


# Testimonial Injustice: Linguistic Bias in the Medical Records of Black Patients and Women



Mary Catherine Beach, MD, MPH<sup>1,2,3,4</sup>, Somnath Saha, MD, MPH<sup>2,5,6</sup> , Jenny Park<sup>2,7</sup>, Janiece Taylor, RN, PhD, FAAN<sup>8</sup>, Paul Drew, PhD<sup>9</sup>, Eve Plank<sup>10</sup>, Lisa A. Cooper, MD, MPH<sup>2,3,4</sup>, and Brant Chee, PhD<sup>11</sup>

<sup>1</sup>Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD, USA; <sup>2</sup>Division of General Internal Medicine, Department of Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA; <sup>3</sup>Department of Health, Behavior & Society, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; <sup>4</sup>Center for Health Equity, Johns Hopkins University, Baltimore, MD, USA; <sup>5</sup>Center to Improve Veteran Involvement in Care, VA Portland Health Care System, 3710 SW U.S. Veterans Hospital Rd. (P3HSRD), Portland, OR, USA; <sup>6</sup>Division of General Internal Medicine & Geriatrics, Oregon Health & Science University, Portland, OR, USA; <sup>7</sup>School of Medicine, Oregon Health & Science University, Portland, OR, USA; <sup>8</sup>School of Nursing, Johns Hopkins University, Baltimore, MD, USA; <sup>9</sup>Department of Language and Linguistic Science, University of York, York, UK; <sup>10</sup>Smith College, Northampton, MA, USA; <sup>11</sup>Applied Physics Laboratory, Johns Hopkins University, Baltimore, MD, USA.

**BACKGROUND:** Black Americans and women report feeling doubted or dismissed by health professionals.

**OBJECTIVE:** To identify linguistic mechanisms by which physicians communicate disbelief of patients in medical records and then to explore racial and gender differences in the use of such language.

**DESIGN:** Cross-sectional.

**SETTING/PARTICIPANTS:** All notes for patients seen in an academic ambulatory internal medicine practice in 2017.

**MAIN MEASURES:** A content analysis of 600 clinic notes revealed three linguistic features suggesting disbelief: (1) quotes (e.g., *had a “reaction” to the medication*); (2) specific “judgment words” that suggest doubt (e.g., “claims” or “insists”); and (3) evidentials, a sentence construction in which patients’ symptoms or experience is reported as hearsay. We used natural language processing to evaluate the prevalence of these features in the remaining notes and tested differences by race and gender, using mixed-effects regression to account for clustering of notes within patients and providers.

**KEY RESULTS:** Our sample included 9251 notes written by 165 physicians about 3374 unique patients. Most patients were identified as Black (74%) and female (58%). Notes written about Black patients had higher odds of containing at least one quote (OR 1.48, 95% CI 1.20–1.83) and at least one judgment word (OR 1.25, 95% CI 1.02–1.53), and used more evidentials ( $\beta$  0.32, 95% CI 0.17–0.47), compared to notes of White patients. Notes about female vs. male patients did not differ in terms of judgment words or evidentials but had a higher odds of containing at least one quote (OR 1.22, 95% CI 1.05–1.44).

**CONCLUSIONS:** Black patients may be subject to systematic bias in physicians’ perceptions of their credibility, a form of testimonial injustice. This is another potential

mechanism for racial disparities in healthcare quality that should be further investigated and addressed.

J Gen Intern Med 36(6):1708–14

DOI: 10.1007/s11606-021-06682-z

©This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2021

## INTRODUCTION

Prior studies indicate that clinicians may be more likely to dismiss, ignore, or downplay the concerns of Black and female patients, compared to White and male patients. For example, focus groups of African Americans overwhelmingly identified the importance of being taken seriously and believed by clinicians as a core component of respect, whereas this theme never arose in focus groups with White participants.<sup>1</sup> In a newspaper article on bias in healthcare, a Black man summarized his experiences as follows, “If there was ever a book on medical racism, it should probably just be called, ‘They Don’t Believe Us.’”<sup>2</sup> Similarly, studies that have examined the experiences of people with sickle cell disease, which disproportionately affects persons of African descent, have found that patients frequently report their pain being treated with suspicion and distrust, and they experience extreme frustration in attempting to convince health professionals of their distress.<sup>3</sup> In terms of gender, there are multiple accounts of women’s symptoms being first misunderstood as psychosomatic before they get needed treatment.<sup>4</sup> Black women, in particular, may be at higher risk of being disbelieved, which is thought to potentially contribute to racial disparities in maternal and infant mortality.<sup>5</sup>

