

*Diversity, Equity, Inclusion, and Justice*

# Race and Gender Bias in Internal Medicine Program Director Letters of Recommendation

Neil Zhang, MD, MS  
 Sarah Blissett, MD, MHPE  
 David Anderson, MD  
 Patricia O'Sullivan, EdD  
 Atif Qasim, MD, MSCE

## ABSTRACT

**Background** While program director (PD) letters of recommendation (LOR) are subject to bias, especially against those underrepresented in medicine, these letters are one of the most important factors in fellowship selection. Bias manifests in LOR in a number of ways, including biased use of agentic and communal terms, doubt raising language, and description of career trajectory. To reduce bias, specialty organizations have recommended standardized PD LOR.

**Objective** This study examined PD LOR for applicants to a cardiology fellowship program to determine the mechanism of how bias is expressed and whether the 2017 Alliance for Academic Internal Medicine (AAIM) guidelines reduce bias.

**Methods** Fifty-six LOR from applicants selected to interview at a cardiology fellowship during the 2019 and 2020 application cycles were selected using convenience sampling. LOR for underrepresented (Black, Latinx, women) and non-underrepresented applicants were analyzed using directed qualitative content analysis. Two coders used an iteratively refined codebook to code the transcripts. Data were analyzed using outputs from these codes, analytical memos were maintained, and themes summarized.

**Results** With AAIM guidelines, there appeared to be reduced use of communal language for underrepresented applicants, which may represent less bias. However, in both LOR adherent and not adherent to the guidelines, underrepresented applicants were still more likely to be described using communal language, doubt raising language, and career trajectory bias.

**Conclusions** PDs used language in a biased way to describe underrepresented applicants in LOR. The AAIM guidelines reduced but did not eliminate this bias. We provide recommendations to PDs and the AAIM on how to continue to work to reduce this bias.

## Introduction

Despite the well-documented bias in program director (PD) letters of recommendation (LOR) against people who are underrepresented in medicine,<sup>1-4</sup> these letters persist as an important factor in fellowship selection.<sup>5-7</sup> The Alliance for Academic Internal Medicine (AAIM) put forth guidelines in 2017 to standardize the PD LOR to decrease bias and increase quality, but it remains unclear whether these standardizations reduce bias (provided as online supplementary data).<sup>8</sup> In-depth analyses could determine how and where the language of bias appears in LOR and whether standardization mitigates this bias.

The presence of bias through language in LOR prior to the AAIM guidelines is well documented. In medical student LOR, descriptive words differ based on race

and gender. Men and White applicants are more likely to be described as “exceptional” or “leaders,” Black applicants as “competent,” and women as “empathetic” or “compassionate.”<sup>9-11</sup> These patterns persist when applicants apply to fellowship.<sup>12</sup>

Bias can manifest in many ways in LOR. First, authors use *agentic* and *communal* terms differently to describe applicants. These describe 2 interconnected fundamental qualities of human existence. Agency reflects concerns about meeting one's own needs (eg, behaviors of leadership and confidence), while communalism reflects concerns with interpersonal issues (eg, behaviors of empathy and interpersonal skills).<sup>13</sup> In LOR, men and White applicants tend to be described using agentic terms, while women and people of color tend to be described using communal terms. Communal terms used in LOR negatively affect hiring in academia, despite controlling for objective measures of productivity and performance.<sup>14,15</sup> Second, bias manifests through the use of *doubt raising* language to describe underrepresented applicants. Examples of doubt raising include

DOI: <http://dx.doi.org/10.4300/JGME-D-20-00929.1>

*Editor's Note: The online version of this article contains the 2017 AAIM guidelines for standardized LOR and the codes for LOR-NonAAIM and LOR-AAIM.*

negative language, hedges (eg, he *appears* to be motivated), and faint praise (eg, she is better than *average*).<sup>13</sup> Third, bias can manifest through *career trajectory bias*, where non-underrepresented groups are described as researchers or professionals, while underrepresented groups are described as students.<sup>13</sup> These patterns reflect current societal racial and gender stereotypes as well as a long history of highly prevalent bias against women and people of color entering the health professions.<sup>16–19</sup>

