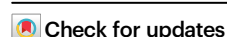


Ethnoracial disparities in emergency department triage are more pronounced in lower-risk visits: a multicenter study

Received: 2 February 2026

Accepted: 28 September 2026

Published online: 21 October 2026



Blanca Romero Milà^{1,2,3,10}, Helena Coggan^{1,2,10}, Andrew M. Fine^{2,4}, Yuval Barak-Corren^{1,2,5}, Ben Y. Reis^{1,2}, Jaya Aysola^{6,7,8}, Pradip P. Chaudhari⁹ & William G. La Cava^{1,2} 

Ethnoracial disparities in emergency department triage decisions persist, despite efforts to standardize protocols. Here, we analyze 1.35 million encounters at four U.S. emergency departments that all use the Emergency Severity Index triage protocol. We independently score visits as emergent (level 2) or urgent (level 3) by applying the criteria computationally and compare to nurse assignments. We show that, matched on clinical presentation and hospital context, Hispanic patients are less often triaged emergent than Non-Hispanic White (NHW) patients at every site (odds ratios (ORs): 0.75 [95% CI 0.72–0.78] to 0.90 [0.86–0.94]), and Non-Hispanic Black patients at two (0.73 [0.71–0.74], 0.89 [0.86–0.92]). Disparities increase where the algorithm scores patients lower-risk (Hispanic-NHW ORs: 0.73 [0.69–0.77] to 0.88 [0.83–0.94]), and attenuate where it scores them higher-risk (0.83 [0.79–0.89] to 0.95 [0.90–1.00]), suggesting a disparate “benefit-of-the-doubt” pattern of clinical discretion. Standardized decision support may reduce these disparities, and we estimate a worst-case increase on ED lengths of stay to be 1.0–3.8% across sites.

Nearly two decades after the Institute of Medicine declared emergency departments (EDs) to be at the breaking point¹, overcrowding continues to strain U.S. EDs^{2–4}. There were 155.4 million visits to U.S. EDs in 2022, roughly 47 per 100 people⁵. To manage the flow of patients through EDs, all visits begin with nurse triage, a process which evaluates the urgency of each visit and shapes downstream care (e.g., admission, length of stay). Patients with high-risk conditions experience delays in accessing life-saving care when assigned an incorrectly low triage score⁶. Consequently, even small unwarranted variations in

triage assignment may affect the timeliness of a patient’s care, the length of their stay, and their mortality risk⁷.

To reduce such variation, healthcare systems have adopted standardized triage protocols intended to support objective, repeatable assignments. The most widely used is the Emergency Severity Index (ESI)^{8,9}, developed in 1998 and used in 94% of U.S. EDs as of 2019⁹. ESI is a five-level scale intended to reflect urgency and risk of deterioration (1- immediate, 2- emergent, 3- urgent, 4- semi-urgent, 5- non-urgent). ESI scores are assigned by a triage nurse according to a protocol (in the sense of a strict protocol, not a computational tool)

¹Computational Health Informatics Program, Boston Children’s Hospital, Boston, MA, USA. ²Harvard Medical School, Boston, MA, USA. ³Department of Biomedical Engineering, University of California, Irvine, Irvine, CA, USA. ⁴Division of Emergency Medicine, Boston Children’s Hospital, Boston, MA, USA. ⁵Department of Pediatric Cardiology, Schneider Children’s Medical Center, Affiliated to Tel Aviv University Faculty of Medical and Health Sciences, Petach Tikvah, Israel. ⁶Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, USA. ⁷Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁸Penn Medicine Center for Health Equity Advancement, Philadelphia, PA, USA. ⁹Division of Emergency and Transport Medicine, Children’s Hospital Los Angeles and Department of Pediatrics, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA. ¹⁰These authors contributed equally: Blanca Romero Milà, Helena Coggan. ✉e-mail: william.lacava@childrens.harvard.edu

which relies on clinical characteristics, including high-risk symptoms and danger-zone vital signs, and does not explicitly incorporate sex or race/ethnicity.

The adoption of the ESI score was intended to standardize triage assessments, but studies show frequent misapplication (59% accuracy¹⁰) and high inter-rater variability¹¹. Assignment practices also vary across sites, and the questions used to elicit and document patient status vary between triage nurses¹². Previous studies have found that a patient's risk of mis-triage is linked to their race and sex, and that Black and Hispanic patients may be assigned less urgent scores than White patients^{13–17}. These differences in triage assignment are likely to contribute to downstream racial disparities in ED care: Black and Hispanic patients are less likely to be imaged¹⁸, are more likely to be passed over in favor of patients who arrived later and/or with less urgent triage scores¹⁹, and wait longer when not seen promptly²⁰. The precise mechanisms of these racial disparities are not known, but previous studies have suggested that implicit bias is a contributing factor²¹ and that clinicians show greater social engagement with patients with whom they share racial identity^{22,23}. Eliminating disparities in ED triage assignments could ensure more comprehensive evaluations of minority patients, fairer allocation of resources, and narrower disparities in hospital admissions.

To address disparities in ESI scoring, it is critical to identify where bias may enter the triage decision pathway, particularly at the boundary separating high-acuity patients (ESI 1–2), who are understood to require urgent intervention, from lower-acuity patients (ESI 3–5), who can generally wait to be seen.

In this study, we sought to identify factors contributing to mistriage, and in particular to systematic racial disparities in triage score assignment. We defined mistriage as taking place whenever the triage score determined by retrospective application of the ESI algorithm—i.e., the score a patient should have received according to their clinical risk factors and vital signs—differed from the score that was actually assigned by the triage nurse. We hypothesized that ethnoracial disparities would be enriched in pathways that (1) carry less apparent clinical risk and (2) rely more heavily on subjective clinical judgment. We tested these hypotheses by implementing an automated ESI algorithm and generating handbook scores for approximately 1.35M encounters from four U.S. EDs over a 15-year period (Fig. 1). This enabled stratification of nurse-assigned triage by agreement with handbook criteria and finer separation of apparent high- versus low-risk visits. We then used causal inference methods to estimate the association between ethnoracial identity and the odds of receiving an emergency (ESI 2) nurse triage assignment, stratified by algorithmic ESI assessment.

Results

Study cohort and clinical setting

We analyzed 1,348,436 encounters at four EDs serving adults on the East (AE: 398,661 encounters) and West coasts (AW: 116,063 encounters) and children on the East (PE: 305,963 encounters) and West coasts (PW: 527,749 encounters). The cohort was diverse (30.7% Non-Hispanic White, 12.5% Non-Hispanic Black, 36.9% Hispanic, 5.1% Asian, and 10.5% Other); see Table 1.

We focused on encounters assigned ESI levels 2 and 3, which comprise the majority of visits and represent the critical boundary between emergent (ESI 2) and urgent (ESI 3) care (Fig. 1E). This boundary is also strongly associated with admission likelihood (Fig. 1E) and is therefore a setting in which clinician judgment can meaningfully shape downstream care.

Accordingly, the final cohort included encounters assigned ESI level 2 (AE: 33.2%, AW: 23.6%, PE: 24.3%, PW: 15.3%) or level 3 (AE: 53.9%, AW: 65.5%, PE: 46.8%, PW: 27.7%) at triage, totaling 342,232 (AE), 103,029 (AW), 217,320 (PE), and 226,864 (PW) encounters. Table 1

provides detailed cohort information for the four medical center EDs analyzed.