The phenomenon of clinicians disbelieving certain patients may be a manifestation of unconscious biases and stereotypes of women and minorities as lacking credibility.<sup>6</sup> Studies have shown that race and gender biases are as prevalent in healthcare as in other settings.<sup>7, 8</sup> Because these biases are typically unconscious and subtle, their potential impact on clinical care can be difficult to detect. One place where biases may be detectable is in the medical record. Literature from the

---

*The findings of this study were presented in a plenary session of SGIM20, the virtual 2020 Annual Meeting of the Society of General Internal Medicine.*

---

Received January 12, 2021

Accepted February 23, 2021

Published online March 22, 2021

field of social psychology finds that attitudes can be reflected through people's language.<sup>9–12</sup> Thus, unconscious biases and stereotypes may reveal themselves in the language used to describe patients, including women and minorities, in clinical notes.

There have been few studies critically examining the language used in medical records to detect stigmatizing language. One set of studies demonstrated that physicians who read a vignette with the term “substance abuser,” as opposed to “having a substance use disorder,” agreed more that the person was personally culpable and should be punished and agreed less that the person needed treatment.<sup>13, 14</sup> A qualitative analysis of medical records of patients with acute pain from sickle cell disease found three forms of negative language about patients: language that perpetuated negative stereotypes, blamed patients for their symptoms, and cast doubt on patients' reports and experiences.<sup>15</sup> A randomized vignette study based on those findings found that physicians who read a note with stigmatizing versus neutral language about a hypothetical patient had more negative attitudes towards the patient and prescribed less pain medication.<sup>15</sup> This suggests that the use of stigmatizing language can adversely influence future healthcare quality by propagating bias.

To further explore the extent to which race and gender bias may manifest in medical records, we sought examples of language suggesting disbelief of patients and then explored racial and gender differences in the use of such language within medical records.

## METHODS

### Study Subjects and Setting

With approval from our institutional review board, we studied all physician notes written within the electronic medical record in 2017 about patients seen in an ambulatory internal medicine clinic at an academic medical center. Although the electronic record does contain some template text, encounter notes in this setting also contain free text written by each physician at the time of the encounter. Physician notes included notes written by both attending and resident physicians. Patients are identified in the medical record by race/ethnicity and by gender (binary designation as male or female). Because 88% of the sample identified as either White/Caucasian or Black/African American, we restricted the analysis to patients who were identified in the medical record as being from one of those two racial groups.

### Linguistic Features

To identify relevant language, we applied the linguistic concepts of epistemic modality and evidentiality to a content analysis of 600 randomly selected ambulatory internal medicine physician notes. Epistemic modality and evidentiality refer to “how speakers use a variety of linguistic resources to

express commitment to the propositional content of their assertions.”<sup>16</sup> By conveying their commitment, or otherwise, to what is being stated, speakers (and writers) further convey their endorsement or trust, or lack thereof, in the source of their knowledge, and hence the likely veracity or truth (or otherwise) of what is being stated.<sup>17, 18</sup> This analysis revealed the three linguistic features described below.

**Evidentials.** Evidentials are a grammatical element that indicates the source of one's knowledge.<sup>16, 19–21</sup> For example, a straight declarative statement (“it will rain later”) indicates certainty. As soon as the speaker adds an evidential (“I heard that it will rain later”), they attribute the information to some other source, declining to endorse its veracity. Physicians use evidentials often: for example, “the patient reports that the headache started yesterday.” These evidentials do not necessarily cast explicit doubt on the truthfulness of the information, but the choice to use them allows the speaker to be agnostic about whether the statement is true. Therefore, we hypothesized that the overall number of times the physician uses evidentials might reflect greater doubt of the patient's word. Further information about the process of identifying evidentials is included in the Appendix in the [Supplementary Information](#).

