Context-Dependent Discrimination: Evidence from Emergency Triage Decisions

Ashley Litwin*

September 11, 2026

Abstract

Group disparities are often treated as stable features of decision-making, reflecting persistent differences in beliefs or preferences. I show instead that they depend on context: extraneous features that happen to vary from one decision to the next. Using 2 million emergency department (ED) triage decisions, I document that non-White patients are 5.4 percentage points (21%) less likely than observably identical White patients to receive top triage priority within nurse-shift. At least 50% of this gap is explained by nurses' local decision context: prior patient exposure, time pressure, and their interaction. I exploit encounters with frequent ED users, a recognizable, lower-risk group that is disproportionately non-White. Among non-frequent users, the racial triage gap widens by 2.3 pp following a non-White frequent user but narrows by 1.2 pp following a White one. Longer-run exposure produces the same pattern, with the gap reversing at high White-exposure levels. Quasi-random increases in time pressure amplify exposure effects. The same effects emerge relative to a race-blind machine-learning benchmark of clinical risk. These results point to similarity-based assessment: exposure shapes which traits become associated with low risk, while time pressure increases reliance on those associations. Because frequent users are disproportionately non-White, this favors White patients on average. More broadly, decision context shapes both the magnitude and direction of disparities, making organizational choices over case assignment, workload, and the sequencing of tasks potential levers for reducing unequal treatment.

*University of Pennsylvania, alitw@upenn.edu. For their support and guidance, I thank Judd Kessler, Corinne Low, and Alex Rees-Jones. For helpful comments and suggestions, I thank Diane Alexander, Nina Buchmann, Alex Chan, John Conlon, Christine Exley, Marius Guenzel, Atul Gupta, Ben Handel, Alex Imas, Sasha Indarte, Supreet Kaur, Ryan Oprea, Pietro Ortoleva, Heather Schofield, Manasvini Singh, Colin Sullivan, and Kevin Volpp, as well as seminar and conference participants at SITE Economics of Health Inequality, the University of Pittsburgh, Princeton (SWEET), the University of Pennsylvania, and the ESA World Meeting. Stephanie Dodd provided excellent research assistance. I gratefully acknowledge funding from the Wharton Behavioral Lab, the Wharton Coalition for Equity and Opportunity, and the Wharton Environmental, Social, and Governance Initiative. This project received approval from the University of Pennsylvania IRB.

1 Introduction

Group disparities are a persistent feature of many high-stakes decisions. Women continue to earn less than men even within narrowly defined occupations, Black defendants receive longer sentences than observably similar White defendants, and minority borrowers are less likely to be approved for credit than otherwise comparable applicants.[1] Given their persistence, these gaps are often attributed to slow-moving determinants, such as entrenched beliefs, discriminatory preferences, or group differences in underlying characteristics, all of which imply limited scope for near-term change. Yet many consequential decisions are made repeatedly by the same decision-maker as the local decision context changes. In principle, features of that context — who was evaluated just before, how many decisions are pending — are irrelevant to the case at hand. In practice, they may influence how decision-makers process information and evaluate cases. If so, how decisions are sequenced, time is allocated, and caseloads are structured become potential policy levers. But to what extent do observed disparities reflect the context in which decisions are made, rather than the slow-moving determinants to which they are often attributed?

In this paper, I show that decision context explains a substantial share of observed group disparities. Holding fixed the decision-maker and the decision problem, I ask how disparities vary with two features of decision context: the cases previously encountered and the time available to decide. I study this question in a setting that makes both features tractable and consequential — emergency department (ED) triage — using detailed, timestamped administrative records on approximately 2 million unique triage decisions by 1,600 nurses across six hospitals from 2017 to 2025. Because ED arrivals are stochastic and patients are quasi-randomly assigned to the next available nurse, the same nurse faces plausibly exogenous variation in both prior case exposure and available decision time. I exploit this within-nurse variation to isolate the effect of decision context on disparities while holding fixed rich observable characteristics of each case.

I first document a large baseline racial gap in emergency triage decisions. Among patients with the same chief complaint evaluated by the same nurse during the same shift, and conditional on detailed clinical information recorded in the electronic medical record,[2]

[1] See, for example, Goldin (2014) and Blau (2025) for the gender earnings gap; Rehavi and Starr (2014) and Cohen and Yang (2019) for the racial sentencing gap; and Butler, Mayer and Weston (2023) and Hurtado and Sakong (2024) for the minority credit gap. For recent overviews of the discrimination literature, see Bertrand and Duflo (2017) and Litwin and Low (2026).

[2] Clinical covariates capture the universe of information recorded in the electronic health record prior to and during triage, including age, sex, visit history, chronic conditions, primary care history, arrival mode, insurance type, and chief complaint, as well as detailed vital signs, pain scores, free-text triage notes, and screeners such as infectious disease, coma, and fall risk.

non-White patients are 5.4 percentage points (pp), or 21 percent, less likely to receive the highest triage priority. The gap is not explained by correlates of race such as primary care access or insurance coverage, nor by proxies for differences in communication or self-advocacy, suggesting that access or patient presentation alone cannot account for the disparity. Similar patterns appear in nationally representative survey data, indicating that the disparity is not unique to the hospitals I study.[3]

I then show that nurses' local — and time-varying — decision context explains at least 50.6% of this large baseline disparity. The analysis is motivated by a simple idea: because emergency triage is fast-paced and requires nurses to make decisions in a matter of minutes,[4] they may assess new patients partly by their resemblance to previously encountered cases rather than evaluating each patient from scratch. I study this idea by exploiting nurses' encounters with a particularly salient and recognizable subset of the patient population: "frequent flyers," patients who repeatedly visit the ED, are typically lower risk, and are disproportionately non-White, yet are similar to the general ED population along other demographics including gender and age.

Exploiting the quasi-random sequencing of patients within nurse shifts, I find that frequent-flyer encounters substantially shift subsequent racial disparities in triage priority, with the direction of the effect depending on the frequent flyer's race. Following a non-White frequent flyer, the White–non-White gap in top triage priority increases immediately by 2.3 pp (36%) among subsequent non-frequent-flyer patients. Following a White frequent flyer, the pattern reverses: the racial gap falls by 1.2 pp (25%). The effect is largest for the next patient, remains relatively stable for at least five subsequent patients, and then begins to decay. Consistent with random sequencing, I find no evidence of pre-trends in the racial triage gap prior to frequent-flyer encounters of either race.

These short-run responses are consistent with a recent frequent-flyer encounter making race-specific associations with low risk more salient when the nurse evaluates subsequent patients. But such associations must first be formed through experience. I therefore next ask whether nurses' accumulated exposure to non-White and White frequent flyers shapes racial disparities over longer horizons. To this end, I measure each nurse's exposure to White and non-White frequent flyers over the preceding year. The analysis exploits changes in exposure within the same nurse over time, controlling flexibly for patient characteristics, hospital-specific time effects, and overall nurse experience, so the results do not simply reflect

[3]The nationally representative data provide external validity but contain substantially less detailed clinical information and do not allow for the high-frequency, within-provider variation in prior exposure and time pressure exploited in the administrative data. They are therefore used to assess generalizability rather than to identify causal effects.

[4]The median triage time is 3 minutes.

persistent differences across nurses who encounter different patient populations. Indeed, greater accumulated exposure to non-White frequent flyers systematically widens the racial gap among non-frequent-flyer patients: moving from the lowest to highest exposure quartile increases the White–non-White gap by 8.9 pp (231%). Accumulated exposure to White frequent flyers produces the mirror pattern, narrowing the gap by 7.4 pp (193%). At sufficiently high levels of White exposure, the racial gap reverses entirely in favor of non-White patients.

Together, the short- and longer-run exposure results point to a form of similarity-based assessment in which prior cases influence how subsequent patients are evaluated. Several mechanisms could generate similarity-based assessment in this setting, including categorical thinking (Mullainathan, 2002; Fryer and Jackson, 2008; Mullainathan, Schwartzstein and Shleifer, 2008; Bordalo et al., 2026) and selective recall (Bordalo, Gennaioli and Shleifer, 2020; Bordalo et al., 2023).[5] Rather than distinguish among these microfoundations, I develop a simple framework that nests their common prediction: experience with similar prior cases can shape the assessment of a new one. I then extend the framework to incorporate time pressure. When time is scarce, decision-makers may collect or process fewer features of the case at hand, making assessment less individualized and increasing reliance on associations formed through prior exposure. This extension yields a new testable prediction: time pressure should amplify the effect of prior exposure on racial disparities.

I test this prediction using two sources of quasi-random variation in time pressure: the number of cases waiting to be evaluated behind the current case, and the wait time of the next case conditional on queue length. Empirically, neither shock is associated with economically meaningful differences in observable pre-triage patient characteristics, supporting their interpretation as quasi-random variation in the time available for each decision. Both shocks also compress the time nurses spend on each evaluation, validating them as measures of acute time pressure.

I first show that greater time pressure widens the racial gap on average: moving from low to high pressure increases the White–non-White gap by 2–4 pp (30–73% of the baseline gap) across the two shocks. Under time pressure, nurses also collect less diagnostic information and write shorter clinical notes, consistent with less individualized assessment based on fewer case-specific features. Importantly, the framework does not interpret this average effect as an independent tendency for time pressure to disadvantage non-White patients. Rather, because nurses are disproportionately exposed to non-White frequent flyers on average, greater reliance on associations formed through prior exposure under time pressure should widen the racial gap.

[5]Appendix C shows formally how the categorization model of Bordalo et al. (2026) provides one microfoundation for the simplified reduced-form framework used here.

The interaction between time pressure and prior exposure provides a direct test of this mechanism. If time pressure increases reliance on associations formed through prior exposure, its effect on racial disparities should depend on the extent to which nurses have been exposed to non-White or White frequent flyers. I find exactly this pattern. At the highest level of non-White frequent flyer exposure, moving from low to high time pressure increases the racial gap by 10.3 pp, compared with 4.4 pp at the lowest level. The pattern reverses for White frequent flyer exposure: at the highest level of White exposure, moving from low to high time pressure reduces the racial gap by 1.7 pp and ultimately reverses it in favor of non-White patients. Thus, prior exposure determines the direction in which time pressure shifts disparities, while time pressure determines how strongly prior exposure shapes assessment. That nurses rely less on prior exposure when they have more time to assess the current patient suggests that its greater influence under time pressure is generated by the constraint, rather than simply reflecting information that is generally useful for assessment.

I next consider statistical learning as an alternative interpretation of the results. Generally speaking, dependence on prior cases need not reflect similarity-based assessment: in sequential decision settings, past observations may provide information about the distribution of underlying risk. In my setting, such statistical learning would imply updating race-specific beliefs about clinical risk as nurses encounter new patients, including frequent flyers (Phelps, 1972; Aigner and Cain, 1977; Bohren et al., 2025).[6] This explanation, however, places restrictions on how responses to new information should vary with prior exposure. As nurses accumulate exposure and their beliefs become more precise, the marginal effect of an additional frequent-flyer encounter should diminish. I find no such attenuation: recent frequent-flyer encounters shift racial disparities by similar amounts at low and high levels of prior exposure, even as accumulated exposure itself shifts the level of the racial gap. This pattern is difficult to reconcile with standard Bayesian updating, but is consistent with similarity-based assessment in which accumulated exposure shapes how nurses evaluate new cases without diminishing the influence of recent encounters.

The preceding results exploit high-frequency, quasi-random variation in recent frequent-flyer encounters and acute time pressure, alongside within-nurse variation in longer-run exposure. Under the identifying assumptions, these changes in context are orthogonal to the underlying clinical risk of the current patient, so the resulting shifts in racial disparities

[6]Similarity-based assessment and statistical discrimination differ in why prior experience matters. Under the former, prior cases shape which features of past experience are used to assess the current case; under the latter, prior cases provide information used to form beliefs about group-level risk. Both can nevertheless generate the reduced-form relationship captured by the framework: greater experience with lower-risk patients of a given race can reduce the relative assessment of subsequent patients of that race. Section 5.6 uses additional predictions to assess whether standard belief updating can account for the exposure effects.

reflect the effect of context rather than changes in patient risk. I next ask whether the same context-induced disparities emerge relative to a patient-specific machine-learning benchmark of clinical risk informed by patients' eventual diagnoses, which I also observe in my data. I construct this benchmark by training a random forest to predict diagnosis severity from information available before triage, excluding race. Because final diagnosis is a downstream outcome that may itself be affected by the initial triage decision, I train the model only on visits occurring under low prior nurse exposure and low time pressure, where the context effects documented above are smallest. I then use predicted risk to construct a benchmark triage classification and define the "triage wedge" as the difference between actual and predicted top-triage assignment.

The resulting triage wedge responds to context in the same direction as the racial gap in actual triage decisions: greater non-White frequent-flyer exposure increases it, greater White exposure reduces it, and greater time pressure widens it. Thus, the same contextual forces that shift racial disparities in triage also generate racial disparities in deviations from predicted clinical risk.

I conclude with reduced-form counterfactual exercises that quantify the scope for reducing disparities by changing the decision context. Setting longer-run frequent-flyer exposure to its lowest quartile—approximating the assignment of frequent flyers to dedicated providers—reduces the racial triage gap by 33%. Setting time pressure to its lowest observed level—approximating reduced queue congestion or increased staffing—reduces the gap by 13%. Implementing both jointly reduces the baseline gap by 50%.[7] These exercises illustrate that a substantial share of seemingly stable disparities reflects modifiable features of the decision context, making the conditions under which decisions are made—not decision-makers' beliefs or preferences alone—a first-order policy lever.

More broadly, the results show that disparities can arise not only from stable differences in decision-makers' beliefs or preferences, but from the experiences generated by the decision environment itself. Population composition shapes the cases decision-makers encounter, those experiences shape how subsequent cases are assessed, and decision constraints determine how strongly prior experience influences current decisions. The same decision-maker can therefore generate substantially different disparities—and even disadvantage different groups—as prior exposure and the conditions surrounding the current decision change. This insight extends to many other settings characterized by repeated, sequential decision-making, ranging from hiring and lending to courts and policing, where prior case exposure and the constraints

[7]The individual counterfactuals change one feature of context at a time, holding the other at its observed level. Because the estimated interaction between exposure and time pressure is therefore embedded in each individual counterfactual, the difference between the joint effect and the sum of the individual effects does not measure the importance of the interaction.

surrounding current decisions may similarly shape both the magnitude and direction of group disparities.

This paper makes four main contributions to the literature. First, it contributes to the health economics literature on clinical decision-making and disparities in care. Prior work shows that clinical decisions respond to fatigue, experience, peer behavior, decision time, and case complexity (Jin et al., 2020; Freedman et al., 2021; Jin et al., 2023; Silver, 2021; Singh, 2021; Handel et al., 2026). A related literature documents unequal treatment across patient groups and examines how disparities depend on clinician identity and patient–clinician concordance (Alsan, Garrick and Graziani, 2019), bias embedded in clinical decision tools (Alsan et al., 2026), and organizational capacity strain (Singh and Venkataramani, 2026).[8] The present paper bridges these literatures by showing that features of the clinical decision environment not only affect decisions on average, but also systematically shape disparities across patient groups. A clinician's prior case mix shapes the associations brought to subsequent cases, while the constraints surrounding the current decision determine how strongly those associations influence judgment.[9] Because patient groups are differently represented in the cases clinicians repeatedly encounter, this process can generate disparities even holding fixed both the clinician and the characteristics of the current patient. The findings therefore identify a new margin for reducing unequal care: institutions can shape the decision contexts clinicians face through case assignment, specialization and rotation, queue management, and staffing policies.

Second, the paper contributes to the literature on the economics of discrimination more broadly by showing that decision context is an important and previously underappreciated source of unequal treatment. Canonical models attribute disparities to stable differences in beliefs or preferences (Aigner and Cain, 1977; Becker, 1957; Phelps, 1972), and much of the empirical literature focuses on distinguishing between these two channels (see Bertrand and Duflo (2017); Lang and Kahn-Lang Spitzer (2020); Litwin and Low (2026) for recent reviews).[10] The results here show that just over half of a large racial disparity is explained by two features of decision context — prior case exposure and time pressure — that are

[8]Singh and Venkataramani (2026) show that hospital-level capacity strain raises in-hospital mortality for Black but not White patients. Such disparities under organizational capacity constraints could arise through multiple channels. The present paper shifts attention to the local context surrounding individual decisions, showing how time pressure and prior experience jointly shape group disparities: time pressure reduces individualized assessment and increases reliance on associations formed through prior experience.

[9]This aspect of the analysis also relates to laboratory evidence showing that time constraints affect, for example, risk-taking, strategic reasoning, generosity, and decision quality (Kocher, Pahlke and Trautmann, 2013; Imas, Kuhn and Mironova, 2022; Kessler et al., 2025), but large-scale field evidence from high-stakes professional settings is scarce.

[10]A review by Bohren et al. (2025) finds that 62% of discrimination papers published in top economics journals from 1990–2018 discuss this distinction.

orthogonal to the characteristics of the case at hand. This reorients where to look for sources of discrimination and, crucially, for policy levers: if disparities reflect modifiable features of the decision context rather than entrenched beliefs or preferences, then institutional and job design become first-order tools for reducing unequal treatment. The findings may also help account for disparities documented in other high-stakes settings with time-constrained, sequential decision-making, including hiring, policing, and consumer finance, as well as other clinical settings (Singh and Venkataramani, 2026).

Third, the paper contributes to labor and organizational economics by identifying workers' own task histories as an important determinant of subsequent decisions. Existing work shows that workplace behavior responds to incentives, managers, peers, workload, and accumulated experience.[11] The results here identify a distinct form of within-worker state dependence: the composition and sequence of cases a worker has previously handled shape how the worker evaluates the next case, even holding fixed both the worker and the characteristics of the current case. This finding gives routine organizational choices potentially important behavioral and distributional consequences. Case-assignment rules determine which experiences workers accumulate, specialization and rotation determine how concentrated those experiences become, and queue design and staffing determine how strongly workers rely on them. Organizations therefore shape decisions not only by selecting, incentivizing, and monitoring workers, but also by constructing the stream of tasks through which workers interpret each new case. In settings involving repeated judgments about people—including hiring, lending, policing, and healthcare—organizational design may consequently affect not only average performance, but also which groups receive favorable treatment.