Ethnoracial disparities in nurse triage assignment

Comparison of nurse-assigned and algorithmic ESI scores showed substantial discordance (Fig. 1F), particularly for algorithmic ESI level 2, consistent with subjectivity at the level-2/3 boundary. To assess the validity of algorithmic assignments, we compared several patient outcomes conditional on nurse-assigned and algorithmic ESI scores. Our primary comparison was admission to the hospital (Fig. 1G). As admission decisions may also be affected by subconscious biases, we additionally compared the following post-disposition clinical outcomes: death in hospital, long (greater than median) hospital stay, 30-day bounceback visits post-discharge, and 30-day bounceback admissions post-discharge (Supplementary Fig. 2).

Encounters assigned nurse level 2 were less likely to be admitted when the algorithm assigned level 3, and conversely, encounters assigned nurse level 3 were more likely to be admitted when the algorithm assigned level 2 ($p < 1.885 \times 10^{-8}$, χ^2 test). Higher algorithm-assigned ESI scores were also associated with significantly higher rates of in-hospital mortality, longer hospital stays, and increased risk of bounceback visits or admissions within 30 days of the ED visit (Supplementary Fig. 2). These results support that the ESI handbook algorithm captures clinically meaningful severity information beyond nurse-assigned triage.

We estimated the odds of receiving a nurse-assigned ESI level 2, relative to propensity-matched Non-Hispanic White encounters, stratified by ethnoracial group and algorithmic triage level (Fig. 2). Across centers, Hispanic and Non-Hispanic Black patients generally had lower odds of nurse-assigned level 2 than matched Non-Hispanic White patients.

For Hispanic patients, odds were lower at all centers (OR, AW: 0.90 [0.86–0.94]; AE: 0.75 [0.72–0.78]; PW: 0.87 [0.85–0.88]; PE: 0.80 [0.78–0.82]). Similarly, Non-Hispanic Black patients had lower odds at two of four centers (AE: 0.73 [0.71–0.74]; PE: 0.89 [0.86–0.92]).

Disparities stratified by algorithmic severity

When stratified by algorithmic triage level, disparities were more pronounced among encounters assessed as lower risk, i.e., algorithmic level 3. In this setting, Hispanic patients had reduced odds of level-2 nurse triage compared to matched Non-Hispanic White patients at all centers (AW: 0.88 [0.83–0.94]; AE: 0.73 [0.69–0.77]; PW: 0.76 [0.74–0.78]; PE: 0.76 [0.73–0.79]). Non-Hispanic Black patients showed similar patterns at three of four centers (AW: 0.77 [0.67–0.89]; AE: 0.75 [0.73–0.77]; PE: 0.88 [0.83–0.93]).

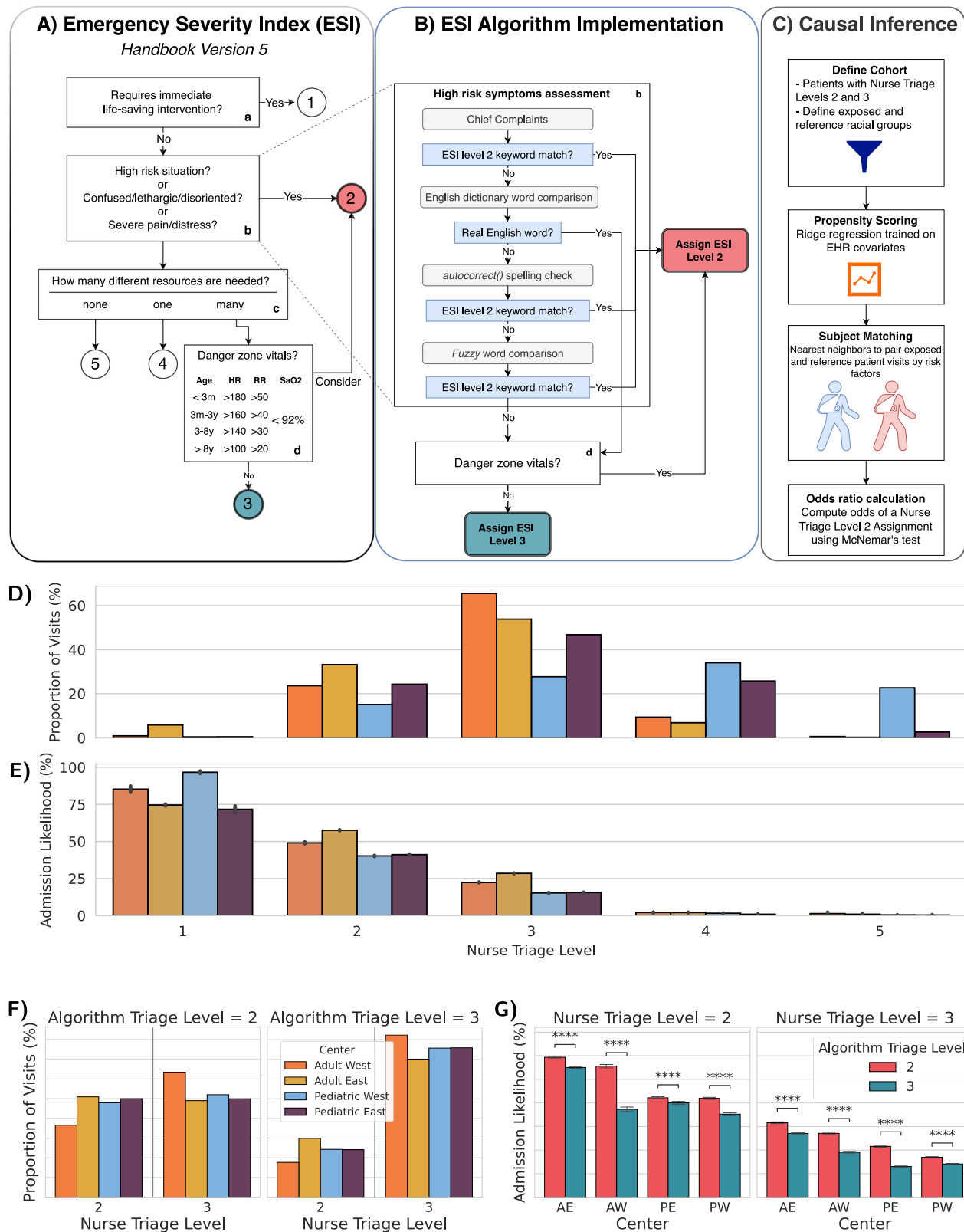
By contrast, among algorithmic level-2 encounters, disparities persisted for Hispanic patients at three centers but were generally attenuated (AW: 0.95 [0.90–1.00]; AE: 0.83 [0.79–0.89]; PW: 0.91 [0.90–0.93]; PE: 0.84 [0.81–0.88]). Non-Hispanic Black patients likewise showed reduced but persistent disparities at two centers (AE: 0.81 [0.78–0.84]; PE: 0.91 [0.86–0.96]).

Overall, patients assessed as lower risk by the algorithm (level 3) exhibited larger disparities than those assessed as higher risk (level 2) across multiple centers (Hispanic: AE, PW and PE $p < 0.01$; Non-Hispanic Black: AE $p < 0.01$, AW $p < 0.05$).

Site-specific variation in triage decision rules

The ESI algorithm contains two paths that arrive at triage level 2 (see Fig. 1). The first is through the presentation of high-risk symptoms. The second is through the presence of concerning (i.e., danger zone) vitals, in which case the handbook recommends reconsidering a triage level 2 assignment, even absent high-risk symptoms.