Specialty organizations have recognized issues with traditional open narrative LOR and have recommended standardized LOR with predetermined elements.<sup>20–22</sup> We have observed variable adherence to these guidelines by internal medicine PDs, despite the AAIM guidelines for standardized LOR.<sup>8</sup> This study examined PD LOR for applicants to a highly regarded cardiology fellowship program to explore how bias is expressed, including mechanisms for the expression of bias and potential mitigation of bias by the AAIM guidelines.

## Methods

### Study Design

This was a directed qualitative content analysis,<sup>23</sup> which we chose because the study involved interpreting meaning from text data and because theory and prior research already existed about bias. We examined LOR for applicants selected to interview in 2019 and 2020 at a cardiology fellowship program ranked within the top 20 in the world by *US News & World Report*. We chose this program in a single quaternary care teaching hospital to reduce variability and to provide a large pool of LOR for underrepresented in cardiology (URC) applicants and non-URC applicants. URC was defined as self-identified Black, Latinx, and female applicants (as extracted from their ERAS application). We included women in our definition of URC applicants because in the United States in 2018, only 25% of first-year cardiology fellows were women, and 11.6% of cardiology fellows self-identified as underrepresented in medicine by race/ethnicity.<sup>24,25</sup>

### Data Source

We selected a convenience sample of LOR to obtain an even distribution of URC applicants, as well as LOR adhering to AAIM guidelines (LOR-AAIM) and LOR not adhering to AAIM guidelines (LOR-NonAAIM). One author (A.Q.) reviewed all letters for applicants chosen to interview and categorized them as LOR-AAIM based on the presence of key sections from the AAIM guidelines. LOR were not considered if they inconsistently completed recommended sections by AAIM guidelines of program description, achievement in core competencies, and

### Objectives

We examined program director (PD) letters of recommendation (LOR) for applicants to a cardiology fellowship program to determine the mechanism of how bias is expressed and whether the 2017 Alliance for Academic Internal Medicine (AAIM) guidelines reduce bias.

### Findings

Bias against underrepresented in cardiology (URC) applicants was expressed in all types of LOR through different forms of language use, but letters following AAIM guidelines appeared to have reduced use of communal language, possibly representing less bias.

### Limitations

This was a study at a single cardiology fellowship program, and the coders were not blind to race or gender.

