightened scrutiny are involved, there's likely always going to be more leeway—notably, including socioeconomic factors. Even in the criminal justice context, the *Griffin* line of cases concerns protection of indigent defendants. The concern is unidirectional; no case in that line evinces any worry about "reverse discrimination" against wealthy defendants, and poverty-related factors have long been raised in mitigation. Many racial disparities are mediated by socioeconomic factors—not just poverty itself, but other potentially measurable predictors (e.g., local air quality or access to medical care). Algorithms that seek to redress the effects of those factors directly will tend to mitigate racial disparities, and can do so without much risk of legal challenge.

Third, even if courts were to flatly bar the use, in decisions directly affecting individual treatment, of metrics incorporating racial classifications, this would not itself imply a bar on the *consideration of racially disparate impacts* (or disparities in predictive accuracy) when *designing* algorithms. And this is what many of the proposals to use algorithms as antidiscrimination tools actually entail. That is, the designers do not put race in the ultimate algorithm that decision-makers will use; they *do* consider the racial impacts of other choices in algorithm design, and might alter the algorithm to minimize adverse impacts.[417]

If courts were to strike this kind of approach down, it would reflect a shift in the law far more profound than the elimination of affirmative action. It is true, of course, that facially neutral actions with a discriminatory purpose can be subject to strict scrutiny, and one could imagine a court wholeheartedly committed to "colorblindness" characterizing this kind of algorithmic engineering in that way. But it is routine, and sometimes legally required, for government and private actors to consider the disparate impacts of their policy choices, and to notice and react to racial disparities in society. For the most part, courts have not yet suggested that this consideration evinces an impermissible discriminatory purpose that would invalidate even a facially race-neutral policy choice.

In a forthcoming piece, I argue in detail that even if the Supreme Court completely bars affirmative action, this holding would not imply a complete bar on all policy with race-conscious ends, or even that such policies should

---

[416] *See* Lisa Friedman, *White House Takes Aim at Environmental Racism, but Won't Mention Race*, N.Y. Times (Feb. 15, 2022), https://www.nytimes.com/2022/02/15/climate/biden-environment-race-pollution.html [https://perma.cc/V2RY-NSEC].

[417] *See, e.g.*, Ludwig & Mullainathan, *supra* note 246, at 89–91 (describing "equity knobs" and equity "calibration tests" in algorithm design).

be routinely subjected to strict scrutiny. [418] The Supreme Court traditionally applies strict scrutiny to all policies that classify individuals based on race (even if their purposes are "benign," like affirmative action), and also to policies with *invidious* race-related purposes, even if they are facially neutral.[419] But it has never applied strict scrutiny to facially-neutral policies that have benign racial purposes: those that seek to reduce disparities or promote diversity and integration, rather than doing the opposite.[420] In my article, I highlight a new wave of litigation that seeks to get courts to do so, and explain why this extension of colorblindness doctrine would be quite radical. To treat goals like reduction of racial disparity as constitutionally suspect would, I argue, be inconsistent with longstanding doctrine, with the normative justifications underlying the "colorblindness" approach as applied to racial classifications like affirmative action, and with the best evidence of the Fourteenth Amendment's original meaning.[421]

Some of the Supreme Court's past cases have not only permitted but affirmatively encouraged the use of race-neutral means to promote diversity; in past affirmative-action case law, these have been identified as less restrictive alternatives to the explicit use of race. For example, in the course of Abigail Fisher's lengthy litigation, the Supreme Court twice fractured over the University of Texas's consideration of race when evaluating applicants who did not qualify under the Texas Ten Percent Plan (TPP), which admitted the top ten percent of every public high school class.[422] But no justice ever called into doubt the permissibility of the TPP itself, which was openly adopted as an alternative racial-diversity strategy after the Fifth Circuit struck down UT's prior affirmative action program