Both routes can be seen as relying on some degree of clinical judgment. High-risk symptoms are interpreted by the clinician from the patient or their guardian, which creates the potential for subjective



translation. Meanwhile, the ESI version 5 handbook explicitly instructs clinicians to use their judgment when considering whether to classify patients as high risk given their vitals (earlier ESI versions always recommend assigning triage 2). Thus, an open question we sought to answer was which route was more often used to assign triage level 2, and whether the routes exhibited markedly different disparities in triage assignment.

We found that there was a significant West coast/East coast divide in which paths led to nurse triage level 2 assignments (see Fig. 3). The adult and pediatric west coast sites were significantly more likely to assign triage 2 when patients had danger zone vitals (median [IQR] LR+, AW: 1.70 [1.62 - 1.78]; PW: 1.48 [1.46 - 1.50]), versus when patients had high risk symptoms (LR+, AW: 1.00 [0.92 - 1.06]; PW: 1.40 [1.17 - 1.48]). Conversely, adult and pediatric east coast sites were much more

Fig. 1 | Emergency Severity Index (ESI) algorithm implementation and validation. **A** Original ESI algorithm flowchart with triage levels 2 and 3 highlighted in red and green, respectively. **B** Computational implementation of ESI algorithm levels 2 and 3, including high-risk symptom detection (box b) and danger-zone vital sign detection (box d). Algorithm ESI levels 2 and 3 are highlighted in red and green. **C** Propensity score matching pipeline workflow. **D** Proportion of visits with nurse-assigned triage levels at each site. Bar heights represent the proportion of visits (%) assigned to each triage level. Levels 2 and 3 are most common overall. **E** Patients assigned high acuity (levels 1–2) are more likely to be admitted than those assigned levels 3–5. Bar heights show the mean likelihood of admission at the site and triage level, and error bars show 95% confidence intervals. Data for **(D, E)** include all encounters assigned ESI levels 1–5 across four centers (AW: 116,063; AE: 398,661; PW: 527,749; PE: 305,963 encounters). **F** Concordance between algorithmic and nurse-assigned triage at levels 2 and 3. Data include encounters assigned ESI levels 2 or 3 across four centers (AW: 103,029; AE: 342,232; PW: 226,864; PE: 217,320 encounters). Groups compared are encounters where the algorithmic and nurse-assigned scores agree versus disagree. Each data point represents an independent

patient encounter. Roughly half of algorithmic level-2 encounters are assigned nurse level 3, whereas algorithmic level-3 encounters more closely agree with nurse assignments. **G** Algorithmic ESI provides diagnostic information beyond nurse-assigned triage. Encounters assigned a more urgent algorithmic score than the nurse-assigned score are more likely to be admitted (right panel), and vice versa (left panel), across sites. Two-sided Pearson χ^2 tests with continuity correction compared admission proportions within each site and nurse-triage stratum. Exact P -values for nurse ESI 2 were AW, 4.693×10^{-183} ; AE, 1.680×10^{-55} ; PW, 1.781×10^{-74} ; and PE, 1.885×10^{-8} . For nurse ESI 3, they were AW, 1.022×10^{-146} ; AE, 7.795×10^{-99} ; PW, 2.865×10^{-46} ; and PE, $P < 2.2 \times 10^{-308}$ (numerical underflow). p -value annotation legend: ns: $0.05 < p \leq 1.0$; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$. Data include encounters assigned ESI levels 2 or 3 across the same four centers (AW: 103,029; AE: 342,232; PW: 226,864; PE: 217,320 encounters). Groups compared are encounters where the algorithmic score is more urgent than the nurse-assigned score, versus less urgent within each center. Each data point represents an independent patient encounter. Arm sling person icon adapted from drawio under Apache-2.0 license.

Table 1 | Descriptive statistics for the Emergency Departments used in the study

		Grouped by Center				
		Overall	Adult East	Adult West	Pediatric East	Pediatric West
Number of visits, n	–	1,348,678	398,661	116,063	305,963	527,991
Unique patients, n	–	670,493	196,421	69,402	164,847	239,823
Time frame	–	2011–2025	2011–2019	2020–2022	2019–2024	2018–2025
Age Group, n (%)	< 3 mo.	33,588 (2.5)	–	–	12,680 (4.1)	20,908 (4.0)
	3–6 mo.	20,977 (1.6)	–	–	8440 (2.8)	12,537 (2.4)
	6–12 mo.	54,355 (4.0)	–	–	18,118 (5.9)	36,237 (6.9)
	12–18 mo.	50,280 (3.7)	–	–	16,513 (5.4)	33,767 (6.4)
	1.5–3	111,234 (8.2)	–	–	37,174 (12.1)	74,060 (14.0)
	3–5	111,745 (8.3)	–	–	36,301 (11.9)	75,444 (14.3)
	5–10	189,972 (14.1)	–	–	62,118 (20.3)	127,854 (24.2)
	10–15	136,541 (10.1)	–	–	52,585 (17.2)	83,956 (15.9)
	15 +	125,262 (9.3)	–	–	62,034 (20.3)	63,228 (12.0)
	18–29	88,749 (6.6)	70,525 (17.7)	18,224 (15.7)	–	–
	30–39	68,095 (5.0)	50,214 (12.6)	17,881 (15.4)	–	–
	60–69	81,313 (6.0)	63,552 (15.9)	17,761 (15.3)	–	–
	70–79	63,636 (4.7)	48,618 (12.2)	15,018 (12.9)	–	–
	40–49	64,769 (4.8)	49,346 (12.4)	15,423 (13.3)	–	–
	50–59	84,787 (6.3)	67,493 (16.9)	17,294 (14.9)	–	–
Sex, n (%)	80 +	63,375 (4.7)	48,913 (12.3)	14,462 (12.5)	–	–
	F	672,010 (49.8)	216,552 (54.3)	63,042 (54.3)	146,306 (47.8)	246,110 (46.6)
	M	676,668 (50.2)	182,109 (45.7)	53,021 (45.7)	159,657 (52.2)	281,881 (53.4)
Race, n (%)	Asian	68,875 (5.1)	17,661 (4.4)	18,945 (16.3)	11,834 (3.9)	20,435 (3.9)
	Hispanic	498,314 (36.9)	32,458 (8.1)	31,943 (27.5)	80,026 (26.2)	353,887 (67.0)
	NH Black	169,062 (12.5)	86,921 (21.8)	7296 (6.3)	45,859 (15.0)	28,986 (5.5)
	NH White	414,570 (30.7)	232,435 (58.3)	43,101 (37.1)	105,812 (34.6)	33,222 (6.3)
	Other	141,390 (10.5)	23,622 (5.9)	13,839 (11.9)	30,702 (10.0)	73,227 (13.9)
	Unknown	56,467 (4.2)	5,564 (1.4)	939 (0.8)	31,730 (10.4)	18,234 (3.5)
Triage Acuity, n (%)	1.0	27,790 (2.1)	22,975 (5.8)	1029 (0.9)	1436 (0.5)	2350 (0.4)
	2.0	313,208 (23.3)	130,564 (33.2)	27,312 (23.6)	74,356 (24.3)	80,976 (15.3)
	3.0	576,469 (43.0)	211,668 (53.9)	75,717 (65.5)	142,964 (46.8)	146,120 (27.7)
	4.0	295,528 (22.0)	26,855 (6.8)	10,830 (9.4)	78,814 (25.8)	179,029 (33.9)
	5.0	128,921 (9.6)	955 (0.2)	668 (0.6)	8,059 (2.6)	119,239 (22.6)
Disposition, n (%)	Admitted	303,701 (22.5)	157,514 (39.5)	31,574 (27.2)	54,656 (17.9)	59,957 (11.4)
	Discharged	1,042,709 (77.3)	241,147 (60.5)	84,489 (72.8)	251,307 (82.1)	465,766 (88.2)
	Transfer	2268 (0.2)				2268 (0.4)