**Judgment Words.** Whereas an evidential indicates the source of information, a statement classified as a judgment evidential goes further to distance the physician from the information and question the credibility more directly. We identified a list of specific words that, when used in the medical record describing a patient's experience, convey a sense of doubt or negative judgment on the part of the physician. The list of judgment words includes *adamant* and *apparently* and various tenses of the verbs *claims*, *insists*, and *states*.

**Quotes.** Quotes are a complicated grammatical element. The original intent of quoting a source is to promote accuracy: by quoting the source directly, there should be nothing lost in its interpretation.<sup>22</sup> Indeed, quoting patients is encouraged in medical training to capture the patient's voice and presumably make the medical record more patient-centered.<sup>23, 24</sup> However, the use of quotes has evolved societally, such that they no longer simply convey that the words have been spoken but are often an indication that the words are to be doubted.<sup>25</sup> When physicians make the choice to write, “*the patient reports she had a ‘reaction’ to the medication,*” they may be trying to indicate that they do not necessarily believe that the reaction occurred or should be attributed to the medication. These are known in popular culture as scare quotes.<sup>25</sup>

### Analysis

We used natural language processing (NLP) methods to identify these three linguistic features—evidentials, judgment

words, and quotes—and to generate variables indicating the number of times each feature appeared in each note. This process involved the following steps. First, we prepared and processed all notes by using custom routines to normalize notes in terms of formatting (such as removal of extraneous spaces) and spaCy to perform sentence and word tokenization along with part of speech tagging and dependency parsing.<sup>26</sup>

In order to count evidentials, we used a custom implementation of Aho-Corasick algorithm with a Trie-based data structure implemented in FlashText<sup>27</sup> to search for each word physicians use to attribute information to the patient (complains of, denies, endorses, feels, says, reports, tells me), in simple present, simple past, and present participle tenses (e.g., endorses, endorsed, endorsing). To ensure validity of this process (i.e., that we were accurately counting evidential use), we abstracted 100 instances of each word usage and had 2 team members (physician-investigator MCB and student EP) independently code whether the use of the word represented an evidential. We eliminated all words which incorrectly identified evidentials >20% of the time (which resulted in elimination of all forms of “feels”). Most of the remaining words had accuracy rates of 94–100% except “notes” which had an accuracy rate of 84%. Intercoder reliability was 99–100% for all words. Further information about this process is included in the Appendix in the [Supplementary Information](#).

We then counted specific judgment words using a custom implementation of Aho-Corasick algorithm with a Trie-based data structure implemented in FlashText<sup>27</sup> enabling fast lookup of words across notes. Finally, we counted quotes using the regular expression:  $(\backslash B[\" ][\\" ]\backslash B)$ . We included both single quotes (') and double quotes (") as we found that both forms are used to quote patients within notes. The  $\backslash B$  term denotes word boundaries and is employed to prevent identifying apostrophes as single quotes (e.g., doesn't, wouldn't).

After abstracting counts of each of these three linguistic features, we used descriptive statistics to explore their distributions by race and gender. Because quotes and judgment words were less common and because we hypothesized that the presence of one such linguistic feature in a note was significant, we created binary variables to indicate whether that feature appeared at all in the note. Because evidentials were substantially more frequent, and because we hypothesized that it was the number of evidentials rather than the presence of one that might convey more doubt of the patient, we modeled evidentials as a continuous variable, representing the number of times that the evidential was used in each note.

We conducted unadjusted analyses examining differences by race and gender using simple linear (for evidentials) or logistic (for quotes and judgment words) regression models. Because there was often more than one note for each patient, and physicians wrote notes about many different patients, we subsequently examined race and gender differences using

mixed-effects regression models accounting for the two levels of clustering.

## RESULTS

### Study Sample

Our sample included 9251 notes written by 165 physicians about 3374 unique patients. Most (74%) of the patients were identified as Black and most (58%) as female. Table 1 displays the number of patients and notes by race and gender.