Fourth, the paper contributes to the literature on coarse decision-making. A growing body of work shows that even experienced professionals rely on simplified assessments, including judges (Chen, Moskowitz and Shue, 2016), hiring managers (Radbruch and Schiprowski, 2025), and institutional investors (Akepanidtaworn et al., 2023). Related theoretical work formalizes how decision-makers simplify complex problems by grouping them into categories or relating them to similar problems encountered in the past (Mullainathan, Schwartzstein and Shleifer, 2008; Bordalo et al., 2026). This paper provides field evidence on how the content of such simplified assessments is shaped by experience, when decision-makers rely on them, and how they generate group disparities. The exposure results show that prior cases systematically change subsequent disparities, with recently encountered cases exerting particularly strong effects. The time-pressure results show that reliance on prior experience increases when individualized assessment is constrained. And the interaction results show

[11] See, e.g., Bandiera, Barankay and Rasul (2007), Mas and Moretti (2009), Levitt, List and Syverson (2013), Coviello, Ichino and Persico (2014), Kaur, Kremer and Mullainathan (2015), Minni (2026).

that these features operate jointly: time pressure magnifies the influence of prior exposure, while prior exposure determines which group is disadvantaged. More broadly, the results show that similarity-based assessment can have systematic distributional consequences that depend on who is overrepresented in the decision-maker's prior case mix.

The remainder of the paper is organized as follows. Section 2 describes the institutional setting and data, and Section 3 presents the conceptual framework. Section 4 outlines the empirical strategy. Section 5 presents the main effects of decision context on triage and uses additional predictions to distinguish similarity-based assessment from standard belief updating. Section 6 benchmarks triage decisions against predicted clinical risk. Section 7 presents policy counterfactuals, and Section 8 concludes.

2 Institutional Setting and Data

I study triage decisions for the universe of ED visits across six hospitals within a large U.S. health system between January 2017 and April 2025.[12] The system's hospitals are non-profit and include urban teaching hospitals as well as other acute-care hospitals serving nearby suburban communities, all of which use a unified electronic health record and standardized triage protocol. The analysis draws on detailed administrative data extracted from the electronic health record, which captures every clinical and operational data point entered by staff in real time, providing a near-census of ED activity.

My dataset comprises approximately 1.9 million triage assessments conducted by roughly 1,600 unique nurses. Each visit record includes pre-triage patient characteristics (age, sex, race and ethnicity, arrival mode, and chief complaint), detailed information collected during triage (vital signs, triage notes, pain score, and the assigned priority level), and post-triage outcomes (waiting times, diagnostics, treatments, and discharge diagnosis). Crucially, the data contain precise timestamps for every key event—including patient arrivals, triage start and end times, and downstream encounters—which allow me to reconstruct the complete sequence of arrivals and triage decisions for each nurse and shift. This rich temporal structure links each decision to preceding and subsequent activity in the department and enables me to construct high-frequency measures of variation in the local decision context.

The triage decision is the first formal point of clinical judgment and directly affects how quickly patients leave the waiting room, are evaluated by a physician, and begin diagnostic testing and treatment. Nurses assign each patient an Emergency Severity Index (ESI) score from 1 to 5, with lower scores indicating greater urgency. ESI 1 is reserved for patients

[12]Use of the de-identified administrative data was approved by the University of Pennsylvania Institutional Review Board.

requiring immediate life-saving intervention, while ESI 2 identifies high-risk patients requiring rapid evaluation; ESI 3–5 designate progressively lower-priority cases.[13] The assessment typically lasts only a few minutes and occurs continuously throughout a 12-hour shift as patients arrive. Because patient arrivals are inherently stochastic, the decision context nurses face varies naturally both within shifts and over longer horizons. I leverage this variation in the empirical analysis outlined in Section 4.

2.1 Sample Construction and Key Variables

The unit of observation is a patient visit to the ED. Encounters with missing nurse or hospital identifiers, implausible timestamps, or incomplete information on patient race or gender are excluded. After applying these restrictions, the analysis sample includes approximately 1.9 million visits conducted by 1,600 triage nurses across six hospitals.[14]

Top Triage Priority. The main outcome variable is an indicator for receiving top triage priority (ESI level 1 or 2), which designates that a patient either requires immediate life-saving intervention (ESI 1) or is considered high-risk and warrants rapid evaluation (ESI 2). This is to distinguish between cases the nurse perceived as urgent versus non-urgent. I use it first to measure the baseline racial disparity in triage and then to examine how that disparity changes with nurses' local decision context.

Patient Race. Race and ethnicity are defined using self-reported information in the electronic health record. I collapse these categories into a binary indicator for non-Hispanic White versus non-White because several racial and ethnic subgroups appear infrequently in the data, making disaggregated estimates imprecise. This approach aligns with common practice in the health disparities literature and provides sufficient power for within-nurse-by-shift comparisons.[15]

Frequent Flyers. Studying how prior experience shapes decisions requires identifying not only which cases nurses previously encountered, but also what those cases plausibly represented to the nurse. For an arbitrary prior patient, I observe a rich set of characteristics but it is impossible to know which features the nurse noticed or associated with clinical risk. Frequent flyers provide a useful exception: they are a recognizable patient group that nurses

[13] Approximately 94% of hospitals in the U.S. use the ESI for triage (Worth et al., 2019).
[14] Less than 5 percent of observations are excluded due to missing demographic information.
[15] See, for example, Alsan, Ianni and Peterson (2024) and Singh and Venkataramani (2026).

themselves associate with lower clinical risk.[16]

To capture this patient type in the administrative data, I define frequent flyers as patients with three or more ED visits in the prior twelve months who present with lower-risk chief complaints.[17] To classify chief complaints by risk, I use the full sample of patients with fewer than three ED visits in the prior year. For each chief complaint, I calculate the share of visits receiving top triage priority and rank complaints into quartiles based on this rate. I classify complaints in the highest-risk quartile as higher risk and those in the remaining three quartiles as lower risk. Thus, the definition identifies high-utilization patients presenting with complaints that are not among the highest-risk complaints in the broader patient population. Results are robust to alternative definitions, including defining frequent flyers solely based on having three or more prior-year ED visits and defining lower-risk repeat visitors based on the severity of their own prior visits.

Table 1 compares frequent flyers to non-frequent flyers. The two groups are similar along several basic dimensions: frequent flyers are roughly the same age, only modestly more likely to be male or arrive by ambulance, and are slightly more likely to report having a primary-care provider. The striking demographic difference is race: only 25% of frequent flyers are non-Hispanic White, compared with 45% of non-frequent flyers. Frequent flyers are also substantially more engaged with the healthcare system, averaging 6.8 ED visits, 1.5 hospitalizations, and 20.9 outpatient visits in the prior year, compared with 0.6, 0.2, and 6.5 among non-frequent flyers. Despite this substantially greater healthcare utilization, frequent flyers are less likely to receive top triage priority: 17% receive top triage compared with 23% of non-frequent flyers. They are also a common feature of nurses' day-to-day experience: a typical 24-patient shift includes roughly four frequent flyers, three of whom are non-White.

Recent Exposure. The first exposure measure indexes non-frequent-flyer patients by their position immediately before and after a non-White or White frequent-flyer encounter with the same nurse. This event-time measure traces how racial disparities change following a frequent-flyer encounter and as the encounter becomes more distant in the nurse's patient sequence. Because patient arrivals and assignment to the next available nurse are quasi-random, the timing and race of frequent-flyer encounters vary within nurse-shift, providing plausibly exogenous variation in recent exposure.

[16] During site visits, I discussed the triage process with nurses from hospitals in my sample, who independently described frequent flyers as a recognizable patient group that they generally associate with lower clinical risk.

[17] Three or more prior ED visits in the last year is a commonly used threshold for frequent utilization in the medical literature (Zuckerman and Shen, 2004; Hansagi et al., 2008; Navratil-Strawn et al., 2014).

Table 1: Summary Statistics by Frequent Flyer Status

	Non-Frequent Flyer		Frequent Flyer	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
Age	46.973	(22.271)	47.881	(19.109)
Male	0.433	(0.496)	0.474	(0.499)
Non-Hispanic White	0.448	(0.497)	0.250	(0.433)
Medicaid	0.260	(0.439)	0.501	(0.500)
Has PCP	0.719	(0.449)	0.790	(0.407)
Prior ED Visits (12mo)	0.601	(0.961)	6.844	(3.227)
Prior Hospitalizations (12mo)	0.178	(0.523)	1.487	(2.112)
Prior Outpatient Visits (12mo)	6.513	(12.343)	20.863	(19.419)
Arrived by Ambulance	0.193	(0.395)	0.219	(0.414)
Received Top Triage	0.234	(0.423)	0.169	(0.374)
Observations	1,556,638		266,695	

Note: Each observation is an ED visit. Columns (1)–(2) report means and standard deviations for non-frequent flyers, defined as patients with fewer than three ED visits in the prior 365 days. Columns (3)–(4) report statistics for frequent flyers, defined as patients with three or more ED visits in the prior 365 days who present with lower-risk chief complaints. To classify chief complaints by risk, I calculate the mean top-triage rate for each chief complaint among non-frequent-flyer visits, restricting to chief complaints observed at least 500 times, and rank complaints into quartiles. Lower-risk chief complaints are those in the bottom three quartiles; complaints in the top quartile are classified as higher risk. *Age* is measured in years. *Non-Hispanic White*, *Male*, *Medicaid*, *Has PCP*, and *Arrived by Ambulance* are binary indicators. *Prior ED Visits*, *Prior Hospitalizations*, and *Prior Outpatient Visits* are counts over the prior 12 months, winsorized at the 99th percentile. *Received Top Triage* is an indicator equal to one if the patient received an Emergency Severity Index (ESI) score of 1 or 2.

Longer-Run Exposure. The second exposure measure captures the nurse's longer-run exposure to non-White and White frequent flyers. For each patient encounter, I calculate the number of non-White and White frequent flyers the nurse encountered over the preceding twelve months, separately by race, and bin each measure into quartiles. All else equal, greater exposure to frequent flyers of a given race therefore increases the representation of that race among the lower-risk repeat visitors in the nurse's prior case history.

Time Pressure. I define time pressure as the urgency to complete the current triage assessment quickly. My primary measure is queue length, or the number of patients waiting

to be triaged behind the focal patient at the moment when triage begins.[18] Because patient arrivals vary unpredictably, queue length fluctuates sharply within shifts and generates high-frequency variation in the urgency to complete the current assessment.

As a robustness check, I construct a second time-pressure measure based on the waiting time of the next patient in line, conditional on queue length. Queue length captures how many patients are waiting, but not how long they have been waiting. For a given queue length, the pressure to complete the current assessment may be greater when the next patient has already been waiting longer. Because this waiting time depends on the timing of prior patient arrivals, it provides an alternative proxy for high-frequency variation in time pressure while holding the number of patients waiting fixed. Appendix Figures A.3, A.4, and A.5 show that the main results are robust to using this alternative measure.

Information Acquisition and Processing. I measure how time pressure changes the extent of individualized assessment along three dimensions. First, I use the duration of the triage encounter, calculated from the precise start and end timestamps and expressed in minutes, which captures the time available to gather and process patient-specific information and provides a behavioral validation of the time-pressure measure. Second, I capture the collection of structured clinical information using a set of core triage signals. I define core signals as vital signs and standardized screening measures recorded in at least 75% of encounters when no patients are waiting to be triaged, yielding eleven distinct measures. This restriction identifies information that is routinely collected when nurses face minimal time pressure, allowing me to test whether greater pressure leads nurses to omit information they would ordinarily collect. For each encounter, I count the number of these signals recorded, capturing the extent of structured information collected during the assessment. Third, I measure narrative documentation using the log number of words in the nurse's free-text triage note. Together, these measures capture complementary margins through which time pressure may reduce individualized assessment: the time devoted to the encounter, the structured clinical information collected, and the amount of narrative information recorded.

Diagnosis Severity. To benchmark triage decisions against underlying clinical risk, I construct a measure of diagnosis severity based on the patient's final ED diagnosis. Unlike the triage assessment, which is made within minutes of arrival, the final diagnosis reflects information revealed over the course of the visit through physician evaluation, diagnostic testing, and treatment. I classify diagnoses using the NYU ED algorithm, which assigns each

[18] Measuring queue length at the start of triage ensures that a longer or more complex assessment cannot mechanically increase the measured queue by allowing additional patients to arrive while the nurse is evaluating the focal patient.

diagnosis a probability of representing an urgent condition, and define a visit as severe when its probability of being urgent exceeds 0.5.[19] In Section 6, I use this measure to ask whether the context effects on triage also generate racial disparities in deviations from predicted clinical risk.

3 Conceptual Framework

I use a simple similarity-based framework to organize how prior experience and time pressure can shape racial disparities in triage decisions. Its central premise is that the evaluation of a current patient can depend on the nurse's experience with similar patients. The framework is intentionally agnostic about the cognitive process through which prior experience affects current decisions. Appendix C shows how the model of Bordalo et al. (2026) provides one possible micro-foundation for this simplified framework.

Prior and Recent Experience. Let E_{NW} and E_W denote a nurse's prior exposure to non-White and White frequent users. Because frequent users tend to present with lower-risk complaints, greater exposure to frequent users of a given race increases the nurse's experience with lower-risk patients of that race. If current decisions depend on prior similar cases, greater exposure to non-White frequent users should therefore reduce triage priority for subsequent non-White patients relative to White patients. Conversely, greater exposure to White frequent users should shift the racial gap in the opposite direction.

Let $S(E, r)$ summarize the extent to which prior experience links non-White rather than White patients to lower-risk cases, where $E = (E_{NW}, E_W)$ denotes longer-run exposure and $r = (r_{NW}, r_W)$ captures the influence of recent experiences with each group. Greater exposure to non-White frequent users strengthens this link for non-White patients, while greater exposure to White frequent users strengthens it for White patients. Thus,

$$\frac{\partial S}{\partial E_{NW}} > 0, \qquad \frac{\partial S}{\partial E_W} < 0. \tag{1}$$

Because frequent users are disproportionately non-White, nurses are more often exposed to non-White patients among the repeated lower-risk cases they encounter. Similarity-based reasoning can therefore generate an average disparity favoring White patients.

Prior experiences may also differ in their influence over time. When deciding what a new patient is similar to, cases seen recently provide a more immediate point of comparison than cases seen further in the past. I capture this intuition by assuming that r_g declines with the

[19] See Appendix B for additional details on the NYU ED Algorithm.

number of intervening cases since a frequent user encounter. Thus,

$$\frac{\partial S}{\partial r_{NW}} > 0, \qquad \frac{\partial S}{\partial r_W} < 0. \tag{2}$$

The framework therefore predicts that the effect of a frequent-user encounter should be strongest immediately and diminish as intervening cases accumulate.

Time Pressure. The influence of prior similar cases may also depend on the time available to evaluate the current patient. When time is limited, decision-makers have less scope to process the full set of patient-specific information and may place greater weight on readily observed dimensions of similarity, including race. Let T denote time pressure and let $\omega(T)$ capture the influence of race-based similarity on current decisions, with $\omega(T) \geq 0$ and $\omega'(T) > 0$.

Let G denote the White–non-White gap in top triage priority and G^0 the component of that gap arising through channels unrelated to similarity-based reasoning. The racial gap can then be summarized as

$$G(E, r, T) = G^0 + \omega(T)S(E, r). \tag{3}$$

Here, $S(E, r)$ captures the direction and strength of the race-specific link generated by prior experience: it is larger when prior experience more strongly links non-White rather than white patients to lower-risk cases. The term $\omega(T) \geq 0$ captures how strongly that similarity-based link influences current decisions, and $\omega'(T) > 0$ allows its influence to increase with time pressure. Thus, prior experience determines the direction of the similarity-based disparities, while time pressure determines how strongly those disparities affect observed decisions.

The framework's key prediction concerns how the effect of prior exposure changes with time pressure. Differentiating Equation 3 with respect to non-White exposure gives

$$\frac{\partial G}{\partial E_{NW}} = \omega(T)\frac{\partial S}{\partial E_{NW}} \geq 0. \tag{4}$$

Greater exposure to non-White frequent users strengthens the link between non-White patients and prior lower-risk cases, thereby widening the White–non-White triage gap.

Differentiating this exposure effect with respect to time pressure gives

$$\frac{\partial^2 G}{\partial T\, \partial E_{NW}} = \omega'(T)\frac{\partial S}{\partial E_{NW}} > 0. \tag{5}$$

Intuitively, non-White exposure widens the racial gap, and greater time pressure increases the influence of similarity on current decisions. The gap-widening effect of non-White exposure therefore becomes stronger as time pressure increases.

The same logic applies to White exposure. Differentiating first with respect to White exposure and then with respect to time pressure gives

$$\frac{\partial G}{\partial E_W} = \omega(T)\frac{\partial S}{\partial E_W} \leq 0, \qquad \frac{\partial^2 G}{\partial T\, \partial E_W} = \omega'(T)\frac{\partial S}{\partial E_W} < 0. \tag{6}$$

Greater exposure to White frequent users shifts the racial gap in the opposite direction, and time pressure strengthens that shift. Thus, time pressure does not by itself determine the direction of racial disparities; it amplifies the influence of the race-specific pattern generated by prior experience. Because exposure to frequent users is disproportionately non-White, this mechanism can also generate an average widening of the racial gap under time pressure.

Relation to Belief Updating. The basic exposure patterns predicted by the framework could also arise if exposure updates nurses' beliefs about group risk, rather than generating similarity-based comparisons across patients. The two explanations, however, differ in additional predictions. Under standard belief updating, the influence of a new encounter should depend on the information it provides about underlying risk and diminish as accumulated experience makes beliefs more precise. Under similarity-based reasoning, by contrast, a recent encounter can affect subsequent decisions because similar prior cases receive greater weight, even when the encounter provides little new information about underlying risk. Its influence therefore need not diminish with accumulated experience.

These predictions motivate the empirical analyses that follow: short-run effects of recent frequent-user encounters, longer-run effects of accumulated exposure, interactions between prior exposure and time pressure, and a test distinguishing similarity-based reasoning from belief updating.