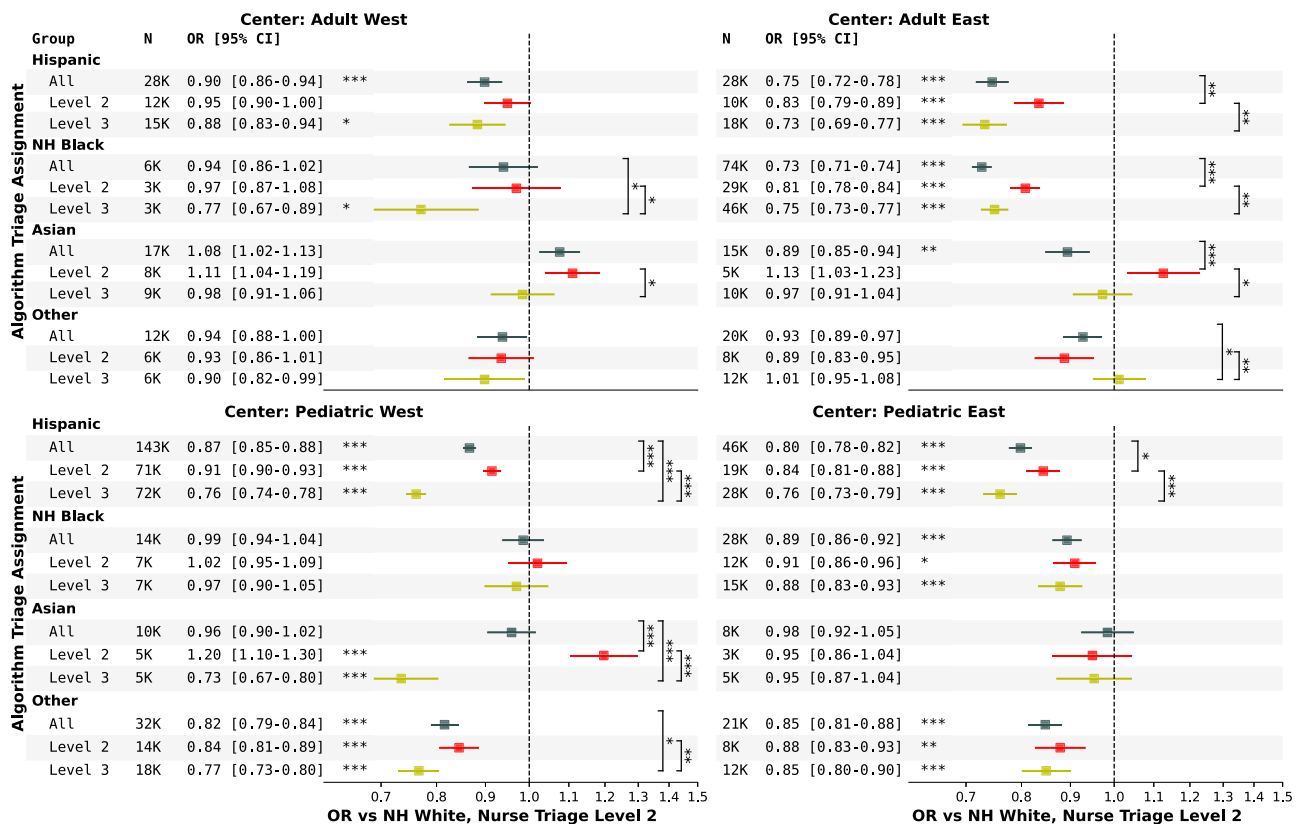


Fig. 2 | Ethnoracial disparities in emergency triage assignment by algorithmic ESI assessment. Points show propensity-score-matched odds ratios for nurse assignment to ESI level 2 (vs level 3) for each racial/ethnic group relative to non-Hispanic White patients within each center and handbook-ESI stratum. Error bars represent 95% Wald confidence intervals based on discordant matched pairs; the horizontal reference line denotes an odds ratio of 1. Each subplot represents a care center (AW, AE, PW, PE). The y-axis denotes Hispanic, Non-Hispanic Black, Asian, and Other groups, and further stratifies by algorithmic triage assignment (ESI 2 or ESI 3). Statistical significance of differences relative to Non-Hispanic White patients

is denoted by asterisks in the test column accompanying the ORs. In addition, within-group tests compare ORs across algorithmic triage levels and are denoted by vertical asterisks. Group-versus-reference tests used two-sided McNemar tests with continuity correction; their *P* values were Bonferroni-adjusted across 48 comparisons. Vertical comparisons used two-sided z-tests of log odds-ratio differences. Asterisks indicate the corresponding *P*-value ranges; the exact values are retained in the analysis outputs used to generate the figure. *p*-value annotation legend: ns: 0.05 < *p* <= 1.0; *: 0.01 < *p* <= 0.05; **: 0.001 < *p* <= 0.01; ***: 0.0001 < *p* <= 0.001.

likely to assign triage 2 for patients with high risk symptoms (LR +, AE: 1.68 [1.62 - 1.85]; PE: 1.65 [1.60 - 1.78]); on Eastern sites, danger zone vitals alone had a weaker effect on the likelihood of assignment, with essentially no effect at AE (LR +, AE: 1.10 [1.07 - 1.12]; PE: 1.41 [1.35 - 1.46]).

We observed small differences in likelihood ratios by patient ethnoracial group but, in general, they were consistent. High-risk complaint frequencies were also broadly similar across ethnoracial groups and centers (Supplementary Fig. 3).

In all sites, patients with both high-risk symptoms and danger zone vitals had the highest likelihood ratio of triage 2 assignments compared to patients with one or the other (median LR +: 1.79 - 2.14). This result suggests, in all sites, that the ESI algorithm is being used in a different manner than the one for which it was designed. By following the algorithm, a priori, we would expect danger zone vitals to only impact the likelihood of triage 2 assignments when patients do not express high-risk symptoms. Furthermore, we would expect high-risk symptoms, regardless of danger zone vitals, to independently increase the likelihood of triage 2 assignment. In clinical practice, we do not observe these independent effects. In AW, high-risk symptoms alone do not increase the likelihood of triage 2 assignments. Similarly, in AE, danger zone vitals alone barely increase the relative likelihood of triage 2 assignment; in PE, their effect is present but substantially weaker than that of high-risk symptoms. We do, however, observe an interactive effect between danger zone vitals and high-risk symptoms.

Patients meeting both criteria are assigned triage level 2 at higher rates than patients meeting only one, suggesting patient-reported symptoms and vitals are often considered in concert, rather than independently, when making triage decisions. In other words, all sites appear to consider both criteria for triage 2 assignment, but (a) consider them differently depending on the site and (b) often depart from the guidelines by requiring both criteria to be met, rather than one, before assigning a higher triage level. The results are noteworthy because all sites studied adhere to the same ESI triage protocol in their training programs, with minimal site-specific differences, discussed further in the Supplement.

These findings highlight meaningful site-specific differences in triage protocols and complicate efforts to apply a one-size-fits-all solution to the problem of triage assignment disparities.