### Linguistic Differences by Race

Table 2 displays the linguistic features used per note by race and gender. All three linguistic features appeared more often in the medical records of Black compared to White patients. In unadjusted analyses, notes of Black patients had 1.52 (95% CI 1.33–1.70) more evidentials than White patients' notes. The odds of judgment word use were 1.56 (95% CI 1.38–1.75) times higher in Black patients' compared to White patients' notes, and the odds of quotes were 1.94 (95% CI 1.74–2.16) times higher.

In analyses accounting for clustering of notes within patients and of patients within physicians, all of the Black-White differences in linguistic features diminished but remained significant. Notes of Black patients had 0.32 (95% CI 0.17–0.47) more evidentials than White patients' notes. The odds of judgment word use were 1.25 (95% CI 1.02–1.53) times higher in Black patients' compared to White patients' notes, and the odds of quotes were 1.48 (95% CI 1.20–1.83) times higher. In post hoc analyses accounting for the two levels of clustering—notes by patient, and patient by physician—separately, we found that the changes in point estimates from unadjusted to adjusted analyses were entirely attributable to clustering of patients within physicians (data not shown).

### Linguistic Differences by Gender

The findings by gender were less pronounced and consistent (Table 2). We found no evidence of differences in the use of evidentials or judgment words in notes of male compared to female patients. We did find that quotes were more likely to be used in notes for women compared to men, in both unadjusted (OR 1.12, 95% CI 1.03–1.22) and adjusted (OR 1.22, 95% CI 1.05–1.44) analyses.

**Table 1 Patient Characteristics**

Race	Gender	N (%) patients	N (%) notes
Black	Male	1003 (29%)	2705 (29%)
	Female	1495 (44%)	4641 (50%)
White	Male	405 (12%)	917 (10%)
	Female	471 (14%)	988 (11%)
Total		3374	9251

**Table 2 Prevalence of Linguistic Features Used in Medical Records by Race and Gender**

Linguistic feature	Race		Gender		Unadjusted Black-White difference	Adjusted <sup>†</sup> Black-White difference	Unadjusted female-male difference	Adjusted <sup>†</sup> female-male difference
	White	Black	Male	Female				
	<b>Mean (SD)</b>	<b>Mean (SD)</b>	<b>Mean (SD)</b>	<b>Mean (SD)</b>	<b>β (95% CI)</b>		<b>β (95% CI)</b>	<b>p-value</b>
Evidentials	3.29 (3.66)	4.81 (3.74)	4.41 (3.89)	4.56 (3.70)	1.52*** (1.33, 1.70)	0.32*** (0.17, 0.47)	0.15 (-0.01, 0.31)	0.10 (-0.02, 0.22)
	<b>N (%)</b>	<b>N (%)</b>	<b>N (%)</b>	<b>N (%)</b>	<b>OR (95% CI)</b>		<b>OR (95% CI)</b>	
Judgment words	403 (21%)	2164 (29%)	1037 (29%)	1530 (27%)	1.56*** (1.38, 1.75)	1.25* (1.02, 1.53)	0.93 (0.85, 1.02)	0.96 (0.82, 1.11)
Quotes	572 (30%)	3337 (45%)	1470 (41%)	2439 (43%)	1.94*** (1.74, 2.16)	1.48*** (1.20, 1.83)	1.12** (1.03–1.22)	1.22* (1.05, 1.44)

\**p*<0.05

\*\**p*<0.01

\*\*\**p*<0.001

<sup>†</sup>Mixed-effects modeling to account for clustering of notes within patients and patients within clinicians

### Race-Gender Interactions

We found a statistically significant interaction (*p*=0.007) between race and gender for the use of evidentials. In stratified analyses (Table 3), we found that this interaction was largely explained by the greater use of evidentials in notes of White women compared to White men. In unadjusted analyses, there were 0.44 (95% CI 0.11–0.77) more evidentials in the notes of White women compared to White men, though after accounting for clustering, this difference was no longer observed (*β* 0.13, 95% CI -0.98–0.36). Because there tended to be more evidentials in the notes of White women vs. men, the racial difference in evidential use was greater when comparing Black and White men (unadjusted *β* 1.80, 95% CI 1.51, 2.08; adjusted *β* 0.42, 95% CI 0.17–0.67) than when comparing Black and White women (unadjusted *β* 1.28, 95% CI 1.03–1.53; adjusted *β* 0.32, 95% CI 0.12–0.51). We did not find significant

race by gender interactions for the use of judgment words or quotes.