4 Empirical Strategy

The empirical strategy exploits variation in two features of the local decision context: the cases nurses have previously encountered and the time available to make each decision. I study prior exposure at two horizons: recent exposure, using high-frequency variation in encounters with non-White and White frequent flyers within nurse shifts, and accumulated exposure, using within-nurse changes in exposure to non-White and White frequent flyers over the preceding year. I study time pressure using minute-to-minute fluctuations in the number

of patients waiting to be triaged. The identifying variation for each measure is described below.

4.1 Identifying Variation in Local Decision Context

Exposure. I exploit frequent-flyer exposure at two horizons. For recent exposure, identification comes from the timing and race of individual frequent-flyer encounters within a nurse's sequence of patients. Patient arrivals are stochastic and patients are assigned to the next available nurse, so whether a nurse encounters a non-White or White frequent flyer at a particular point in the shift is plausibly unrelated to the characteristics of the patients immediately surrounding that encounter. I use an event-study design to compare the White–non-White triage gap among non-frequent-flyer patients immediately before and after the encounter, within the same nurse-shift. The identifying assumption is that, absent the frequent-flyer encounter, the racial gap would have evolved smoothly across adjacent patients. While this assumption is fundamentally untestable, I assess its plausibility by testing for differential trends in the racial gap prior to the frequent-flyer encounter.

For longer-run exposure, I exploit changes in the same nurse's accumulated frequent-flyer exposure over time. For example, I compare the racial gap generated by a given nurse when her prior-year exposure to non-White frequent flyers is relatively low with the gap generated by that same nurse at another point in time when her exposure is relatively high. Nurse fixed effects absorb persistent differences across nurses, while hospital-by-year-month fixed effects absorb common changes in the patient population and clinical environment within each hospital over time. I additionally control for the nurse's overall experience and the focal patient's characteristics. I estimate the effects of White and non-White frequent-flyer exposure jointly in the same model, so variation in each captures changes in exposure to frequent flyers of one race holding exposure to frequent flyers of the other race fixed.

The identifying assumption for longer-run exposure is that, conditional on these controls, changes in a nurse's race-specific frequent-flyer exposure are not accompanied by other changes that independently affect her relative treatment of White and non-White patients. This assumption would be violated, for example, if a nurse's exposure to non-White frequent flyers increased at the same time that her non-frequent-flyer patients changed in ways that independently widened the racial gap. The hospital-by-year-month fixed effects absorb changes common to the hospital over time, while the patient controls and nurse-experience fixed effects account for observable changes in case mix and the nurse's overall experience.

Time Pressure. Identification comes from minute-to-minute fluctuations in queue length within the same nurse-shift. Because patient arrivals are stochastic, otherwise similar patients

evaluated by the same nurse can face different queue lengths depending on how many other patients happen to arrive around the time of their assessment. Nurse-by-shift fixed effects hold fixed the nurse and broader conditions over the shift, while hospital-hour fixed effects absorb systematic changes in patient volume and composition across times of day. Identification therefore comes from high-frequency fluctuations in time pressure within a nurse-shift, rather than from persistent differences across nurses, shifts, or predictable periods of high demand.

The identifying assumption is that, conditional on these fixed effects and patient characteristics, short-run fluctuations in queue length affect triage decisions through the time pressure they create rather than through correlated changes in the focal patient or the decision environment. There are two potential threats. First, bursts of arrivals that increase the queue could systematically change the race-specific clinical risk of patients being triaged. Appendix Figure A.1 provides evidence against this concern: queue length is not associated with economically meaningful differences in observable pre-triage patient characteristics. Second, other contemporaneous shocks could both increase queue length and independently affect nurse behavior. The fixed effects substantially narrow the scope for such confounding: nurse-by-shift fixed effects absorb conditions common to the nurse's shift, while hospital-hour fixed effects absorb hospital-wide changes in patient volume and the clinical environment within each hour. Any remaining confound would therefore need to vary at a finer frequency within these cells and simultaneously affect queue length and race-specific triage decisions.

4.2 Estimating Equations

Baseline Racial Gap. I first estimate the average within-nurse-shift racial gap in top triage priority using a linear probability model:

$$\text{TopTriage}_{ijst} = \alpha + \beta_W \,\text{White}_i + X'_{ijst}\gamma + \mu_{js} + \kappa_{c(i)} + \lambda_{ht} + \varepsilon_{ijst}, \tag{7}$$

where TopTriage_{ijst} indicates top triage priority (ESI 1 or 2) for patient i evaluated by nurse j during shift s at hospital h and time t; White_i indicates non-Hispanic White; X_{ijst} includes a male indicator, a cubic in age, vital signs with missingness indicators, arrival mode, and visit number fixed effects. Visit-number fixed effects compare patients at the same point in their own ED utilization history, absorbing systematic differences in triage associated with prior ED use, which differs substantially by race. μ_{js} are nurse-by-shift fixed effects; $\kappa_{c(i)}$ are chief-complaint fixed effects; and λ_{ht} are hospital-hour fixed effects. Standard errors are clustered at the nurse level. The coefficient β_W is the within-nurse-shift difference in the probability of top triage for White versus non-White patients, conditional on observable

clinical characteristics.[20]

Recent Exposure Effects. To estimate the short-run effects of frequent-flyer encounters on racial disparities, I restrict the sample to non-frequent-flyer patients and estimate event-study specifications separately around encounters with non-White and White frequent flyers:

$$\begin{aligned}\text{TopTriage}_{ijst} = \alpha + \beta_W \,\text{White}_i + \sum_k \beta_k \left(\text{White}_i \times R_{itk}\right) + \sum_b \pi_b \left(\text{White}_i \times P_{jtb}\right) \\ + \sum_d \phi_d \left(\text{White}_i \times D_{jtd}\right) + \sum_k \delta_k R_{itk} + \sum_b \lambda_b P_{jtb} \\ + \sum_d \psi_d D_{jtd} + X_i'\gamma + \text{FE} + \varepsilon_{ijst}, \qquad (8)\end{aligned}$$

where R_{itk} indicates patient i's position k relative to the focal frequent-flyer encounter; P_{jtb} indicates the number of cases since nurse j's previous frequent flyer of the focal race and is included for pre-exposure observations; and D_{jtd} indicates the number of cases since nurse j last encountered a frequent flyer of the non-focal race. These additional exposure controls account for the fact that frequent-flyer encounters can occur close together in the patient sequence. In particular, they ensure that changes around the focal encounter are not driven by lingering exposure to an earlier frequent flyer of the same race or by contemporaneous exposure to frequent flyers of the opposite race.

I estimate Equation 8 separately around non-White and White frequent-flyer encounters. Event time spans the five non-frequent-flyer patients immediately before and after the focal frequent-flyer encounter, with the final pre-exposure patient ($k = -1$) omitted as the reference category. X_i includes a male indicator and a cubic in age, and FE includes nurse-by-shift, hour-of-day, chief complaint, and arrival-mode fixed effects. Standard errors are clustered at the nurse level. The coefficients β_k trace how the within-nurse-shift White–non-White racial gap at each position k relative to the focal frequent-flyer encounter differs from the gap immediately before the encounter.

Longer-Run Exposure Effects. To estimate how the racial gap varies with nurses' longer-run exposure to frequent flyers, I restrict the sample to non-frequent-flyer patients

[20] In robustness checks, I additionally control for insurance status and primary care access, which may themselves reflect racial disparities in healthcare access.

and estimate:

$$\text{TopTriage}_{ijt} = \alpha + \beta_W \,\text{White}_i + \sum_b \beta_b^{NW} \left(\text{White}_i \times E_{jtb}^{NW}\right) + \sum_b \beta_b^{W} \left(\text{White}_i \times E_{jtb}^{W}\right) + \sum_b \delta_b^{NW} E_{jtb}^{NW} + \sum_b \delta_b^{W} E_{jtb}^{W} + X_i'\gamma + \text{FE} + \varepsilon_{ijt}, \tag{9}$$

where E_{jtb}^{NW} indicates whether nurse j's exposure to non-White frequent flyers over the preceding twelve months falls in quartile b, and E_{jtb}^{W} is defined analogously for White frequent flyers. The two exposure measures enter jointly, so variation in each captures changes in exposure to frequent flyers of one race holding exposure to frequent flyers of the other race fixed. X_i includes a male indicator and a cubic in age, and FE includes nurse, hospital-year-month, hospital-hour, chief complaint, arrival-mode, and nurse-experience fixed effects. Nurse experience is measured by the number of patients the nurse has evaluated over the preceding twelve months and enters flexibly using ventiles of this distribution. Standard errors are clustered at the nurse level. The coefficients β_b^{NW} and β_b^{W} capture how the within-nurse White–non-White racial gap varies across quartiles of non-White and White frequent-flyer exposure, respectively, with the lowest quartile of each omitted as the reference category.

Time Pressure Effects. To estimate how the racial gap varies with time pressure, I estimate:

$$\text{TopTriage}_{ijst} = \alpha + \beta_W \,\text{White}_i + \sum_k \theta_k \,\mathbf{1}\{TP_{ijst} \in k\} + \sum_k \phi_k \left(\text{White}_i \times \mathbf{1}\{TP_{ijst} \in k\}\right) + X_i'\gamma + \text{FE} + \varepsilon_{ijst}, \tag{10}$$

where $\mathbf{1}\{TP_{ijst} \in k\}$ indicates that the number of patients waiting to be triaged behind the focal patient falls in bin k, with zero patients waiting omitted as the reference category. X_i includes a male indicator and a cubic in age, and FE includes nurse-by-shift, hospital-hour, chief-complaint, ambulance-arrival, and visit number fixed effects. Standard errors are clustered at the nurse level. The coefficients ϕ_k capture how the White–non-White racial gap changes with queue length relative to the zero-queue state. As a robustness test, I re-estimate Equation 10 using next-patient wait time, conditional on queue length, as an alternative measure of time pressure. These results are reported in Appendix Figure A.4.

To validate queue length as a measure of time pressure and examine how it affects individualized assessment, I separately regress triage duration, log note length, and the

number of core triage signals collected on queue-length indicators, using the same high-frequency identifying variation. See Section 2.1 for details on how these variables are constructed.

Interaction Between Exposure and Time Pressure. To test whether time pressure changes how strongly prior exposure shapes racial disparities, I estimate a fully interacted specification for non-frequent-flyer patients:

$$\begin{aligned}\text{TopTriage}_{ijst} = \alpha + \beta_W \,\text{White}_i &+ \sum_b \sum_k \beta^{NW}_{bk} \left(\text{White}_i \times E^{NW}_{jtb} \times \mathbf{1}\{TP_{ijst} \in k\}\right) \\ &+ \sum_b \sum_k \beta^{W}_{bk} \left(\text{White}_i \times E^{W}_{jtb} \times \mathbf{1}\{TP_{ijst} \in k\}\right) \\ &+ X_i'\gamma + \text{FE} + \varepsilon_{ijst}, \end{aligned} \tag{11}$$

where $E^{NW}jtb$ and $E^{W}jtb$ denote quartiles of nurses' prior-year exposure to non-White and White frequent flyers, respectively, and $\mathbf{1}\{TP_{ijst} \in k\}$ denotes bins of queue length, with zero patients waiting omitted as the reference category. All main effects and two-way interactions are included but suppressed from Equation 11 for readability. X_i includes a male indicator and a cubic in age, and FE includes nurse, hospital-year-month, hospital-hour, chief-complaint, ambulance-arrival, and nurse-experience fixed effects, as defined above. Standard errors are clustered at the nurse level. As a robustness check, I re-estimate the specification using next-patient wait time, conditional on queue length, as an alternative measure of time pressure; these results are reported in Appendix Figure A.5.

The coefficients $\beta^{NW}bk$ and $\beta^{W}bk$ characterize how the White–non-White racial gap varies jointly with time pressure and prior exposure. The conceptual framework predicts that greater time pressure should magnify the effect of prior exposure on the racial gap: as non-White frequent-flyer exposure increases, time pressure should increasingly widen the racial gap, whereas as White frequent-flyer exposure increases, time pressure should increasingly narrow or reverse it.

4.3 Benchmarking Triage Against ML-Predicted Clinical Risk

The main analysis identifies how prior exposure and time pressure affect racial disparities in triage decisions. To assess whether these context effects also generate disparities relative to predicted clinical risk, I construct a benchmark triage classification based on machine-learning-predicted diagnosis severity. The benchmark asks which patients would receive top triage if priority were allocated according to predicted clinical risk rather than the nurse's

observed triage assessment.

I use the patient's final diagnosis as a downstream measure of clinical severity, as defined in Section 2. Unlike the triage assessment, which is made within minutes of arrival, the final diagnosis reflects information revealed over the course of the ED visit through physician evaluation, diagnostic testing, and treatment. I use a random forest to predict the severity of this final diagnosis from information available before triage. Restricting predictors to pre-triage information prevents the benchmark from incorporating nurses' own potentially context-dependent information collection. Predictors include chief complaint, arrival mode, insurance type, age, sex, primary-care status, and prior healthcare utilization; race and ethnicity are excluded.

A limitation is that final diagnosis is itself downstream of triage and may therefore be affected by the initial triage decision. To mitigate this concern, I estimate the prediction model using only visits in which the nurse is in the lowest quartile of prior-year exposure to both non-White and White frequent users and no patients are waiting to be triaged behind the focal patient. This restriction does not eliminate the potential endogeneity of downstream diagnosis, but reduces the extent to which the relationship used to construct the benchmark is estimated from encounters in which triage is most affected by context. Appendix B provides details on the diagnosis-severity measure, prediction procedure, and model validation.

I then translate predicted clinical risk into a benchmark triage assignment. Within each nurse-shift, I rank patients by predicted severity and assign predicted top-triage status to the highest-risk patients until the benchmark assigns the same number of patients to top triage as the nurse actually did. Thus, the exercise holds fixed the amount of top-triage priority allocated within each nurse-shift and changes only which patients receive it. Let $\widehat{Y}_i^{ML}$ denote this predicted top-triage assignment. I define the *triage wedge* as

$$W_i = Y_i - \widehat{Y}_i^{ML}, \tag{12}$$

where Y_i is the nurse's actual top-triage assignment. A positive wedge indicates that a patient receives top triage when the clinical-risk benchmark would not assign it, while a negative wedge indicates that a patient does not receive top triage when the benchmark would.

I use this wedge to examine whether racial disparities remain after benchmarking decisions against predicted clinical risk and whether those disparities respond to the same features of local context as actual triage decisions. Specifically, I re-estimate the baseline, longer-run exposure, and time-pressure specifications using the triage wedge as the outcome. Section 6 presents these results.

5 Results

5.1 Racial Disparities in Top Triage Priority

Table 2 documents a large baseline racial disparity in emergency triage decisions. In the unconditional comparison in Column (1), non-White patients are 9.2 percentage points (pp) less likely than White patients to receive top triage priority.[21]

I add clinical controls in two stages, distinguishing information available before the triage assessment from information collected by the nurse during the assessment. This distinction matters because the latter may itself be an outcome of the triage process. If nurses differ in whether or how they collect clinical information across patients, then patients with the same recorded information need not have entered triage with the same underlying clinical state. Conditioning on these measures can therefore select patients based partly on the nurse's assessment itself, rather than simply holding fixed clinical need at the start of triage. Column (2) conditions only on information determined prior to triage, including chief complaint, age, arrival mode, and location and time of arrival. The racial gap falls to 5.7 pp but remains large and statistically significant.

Column (3) additionally conditions on clinical information collected during triage, including vital signs. These measures provide substantially richer information about clinical urgency, but are potentially endogenous to the assessment process for the reasons above. I therefore add them separately rather than treating them as predetermined patient characteristics. Even after conditioning on this richer clinical information, the racial gap remains similar at 5.2 pp.[22]

In the preferred specification with nurse-by-shift fixed effects in Column (4), which compares patients evaluated by the same nurse during the same 12-hour shift, the racial gap is 5.4 pp. Thus, conditional on detailed observable clinical information and holding fixed the decision-maker and shift, non-White patients are 5.4 pp, or 21%, less likely than White patients to receive top triage priority.

Column (5) additionally controls for insurance type and primary-care-provider status, two important correlates of race and healthcare access. Including these measures may help account for otherwise unobserved differences in patient health that are correlated with access. At the same time, insurance and primary-care access may themselves reflect prior racial differences

[21] I always include a male indicator, so racial disparities are estimated within-gender, pooling across male and female patients.

[22] For this reason, the main analysis focused on decision context conditions only on patient characteristics determined prior to triage. In particular, I do not condition on information collected during the triage assessment because information acquisition may itself respond to the decision context; Section 5.4 examines this directly.

Table 2: Racial Disparities in Top Triage Probability

	(1) Unconditional	(2) Pre-Triage	(3) Clinical Urgency	(4) Nurse-Shift FE	(5) Access
Non-Hispanic White	0.092***	0.057***	0.052***	0.054***	0.048***
	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Observations	1,896,898	1,896,772	1,896,376	1,800,440	1,800,440
Nurses	1,815	1,815	1,815	1,661	1,661
R^2	0.013	0.276	0.303	0.408	0.410
Chief complaint FE	No	Yes	Yes	Yes	Yes
Hospital-Year FE	No	Yes	Yes	No	No
Month FE	No	Yes	Yes	No	No
Day-of-week FE	No	Yes	Yes	No	No
Hour-of-day FE	No	Yes	Yes	Yes	Yes
Visit number FE	No	No	Yes	Yes	Yes
Nurse-Shift FE	No	No	No	Yes	Yes
Ambulance FE	No	Yes	Yes	Yes	Yes
PCP FE	No	No	No	No	Yes
Insurance Type FE	No	No	No	No	Yes

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include a male indicator so racial gaps are estimated within gender. Unconditional probability of top triage for non-White patients is 19.9%. Column (1) reports the unconditional gap. Column (2) adds pre-triage controls: a cubic in age, ambulance arrival, chief complaint fixed effects, and hospital-by-year, month, day-of-week, and hour-of-day fixed effects. Column (3) additionally controls for clinical urgency, including temperature, pulse, respiratory rate, oxygen saturation, systolic and diastolic blood pressure (with missingness indicators), and patient visit number fixed effects. Column (4) introduces nurse-by-shift fixed effects (preferred specification), comparing patients seen by the same nurse during the same 12-hour shift. This specification retains the full set of clinical urgency controls, along with hour-of-day, visit number, ambulance, and chief complaint FE. Column (5) further absorbs access factors: primary care provider status and health insurance type. All specifications cluster standard errors at the nurse level.

in access to healthcare. Conditioning on them would then remove from the estimated racial gap any differences in triage associated with those prior access disparities. I therefore exclude them from the preferred specification and report their inclusion as a robustness check. The estimated racial gap remains similar at 4.8 pp, indicating that observable differences in healthcare access do not account for the baseline disparity.