Impact of algorithmic triage on emergency department flow

A key translational question is whether algorithmic ESI scores could serve as clinical decision support to reduce disparities while preserving patient flow. To probe this, we estimated the impact on ED length of stay (LOS) if algorithmic ESI scores were used in place of nurse-assigned triage scores. We used an accelerated failure time model to estimate the expected change in ED LOS under the replacement of nurse-assigned with algorithmic scores, evaluated at triage.

At sites AW, AE, PW, and PE, algorithmic ESI would change the proportion of level-2 assignments by 75.9%, 0.7%, 35.4%, and 15.6%, and

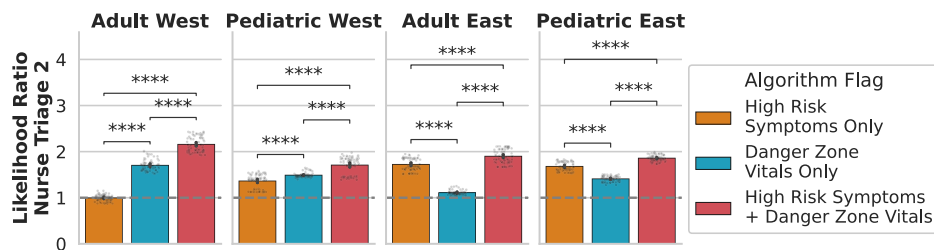


Fig. 3 | Variation in factors driving Nurse Triage Assignment among Centers.

The ESI triage algorithm relies first on high-risk symptoms, then danger zone vitals, to assign triage level 2. The Western and Eastern centers differ in the likelihood of a Nurse Triage 2 assignment based on these factors. In the adult and pediatric western centers, Nurse Triage Level 2 assignments are more likely when patients have danger zone vitals than when they have high-risk symptoms; in the Eastern centers, we observe the opposite. Asterisks denote two-sided Mann-Whitney-Wilcoxon tests comparing the 10 bootstrap replicates for each pair of flag types (three comparisons per site). Exact *P* values for high-risk symptoms only versus danger-zone vitals only, danger-zone vitals only versus both flags, and high-risk symptoms only versus both flags were, respectively: AW, 1.809×10^{-24} , 8.395×10^{-24} , and

1.809×10^{-24} ; PW, 3.563×10^{-7} , 3.122×10^{-7} , and 5.495×10^{-16} ; AE, 1.809×10^{-24} , 1.809×10^{-24} , and 1.497×10^{-8} ; PE, 1.809×10^{-24} , 1.809×10^{-24} , and 2.722×10^{-19} . *p*-value annotation legend: ns: $0.05 < p \leq 1.0$; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$. Data include encounters assigned ESI levels 2 or 3 across four centers (AW: 103,029; AE: 342,232; PW: 226,864; PE: 217,320 encounters). The bar estimates, and error bars summarize 10 bootstrap replicates; patient encounters are the underlying observations. Groups compared are encounters flagged for high-risk symptoms only, danger zone vitals only, or both, within each center. Bars and error bars represent the bootstrapped mean and 95% confidence interval of the likelihood ratio of Nurse Triage 2 assignment for each group.

change the proportion of level-3 assignments by -27.4% , -0.4% , -19.6% , and -8.1% , respectively. The estimated effect of replacing nurse-assigned scores with algorithmic scores was to increase mean ED LOS by 2.0% [2.0–2.1], 1.0% [1.0–1.1], 3.8% [3.7–3.8], and 1.5% [1.4–1.6] at sites AW, AE, PW, and PE, respectively.

In AW, where the time to admit/discharge decision is available, the estimated change in time to disposition was negligible (-0.1% [-0.1 – 0.0]). Given the high prevalence of level-3 encounters, a large relative increase in level-2 assignments can be partially offset by a small relative decrease in level-3 assignments. In practice, an ESI algorithm would act as a tool for decision support rather than a replacement for clinician judgment, and these results suggest potential to reduce disparities with modest or negligible impact on ED throughput.

Discussion

In summary, we observed significant ethnoracial disparities in triage assignment across four diverse emergency departments, with under-prioritization of minoritized patients relative to Non-Hispanic White patients across centers. These disparities persisted after adjustment for clinical, demographic, and hospital factors available at triage. Notably, disparities were most pronounced among encounters assessed as lower urgency by the ESI algorithm. These findings indicate that disparities are concentrated in discretionary triage decisions where clinicians may be more willing to (differentially) extend the benefit of the doubt, particularly for patients assessed as lower risk by standardized criteria.

Beyond documenting disparities, these findings have important implications for emergency care practice and quality improvement. Our results suggest that the ESI framework, while standardized in principle, is operationalized in ways that allow subjective judgment to differentially influence triage decisions at key decision boundaries.

Potential mechanisms

Several mechanisms could plausibly account for the observed disparities, and our analysis permits some discrimination among them. Language and communication barriers are an often-cited contributor to inequitable ED care, yet studies of triage specifically have found that non-English language preference does not consistently predict lower triage scores²⁴. In our analysis, disparities persisted in the pediatric cohorts even after adjustment for patients' preferred language, making language alone an insufficient explanation. Similarly, socio-economic factors (e.g., insurance status and social deprivation index) were available and adjusted for in the pediatric sites, and their inclusion did not remove disparities. Differential staffing or crowding at the

time of visit is another structural candidate. Although precise crowding measures were unavailable for all sites, we controlled for time of arrival and total patient volume at triage, where recorded, yet disparities remained. Crucially, our analysis also adjusted directly for the primary inputs to the triage decision—vital signs and chief complaint category—meaning that systematic differences in objective clinical severity presentation are unlikely to account for the residual disparity.

The concentration of disparities in algorithmic level-3 encounters, precisely those where ESI criteria are less decisive and clinical discretion is greatest, is most consistent with differential application of subjective judgment rather than any of the preceding structural or communication explanations. The remaining most plausible mechanism is implicit bias operating at moments of clinical ambiguity, in which the same presentation leads to meaningfully different acuity assignments depending on patient race or ethnicity. Provider–patient racial concordance, which has been associated with improved communication and trust^{25,26}, could not be assessed here due to the absence of provider demographic data, and may represent an additional contributing pathway. Prospective, qualitative studies that observe triage encounters in real time will be essential to isolate the specific interactional moments at which disparities emerge.

Limitations

As a multi-site observational study, our results share the general limitations of retrospective, cross-sectional analyses. Namely, causal inference approaches such as the ones employed here cannot control for unmeasured confounders, which would require randomized controlled trials to fully address. This limitation may impact our conclusions, though we note that racial disparities persisted across datasets even when the available visit information (i.e., controls for confounding) varied quite widely. For example, disparities observed in the adult sites were generally confirmed in the pediatric context, even though at these sites we were additionally able to adjust for social risk factors such as socio-economic deprivation and insurance. A noted limitation of the de-identification process in the public, adult datasets is the lack of precise temporal information at the population level, meaning we could not analyze whether disparities changed over time.

There were some variables which were not available in any dataset and which may have affected our results across sites. In particular, we note the absence of demographic information about providers, which prevented assessment of provider–patient racial concordance and its potential contribution to observed disparities. Prior studies have suggested that racial concordance may influence patient–physician communication, trust, and perceived quality of care^{22,25–27}. In addition,

as our analysis was performed on EHRs, we could not account for possible racial disparities in the completeness of documentation of a patient's chief complaint, possibly due to communication barriers or differences in the ways patients communicate pain or distress. If, for example, minoritized patients were less likely to have high-risk symptoms documented at triage, this could lead to underestimation of their clinical severity and contribute to algorithmic under-triage. This would require disparities to impact the rates at which high-risk symptoms, in particular, were noted in a patient's chart. This seems unlikely given their clinical relevance. If it were to occur, given the propensity matching process employed here, we would expect our results to underestimate the true degree of disparity, as patients with similar documented symptoms but different race/ethnicity are paired to compute effect sizes. In short, we cannot exclude the possibility that our results are affected by underlying differences in clinical severity which are not captured in the health record.