### DISCUSSION

Our study found more markers of disbelief in the medical records of Black compared to White patients, suggesting that Black patients may be subject to systematic bias in physicians’ perceptions of their credibility, a form of testimonial injustice. This injustice, especially when persistent, can cause harm to persons whose self-knowledge is doubted, and may have adverse downstream effects such as undermining the patient’s ability to trust their clinicians and ultimately reducing healthcare quality and outcomes.

Testimonial injustice is one of a broader category of epistemic injustices, first described by philosopher Miranda Fricker, who defines testimonial injustice as that which occurs when a speaker receives an unfair deficit of credibility due to prejudice on the part of the hearer.<sup>28</sup> Many of the examples used by Fricker draw on interactions of Black Americans with law enforcement, where credibility bias contributes to substantial harms in the Black community, including mass incarceration, disproportionate use of the death penalty, and the murder of innocent people by police officers.<sup>28</sup> In healthcare settings, there are also very real harms that can occur when people are not believed, such as delayed diagnosis, inappropriate treatments, unnecessary pain and suffering, and even death.<sup>2, 5, 7, 8</sup>

In addition to the consequential harms that befall those who are not viewed as credible by law enforcement and health professionals, there are substantive harms, similar to the harms of microaggressions,<sup>29, 30</sup> to the persistent experience of being disbelieved in the first place. When a person is wrongfully discredited, they are dishonored as a human. It is not merely symbolic and not merely consequential—it is a core epistemic insult.<sup>28</sup>

There are two possible reasons for doubting a person’s credibility: concerns about competency (inability to interpret a situation and convey it accurately) and/or concerns about

**Table 3 Race by Gender Differences in the Use of Evidentials in Medical Records**

	White	Black	Unadjusted Black-White difference	Adjusted <sup>†</sup> Black-White difference
	Mean (SD)	Mean (SD)	β (95% CI)	β (95% CI)
Males	3.07 (3.59)	4.86 (3.89)	1.80 (1.51, 2.08)***	0.42 (0.17, 0.67)***
Females	3.51 (3.70)	4.78 (3.66)	1.28 (1.03, 1.53)***	0.32 (0.12, 0.51)**
Unadjusted female-male difference	0.44 (0.11, 0.77)**	-0.08 (-0.26, 0.10)		
Adjusted <sup>†</sup> female-male difference	0.13 (-0.98, 0.36)	0.08 (-0.06, 0.22)		

\**p*<0.05

\*\**p*<0.01

\*\*\**p*<0.001

<sup>†</sup>Mixed-effects modeling to account for clustering of notes within patients and patients within clinicians

sincerity (deliberate deception). In the setting of healthcare, either or both of these may be operating and could explain the race and gender differences we observed. Words such as “claims” are more direct expressions questioning sincerity, whereas quotes may be more suggestive of incompetency (e.g., perhaps implying irrationality). If so, then the differential use of quotes in the notes written about women may align with common gender prejudices.

It is worth noting that the linguistic features examined in this study may not be precise markers of testimonial injustice. In particular, the use of evidentials to describe patients’ experiences is not inherently disparaging and may be helpful for clinical reasoning. Because the use of an evidential does not explicitly indicate doubt, future work should explore whether racial differences in the use of evidentials represent true testimonial injustice. Judgment words, by definition, are more explicit markers of doubt, but the list of judgment words we focused on in this study may also insinuate that the patient is argumentative and perhaps represents a different kind of bias or stigma. Future research could employ experimental vignette designs to test the impact of these different linguistic features on the attitudes and decision-making of clinicians reading those notes.<sup>15</sup>