Alternative Explanations for the Baseline Gap. One possibility is that the racial disparity reflects differences in patient behavior rather than differences in how nurses assess otherwise similar patients. I first examine whether non-White patients communicate less severe symptoms or are less able to advocate for care. Appendix Table A.1 tests whether the racial gap varies with socioeconomic status (SES), using insurance type as a proxy for SES.[23] If higher-SES patients are better able to communicate symptoms or advocate for care,

[23]Insurance type is strongly correlated with income, employment status, and underlying access to healthcare, and is a common proxy for SES (Fischer et al., 2020; Snyder and Chang, 2020; Monuteaux, Du and Neuman,

this explanation would predict a smaller racial gap at higher SES. Instead, the racial gap is nearly identical across SES groups.

Appendix Table A.2 provides complementary evidence using reported pain, a standardized measure of patients' expressed symptom severity. Conditional on chief complaint, clinical characteristics, and nurse-by-shift fixed effects, there is little racial difference in whether a pain score is recorded. Among patients with a recorded score, however, non-White patients report approximately 0.7 points more, rather than less, pain on a 0–10 scale than otherwise similar White patients. Thus, the lower triage priority assigned to non-White patients does not reflect lower expressed pain. This result does not rule out differences in how nurses interpret or respond to reported symptoms across races.

A distinct possibility is that nurses perceive patients of different races as differing in their likelihood of complaining or pursuing legal action and respond defensively to those perceived differences. Existing evidence on racial differences in complaint and litigation behavior is mixed: some studies find that White patients are more likely to file complaints or malpractice claims, while other evidence finds no racial difference in actual lawsuits (Carroll et al., 2005; Thomas, Pandit and Krevat, 2023). It is impossible to observe nurses' perceptions of complaint or litigation risk, so this mechanism may partly contribute to the baseline racial gap. However, a race-specific perception of complaint or litigation risk is not easily reconcilable with the high-frequency variation in the size – and direction – of the racial gap induced by context that I document below.

Coefficient-Stability Bounds (Oster, 2019). As an additional check for omitted-variable bias, I apply the coefficient-stability method from Oster (2019), which uses movements in the race coefficient and model fit as controls are added to assess sensitivity to selection on unobservables. Assuming equal selection on observables and unobservables, the bias-adjusted White coefficient is 0.042, compared with 0.054 in the preferred specification.[24] Moreover, reducing the coefficient to zero would require selection on unobservables to be approximately 4.4 times as strong as selection on observables. These bounds indicate that the baseline racial gap is robust to substantial selection on unobserved characteristics.

External Validity of the Baseline Gap. Finally, I assess whether the racial disparity is specific to this health system or also appears in U.S. EDs more broadly. I replicate the baseline analysis using the National Hospital Ambulatory Medical Care Survey (NHAMCS),

2024).

[24]Following Oster (2019), I set $R_{\max} = 1.3 \times R^2_{\text{full}}$ and assume equal selection on observables and unobservables ($\delta = 1$).

a nationally representative sample of U.S. ED visits from 2012–2022.[25] Appendix Figure A.2 shows a racial gap in the same direction in the national data. The unconditional White–non-White gap in top triage priority is 2.9 pp, and the adjusted gap is 1.8 pp. Although smaller than the disparity in my setting, the national estimates indicate that the racial gap in top triage priority is not unique to the hospitals I study.

5.2 Recent Exposure Shifts Racial Triage Gaps

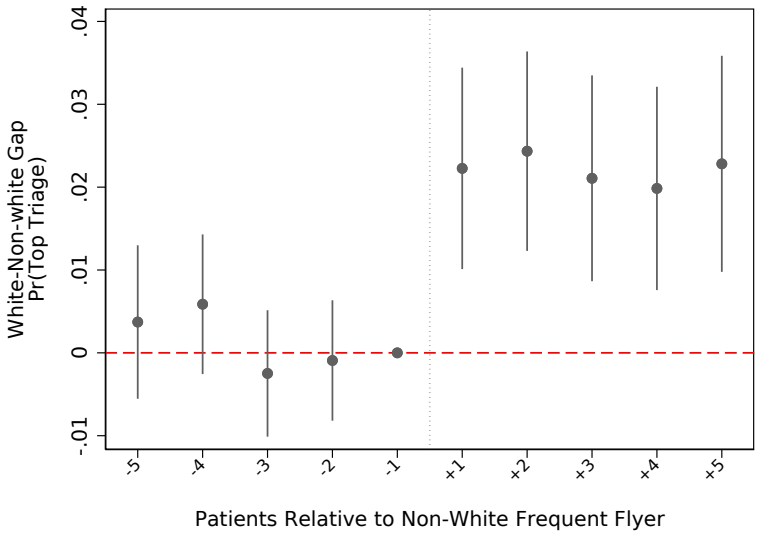
Having established the baseline racial gap, I next ask how it changes with nurses' local decision context. I begin with prior patient exposure. The analysis is motivated by a simple idea: because triage decisions are made in a matter of minutes, nurses may assess new patients partly by their similarity to previously encountered cases rather than evaluating each patient from scratch. I study this idea using nurses' encounters with frequent flyers, a recognizable, lower-risk group that is disproportionately non-White.

I first examine recent exposure, exploiting the quasi-random sequencing of patients within nurse shifts to compare racial disparities immediately before and after nurses encounter a non-White or White frequent flyer. I restrict the outcome sample to non-frequent-flyer patients, so the analysis asks whether exposure to this subset of patients changes the subsequent evaluation of patients outside that group.

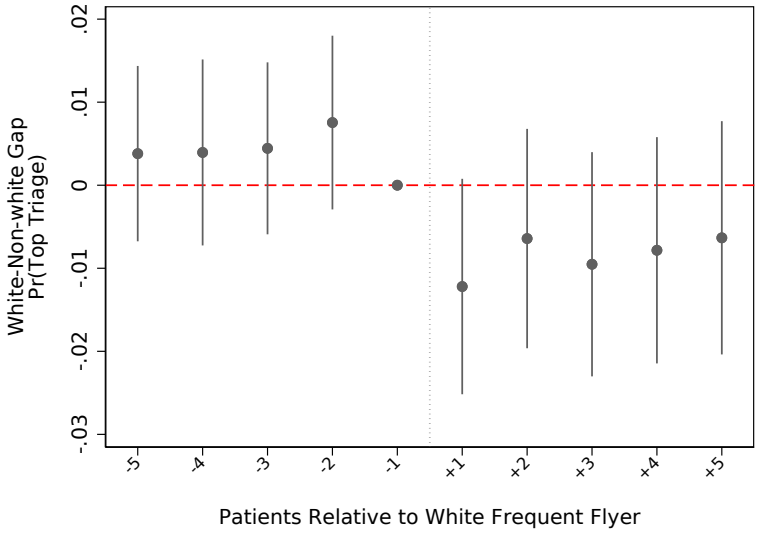
Figure 1 shows that frequent-flyer encounters shift subsequent racial disparities in triage priority sharply, with the direction of the effect depending on the frequent flyer's race. Following a non-White frequent flyer, the White–non-White gap in top triage priority increases by 2.3 pp (36% of the average gap over the five preceding patients). Following a White frequent flyer, the pattern reverses: the racial gap falls by 1.2 pp (25%). In both cases, the shift persists across the five subsequent non-frequent-flyer patients. Consistent with the quasi-random sequencing of patients, there is little systematic movement in the racial gap prior to frequent-flyer encounters of either race.

The direction of these responses is consistent with similarity-based assessment. A recent frequent-flyer encounter makes race-specific associations with low risk more salient when the nurse evaluates subsequent patients. Following a non-White frequent flyer, this shifts the relative assessment of subsequent non-White patients downward, widening the racial gap. Following a White frequent flyer, the same pattern operates in the opposite direction, narrowing the gap. Thus, the race of the recently encountered lower-risk patient determines the direction in which the racial disparity shifts.

[25]The available controls in NHAMCS most closely correspond to those in Column (2) of Table 2.



(a) Non-White Frequent Flyer Exposure



(b) White Frequent Flyer Exposure

Figure 1: Recent Frequent Flyer Exposure and the Racial Gap in Triage Priority

Note: Figures plot the White–non-White gap in the probability of receiving top triage priority across the five non-frequent-flyer patients immediately before and after the nurse encounters a non-White frequent flyer (Panel (a)) or a White frequent flyer (Panel (b)), estimated from Equation 8. The sample is restricted to non-frequent-flyer patients. Event time is measured by the focal patient's position relative to the frequent-flyer encounter, with the final pre-exposure case ($k = -1$) omitted as the reference category. The specification controls for distance to the nurse's previous frequent flyer of the focal race during the pre-period and distance to the most recent frequent flyer of the non-focal race. Estimates include nurse-by-shift, hour-of-day, chief complaint, and arrival-mode fixed effects. Patient controls include a male indicator and a cubic in age. Standard errors are clustered at the nurse level. Frequent flyers are defined as patients with three or more ED visits in the prior twelve months who present with lower-risk chief complaints. Vertical lines denote 95% confidence intervals.

5.3 Longer-Run Exposure Shapes Racial Triage Gaps

These short-run responses are consistent with a recent frequent-flyer encounter making race-specific associations with low risk more salient when the nurse evaluates subsequent patients. But such associations must first be formed through experience. I therefore next ask whether nurses' accumulated exposure to non-White and White frequent flyers shapes racial disparities over longer horizons.

As described above, for each encounter, I measure the nurse's exposure to non-White and White frequent flyers over the preceding twelve months. The analysis exploits changes in exposure within the same nurse over time, controlling flexibly for patient characteristics, hospital-specific time effects, and the nurse's overall experience. Non-White and White frequent-flyer exposure enter jointly, so changes in exposure to frequent flyers of one race are estimated holding exposure to frequent flyers of the other race fixed.

Figure 2 shows that within-nurse variation in longer-run exposure systematically shapes both the magnitude and direction of the racial gap. Moving from the lowest to highest quartile of non-White frequent-flyer exposure widens the White–non-White gap by 8.9 pp (231%). Moving from the lowest to highest quartile of White frequent-flyer exposure produces the mirror pattern, reducing the gap by 7.4 pp (193%). At sufficiently high levels of White exposure, the racial gap reverses entirely in favor of non-White patients. The reversal shows that local decision context shapes not only the magnitude of racial disparities, but which group is disadvantaged.

Together, the recent and accumulated exposure results point to similarity-based assessment in which prior cases influence how subsequent patients are evaluated. The fact that the same mirror-image pattern emerges at both short and longer horizons suggests that prior experience systematically shapes racial disparities in subsequent decisions.

5.4 The Effect of Time Pressure on Triage Decisions

I next turn to the second feature of local decision context: time pressure. When time is scarce, nurses may collect or process fewer features of the case at hand, making assessment less individualized and increasing reliance on associations generated by prior exposure.

I measure acute time pressure using the number of patients waiting to be triaged at the moment the current patient's triage assessment begins. Because patient arrivals are stochastic, queue length fluctuates sharply within shifts, generating high-frequency variation in the time pressure facing the nurse.

I first validate that greater queue length compresses the time nurses spend on each assessment. Figure 3 shows that triage duration falls monotonically as the number of patients

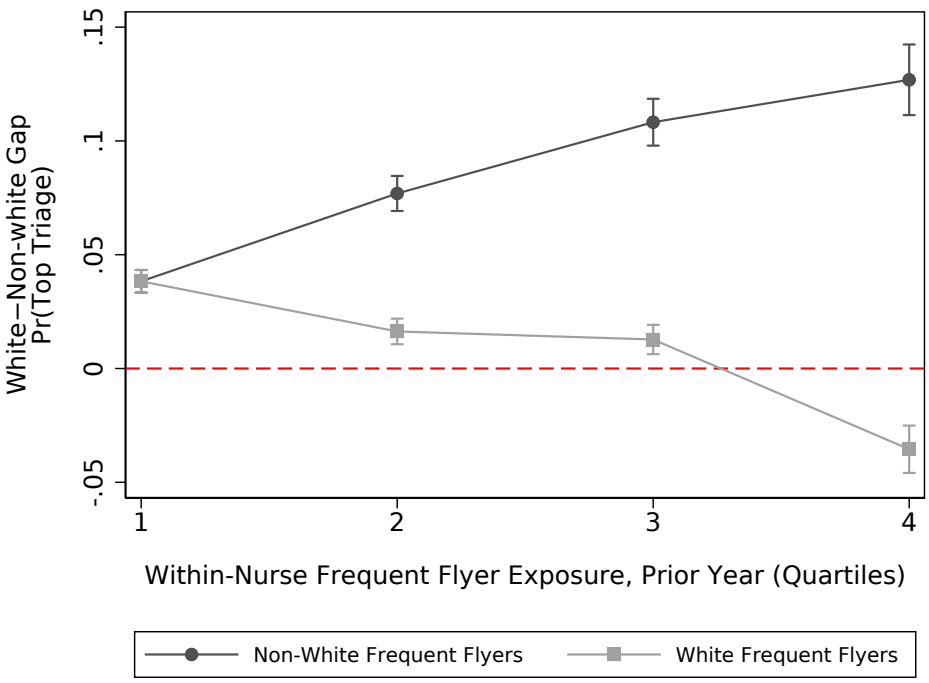


Figure 2: Within-Nurse Frequent Flyer Exposure Shapes the Racial Gap in Triage Priority

Note: Figure plots the predicted White–non-White gap in the probability of receiving top triage priority across quartiles of nurses' rolling prior-year exposure to non-White and White frequent flyers, estimated jointly from Equation 9. The sample is restricted to non-frequent-flyer patients. The omitted reference category is the lowest quartile of both non-White and White frequent-flyer exposure. Estimates are from a linear probability model with nurse, hospital-year-month, hospital-hour, chief complaint, and ambulance arrival fixed effects, as well as nurse experience fixed effects defined as ventiles of the cumulative case count over the prior twelve months. Standard errors are clustered at the nurse level. Patient controls include a male indicator and a cubic in age. Frequent flyers are defined as patients with three or more ED visits in the prior twelve months who present with lower-risk chief complaints. Exposure quartiles are constructed separately for non-White and White frequent flyers using the cumulative count of encounters over the rolling prior twelve months. Vertical lines denote 95% confidence intervals.

waiting increases. Moving from zero to five or more patients waiting reduces triage duration by 13.6% relative to a baseline of 5.5 minutes. Thus, nurses respond to greater time pressure by completing the current assessment more quickly.

I next examine whether time pressure changes racial disparities in triage decisions. Figure 4 shows that the average White–non-White gap in top triage priority widens monotonically with queue length. Moving from zero to five or more patients waiting widens the racial gap by 3.9 pp (73%). Thus, quasi-random increases in time pressure substantially widen racial disparities on average.[26]

One potential reason that time pressure amplifies the effect of prior exposure is that time

[26]Results are robust to using the wait time of the next patient conditional on queue length as an alternative measure of acute time pressure; see Appendix Figures A.3 and A.4.

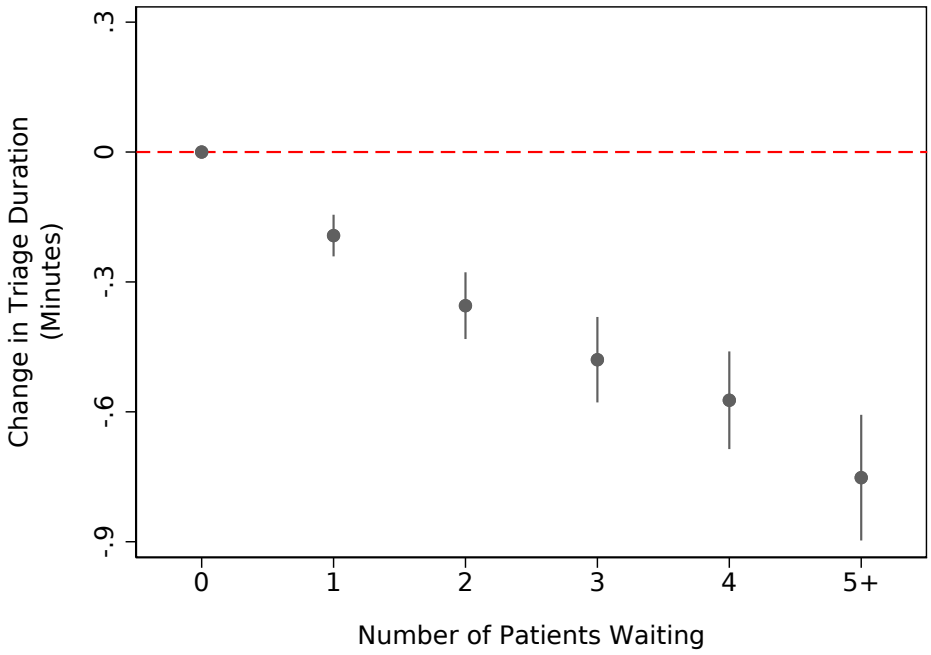


Figure 3: Queue Length and Time Spent in Triage

Note: Figure plots the estimated effect of the number of patients waiting to be triaged on the triage duration of the focal patient. The omitted reference category is zero patients waiting. Estimates are from a linear regression with nurse-by-shift, hospital-hour, visit number, chief complaint, and arrival mode fixed effects. Patient controls include indicators for race and gender and a cubic in age. Standard errors are clustered at the nurse level. The sample includes all triage encounters. Vertical lines denote 95% confidence intervals.

pressure reduces the amount of patient-specific information incorporated into the assessment. I therefore examine how queue length affects the information nurses collect and record during triage. Figure 5 shows that as queue length increases, nurses write shorter triage notes and record fewer core triage signals—clinical measures that they routinely collect under low time pressure.[27] Moving from zero to five or more patients waiting reduces triage-note length by 5.1% and the number of core triage signals recorded by 1.8%. For reference, when no patients are waiting, the median triage note contains 30 words and the mean encounter records 9 core triage signals. Thus, under greater time pressure, nurses omit information that they typically record when less constrained. Together with the reduction in triage duration, these results indicate that assessment becomes less individualized: nurses spend less time evaluating each patient and obtain and record less patient-specific information.