Hospital sites differed widely in their approaches to recording and categorizing race and ethnicity. In order to conduct a multi-site analysis, we were forced to coarse-grain race/ethnicity into a small number of categories; prior work suggests this may conceal disparities among more granular racial/ethnic categories²⁸. We note that not all patients recorded, for example, as Asian or Non-Hispanic Black are subject to the same degree of discrimination or structural biases, which may be affected by other factors like accent, national origin, or skin color. Qualitative and prospective studies may help in understanding the specific clinical subpopulations facing structural discrimination at triage. Finally, we note that all four sites under study were large academic medical centers in urban areas on the East and West Coasts of the United States, and that our results may not extend to other settings (for example, to rural hospitals or those whose populations are majority-Black).

All sites in our study use the ESI triage system, yet site-specific differences are apparent in how the ESI algorithm is applied in practice, as we discuss in the results. These results warrant further investigation into site-specific training or complementary operational workflows that may contribute to or impact how ESI is operationalized for patient care. As an example, sites may develop supplemental operational workflows layered onto standard ESI practice to address specific patient populations or clinical presentations, which may inadvertently impact how ESI criteria are applied in practice. By controlling for clinical presentation and hospital factors, our analysis may appropriately adjust for some, but not all, of these site-specific differences in triage practice.

Clinical implications

Our results suggest that racial disparities in emergency triage persist even after adjustment for preferred language, socio-economic deprivation, insurance, and diagnoses. Eliminating these disparities will require a two-pronged approach. Firstly, educational initiatives should aim to make triage nurses aware of these disparities. Experienced triage nurses may be inclined to believe that their own processes are fully objective, and may benefit from being shown quantitative evidence of their own susceptibility to subconscious biases. Qualitative analyses, such as interviews with patients and clinicians²⁹ and video documentation of triage encounters, may be helpful in isolating specific points in the nurse-patient interaction at which unwarranted variations in care arise. Secondly, algorithmic or rules-based decision support tools that explicitly encode ESI criteria could promote consistency in triage assignment, particularly for patients whose presentations fall near acuity thresholds. Such tools could be integrated into triage workflows as decision aids rather than replacements for clinical judgment, and may inform updates to triage training, audit processes, and quality assurance programs by identifying systematic deviations from guideline-based criteria. More broadly, our findings highlight the need for triage systems and clinical practice guidelines to

be evaluated not only for overall accuracy and efficiency, but also for their equity impacts in real-world implementation. Careful design, monitoring, and governance of decision support tools will be essential to ensure that efforts to standardize triage reduce—rather than entrench—existing disparities at the point of first contact in emergency departments.

Methods

This research was approved by the Institutional Review Boards (IRBs) at Boston Children's Hospital (IRB #P00048571) and Children's Hospital Los Angeles (IRB #CHLA-24-00151) under a waiver of informed consent. The waiver of informed consent was granted because it was assessed that (1) the proposed use of this data presents no more than minimal risk to the privacy of individuals; (2) the research could not practicably be conducted without the waiver of informed consent and authorization; (3) the research could not practicably be conducted without access to and use of protected health information with identifiers; and (4) waiving informed consent will not adversely affect the subject's rights or welfare. Participants were not compensated for inclusion. The MIMIC-IV-ED and MC-MED datasets were used in compliance with the data use agreements on Physionet. Data from Beth Israel Deaconess Medical Center (MIMIC-IV-ED) was reviewed by the Beth Israel Deaconess Medical Center IRB, who granted a waiver of informed consent and approved the data sharing initiative (IRB #2001P001699). Data collection from Stanford Medical Center was reviewed by the Stanford University Institutional Review Board and Secondary Data Use Committee. The MC-MED project was approved by the Stanford University Institutional Review Board (58581) with waiver of consent for retrospective research on de-identified, routinely collected data.

Here, we describe ESI algorithm development, the causal inference approach to estimating disparities in triage assignment, and data preparation for the study.

ESI Algorithm development

The four sites in our study implement the ESI triage system^{8,9}. Prior to 2023, all sites followed the ESI version 4 protocol; ESI version 5 was released in 2023 and subsequently adopted by the pediatric sites (PE and PW) the same year, within the study period. The differences between these versions are minor and discussed below.

High-risk patient identification. According to the ESI handbook⁹, level-2 criteria include conditions that may rapidly deteriorate or require time-sensitive treatment (Fig. 1A). To computationally replicate this, we extracted 104 high-risk keywords from the ESI handbook and matched them to ED chief complaints (Fig. 1B).

Danger zone vitals identification. When high-risk symptoms are absent but multiple resources are anticipated, ESI next considers vital signs (Fig. 1A). This assessment includes heart rate, respiratory rate, and oxygen saturation. If any vital sign exceeds the danger-zone threshold and multiple resources are anticipated, ESI recommends escalation to level 2⁹. To implement this component, we stratified patients into age groups and compared triage vital signs with age-specific danger-zone thresholds. If any of the three vital signs exceeded the threshold, the algorithm assigned level 2.

Version 5 ESI handbook introduces a consideration factor factor⁹, allowing clinician discretion when danger-zone vital signs are present, whereas earlier ESI versions did not⁸. In these cases, we defaulted to the ESI version 4 behavior, in which any danger-zone vital sign results in automatic assignment to level 2. We further decompose these assignments by site in Fig. 3. Notably, all sites studied here adhere to the ESI version 4 protocol for the majority of the study period (2011–2023). Both pediatric sites implemented version 5 in 2023.

Table 3 | A description of the categorization of variables in each of the four datasets

Variable	Adult West	Adult East	Pediatric West	Pediatric East
Sex	M, F	M, F	M, F	M, F (trans or NB gender identity separately recorded)
Race	Asian, HB, HW, H, NHB, NHW, Other, Unknown	Asian, H, NHB, NHW, Other, Unknown	Asian, H, HW, NHB, NHW, Other, Unknown	Asian, H, HW, NHB, NHW, Other, Unknown
Age	18–29, 30–39, ... 70–79, 80+	Same as AW	0–3m, 3–6m, 6–12m, 12–18m, 18–36m, 3–5y, 5–10y, 10–15y, 15y+	Same as PW
Insurance	Medicaid, Medicare, Other/Unknown	N/A	Public, Private, Unknown	Public, Private
Triage acuity (ESI score)	1, 2, 3, 4, 5, unknown			
Mode of arrival	EMS, Self, Other/Unknown			
Visit history	Number of admissions/discharges from the ED in the last 30 days			
Primary language	N/A	N/A	Armenian, English, Mandarin, Russian, Spanish, Other/Unknown	Arabic, Cape Verdean, English, Haitian Creole, Mandarin, Portuguese, Spanish, Other
Socio-economic deprivation	N/A	N/A	Q1, Q2, Q3, Q4, unknown	Q1, Q2, Q3, Q4, unknown
Miles traveled	N/A	N/A	Q1, Q2, Q3, Q4, unknown	Q1, Q2, Q3, Q4, unknown
State of origin	N/A	N/A	In-state, out-of-state, unknown	In-state, out-of-state, unknown
Chief complaint	Categorized per ³³	Categorized per ³³	Categorized per ³⁴	Categorized per ³⁴
Comorbidities	Based on CCI	Taken from common medications	Based on PCI	Based on PCI
Visit diagnosis	Based on CCI	Based on CCI	Based on PCI	Based on PCI
Weight	N/A	N/A	Number of st. devs. from age-group mean	Same as PW
Time of arrival	N/A	N/A	Weekend (yes/no), time of day at arrival (0–6h, 6–12h, ...), season (Dec-Feb, Mar-May...)	Same as PW
Crowding	N/A	N/A	Number of other patients waiting in the ED at the time of arrival	Same as PW

H Hispanic, NH Non-Hispanic, W White, B Black; CCI Charlson comorbidity index; PCI pediatric comorbidity index; st. dev. standard deviation. Q1, Q4, etc., refer to dataset-specific quartiles.