Decisions to use quotes also may have different motivations in different instances. Quoting a patient in their record is not necessarily wrong and may even have a benevolent intention, beneficial effect, or both. Quotes are sometimes used when the words spoken by the patient are not the typical words used to describe the phenomenon by clinicians, which may happen more commonly in situations of greater cultural distance between clinician and patient. On the other hand, we have found in our formative work that quotes serve several other negative functions, in addition to casting doubt on what the patient has said. For example, we have found that quotes that highlight irrational behavior (*reports that if she were to fall, she would just “lay there” until someone found her*), colloquial language (“*it busted open*”), or African American Vernacular English (*Chief Complaint: “I stay tired”*). Therefore, some of the racial differences in quoting patients may be due to other forms of stigma or bias in addition to credibility bias. Even when quotes are intended to be patient-centered, they carry the risk of being misinterpreted because of the increased use of scare quotes in society at large.

The fact that the use of the linguistic features we examined may, in some instances, be non-prejudicial (or a different sort of prejudicial) may raise concern as to whether our results reflect true testimonial injustice. However, the fact that we found racial and gender differences in the use of this language suggests that there is a pernicious influence behind their use. To the extent that there were instances of non-prejudicial use of these linguistic features, it would introduce misclassification bias into our study. However, the misclassification would be non-differential; there are no compelling reasons for systematically greater use of evidentials, judgment words, and quotes for Black and female patients, other than race or gender bias. Non-differential misclassification of outcomes generally

biases results towards the null,<sup>31</sup> which in our study would make it more difficult to detect race and gender differences. The fact that we found race and gender differences, despite this potential for misclassification, suggests that our findings represent a conservative estimate of testimonial injustice in medical records.

It is also important to consider that the use of these linguistic features is not likely, in most cases, to represent consciously prejudicial attitudes. More likely is that the use of these linguistic features, and the doubt they cast on patients’ testimonials, reflects unconscious race and gender bias. That is, the clinicians using more of these linguistic elements, especially evidentials, may not be doing so deliberately. It is more likely that they are doing so without realizing it. There has been much speculation about the role of clinicians’ implicit bias in contributing to racial and gender disparities in healthcare, but little evidence about the pathways by which implicit bias affects healthcare decisions and delivery. Our findings elucidate one potential pathway that may serve as a target for interventions to limit the negative impact of implicit bias.

Our analyses demonstrated that the racial differences we observed in the use of linguistic features potentially casting doubt on patients’ testimonials diminished, in many cases substantially, when we accounted for clustering of patients by physician. This change suggests that some physicians used those linguistic features in their notes—for all patients, Black and non-Black—more commonly than other physicians and that Black patients are more likely than White patients to see those physicians. This result could indicate that the differences in the use of doubt-casting language are less related to patient race and more related to the habits of physicians who happen to see more Black patients. That possibility, however, leads one to wonder why physicians seeing more Black patients are adopting these linguistic features more than other physicians. It seems plausible that the underlying cause may be racial bias leading to testimonial injustice and that the habit of using doubt-casting language then spills over into their notes for non-Black patients as well. Regardless of the underlying cause, the fact that this type of language is being used more commonly in Black patients’ notes indicates that they are systematically subjected to testimonial injustice. Adjustment did not substantially affect our findings for women, suggesting that those findings are not attributable to the habits of specific physicians.

There are potential methodological limitations to our study. First, this study was conducted in a single clinical setting from one academic medical center. We therefore do not know whether these findings are generalizable to other settings. Future studies should attempt to reproduce these findings. Second, our analysis did not account for socioeconomic status, as data on income and/or education are not readily available in medical records. Future studies could address this by obtaining zip code or insurance data as a proxy. Third, we did not have data on the demographic or training-level (attending vs. resident) characteristics of the physicians writing the notes we

analyzed. As such, we could not determine whether the use of the linguistic features we examined varied by physician characteristics such as race, gender, or training status. Fourth, we did not explicitly study the prevalence of text that might have been carried forward, meaning copied and pasted from previous notes, which may be important to explore in the future. Although the notes did contain text that was part of the note template (and therefore not written de novo by the physician), the template text did not contain any of the linguistic features that were the focus of our analysis, and therefore, we do not believe it would have affected our analysis. Finally, our population of patients nearly all identified as either Black/African American or White/Caucasian, and we were therefore not able to evaluate language for other racial/ethnic groups.