Importantly, I do not interpret the average widening of the racial gap as an independent tendency for time pressure to disadvantage non-White patients. Instead, nurses are dispro-

[27]Core triage signals are defined as the 11 clinical measures recorded in at least 75% of triage encounters when no patients are waiting; see Section 2 for details.

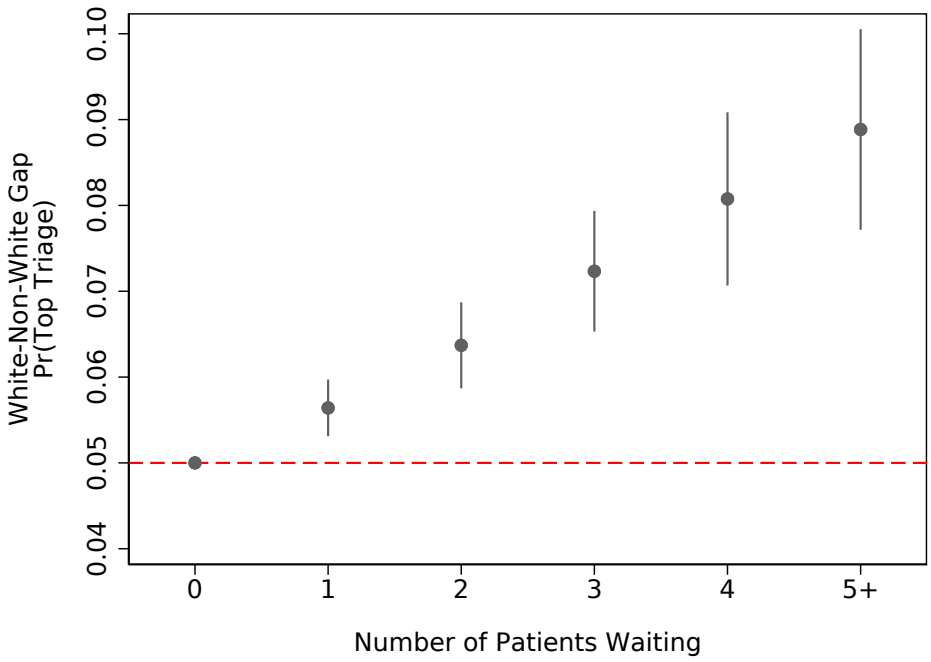


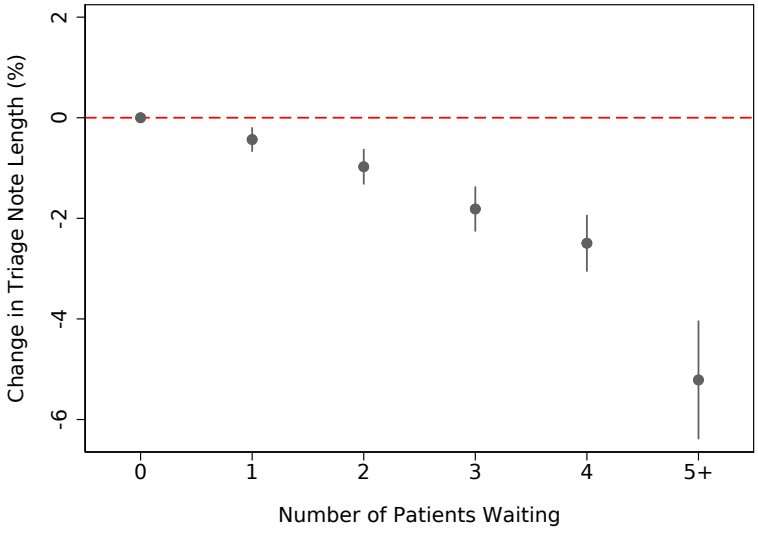
Figure 4: Time Pressure Widens the Racial Gap in Triage Priority

Note: Figure plots the estimated White–non-White gap in the probability of receiving top triage priority by the number of patients waiting to be triaged. The omitted reference category is zero patients waiting. The red dashed line indicates the estimated racial gap when no patients are waiting. Estimates are from a linear probability model with nurse-by-shift, hospital-hour, visit number, chief complaint, and arrival mode fixed effects. Patient controls include a male indicator and a cubic in age. Standard errors are clustered at the nurse level. The dependent variable is an indicator for receiving top triage priority (ESI 1 or 2). The sample includes all triage encounters. Vertical bars denote 95% confidence intervals.

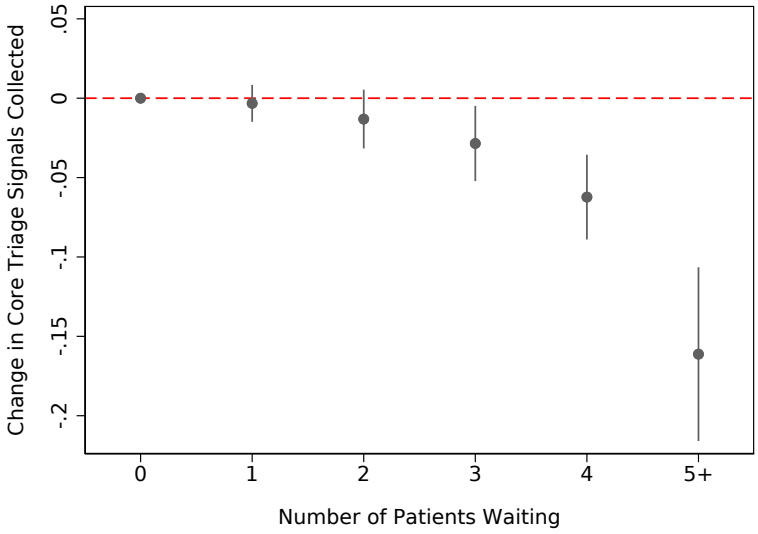
portionately exposed to non-White frequent flyers, so greater reliance on associations formed through prior exposure would widen the racial gap on average. If time pressure increases reliance on these associations, its effect on racial disparities should depend on which patients nurses have previously encountered. I test this prediction in the next section.

5.5 Time Pressure Increases Reliance on Prior Exposure

I next examine the interaction between prior exposure and time pressure. Figure 6 plots the White–non-White gap across queue length separately by quartile of nurses' accumulated exposure to non-White and White frequent flyers over the prior year, holding exposure to frequent flyers of the other race at its lowest quartile. As in the longer-run exposure analysis, the analysis exploits changes in exposure within the same nurse over time, and the sample is restricted to non-frequent-flyer patients.



(a) Triage Note Length



(b) Core Triage Measures

Figure 5: Time Pressure Compresses Individualized Information

Note: Panels plot the estimated effect of the number of patients waiting to be triaged on triage note length (Panel A) and the number of routinely collected triage measures recorded (Panel B). Routinely collected measures are defined as measures recorded in at least 75% of triage encounters when no patients are waiting (11 total). The omitted reference category is zero patients waiting. Estimates include nurse-by-shift, hospital-hour, chief complaint, and arrival mode fixed effects, as well as patient controls for race, gender, and a cubic in age. Standard errors are clustered at the nurse level. Vertical bars denote 95% confidence intervals.

Panel (a) shows that the effect of prior non-White frequent-flyer exposure on the racial gap increases sharply with time pressure. Moving from zero to five or more patients waiting widens the racial gap by 4.4 pp at the lowest exposure quartile. This effect grows monotonically with prior exposure, reaching 10.3 pp at the highest quartile.

Panel (b) shows the mirror pattern for prior White frequent-flyer exposure. At the lowest exposure quartile, moving from zero to five or more patients waiting widens the racial gap by 4.4 pp.[28] As time pressure increases, however, greater White frequent-flyer exposure increasingly narrows the racial gap, ultimately reversing it at high levels of exposure and time pressure.

Taken together, these results show that prior exposure determines the direction in which racial disparities shift, while time pressure determines how strongly prior exposure shapes assessment. When prior exposure more strongly associates non-White patients with lower-risk cases, greater time pressure magnifies the disparity against non-White patients; when prior exposure instead associates White patients with lower-risk cases, greater time pressure shifts the disparity in the opposite direction. Because nurses are disproportionately exposed to non-White frequent flyers, this interaction also explains why time pressure widens the racial gap on average.

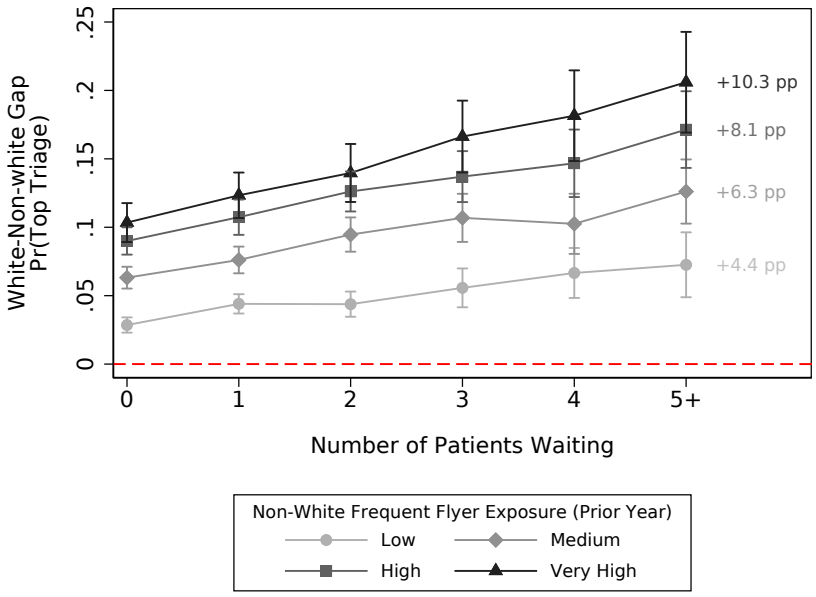
The interaction also helps interpret the nature of the exposure effects. That prior exposure plays a smaller role when nurses have more time to assess the current patient suggests that its greater influence under time pressure is generated by the constraint, rather than simply reflecting information that is generally useful for assessment.

5.6 Similarity-Based Assessment vs. Belief Updating

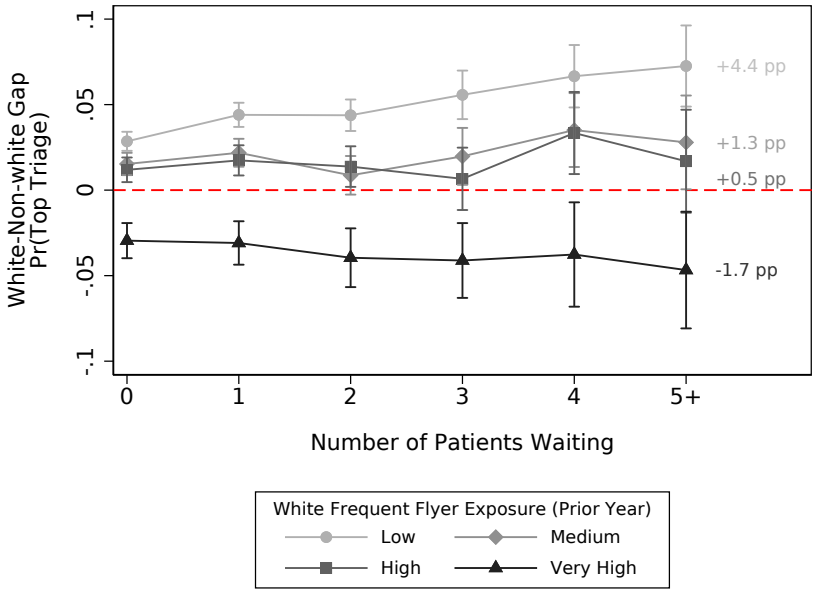
The preceding results show that prior cases systematically influence subsequent racial disparities. One interpretation is the similarity-based assessment described above, but dependence on prior experience could also arise through standard belief updating. Past observations may provide information about underlying clinical risk, leading nurses to update race-specific beliefs as they encounter new patients. This section considers whether learning about persistent race-specific differences in clinical risk can account for the exposure effects.

Suppose nurses are uncertain about race-specific clinical risk and learn from the patients they encounter. A lower-risk non-White frequent user provides information that non-White patients may be lower risk than previously believed, leading a Bayesian nurse to lower her assessment of subsequent non-White patients; a White frequent user produces the analogous

[28]The 4.4 pp estimate is common to both panels because both use the same reference state in which non-White and White frequent-flyer exposure are in their lowest quartiles.



(a) Non-White Exposure × Queue Length



(b) White Exposure × Queue Length

Figure 6: Time Pressure Increases Reliance on Prior Exposure

Note: Figures plot the predicted White–non-White gap in top triage priority (ESI 1 or 2) across queue length, separately by quartile of nurses' prior non-White frequent flyer exposure (Panel A) and White frequent flyer exposure (Panel B). The sample excludes frequent-flyer patients, defined as patients with three or more ED visits in the prior twelve months presenting with lower-risk chief complaints. Exposure is measured as the cumulative number of frequent-flyer encounters of each race over the prior twelve months. Regressions include nurse, hospital-year-month, hospital-hour, chief complaint, and arrival mode fixed effects, as well as gender and a cubic in age. Standard errors are clustered by nurse. Annotations report the change in the racial gap from zero patients waiting to the highest queue-length category within each exposure quartile. Vertical bars denote 95% confidence intervals.

effect for White patients.[29] Note that for this mechanism to generate the exposure effects documented above, frequent users must be informative about the clinical risk of same-race non-frequent users.

The two mechanisms differ in how accumulated experience should affect the influence of a recent encounter. Under Bayesian updating, prior and recent observations are substitutes as sources of information. As a nurse accumulates more experience with frequent users of a given race, her beliefs about that group's clinical risk become more precise, so an additional frequent-user encounter should generate a smaller belief update. The effect of a recent frequent-user encounter on subsequent racial disparities should therefore attenuate as accumulated exposure to frequent users of that race increases. Under similarity-based assessment, by contrast, accumulated exposure shifts the level of the racial gap, while recent encounters should generate similar or stronger shifts around that level as accumulated exposure increases, rather than the systematic attenuation implied by Bayesian updating.

I examine this prediction by returning to the recent-exposure design and asking whether the effect of a frequent-user encounter varies with the nurse's accumulated exposure to frequent users of the focal race. For each focal non-White frequent-user encounter, I measure the nurse's prior-year exposure to non-White frequent users and compare recent-exposure effects across levels of this accumulated exposure; I construct the analogous measure for White frequent-user encounters. The recent-exposure effects continue to be identified from changes in the racial gap around frequent-user encounters within nurse shifts. The exercise then asks whether these within-nurse-shift responses are smaller when nurses have greater accumulated exposure to frequent users of the focal race.

Appendix Figure A.6 shows no evidence that recent-exposure effects attenuate with accumulated exposure to frequent users of the focal race. For non-White frequent users, the effect of a recent encounter remains substantial at high levels of prior exposure, and the difference between the effects at high and low accumulated exposure is statistically insignificant. For White frequent users, the pattern runs in the opposite direction from Bayesian attenuation: the recent-exposure effect is concentrated among nurses with high prior exposure to White frequent users, while there is little effect among nurses with low prior exposure. Greater accumulated exposure therefore does not reduce the influence of a recent encounter, contrary to the attenuation predicted by Bayesian updating about persistent group risk.

[29] Statistical discrimination also involves treating members of the same racial group as similar for prediction. The distinction here is that belief updating operates through changes in beliefs about group-level risk, whereas similarity-based assessment allows prior experience to shape which features of past cases influence the assessment of the current patient.

6 Benchmarking Triage Decisions Against ML-Predicted Clinical Risk

The results thus far show that local decision context substantially changes racial disparities in triage priority. Under the identifying assumptions, the variation in recent exposure and time pressure is orthogonal to the underlying clinical risk of the current patient, so these shifts should not reflect changes in patient risk. To further probe this interpretation, I next ask whether the same context-induced disparities emerge when triage decisions are benchmarked against a patient-specific measure of predicted clinical risk. To do so, I compare nurses' actual triage assignments with those implied by a machine-learning prediction model, as described in Section 4.3 and Appendix B.

6.1 Racial Disparities Relative to Predicted Clinical Risk

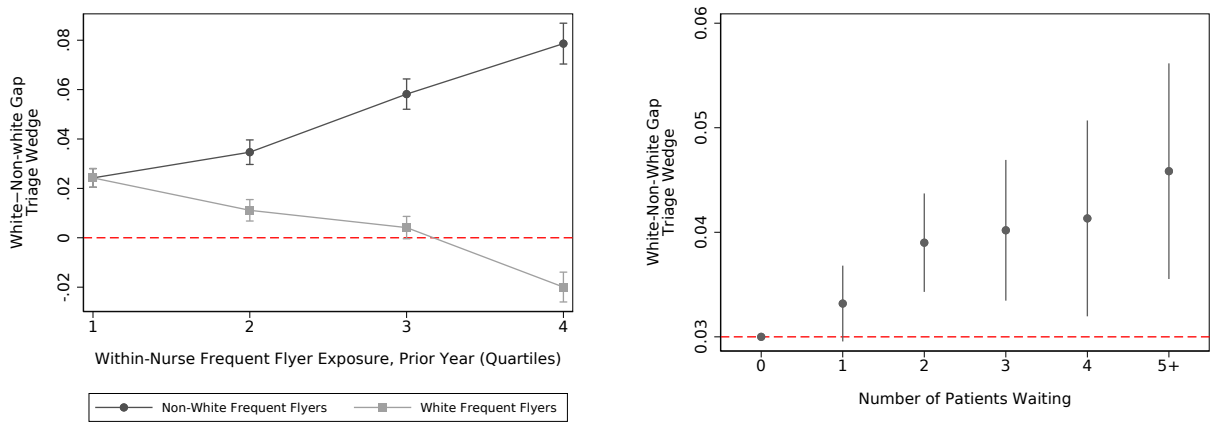
I first assess how closely nurses' triage decisions align with the clinical-risk benchmark. The random forest predicts diagnosis severity well out of sample, with an AUC of 0.78, substantially outperforming linear and LASSO benchmarks. Actual and predicted top-triage assignments agree for 82% of patients, indicating that nurses' decisions align reasonably well with predicted clinical risk.