Statistical analysis

We estimated the association between patient race/ethnicity and the odds of receiving an emergency (ESI 2) nurse triage assignment. We measured the effect of race/ethnicity by matching encounters on clinical presentation and contextual factors before estimating adjusted odds ratios. The approach is described below.

Exposure of interest and outcome measure. The primary exposure was patient race/ethnicity, categorized as Hispanic, Non-Hispanic Black, Asian, and Other. Encounters with race/ethnicity recorded as Unknown were excluded (3.8% on average across datasets) due to uncertainty and potential misclassification. The primary outcome was assignment to nurse-assigned ESI level 2 at triage.

Adjustment variables. We adjusted for variables available at triage and, where available, diagnoses recorded during the visit to support clinically meaningful matching. Our underlying causal assumptions are described by the directed acyclic graph (DAG) provided in the Supplement (Supplementary Fig. 1). All datasets included sex, age, arrival mode, number of prior visits (with and without admission), comorbidity index, triage vital signs (heart rate, respiratory rate, oxygen saturation), and chief complaint. PE and PW additionally included preferred language, social deprivation index, miles traveled, state of origin, number of patients at arrival, and weight. Temporal variables (year and time of arrival) were available for PE and PW, and insurance information was available for PE, PW, and AW.

Odds ratio calculation

To assess disparities in ESI level assignment, we performed propensity score matching and estimated adjusted odds ratios³⁰, using Non-Hispanic White patients as the reference group because it was the largest group overall. For each comparison, analyses were restricted to encounters from the focal ethnoracial group and the reference group. Propensity scores were estimated using logistic regression with ridge regularization after standardizing the covariates described in the adjustment variables. Matching was performed using 1:1 nearest-neighbor matching without replacement based on the fitted propensity scores using a ball-tree algorithm³¹, with a default caliper of 0.2 on the propensity-score scale. This approach achieved covariate balance across groups, with absolute standardized mean differences ≤ 0.1 for all covariates³⁰.

Odds ratios were computed in matched samples using McNemar's test. *P*-values were adjusted with Bonferroni correction for multiple comparisons across the full set of 48 comparisons. This test assumes independence between matched pairs and conditional independence of outcomes within pairs given the matching variables, assumptions that are reasonable given the encounter-level analysis and strict matching criteria. Differences across algorithmic triage strata within an ethnoracial group were assessed using a two-sided *z*-test, which assumes approximate normality of the log-odds ratio estimates and independent variance estimates across strata. Given the large sample sizes within strata, these normal approximation and variance assumptions are expected to hold.

Table 2 | A description of the size of each dataset before and after the filtering process

Site	Number of unique visits		Number of unique patients	
	Before filtering	After filtering (% remaining)	Before filtering	After (% remaining)
Adult West	118,385	116,063 (98.0%)	70,545	69,402 (98.4%)
Adult East	425,087	398,661 (93.8%)	205,504	196,421 (95.6%)
Pediatric West	582,169	527,749 (90.7%)	255,907	239,726 (93.7%)
Pediatric East	352,324	305,963 (86.8%)	184,086	179,944 (97.7%)

Dataset preparation

Each dataset was filtered for quality control (see Table 2). Visit records were discarded if any of the following criteria were met: unknown sex (i.e., sex recorded as U or X); missing demographic information; missing patient identifier; disposition other than admission or discharge (e.g., death, left without being seen, left against medical advice), or missing disposition with no associated vitals (suggesting departure before triage); missing chief complaint or primary visit diagnosis for datasets where these were recorded separately (suggesting data loss rather than inability to report a complaint); implausibly long visit length in the context of the dataset (generally ≥ 2 weeks), suggesting a recording error; irreconcilable timestamps (e.g., departure time earlier than arrival time).

Variables were processed into categories as described in Table 3. We aimed for consistency across datasets but were constrained by site-specific recording practices. For example, PE recorded both sex and gender, PW and AE recorded only sex, and AW recorded only gender, which was not defined. In all cases, we assumed that sex reflected provider-reported sex. Race/ethnicity recording also differed; notably, AE defined Hispanic, White, and Black as mutually exclusive categories, whereas other sites recorded Hispanic ethnicity separately from race. We harmonized categories as Hispanic, Non-Hispanic Black, Non-Hispanic Asian (denoted Asian), Non-Hispanic White, Non-Hispanic Other (denoted Other), and Unknown.

Temporal variables (e.g., month, year and crowding) were unavailable for adult datasets, which obscured visit dates for deidentification. In AW, age was shifted by up to two years (all patients were ≥ 18 years, so this shift does not affect ESI algorithm assignments). Both adult datasets truncated ages above 90 at first visit; because such patients were rare, we coarsened older ages into an 80+ group.

Where available, socio-economic and geographic measures (e.g., miles traveled) were divided into dataset-specific quartiles. Socio-economic deprivation scores were obtained from the Robert Graham Center database (2019; most recent available)³². This resource assigns a score from 1 (least deprived) to 100 (most deprived) to each U.S. zip code. The distance traveled to the hospital was computed assuming travel from the patient's home address.

Chief complaints were grouped using adult and pediatric schemas from ref. 33 and ref. 34, respectively, with site-specific additions where needed to capture common reasons for visit (e.g., shingles for AW; see preprocessing code for details). Primary diagnoses were transformed into binary indicators corresponding to the Charlson Comorbidity Index categories for adult datasets (using the R package *comorbidity*) and the Pediatric Comorbidity Index³⁵ for pediatric datasets. Where available, prior diagnoses (comorbidities) were encoded similarly (e.g., prior diabetes, congestive heart failure, renal disease, cancer). For AE, where prior diagnoses were unavailable, we extracted therapeutic medication classes at the time of visit as proxies for chronic conditions, yielding 85 binary indicators (e.g., thyroid therapy).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data from Beth Israel Deaconess Medical Center and Stanford Medical Center used in this study are available from the MIMIC-IV-ED and MC-MED databases respectively (Adult East: add <https://physionet.org/content/mimic-iv-ed/2.2/>; Adult West: <https://physionet.org/content/mc-med/1.0.0/>). Patient data from Boston Children's Hospital and Children's Hospital Los Angeles are not available under the institutional data usage terms.

Code availability

The analysis code is publicly available in the ESI GitHub repository and archived on Zenodo³⁶.