As a profession, physicians must recognize and neutralize the impact of racial prejudice in credibility assessments by seriously and constantly considering that we might make implicitly biased judgments that are unwarranted. When we have attempted to account for this bias and still have doubts about patient credibility, we must consider respectful ways to document that doubt. Using scare quotes or judgmental language may unfairly put the patient at risk for lower-quality care and disrespect from future providers. Further research is needed to explore the phenomenon of testimonial injustice and other forms of stigmatizing language in patient medical records,<sup>32</sup> and interventions should be developed to reduce their impact.

---

**Corresponding Author:** Somnath Saha, MD, MPH; Center to Improve Veteran Involvement in Care, VA Portland Health Care System, 3710 SW U.S. Veterans Hospital Rd. (P3HSRD), Portland, OR 97239, USA (e-mail: sahas@ohsu.edu).

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11606-021-06682-z>.

**Funding** This work was supported by a grant from the Robert Wood Johnson Foundation's Building Trust and Mutual Respect to Improve Healthcare Program. Dr. Beach was also supported by the National Institute on Drug Abuse (K24 DA037804). Dr. Saha was supported by the Department of Veterans Affairs. Dr. Taylor was supported by the Robert Wood Johnson Harold Amos Medical Faculty Program.

#### Declarations:

**Disclaimer:** The Robert Wood Johnson Foundation had no role in the study's design, conduct, or reporting. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Robert Wood Johnson Foundation, the Department of Veterans Affairs, or the United States government.