Government and private decisionmakers also often are *required* by antidiscrimination statutes to consider the disparate racial impacts of their decisions and adjust their policies to avoid inflicting such impacts unnecessarily. The Supreme Court case that has gone the farthest in casting race-conscious government policymaking of this sort into doubt is *Ricci v. DeStefano*, which held that a fire department could not discard the results of a promotion test on the basis that only white officers had passed it.[423] The Court decided *Ricci* on Title VII disparate-treatment grounds (avoiding the plaintiffs' constitutional claim), and left open the possibility that stronger evidence of disparate impact unjustified by business necessity could permit a government employer to take a similar step. Still, at the time, *Ricci* occa-

---

[418] *See* Sonja Starr, *The Magnet-School Wars and the Future of Colorblindness*, 76 STAN. L. REV. (forthcoming 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354321 [https://perma.cc/K2DN-HE3V].

[419] *See id.*

[420] *See id.*

[421] *See id.*

[422] *See* Fisher v. Univ. of Tex. at Austin, 570 U.S. 297, 297 (2013) ("*Fisher I*"); Fisher v. Univ. of Tex. at Austin, 579 U.S. 365, 365 (2016) ("*Fisher II*").

[423] 557 U.S. 557 (2009).

sioned much worry that the Court had Title VII disparate impact litigation in its crosshairs.[424]

In 2015, however, in *Texas Department of Housing and Community Affairs v. Inclusive Communities Project*, [425] the Supreme Court appeared to reject this possible extension of *Ricci*. Indeed, the Court in *Inclusive Communities* extended disparate impact analysis to the Fair Housing Act notwithstanding constitutional avoidance arguments, and in doing so squarely embraced the permissibility of government actors' pursuit of racial equality and integration as objectives, so long as they use race-neutral means.[426] As I argue in my forthcoming article, a move away from this principle would be profoundly destabilizing to a wide range of laws and practices.[427]

For now, at least, there is no constitutional doctrine barring policymakers from trying to avoid racially disparate impacts, and so the design of algorithms to avoid such impacts certainly remains on the right side of current law. Professor Deborah Hellman has likewise written in favor of race-conscious algorithm design in criminal justice, defending its constitutionality.[428] She argues that *Ricci* is unusual because it involved the retrospective abandonment of a procedure on which "specific, identifiable" people had relied, and that if it were read to bar facially neutral actions simply because they result from race-conscious decision-making processes, policies "in all sorts of areas would be constitutionally in jeopardy." [429] Other scholars have offered similar takes on *Ricci*'s limits.[430] Time will tell, and we may still see a more aggressive judicial approach in the future; as my other work explores, there is a current wave of litigation challenging race-neutral school admissions policies on the basis that they were designed with diversity in mind, and at least one district court has agreed.[431] But I am hopeful that the Supreme Court will not extend colorblindness doctrine so far.

---

[424] Richard Primus, *The Future of Disparate Impact*, 108 Mich. L. Rev. 1341–   87 (raising these concerns, but arguing that *Ricci* should be read more narrowly).

[425] 576 U.S. 519 (2015).

[426] *See id.* at 530; *see also* Starr, *supra* note 418, Part II.

[427] *See* Starr, *supra* note 418, Part IV.D.

[428] *See* Deborah Hellman, *supra* note 240, at 846–64.

[429] *Id.* at 864.

[430] *See, e.g.*, Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. Pa. L. Rev. Online 189, 189 (2017); Richard A. Primus, Title VII of the Civil Rights Act After 50 Years: Proceedings of the New York University 67th Annual Conference on Labor at 295–318 (2015); *see also* Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, & Harlan Yu, *Accountable Algorithms*, 165 U. Pa. L. Rev. 633, 692–95 (2017) (arguing that *Ricci* may bar race-conscious amendments to existing algorithms but does not bar "designing for nondiscrimination").