Despite this general agreement, deviations from the benchmark are systematically related to patient race. Recall that I define the triage wedge as the difference between actual and predicted top-triage assignment. Estimating the baseline racial-gap specification with the triage wedge as the outcome yields a 3.1 pp White–non-White gap. This estimate is conservative because the prediction model is trained under the lowest observed levels of prior exposure and time pressure, rather than in the absence of these contextual forces. Any context-induced racial differences that remain even under these conditions may therefore be incorporated into the benchmark itself.

6.2 Decision Context and the Racial Triage Wedge

I next ask whether the context effects that shape racial disparities in actual triage decisions also shape racial disparities in departures from predicted clinical risk. I re-estimate the longer-run exposure and time-pressure specifications using the triage wedge as the outcome.

Figure 7 shows that they do. Panel (a) plots the White–non-White difference in the triage wedge across levels of longer-run frequent-flyer exposure. Greater exposure to non-White frequent flyers increases the racial wedge, indicating that White patients become increasingly likely to receive top triage relative to non-White patients beyond what predicted clinical risk



(a) Exposure to Frequent Users

(b) Time Pressure

Figure 7: Context Effects and the Racial Triage Wedge

Note: Figures plot the predicted White–non-White difference in the triage wedge, defined as actual minus machine-learning-predicted top-triage assignment. Panel (a) shows the racial triage wedge across levels of nurses' rolling prior-year exposure to non-White and White frequent flyers, estimated jointly. Panel (b) shows the racial triage wedge across levels of queue length, measured by the number of patients waiting to be triaged behind the current patient. The exposure sample is restricted to non-frequent-flyer patients. Estimates use the same controls and fixed effects as the corresponding main specifications in Section 5. Standard errors are clustered at the nurse level. Patient controls include a male indicator and a cubic in age. Frequent flyers are defined as patients with three or more ED visits in the prior twelve months who present with lower-risk chief complaints. Vertical lines denote 95% confidence intervals.

would imply. Greater exposure to White frequent flyers produces the mirror pattern: the wedge narrows and eventually reverses in favor of non-White patients. Thus, the direction of race-specific departures from the clinical-risk benchmark depends systematically on nurses' prior exposure.

Panel (b) shows a similar pattern for time pressure. As the number of patients waiting increases, the White–non-White triage wedge widens monotonically. Under greater time pressure, White patients therefore become increasingly likely relative to non-White patients to receive top triage beyond what their predicted clinical risk would imply. The same overall pattern holds using next-patient wait time as an alternative measure of time pressure, although the point estimates are not monotonic across the two highest pressure bins; the difference between these estimates is not statistically significant (Appendix Figure A.7).

Together, these results show that the context effects documented in Section 5 not only shape how nurses allocate triage priority, but also systematically change the direction and magnitude of racial disparities in departures from predicted clinical risk.

7 Reduced-Form Counterfactuals

To quantify how much of the observed racial disparity is associated with modifiable features of decision context, I conduct reduced-form counterfactual exercises using the estimated effects of prior exposure and time pressure. For each dimension of context, I compare predicted triage outcomes under the observed distribution to a counterfactual in which all encounters are assigned the lowest observed level, holding fixed patient characteristics, the assigned nurse, and all other covariates. These exercises are intended to quantify the potential importance of decision context rather than to evaluate specific policies.

The counterfactual estimates are derived from Equation 11. Setting longer-run frequent-flyer exposure to its lowest quartile for all nurses reduces the predicted racial gap by 33%. Setting time pressure to its lowest observed level for all encounters reduces the gap by 13%. Setting both dimensions of context to their lowest observed levels jointly reduces the baseline racial gap by 50%, suggesting that roughly half of the observed disparity is associated with variation in these two features of decision context.[30]

These counterfactual changes correspond to margins that institutions can influence. Lower exposure to frequent users could arise through changes in case assignment, provider specialization, or reductions in repeat ED utilization, while lower time pressure could arise through changes in staffing or queue management. The exercises do not capture equilibrium responses to such policies: assigning frequent users to dedicated providers, for example, would reallocate rather than eliminate their exposure, potentially changing the decisions of the providers who absorb those cases. Moreover, the counterfactuals move both dimensions only to their most favorable observed levels—exposure remains positive, and triage remains time-constrained even when measured time pressure is lowest. The estimates should therefore be interpreted as reduced-form evidence that the racial gap is highly responsive to modifiable features of decision context, rather than as evaluations of any particular intervention.

8 Conclusion

This paper shows that group disparities reflect not only who makes decisions, but also the context in which those decisions are made. The same decision-maker generates substantially different disparities—and even disparities in opposite directions—as the decision context changes. In emergency department triage, non-White patients are 5.4 pp (21%) less likely

[30] Importantly, each of these counterfactuals already incorporates the estimated interaction between exposure and time pressure, because one dimension is changed while the other remains at its observed level. The difference between the joint effect and the sum of the individual effects therefore does not measure the importance of the interaction.

than observably similar White patients evaluated by the same nurse during the same shift to receive top triage priority. Two features of the local decision context—the cases nurses have previously encountered and the time available to make the current decision—together account for just over half of this gap.

Exposure shapes who is disadvantaged. Prior exposure to lower-risk non-White frequent users widens the racial triage gap among subsequent patients, while exposure to White frequent users shifts the gap in the opposite direction and, at high levels, reverses it. Recent encounters sharply shift disparities over the next several patients, while accumulated exposure generates large differences in the magnitude and direction of the gap within the same nurse over time. That exposure to one salient subgroup changes how patients outside that subgroup are evaluated—and in opposite directions depending on whether exposure is to non-White or White frequent users—is difficult to reconcile with explanations based solely on stable beliefs or preferences.

Time pressure determines how strongly prior exposure shapes decisions. Quasi-random increases in time pressure widen racial disparities on average, but their effect depends sharply on prior exposure: greater non-White frequent-user exposure amplifies this effect, while greater White exposure can reverse it. Together, these patterns are consistent with similarity-based assessment in which experience shapes the associations brought to new cases, while time pressure reduces individualized assessment and increases reliance on those associations. That prior exposure matters less when nurses have more time to assess the current patient suggests that its greater influence under time pressure is generated by the constraint, rather than simply reflecting generally useful information. Consistent with this interpretation, the same contextual forces also generate racial disparities in deviations from a race-blind machine-learning benchmark of predicted clinical risk.

Together, these findings show how unequal treatment can emerge endogenously from the decision environment. Frequent users in this setting are disproportionately non-White, so nurses disproportionately encounter non-White patients among the salient, repeatedly encountered lower-risk cases from which they may generalize. Population composition therefore shapes decision-maker experience, experience shapes the associations brought to subsequent cases, and time pressure determines how strongly those associations influence decisions. Disparities can consequently emerge and persist not only because decision-makers hold different beliefs or preferences, but because groups are differently represented in the experiences from which decision-makers generalize. In this sense, the decision environment can systematically generate unequal treatment even holding fixed both the decision-maker and the characteristics of the current case.

This perspective changes the set of potential policy responses. If disparities depend partly

on decision context, reducing them need not require changing decision-makers' underlying preferences or beliefs. Institutions also shape the contexts in which decisions occur—how cases are assigned, whether workers specialize or rotate across different types of cases, how queues are managed, and how much time is available per case. The counterfactual exercises illustrate the scope of these margins: setting frequent-user exposure to its lowest quartile reduces the racial gap by 33%, setting time pressure to its lowest observed level reduces it by 13%, and implementing both accounts for 50% of the baseline gap. While these partial-equilibrium exercises are not evaluations of specific policies, they demonstrate the quantitative importance of modifiable features of the decision context.

This logic extends even to settings in which time pressure is inherent and difficult to reduce. Policing provides a salient example: officers sometimes make consequential decisions under acute time constraints that cannot simply be relaxed. Yet institutions retain discretion over other dimensions of the decision context, including the cases decision-makers encounter. Police departments determine how officers rotate across assignments and neighborhoods, shaping the experience they bring to subsequent decisions. If prior exposure shapes how new cases are assessed, these organizational choices may have distributional consequences even when the immediate time constraints surrounding a decision cannot be relaxed.

More broadly, the results suggest a different way of thinking about persistent group disparities. They reflect not only who decision-makers are or what they believe, but also the context in which they decide. That context has both a history—the cases encountered before the current decision—and an immediate constraint—the time available to evaluate the case. Because both are shaped in part by institutional design, persistent disparities may nevertheless be highly responsive to the environments in which decisions are made. Understanding group disparities therefore requires looking beyond the decision-maker and the person being evaluated to the decision context that connects them.

References

Aigner, Dennis J., and Glen G. Cain. 1977. "Statistical Theories of Discrimination in Labor Markets." *ILR Review*, 30(2): 175–187.

Akepanidtaworn, Klakow, Rick Di Mascio, Alex Imas, and Lawrence D.W. Schmidt. 2023. "Selling Fast and Buying Slow: Heuristics and Trading Performance of Institutional Investors." *The Journal of Finance*, 78(6): 3055–3098.

Alsan, Marcella, Katherine Ianni, and Graeme Peterson. 2024. "Racial Health Disparities in The United States." *Oxford Review of Economic Policy*, 40(3): 498–517.

Alsan, Marcella, Liran Einav, Amy Finkelstein, and Jonathan Zhang. 2026. "Clinician Behavior When Skin-Tone Affects Test Results." *Proceedings of the National Academy of Sciences*, 123(19): e2530585123.

Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. "Does Diversity Matter for Health? Experimental Evidence from Oakland." *American Economic Review*, 109(12): 4071–4111.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2007. "Incentives for Managers and Inequality Among Workers: Evidence from a Firm-Level Experiment." *The Quarterly Journal of Economics*, 122(2): 729–773.

Becker, Gary. 1957. *The Economics of Discrimination.* University of Chicago Press.

Bertrand, Marianne, and Esther Duflo. 2017. "Chapter 8 – Field Experiments on Discrimination." *Handbook of Economic Field Experiments*, 1: 309–393. eds. Abhijit Vinayak Banerjee and Esther Duflo, North-Holland.

Blau, Francine D. 2025. "Gender Inequality in the Labor Market: Continuing Progress?" *ILR Review*, 78(2): 275–303.

Bohren, J. Aislinn, Kareem Haggag, Alex Imas, and Devin G. Pope. 2025. "Inaccurate Statistical Discrimination: An Identification Problem." *The Review of Economics and Statistics*, 107(3): 605–620.

Bordalo, Pedro, John J. Conlon, Nicola Gennaioli, Spencer Y. Kwon, and Andrei Shleifer. 2023. "Memory and Probability." *The Quarterly Journal of Economics*, 138(1): 265–311.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, Attention, and Choice." *The Quarterly Journal of Economics*, 135(3): 1399–1442.

Bordalo, Pedro, Nicola Gennaioli, Giacomo Lanzani, and Andrei Shleifer. 2026. "A Cognitive Theory of Reasoning and Choice." *The Quarterly Journal of Economics*, 141(3): 1921–1963.

Butler, Alexander W, Erik J Mayer, and James P Weston. 2023. "Racial Disparities in the Auto Loan Market." *The Review of Financial Studies*, 36(1): 1–41.

Carroll, Kecia N., William O. Cooper, Jennifer U. Blackford, and Gerald B. Hickson. 2005. "Characteristics of Families That Complain Following Pediatric Emergency Visits." *Ambulatory Pediatrics*, 5(6): 326–331.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *The Quarterly Journal of Economics*, 131(3): 1181–1242.

Cohen, Alma, and Crystal S. Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy*, 11(1): 160–91.

Coviello, Decio, Andrea Ichino, and Nicola Persico. 2014. "Time Allocation and Task Juggling." *American Economic Review*, 104(2): 609–623.

Fischer, Igor, Hendrik-Jan Mijderwijk, Ulf D Kahlert, Marion Rapp, Michael Sabel, Daniel Hänggi, Hans-Jakob Steiger, Marie-Therese Forster, and Marcel A Kamp. 2020. "Association Between Health Insurance Status and Malignant Glioma." *Neuro-Oncology Practice*, 7(5): 531–540.

Freedman, Seth, Ezra Golberstein, Tsan-Yao Huang, David J Satin, and Laura Barrie Smith. 2021. "Docs With Their Eyes on The Clock? The Effect of Time Pressures on Primary Care Productivity." *Journal of Health Economics*, 77(102442).

Fryer, Roland G., and Matthew O. Jackson. 2008. "A Categorical Model of Cognition and Biased Decision Making." *The B.E. Journal of Theoretical Economics*, 8(1): 1–44.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1091–1119.

Handel, Ben, Louis Heizlsperger, Jonas Knecht, Jonathan Kolstad, Ulrike Malmendier, and Filip Matejka. 2026. "Thinking Versus Doing: Cognitive Capacity, Decision Making and Medical Diagnosis." *Working Paper*.

Hansagi, Helen, Mariann Olsson, Aliasgher Hussain, and Gunnar Ohlén. 2008. "Is Information Sharing Between the Emergency Department and Primary Care Useful to the Care of Frequent Emergency Department Users?" *European Journal of Emergency Medicine*, 15(1): 34–39.

Hurtado, Agustin, and Jung Sakong. 2024. "Racial Disparities in the US Mortgage Market." *AEA Papers and Proceedings*, 114: 201–204.

Imas, Alex, Michael A. Kuhn, and Vera Mironova. 2022. "Waiting to Choose: The Role of Deliberation in Intertemporal Choice." *American Economic Journal: Microeconomics*, 14(3): 414–440.

Jin, Lawrence, Rui Tang, Han Ye, Junjian Yi, and Songfa Zhong. 2020. "Time Dependency in Physician Decision-Making." *AEA Papers and Proceedings*, 110: 284–88.

Jin, Lawrence, Rui Tang, Han Ye, Junjian Yi, and Songfa Zhong. 2023. "Path Dependency in Physician Decisions." *The Review of Economic Studies*, 91(5): 2916–2953.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy*, 123(6): 1227–1277.

Kessler, Judd B., Hannu Kivimaki, Ashley Litwin, and Muriel Niederle. 2025. "Please Take A Minute: How Prosocial Choices Change With Deliberation." *Working Paper*.

Kocher, Martin G, Julius Pahlke, and Stefan T Trautmann. 2013. "Tempus Fugit: Time Pressure In Risky Decisions." *Management Science*, 59(10): 2380–2391.

Lang, Kevin, and Ariella Kahn-Lang Spitzer. 2020. "Race Discrimination: An Economic Perspective." *Journal of Economic Perspectives*, 34(2): 68–89.

Levitt, Steven D., John A. List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." *Journal of Political Economy*, 121(4): 643–681.

Litwin, Ashley, and Corinne Low. 2026. "Measuring Discrimination with Experiments." *Handbook of Experimental Methods in the Social Sciences*. ed. Alex Rees-Jones, Edward Elgar Publishing Ltd.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review*, 99(1): 112–145.

Minni, Virginia. 2026. "Making the Invisible Hand Visible: Managers and the Allocation of Workers to Jobs." *The Quarterly Journal of Economics*, 141(3): 1871–1920.

Monuteaux, Michael C, Michelle Du, and Mark I Neuman. 2024. "Evaluation of Insurance Type As A Proxy For Socioeconomic Status In The Pediatric Emergency Department: A Pilot Study." *Annals of Emergency Medicine*, 83(6): 562–567.

Mullainathan, Sendhil. 2002. "Thinking Through Categories." Working Paper, Harvard University.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer. 2008. "Coarse Thinking and Persuasion." *The Quarterly Journal of Economics*, 123(2): 577–619.

Navratil-Strawn, Jessica L., Kevin Hawkins, Timothy S. Wells, Ronald J. Ozminkowski, Stephen K. Hartley, Richard J. Migliori, and Charlotte S. Yeh. 2014. "An Emergency Room Decision-Support Program That Increased Physician Office Visits, Decreased Emergency Room Visits, and Saved Money." *Population Health Management*, 17(5): 257–264.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204.

Phelps, Edmund. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review*, 62(4): 659–661.

Radbruch, Jonas, and Amelie Schiprowski. 2025. "Interview Sequences and the Formation of Subjective Assessments." *The Review of Economic Studies*, 92(2): 1226–1256.

Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy*, 122(6): 1320–1354.

Silver, David. 2021. "Haste or Waste? Peer Pressure and Productivity in the Emergency Department." *The Review of Economic Studies*, 88(3): 1385–1417.

Singh, Manasvini. 2021. "Heuristics In The Delivery Room." *Science*, 374(6565): 324–329.

Singh, Manasvini, and Atheendar Venkataramani. 2026. "Rationing by Race." *American Economic Review*, 116(9): 3552–85.

Snyder, Rebecca A, and George J Chang. 2020. "Insurance Status As A Surrogate for Social Determinants of Health In Cancer Clinical Trials." *JAMA Network Open*, 3(4).

Thomas, Angela D., Chinmay Pandit, and Seth A. Krevat. 2023. "Race Differences in a Malpractice Event Database in a Large Healthcare System." *Journal of Patient Safety*, 19(2): 67–70.

Worth, Murtis, Leslie L. Davis, Debra C. Wallace, Robin Bartlett, and Debbie Travers. 2019. "Are Emergency Departments In The United States Following Recommendations by The Emergency Severity Index To Promote Quality Triage and Reliability?" *Journal of Emergency Nursing*, 45(6): 677–684.

Zuckerman, Stephen, and Yu-Chu Shen. 2004. "Characteristics of Occasional and Frequent Emergency Department Users: Do Insurance Coverage and Access to Care Matter?" *Medical Care*, 42(2): 176–182.