## REFERENCES

1. Beach MC, Branyon E, Saha S. Diverse patient perspectives on respect in healthcare: A qualitative study. *Patient Educ Couns*. 2017;100(11):2076-2080. <https://doi.org/10.1016/j.pec.2017.05.010>
2. Schencker L. 'Nine times out of 10, I was completely brushed off': Black Chicagoans confront bias in health care, hope for change. *Chicago Tribune*. August 13, 2020. [http://www.chicagotribune.com/news/breaking/ct-chicago-health-care-racism-george-floyd-implicit-bias%2D%2D20200813-nxrjltiyvg37pks2hyf6iw7m-story.html?utm\\_medium=notification&utm\\_source=onesignal](http://www.chicagotribune.com/news/breaking/ct-chicago-health-care-racism-george-floyd-implicit-bias%2D%2D20200813-nxrjltiyvg37pks2hyf6iw7m-story.html?utm_medium=notification&utm_source=onesignal)
3. Haywood C, Beach MC, Lanzkron S, et al. A systematic review of barriers and interventions to improve appropriate use of therapies for sickle cell disease. *J Natl Med Assoc*. 2009;101(10):1022-1033.
4. Macphee M. Hysterical women: Pain bias and epistemic injustice in medicine. *Granite J Postgrad Interdiscip J*. 2019;3(2):12-19.
5. Villarosa L. Why America's Black mothers and babies are in a life-or-death crisis. *The New York Times*. April 11, 2018. <https://www.nytimes.com/2018/04/11/magazine/black-mothers-babies-death-maternal-mortality.html>
6. Beeghly E, Madva A, eds. *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*. New York: Routledge; 2020.
7. Mossey JM. Defining racial and ethnic disparities in pain management. *Clin Orthop Relat Res*. 2011;469(7):1859-1870. <https://doi.org/10.1007/s11999-011-1770-9>
8. Santry HP, Wren SM. The role of unconscious bias in surgical safety and outcomes. *Surg Clin North Am*. 2012;92(1):137-151. <https://doi.org/10.1016/j.suc.2011.11.006>
9. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language use: our words, our selves. *Annu Rev Psychol*. 2003;54:547-577. <https://doi.org/10.1146/annurev.psych.54.101601.145041>
10. Berry DS, Pennebaker JW, Mueller JS, Hiller WS. Linguistic bases of social perception. *Pers Soc Psychol Bull*. 1997;23(5):526-537. <https://doi.org/10.1177/0146167297235008>
11. Lindquist KA, MacCormack JK, Shablack H. The role of language in emotion: Predictions from psychological constructionism. *Front Psychol*. 2015;6. <https://doi.org/10.3389/fpsyg.2015.00444>
12. Beukeboom CJ, Burgers C. Linguistic bias. In: *Oxford Research Encyclopedia of Communication*. Oxford, UK: Oxford University Press; 2017. <https://doi.org/10.1093/acrefore/9780190228613.013.439>
13. Kelly JF, Westerhoff CM. Does it matter how we refer to individuals with substance-related conditions? A randomized study of two commonly used terms. *Int J Drug Policy*. 2010;21(3):202-207. <https://doi.org/10.1016/j.drugpo.2009.10.010>
14. Kelly JF, Dow SJ, Westerhoff C. Does our choice of substance-related terms influence perceptions of treatment need? An empirical investigation with two commonly used terms. *J Drug Issues*. 2010;40(4):805-818. <https://doi.org/10.1177/002204261004000403>
15. Goddu AP, O'Connor KJ, Lanzkron S, et al. Do words matter? Stigmatizing language and the transmission of bias in the medical record. *J Gen Intern Med*. 2018;33(5):685-691. <https://doi.org/10.1007/s11606-017-4289-2>
16. Sidnell J, Stivers T, eds. *The Handbook of Conversation Analysis*. Chichester, UK: Wiley-Blackwell; 2013.
17. Pomerantz AM. Giving a source or basis: The practice in conversation of telling 'how I know'. *J Pragmat*. 1984;8(5-6):607-625. [https://doi.org/10.1016/0378-2166\(84\)90002-X](https://doi.org/10.1016/0378-2166(84)90002-X)
18. Pomerantz A. Constructing skepticism: Four devices used to engender the audience's skepticism. *Res Lang Soc Interact*. 1988;22(1-4):293-313. <https://doi.org/10.1080/08351818809389307>
19. Chafe WL, Nichols J, eds. *Eventuality: The Linguistic Coding of Epistemology*. New York: Ablex Pub. Corp; 1986.
20. Heritage J. The epistemic engine: Sequence organization and territories of knowledge. *Res Lang Soc Interact*. 2012;45(1):30-52. <https://doi.org/10.1080/08351813.2012.646685>
21. Drew P. Epistemics in social interaction. *Discourse Stud*. 2018;20(1):163-187. <https://doi.org/10.1177/1461445617734347>
22. Finnegan RH. *Why Do We Quote? The Culture and History of Quotation*. Cambridge, UK: Open Book Publishers; 2011.
23. JW Ball. Chapter 1: The history and interviewing process. In: *Seidel's Guide to Physical Examination*. St. Louis, MO: Elsevier; 2015.
24. LS Bickley. Chapter 1: Overview: Physical examination and history taking. In: *Bates' Guide to Physical Examination and History Taking*. Philadelphia, PA: Wolters Kluwer Health; 2013.
25. Chait J. Scared yet? *The New Republic*. December 30, 2008. <https://newrepublic.com/article/64702/scared-yet>
26. Honnibal M, Montani I. spacy 2: Natural language understanding with bloom embeddings. *Convolutional Neural Netw Increm Parsing*. 2017;(7.1).

27. Singh V. Replace or retrieve keywords in documents at scale. ArXiv171100046 Cs. Published online November 9, 2017. <http://arxiv.org/abs/1711.00046>
28. Fricker M. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford, UK: Oxford University Press; 2009.
29. Miller LR, Peck BM. A prospective examination of racial microaggressions in the medical encounter. *J Racial Ethn Health Disparities*. 2020;7(3):519-527. <https://doi.org/10.1007/s40615-019-00680-y>
30. Hall JM, Fields B. "It's killing us!" Narratives of Black adults about microaggression experiences and related health stress. *Glob Qual Nurs Res*. 2015;2:2333393615591569. <https://doi.org/10.1177/2333393615591569>
31. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Third edition. Philadelphia, PA: Wolters Kluwer Health, Lippincott Williams & Wilkins; 2008.
32. Blair IV, Steiner JF, Fairclough DL, et al. Clinicians' implicit ethnic/racial bias and perceptions of care among Black and Latino patients. *Ann Fam Med*. 2013;11(1):43-52. <https://doi.org/10.1370/afm.1442>

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