[431] A district court recently struck down a race-blind magnet-school admissions policy because its adopters wanted to increase Black and Hispanic representation. *See* Coal. for TJ v. Fairfax Cnty. Schl. Bd., No. 1:21CV296 at *1–*11 (E.D. Va. Feb. 25, 2022). But the Fourth Circuit stayed that decision, implying likely reversal. *See* Coal. for TJ v. Fairfax Cnty. Schl. Bd., No. 22-1280 at *1 (4th Cir. Mar. 31, 2022). The Supreme Court let this stay stand. *See* Coal. for TJ v. Fairfax Cnty. Schl. Bd., 142 S. Ct. 2672 (2022). *See also* Starr, *supra* note 418, for a detailed exploration of this litigation and other related cases.

Fourth, I don't think it's worth worrying that applying the prohibition of statistical discrimination to the practices this Article critiques will meaningfully alter the likelihood that courts will enforce it to interfere with algorithmic-fairness efforts. As Part II makes clear, extremely strong doctrinal support for that prohibition already exists. Moreover, nothing I've said in this section is specific to the problem of *statistical* discrimination per se. The scope of the prohibition of statistical discrimination is shaped by what kinds of discrimination are considered unconstitutional in the first place. The extent to which the Supreme Court will continue down the "colorblindness" path in defining unconstitutional discrimination remains unknown, but it probably won't be meaningfully shaped by the state of its statistical discrimination doctrine.

And finally, even if there *is* a risk that encouraging courts to take a harder look at statistical discrimination might impede some pro-equity efforts, it might be a risk worth taking. While there are valid counterexamples, the history of how race-norming and similar statistical adjustments have been used shouldn't inspire much confidence that algorithms *will* predominantly be used to promote equity, even if they *could* be. Achieving that outcome will likely not just require development of technical capacity, but something harder: social, political, and moral commitments to that aim. It will require pro-equity uses of algorithms not only to achieve buy-in from key decision-makers, but to avoid the type of political backlash that the race-normed GATB faced. In the meanwhile, between now and that uncertain future, we still have numerous statistical practices that discriminate against disadvantaged groups and that readily available legal strategies can help us to fix.

CONCLUSION

I began this Article with the NFL story on the premise that what happened there illustrates a broader phenomenon in law and healthcare. But in one respect the story is unusual: the practice got reversed relatively quickly. Why? Perhaps it's because NFL players have a more powerful voice than most medical patients, civil plaintiffs, and criminal defendants do. The NFL was already under public pressure on both race issues and concussions, and once players' families brought it to light, the story was a public relations nightmare. Once the League's incentives shifted, the court was able to encourage it to abandon race-norming without even weighing in on its legality, just by ordering mediation. This story doesn't provide much reason for optimism about ending discriminatory practices that are much more entrenched and harm people without the leverage of star athletes.

But the NFL scandal, in combination with the so-called "national racial reckoning" with which it coincided, has nonetheless helped to catalyze conversation about some of those practices—in particular, the use of racialized algorithms in health care. Lawyers and courts should be part of that conver-

sation—participants in it, and subjects of it. Law has much to say about the role of race in medicine. And the legal system itself has embraced numerous statistically discriminatory practices, despite the fact that they run sharply afoul of binding doctrine that seems to apply squarely to them.

U.S. equal protection and statutory antidiscrimination law has many weaknesses, but it *does* provide forceful tools for challenging statistical discrimination, and those tools should be used. The stakes are high. The practices examined here make people's lives and health depend on their race (in the death penalty and medical-practice contexts), their financial well-being and the valuation given to their life and health depend on their race and sex (civil damages), and their liberty depend on how much money they have (criminal justice risk assessment). They do this in the most straightforward disparate-treatment sense that the law is well equipped to address.

Embedding these discriminatory choices in an actuarial prediction, a normed test score, or another algorithmic adjustment may obscure what's happening, giving it a veneer of scientific objectivity. But one does not have to dig deeply to see what's beneath. At the very least, those who support these practices should have to defend their legality, rather than counting on courts and lawyers to nod along or look the other way.