A Additional Results

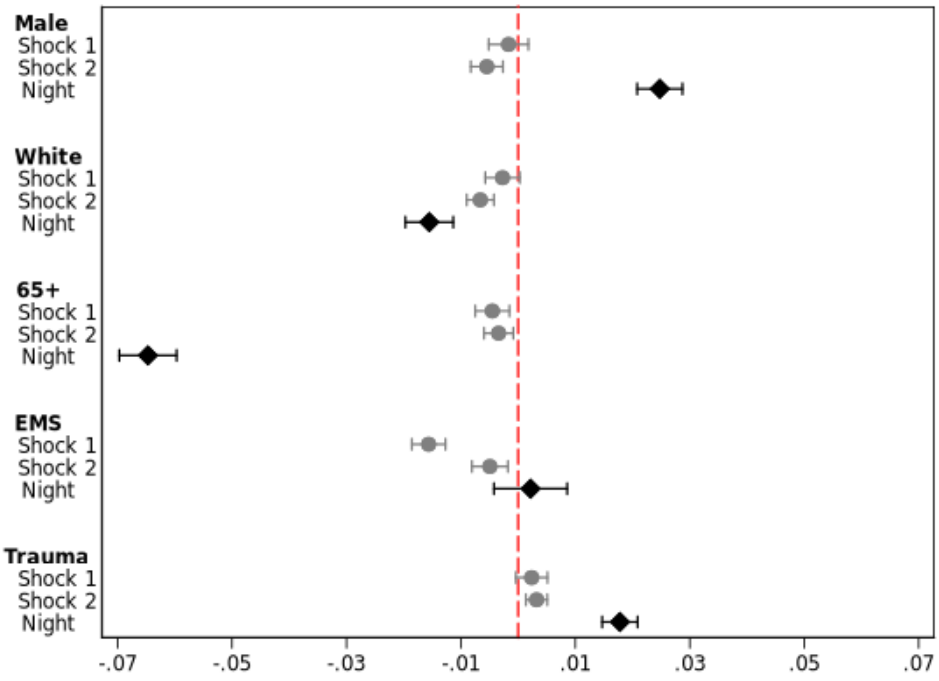


Figure A.1: Balance on Observable Patient Characteristics Across Time-Pressure Shocks

Note: Figure shows balance tests for the two time-pressure shocks. Each point reports the coefficient on an indicator for being in the top two bins of the relevant time-pressure measure from a regression of the indicated patient characteristic on that indicator. All regressions condition on nurse-by-shift, hospital-hour, and visit number fixed effects, include the non-focal time-pressure measure as a control, and cluster standard errors at the nurse level. Shock 1 is an indicator for long queue length (number of patients waiting to be triaged behind the current patient). Shock 2 is an indicator for a long next-patient wait time, conditional on queue length. Night visit indicates the visit occurred during a night shift and is included as an endogenous benchmark. EMS indicates arrival by emergency medical services. Trauma indicates a trauma-related chief complaint. Horizontal lines denote 95% confidence intervals.

Table A.1: Racial Disparities in Top Triage Probability by SES

	(1)
Non-Hispanic White	0.053***
	(0.004)
High-SES	0.015***
	(0.001)
Non-Hispanic White × High-SES	0.001
	(0.002)
Observations	1,188,279
Nurses	1,562
R^2	0.388
Nurse-Shift FE	Yes
Hour-of-day FE	Yes
Chief complaint FE	Yes
Ambulance FE	Yes
Visit number FE	Yes

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression includes a male indicator so racial gaps are estimated within gender. The dependent variable is an indicator for receiving top triage priority. Unadjusted probability of top triage for low-SES, non-white patients is 16.5%. "High-SES" is defined as having commercial or employer-sponsored insurance (Blue Cross, Commercial, or Managed Care). Patients with Medicaid or self-pay are classified as low-SES, and Medicare patients are excluded since eligibility is age-driven. The coefficient on "Non-Hispanic White" captures the racial gap in top triage probability among low-SES patients. The coefficient on "High-SES" captures the difference between high- and low-SES patients among non-Whites. The interaction term reports the difference in the racial gap between high-SES and low-SES patients. Specification controls for a cubic in age and clinical urgency covariates (temperature, pulse, respiratory rate, oxygen saturation, systolic and diastolic blood pressure, with missingness indicators), and absorbs fixed effects for nurse-by-shift, chief complaint, ambulance arrival, visit number, and hour of arrival. Standard errors are clustered at the nurse level.

Table A.2: Pain Score Documentation and Reported Pain by Race

	Missing Pain Score		Pain Score (0-10)	
	(1)	(2)	(3)	(4)
	Unadjusted	Adjusted	Unadjusted	Adjusted
Non-Hispanic White	0.103***	-0.004***	-1.056***	-0.699***
	(0.007)	(0.001)	(0.027)	(0.017)
Observations	1,896,898	1,800,440	1,020,298	940,458
Nurses	1,815	1,661	1,570	1,394
R^2	0.014	0.495	0.044	0.297
Nurse-Shift FE	No	Yes	No	Yes
Hour-of-day FE	No	Yes	No	Yes
Chief complaint FE	No	Yes	No	Yes
Ambulance FE	No	Yes	No	Yes
Visit number FE	No	Yes	No	Yes

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All regressions include a male indicator so racial gaps are estimated within gender. Unadjusted missingness of the pain score among non-White patients is 41.8 percent, and their unadjusted mean reported pain score (0–10 scale) is 4.34. Column (1) reports the unadjusted racial gap in the probability that a pain score is missing. Column (2) adds controls: a cubic in age, clinical urgency covariates (temperature, pulse, respiratory rate, oxygen saturation, systolic and diastolic blood pressure, with missingness indicators), and absorbs fixed effects for ambulance arrival, chief complaint, nurse-by-shift, visit number, and hour of arrival. Column (3) reports the unadjusted racial gap in reported pain score, conditional on the score being non-missing. Column (4) adds the same controls and fixed effects as Column (2). All specifications cluster standard errors at the nurse level.

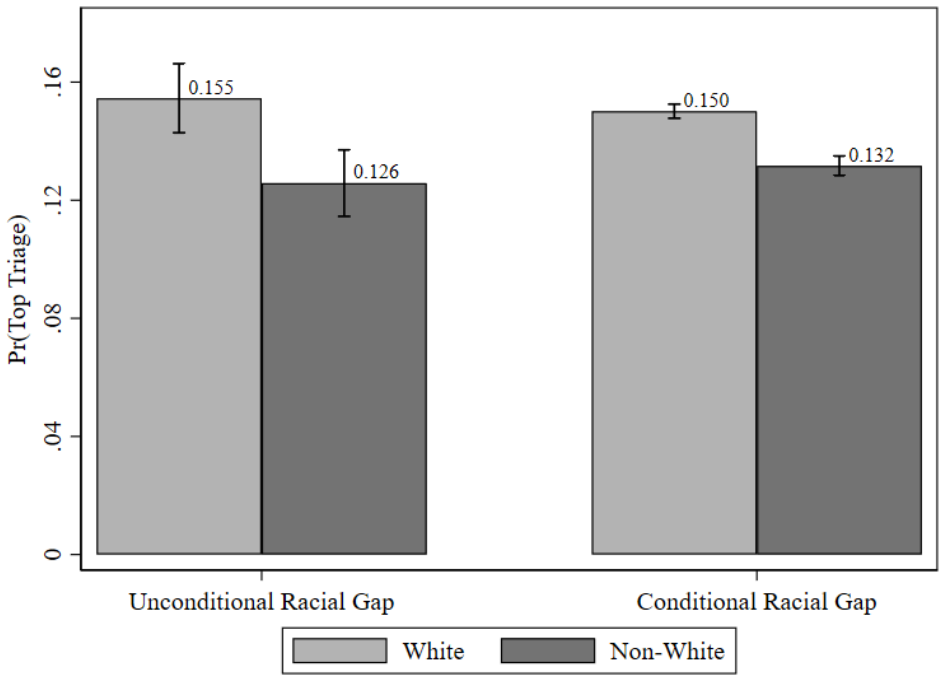


Figure A.2: Racial Disparities in Top Triage Probability in the U.S.

Note: Figure shows estimates from the National Hospital Ambulatory Medical Care Survey (NHAMCS), 2012-2022. All regressions include a male indicator so racial gaps are estimated within gender. The unadjusted White-non-White gap in the probability of receiving top triage is 2.9 pp; the adjusted gap is 1.8 pp. The adjusted specification controls for a cubic in patient age, an indicator for ambulance arrival, chief complaint fixed effects, hospital-by-year fixed effects, day-of-week fixed effects, hour-of-day fixed effects, and month fixed effects. Estimates use NHAMCS patient weights, with standard errors clustered at the primary sampling unit level to account for survey design.

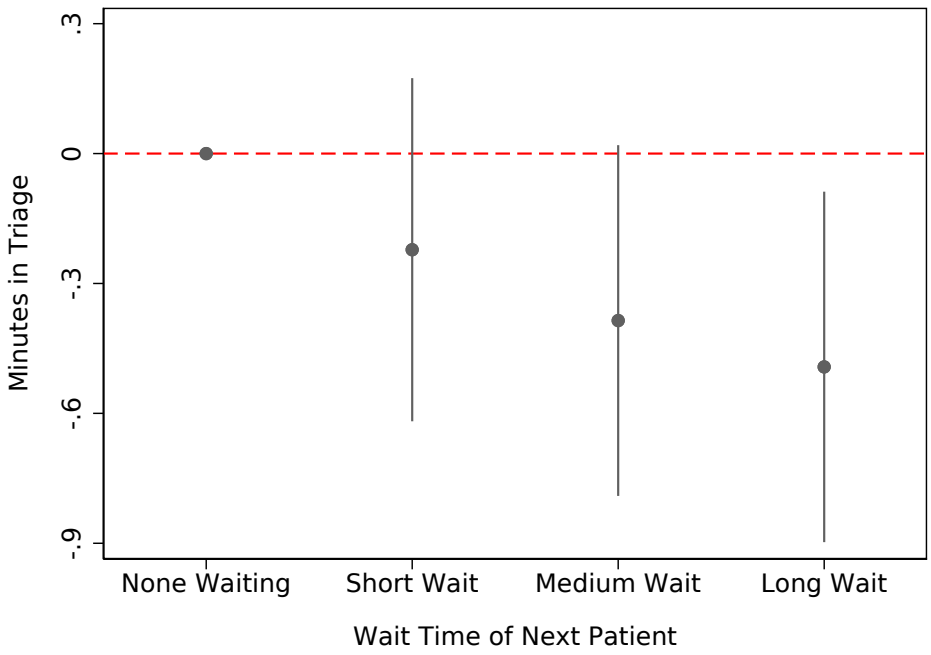


Figure A.3: Next-Patient Wait Time and Time Spent in Triage

Note: Figure plots the estimated effect of the wait time of the next patient in line on the triage duration of the focal patient, conditional on queue length. The omitted reference category is no next patient waiting. Regressions include nurse-by-shift, hospital-hour, chief complaint, and arrival mode fixed effects, as well as controls for patient race, gender, and a cubic in age. Standard errors are clustered by nurse. Vertical bars denote 95% confidence intervals.

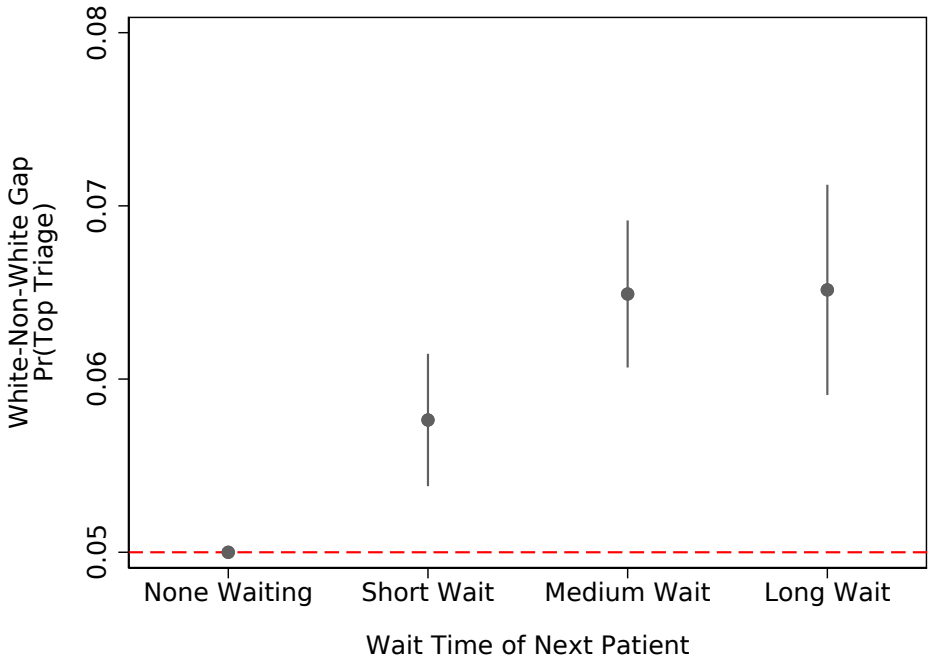
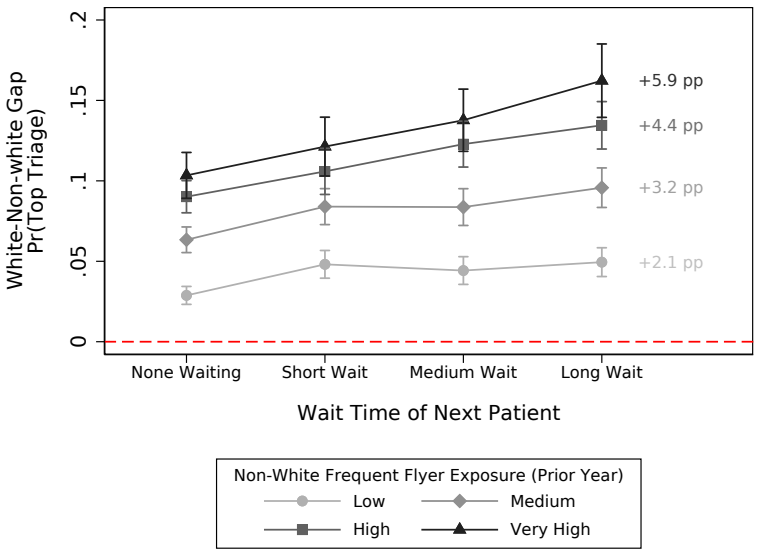
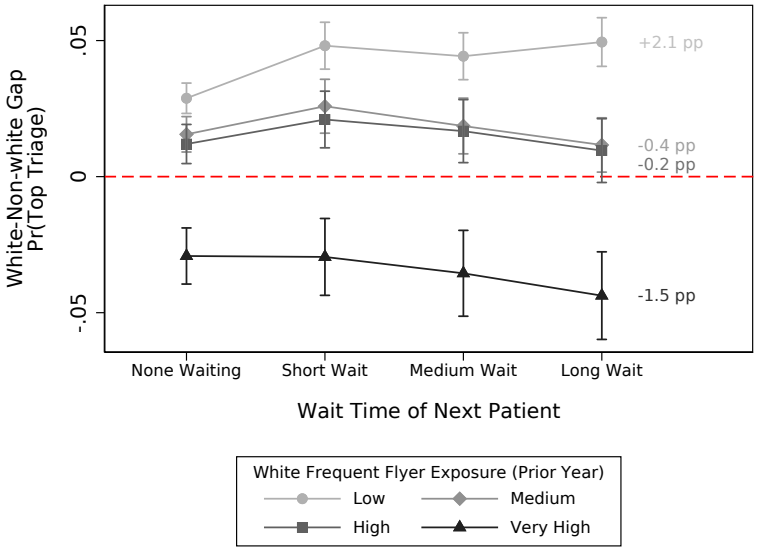


Figure A.4: Time Pressure and the Racial Gap in Triage Priority: Next-Patient Wait Time

Note: Figure plots the predicted White–non-White gap in the probability of receiving top triage priority (ESI 1 or 2) across the wait time of the next patient in line, conditional on queue length. The omitted reference category is no next patient waiting. Regressions include nurse-by-shift, hospital-hour, chief complaint, and arrival mode fixed effects, as well as controls for gender and a cubic in age. Standard errors are clustered by nurse. Vertical bars denote 95% confidence intervals.