References

1. Institute of Medicine of the national academies. *Hospital-Based Emergency Care: At the Breaking Point*. (National Academies Press, Washington, D.C., 2007).
2. Rabin, E. et al. Solutions to emergency department 'boarding' and crowding are underused and may need to be legislated. *Health Aff.* **31**, 1757–1766 (2012).
3. Bees, J. Emergency departments are overcrowded and understaffed. *NEJM Catalyst* **3**, <https://doi.org/10.1056/CAT.22.0011> (2022).
4. Hsia, R. Y. et al. Patterns in patient encounters and emergency Department Capacity in California, 2011–2021. *JAMA Netw. Open* **6**, e2319438 (2023).
5. Cairns, C. & Kang, k. *National Hospital Ambulatory Medical Care Survey: 2022 Emergency Department Summary Tables*. (2022).
6. Sax, D. R., Warton, E. M., Mark, D. G. & Reed, M. E. Emergency department triage accuracy and delays in care for high-risk conditions. *JAMA Netw. Open* **8**, e258498 (2025).
7. Sun, B. C. et al. Effect of emergency department crowding on outcomes of admitted patients. *Ann. Emerg. Med.* **61**, 605–611.e6 (2013).
8. Wuerz, R. C., Milne, L. W., Eitel, D. R., Travers, D. & Gilboy, N. Reliability and validity of a new five-level triage instrument. *Acad. Emerg. Med.* **7**, 236–242 <https://doi.org/10.1111/j.1553-2712.2000.tb01066.x> (2000).
9. Wolf, L., Ceci, K., McCallum, D. & Brecher, D. *Emergency Severity Index Handbook Fifth Edition* (Emergency Nurses Association, 2023).
10. Jordi, K. et al. Nurses' accuracy and self-perceived ability using the Emergency Severity Index triage tool: A cross-sectional study in four Swiss hospitals. *Scand. J. Trauma Resuscitation Emerg. Med.* **23**, 62 (2015).
11. Alpert, E. A., Lipsky, A. M., Hertz, D., Rieck, J. & Or, J. Simulated evaluation of two triage scales in an emergency department in Israel. *Eur. J. Emerg. Med.* **20**, 431 (2013).
12. Wolf, L., Delao, A., Jodelka, F. M. & Simon, C. Determining emergency severity index acuity: Key triage elements identified by emergency nurses. *J. Emerg. Nurs.* **51**, 472–479 (2025).
13. Schrader, C. D. & Lewis, L. M. Racial disparity in emergency department triage. *J. Emerg. Med.* **44**, 511–518 (2013).
14. Sax, D. R. et al. Evaluation of Version 4 of the Emergency Severity Index in US Emergency Departments for the Rate of Mistriage. *JAMA Netw. Open* **6**, e233404 (2023).

15. Zhang, X. et al. Trends of racial/ethnic differences in emergency department care outcomes among adults in the united states from 2005 to 2016. *Front. Med.* **7** <https://doi.org/10.3389/fmed.2020.00300> (2020).
16. Joseph, J. W. et al. Association of race and ethnicity with triage emergency severity index scores and total visit work relative value units for emergency department patients. *JAMA Netw. Open* **5**, e2231769–e2231769 <https://doi.org/10.1001/jamanetworkopen.2022.31769> (2022).
17. Patel, M. D. et al. Patient sex, racial and ethnic disparities in emergency department triage: A multi-site retrospective study. *Am. J. Emerg. Med.* **76**, 29–35 (2024).
18. Shan, A., Baumann, G. & Gholamrezanezhad, A. Patient race/ethnicity and diagnostic imaging utilization in the emergency department: a systematic review. *J. Am. Coll. Radiol.* **18**, 795–808 (2021).
19. Sangal, R. B. et al. Sociodemographic disparities in queue jumping for emergency department care. *JAMA Netw. Open* **6**, E2326338 (2023).
20. Lu, F. Q., Hanchate, A. D. & Paasche-Orlow, M. K. Racial/ethnic disparities in emergency department wait times in the united states, 2013–2017. *Am. J. Emerg. Med.* **47**, 138–144 (2021).
21. Guedj, R. et al. Explicit and implicit bias based on race, ethnicity, and weight among pediatric emergency physicians. *Acad. Emerg. Med.* **28**, 1073–1076 <https://doi.org/10.1111/acem.14301> (2021).
22. Aysola, J. et al. Understanding contributors to racial/ethnic disparities in emergency department throughput times: a sequential mixed methods analysis. *J. Gen. Intern. Med.* **37**, 341–350 (2022).
23. Lowe, C., Beach, M. C., Saha, S., Links, A. R. & Boss, E. F. Parent race and communication during elective pediatric surgery consultations. *JAMA Netw. Open* **8**, e2542758 (2025).
24. Metzger, P., Allum, L., Sullivan, E., Onchiri, F. & Jones, M. Racial and language disparities in pediatric emergency department triage. *Pediatr. Emerg. Care* **38**, e556–e562 (2022).
25. Saha, S., Komaromy, M., Koepsell, T. D. & Bindman, A. B. Patient-physician racial concordance and the perceived quality and use of health care. *Arch. Intern. Med.* **159**, 997 <https://doi.org/10.1001/archinte.159.9.997> (1999).
26. Meghani, S. H. et al. Patient-provider race-concordance: does it matter in improving minority patients' health outcomes? *Ethnicity Health* **14**, 107–130 <https://doi.org/10.1080/13557850802227031> (2009).
27. Shen, M. J. et al. The effects of race and racial concordance on patient-physician communication: a systematic review of the literature. *J. Racial Ethn. Health Disparities* **5**, 117–140 <https://doi.org/10.1007/s40615-017-0350-4> (2018).
28. Movva, R. et al. Coarse race data conceals disparities in clinical risk score performance. In *Proceedings of the 8th Machine Learning for Healthcare Conference*, 443–472 (PMLR, 2023).
29. Agarwal, A. K. et al. Perspectives of black patients on racism within emergency care. *JAMA Health Forum* **5**, e240046 (2024).
30. Austin, P. C. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar. Behav. Res.* **46**, 399–424 (2011).
31. Geldof, T., Popovic, D., Van Damme, N., Huys, I. & Van Dyck, W. Nearest neighbour propensity score matching and bootstrapping for estimating binary patient response in oncology: A monte carlo simulation. *Sci. Rep.* **10** <https://doi.org/10.1038/s41598-020-57799-w> (2020).
32. Robert Graham Center - Policy Studies in Family Medicine & Primary Care. Social deprivation index (SDI). <https://www.graham-center.org/maps-data-tools/social-deprivation-index.html>. (2018).
33. Arya, R. et al. Decreasing length of stay in the emergency department with a split emergency severity index 3 patient flow model. *Acad. Emerg. Med.* **20**, 1171–1179 (2013).
34. Gorelick, M. H., Alpern, E. R. & Alessandrini, E. A. A system for grouping presenting complaints: The pediatric emergency reason for visit clusters. *Acad. Emerg. Med.* **12**, 723–731 (2005).
35. Sun, J. W. et al. Development and validation of a pediatric comorbidity index. *Am. J. Epidemiol.* **190**, 918–927 (2021).
36. Romero Milà, B. et al. Ethnoracial disparities in emergency department triage are enriched in lower-risk visits: a multicenter observational study. ESI GitHub repository. <https://github.com/cavalab/ESI> (2026).

Author contributions

B.R.M. and H.C. contributed equally to this work, including code development, data preprocessing, analysis, and manuscript writing. A.F., Y.B., J.A., and P.C. provided data curation, clinical expertise, and manuscript review and editing. B.R. contributed to data analysis and manuscript review and editing. W.G.L. served as principal investigator, overseeing all aspects of the project, and contributing to code development, data preprocessing, analysis, and manuscript writing.

Funding

Research reported in this publication was supported by the National Library of Medicine of the National Institutes of Health (NIH) Award R01LM014300. The content is the responsibility of the authors and does not necessarily represent the official views of NIH.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-78526-5>.

Correspondence and requests for materials should be addressed to William G. La Cava.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.