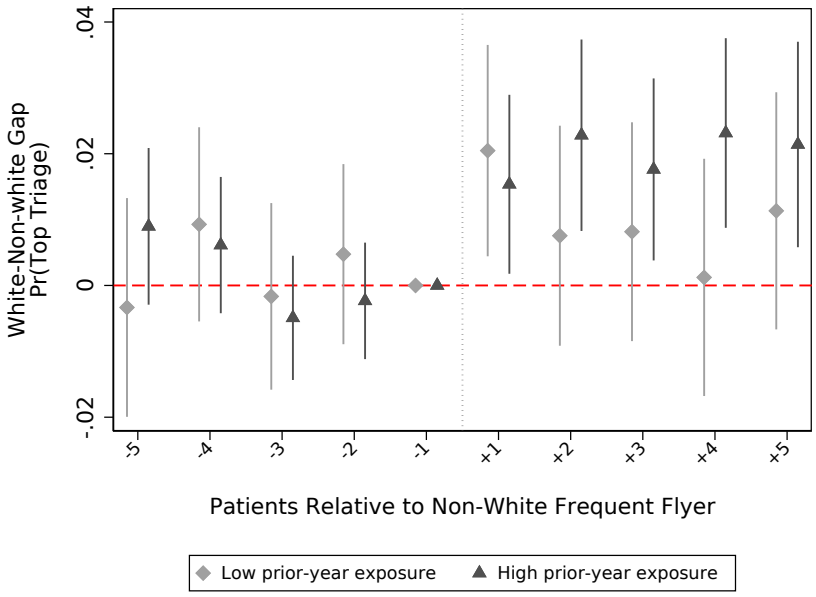
(a) Non-White Exposure × Next Wait Time



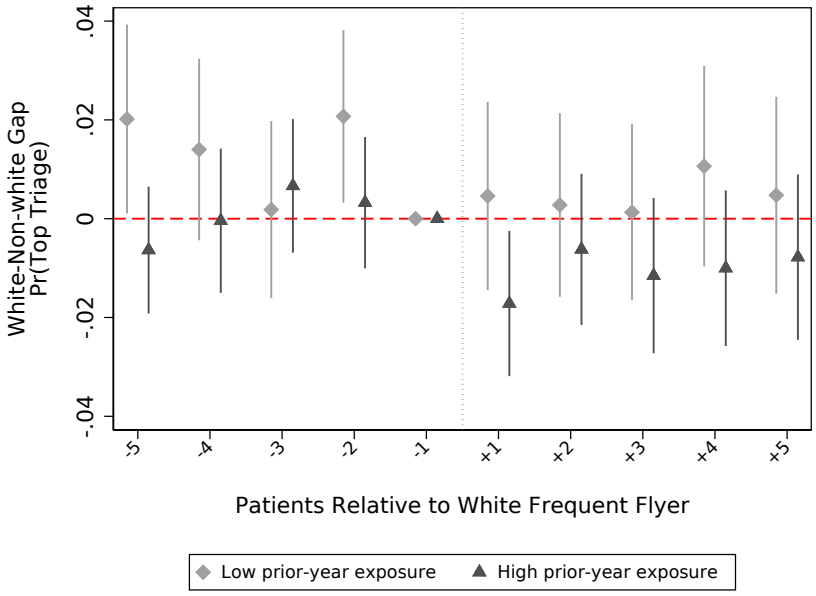
(b) White Exposure × Next Wait Time

Figure A.5: Time Pressure Increases Reliance on Prior Exposure (Robustness)

Note: Figures plot the predicted White–non-White gap in top triage priority (ESI 1 or 2) across next-patient wait time, separately by quartile of nurses' prior non-White frequent flyer exposure (Panel A) and White frequent flyer exposure (Panel B), conditional on queue length. The sample excludes frequent-flyer patients, defined as patients with three or more ED visits in the prior twelve months presenting with lower-risk chief complaints. Exposure is measured as the cumulative number of frequent-flyer encounters of each race over the prior twelve months. Regressions include nurse, hospital-year-month, hospital-hour, chief complaint, and arrival mode fixed effects, as well as gender and a cubic in age. Standard errors are clustered by nurse. Annotations report the change in the racial gap from the lowest to highest next-patient wait-time category within each exposure quartile. Vertical bars denote 95% confidence intervals.



(a) Non-White Frequent User Encounter



(b) White Frequent User Encounter

Figure A.6: Recent Exposure Effects by Longer-Run Frequent-User Exposure

Note: Figures plot changes in the White–non-White gap in top triage priority (ESI 1 or 2) around encounters with a non-White frequent user (Panel A) and a White frequent user (Panel B), separately by the nurse's accumulated exposure to frequent users of the focal race over the prior twelve months. Event time is measured by the focal patient's position relative to the frequent-user encounter, with the final pre-exposure patient omitted as the reference category. The sample is restricted to non-frequent-user patients. Estimates include nurse-by-shift, hospital-hour, chief complaint, and arrival-mode fixed effects, as well as patient controls for gender and a cubic in age. Standard errors are clustered by nurse. Frequent users are defined as patients with three or more ED visits in the prior twelve months who present with lower-risk chief complaints. Vertical bars denote 95% confidence intervals.

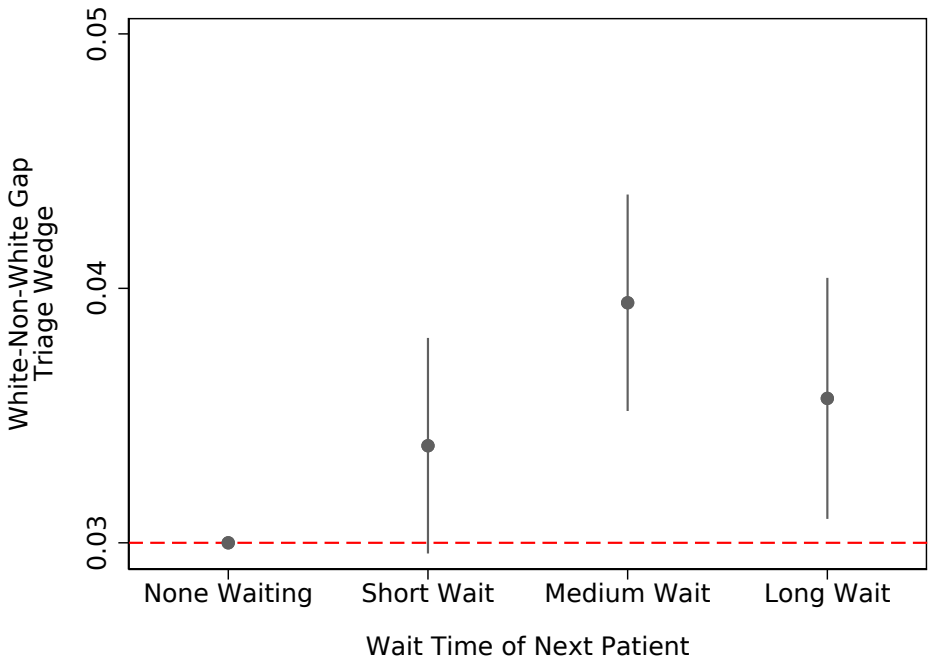


Figure A.7: Time Pressure and the Racial Triage Wedge: Next-Patient Wait Time

Note: Figure plots the predicted White–non-White difference in the triage wedge, defined as actual minus machine-learning-predicted top-triage assignment, across bins of the wait time of the next patient in line. Next-patient wait time is measured at the start of the current patient's triage assessment, conditional on the number of patients waiting. The omitted reference category is no next patient waiting. Estimates are from a linear probability model with nurse-by-shift, hospital-hour, visit number, chief complaint, and ambulance arrival fixed effects. Patient controls include a male indicator and a cubic in age. Standard errors are clustered at the nurse level. Vertical lines denote 95% confidence intervals. The difference between the medium- and long-wait estimates is not statistically significant.

B Machine Learning Benchmark

B.1 NYU ED Algorithm

I measure downstream clinical severity using the NYU ED Algorithm, a diagnosis-based classification system designed to characterize whether an emergency department visit required ED care. The algorithm was developed from a physician panel's classification of approximately 6,000 ED records into four categories:

1. *Non-emergent*: care was not required within 12 hours;
2. *Primary-care treatable*: care was required within 12 hours, but could safely have been provided outside the ED;
3. *Emergent, ED care needed, preventable*: ED care was required, but the condition was potentially preventable or avoidable; and
4. *Emergent, ED care needed, not preventable*: ED care was required and the condition was not preventable or avoidable.

The algorithm assigns each final discharge diagnosis a probability of belonging to each category rather than classifying diagnoses deterministically. For example, a diagnosis of chest pain may be assigned a 32% probability of being primary-care treatable and a 68% probability of requiring ED care.

For the machine-learning analysis, I define a visit as having a *severe diagnosis* when the combined probability assigned to the two categories requiring ED care exceeds 0.5. Thus, the severity outcome indicates that, based on the patient's final diagnosis, the condition was more likely than not to require emergency-department care. I use this binary measure as the prediction target for the clinical-risk benchmark described in Section 6.

B.2 Prediction Model and Training Sample

I use a random forest to predict the probability that a patient's final diagnosis is classified as severe according to the measure described in Appendix B.1. The prediction model uses only information recorded prior to the triage encounter. This restriction is important because information collected during triage, such as vital signs, screening responses, and clinical notes, may itself depend on the nurse's assessment and the decision context. Restricting the feature set to pre-triage information therefore prevents the benchmark from mechanically incorporating the potentially context-dependent behavior it is intended to evaluate. Patient race and ethnicity are excluded from the prediction model.

The predictors include chief complaint, arrival mode, insurance type, a cubic in age, sex, primary-care-provider status, and measures of healthcare utilization over the prior twelve months: primary-care visits, outpatient visits, ED visits, and hospital admissions. Categorical variables are represented by indicator variables, and missing values of continuous predictors are imputed using the median of the relevant training sample.

Because the prediction target is the patient's final diagnosis, a potential concern is that diagnosis severity may itself be affected by the initial triage decision. I therefore train the model on a restricted sample of encounters occurring under the most favorable decision context observed in the data: the nurse is in the lowest quartile of prior-year exposure to both non-White and White frequent flyers, and no patients are waiting to be triaged. As discussed in Section 4.3, this restriction does not eliminate the potential endogeneity of downstream diagnosis. Rather, it estimates the relationship between pre-triage characteristics and downstream severity in the portion of the data where the context effects documented in the paper are smallest.

I estimate the random forest using nested cross-validation. I divide the training sample into five outer folds. For each fold, I use the remaining four folds to select model hyperparameters through an inner three-fold cross-validation procedure, choosing the specification that maximizes the area under the receiver operating characteristic curve (AUC). I then refit the selected model on the full outer training sample and generate predicted probabilities for the held-out fold. Repeating this procedure across all five folds produces an out-of-fold predicted probability of severe diagnosis for every observation in the training sample, ensuring that no observation's prediction is generated by a model estimated using its own realized diagnosis.

To generate predictions for encounters outside the restricted training sample, I select the random-forest hyperparameters most commonly chosen across the five outer folds and re-estimate the model using the full training sample. I then use this model to predict diagnosis severity for all remaining encounters. Thus, predictions are out of sample with respect to the diagnosis outcome used for estimation: observations in the restricted training sample receive out-of-fold predictions, while all remaining observations are predicted using a model estimated exclusively on the restricted training sample.

B.3 Model Performance and Validation

I evaluate the predictive performance of the random forest using the out-of-fold predictions generated by the cross-validation procedure described above. The model achieves an out-of-sample area under the receiver operating characteristic curve (AUC) of 0.78. An AUC of 0.78 means that, for a randomly selected pair of patients in which one ultimately receives a

severe diagnosis and the other does not, the model assigns higher predicted risk to the severe patient 78% of the time. An AUC of 0.5 corresponds to random ranking, while an AUC of 1 indicates perfect discrimination. Thus, the model contains substantial information about patients' downstream clinical severity using only information available before triage.

To benchmark the predictive performance of the random forest, I compare it with two linear prediction models using the same patient characteristics: an OLS linear probability model and a LASSO model with cross-validated regularization. All three models use the same five outer cross-validation folds, so performance is evaluated on identical training and validation observations. The random forest achieves an AUC of 0.78, approximately 16 percentage points higher than either linear benchmark. This comparison indicates that the random forest provides a substantially more informative ranking of patients by downstream diagnosis severity.

I next assess how closely the resulting clinical-risk benchmark corresponds to nurses' actual triage assignments. As described in Section 4.3, within each nurse-shift I rank patients by predicted diagnosis severity and assign predicted top-triage status to the highest-risk patients until the benchmark assigns the same number of patients to top triage as the nurse actually did. Actual and benchmark top-triage assignments agree for 82% of patient encounters. Thus, despite relying only on pre-triage information and excluding race, the benchmark reproduces the large majority of nurses' triage classifications.

Importantly, the purpose of the benchmark is not to establish the machine-learning classification as the uniquely correct triage decision. Rather, it provides a common measure of predicted clinical risk against which to evaluate how nurses allocate a fixed number of top-triage assignments. Section 6 uses deviations from this benchmark to test whether racial disparities, and their response to decision context, extend beyond differences in predicted clinical risk.

C A Micro-foundation for Similarity-Based Reasoning

The conceptual framework in the main text is intentionally agnostic about the cognitive mechanism underlying similarity-based reasoning because the empirical patterns do not uniquely identify the process through which prior similar experiences affect current decisions. This appendix shows that the categorization model of Bordalo et al. (2026) provides one micro-foundation that generates the empirical predictions studied in the main text.

C.1 Prior Exposure and Categorization

Consider a simple distinction between routine, low-risk cases and high-risk cases. Suppose that, when evaluating a new patient, a nurse simplifies the problem by assessing which type of case the patient most resembles. A patient who more closely resembles a routine, low-risk case is less likely to receive top triage priority.

Which features shape this mapping depends on prior experience. Features that repeatedly characterize patients as low risk become cues that a new patient belongs to the same category. Frequent users provide a particularly relevant source of such experience: the same patients return repeatedly with lower-risk complaints, increasing nurses' exposure to these patients as examples of routine cases. Because frequent users are disproportionately non-White, repeated exposure to non-White frequent users makes non-White race more characteristic of the nurse's experience with low-risk cases. Exposure to White frequent users works in the opposite direction, making White race more characteristic of the low-risk category.

To fix ideas, consider two otherwise clinically identical patients, one White and one non-White, evaluated by a nurse whose prior experience with routine cases has disproportionately consisted of non-White frequent users. Although the patients present with the same clinical characteristics, the non-White patient resembles the nurse's prior low-risk cases along one additional dimension—race—and may therefore be more likely to be categorized as low risk.

Let E_g denote prior exposure to frequent users of race $g \in \{NW, W\}$, and let $A_g(E_g)$ capture how strongly race g is associated with the low-risk category after a given exposure. The categorization mechanism implies

$$A'_g(E_g) > 0. \tag{13}$$

Greater exposure to non-White frequent users should strengthen the association between non-White race and the low-risk category and widen the White–non-White gap in triage priority; conversely, greater exposure to White frequent users should strengthen the association between White race and the low-risk category and narrow the racial gap.

C.2 Recent Exposure and Accessibility

Past experiences used to categorize the current problem also differ in how easily they come to mind. In Bordalo et al. (2026), more accessible experiences play a greater role in determining which category a new problem is mapped to. Intuitively, when deciding whether the current patient resembles a routine, low-risk case, a case encountered moments ago is easier to retrieve than one encountered much earlier.

Let $r_g > 0$ capture the accessibility of the association between race g and the low-risk category, with r_g increasing when a frequent user of that race has been encountered more recently. Let $E = (E_{NW}, E_W)$ and $r = (r_{NW}, r_W)$. I summarize the relative categorical association between non-White rather than White race and the low-risk category as

$$C(E, r) = r_{NW} A_{NW}(E_{NW}) - r_W A_W(E_W). \tag{14}$$

A recent non-White frequent-user encounter therefore makes the non-White low-risk association more accessible, widening the racial gap; a recent White frequent-user encounter shifts categorization in the opposite direction.

Accessibility also distinguishes this mechanism from standard statistical discrimination (Phelps, 1972; Aigner and Cain, 1977). Under Bayesian learning, a prior encounter matters because of what it teaches the decision-maker about the underlying risk of similar patients. Its influence should therefore depend on how informative the encounter is and should diminish as prior experience makes beliefs more precise. Under categorization, by contrast, an encounter can matter because it changes which past experiences are easiest to retrieve when interpreting the current case, even if it provides little new information about underlying risk. Recent-exposure effects therefore need not attenuate among nurses with substantial prior experience or vary with the informativeness of the encounter. I test these contrasting predictions empirically in Section 5.6.

C.3 Time Pressure and Comparative Statics

The second dimension of local context is the time available to evaluate the current patient. When time is abundant, a nurse can use a rich set of patient-specific features to determine what kind of case the patient resembles. Suppose that time pressure reduces the set of features incorporated into this assessment, increasing the relative importance of readily observed features such as race.

Let T denote time pressure and let $\omega(T)$ capture how strongly race-based similarity influences the mapping of the current patient to a category. I capture the assumption that

race-based similarity receives greater weight under time pressure by imposing $\omega'(T) > 0$. Let $G^C(E, r, T)$ denote the component of the White–non-White gap in top triage priority generated by this categorization mechanism. I use the parsimonious representation

$$G^C(E, r, T) = \omega(T)C(E, r), \qquad \omega'(T) > 0. \tag{15}$$

Here, $C(E, r)$ captures the direction and strength of the race–low-risk association generated by prior and recent exposure, while $\omega(T)$ captures how strongly race-based similarity influences categorization under the current level of time pressure. Their product therefore captures the contribution of this race-based categorization channel to the racial gap.

Equation 15 gives each dimension of context a distinct role. Prior exposure, E_g, shapes the association between race and the low-risk category, $A_g(E_g)$. Recent exposure affects the accessibility of that association, r_g, which together determine the relative categorical association $C(E, r)$. Time pressure, T, makes the similarity-based mapping coarser, increasing $\omega(T)$ and therefore the influence of race-based similarity on categorization.

Exposure and Time Pressure Interact. The key comparative static concerns the interaction between exposure and time pressure. Equation 15 imposes a complementarity: in this formulation, time pressure does not determine which race is associated with the low-risk category but increases the influence of that association on categorization.

Holding accessibility fixed, the effect of non-White frequent-user exposure on the categorization-generated racial gap is

$$\frac{\partial G^C}{\partial E_{NW}} = \omega(T) r_{NW} A'_{NW}(E_{NW}) > 0. \tag{16}$$

Because $\omega'(T) > 0$, this effect becomes stronger as time pressure increases:

$$\frac{\partial^2 G^C}{\partial T \, \partial E_{NW}} = \omega'(T) r_{NW} A'_{NW}(E_{NW}) > 0. \tag{17}$$

The same logic applies to White frequent-user exposure, which enters the relative categorical association with the opposite sign:

$$\frac{\partial^2 G^C}{\partial T \, \partial E_W} = -\omega'(T) r_W A'_W(E_W) < 0. \tag{18}$$

Greater non-White frequent-user exposure strengthens the association between non-White race and the low-risk category; because race-based similarity matters more when categorization is coarser, the gap-widening effect of this exposure becomes larger under greater time pressure. Greater White frequent-user exposure strengthens the association between White race and

the low-risk category, so greater time pressure instead strengthens the gap-narrowing effect of White exposure. Thus, the categorization mechanism does not require time pressure to have a particular effect on racial disparities on average. Its sharper prediction is that the effect of time pressure depends on prior exposure. The framework predicts the direction of these interactions, but not that they are equal in magnitude.

These predictions map directly to the empirical analyses: E_{NW} and E_W correspond to the longer-run exposure measures, changes in r_g to the recent-exposure event studies, and Equations 17 and 18 to the race-by-exposure-by-time-pressure interactions estimated in Section 4.2.