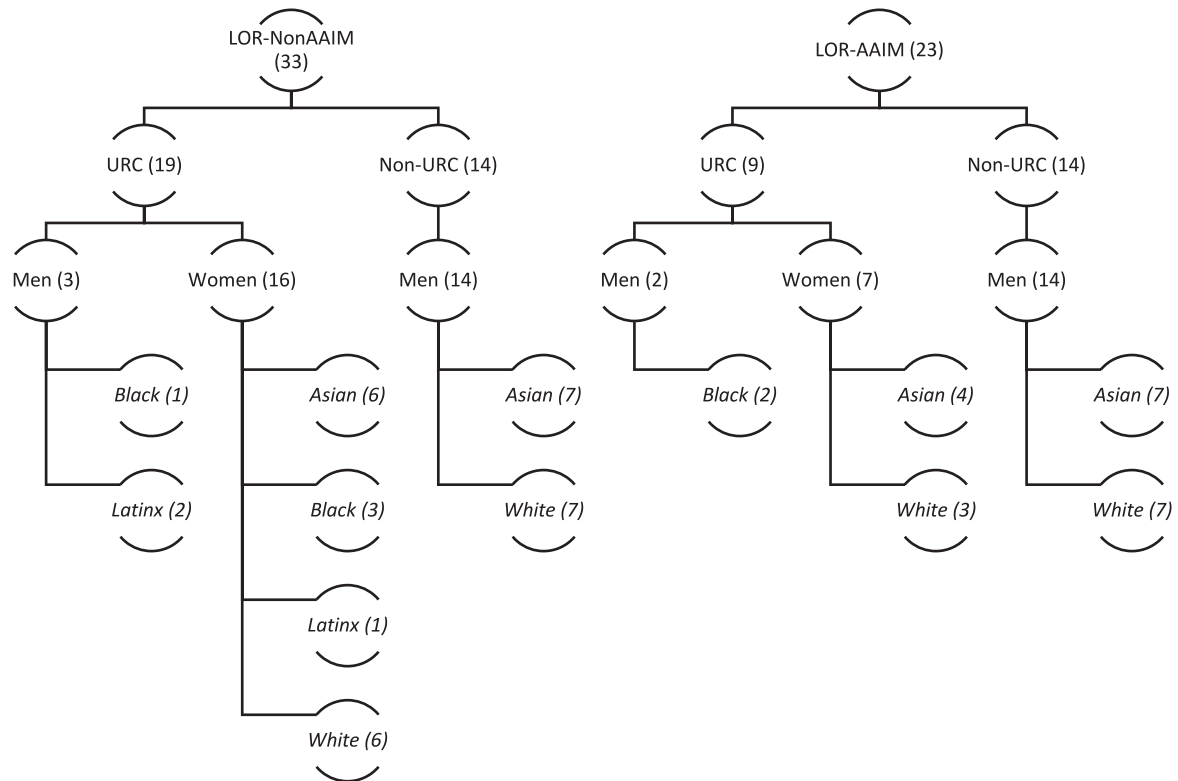
### Bottom Line

While language is used in a biased pattern toward URC applicants in all types of LOR, there are opportunities to reduce this bias, including anti-bias training, expansion of AAIM guidelines, and widespread adoption of AAIM guidelines for PD LOR.

overall assessment. This categorization was reviewed by 2 coders (N.Z., S.B.) without disagreements. All LOR-AAIM were included in the study as well as all LOR-NonAAIM for Black and Latinx applicants. Finally, a comparable number of random LOR-NonAAIM were selected with slight oversampling of letters for URC applicants. Identifying features other than race and gender were anonymized by 2 of the authors (A.Q., D.A.) who were not involved in the coding process. The study was approved by the University of California, San Francisco Institutional Review Board. The 2017 AAIM guidelines for standardized LOR are provided as online supplementary data.<sup>8</sup>

### Analysis

A PubMed literature review for bias in medical LOR identified key concepts for preliminary coding categories (including agentic vs communal terms, doubt raising, and career trajectory bias). Initial exploratory coding was performed reviewing the letters with the generation of new codes until theoretical saturation.<sup>26</sup> A codebook was created and iteratively refined. Codes regarding structure of LOR and format of evaluative comments were developed during coding. Our final codebooks can be found in the online supplementary data. A primary coder (N.Z.), an Asian man and internal medicine resident, and a second coder (S.B.), a White woman and congenital cardiology fellow, who were non-experts in bias, coded all of the transcripts in Dedoose (SocioCultural Research Consultants, Los Angeles, CA). Because transcripts were anonymized, coders could not identify if they had ever interacted with any applicants. Coders were not blind to race/gender and intentionally looked for supporting and



**FIGURE**  
Distribution of Letters by Compliance With Guidelines, Gender, and if Underrepresented in Cardiology

non-supporting evidence of bias. A senior author (P.O.) who analyzed alignment of selected quotes with themes was blinded to race/gender. Disagreements in coding were resolved by consensus. Data were analyzed using outputs from these codes, analytical memos maintained, and themes summarized.

## Results

Fifty-six LOR were studied. FIGURE 1 provides a distribution of letters by compliance with guidelines, gender, and if URC. We had more LOR-NonAAIM than LOR-AAIM due to sample constraints. We purposefully oversampled LOR-NonAAIM for URC applicants. We had more LOR-AAIM for non-URC applicants due to sample constraints, including no letters from Latinx applicants.

In LOR-NonAAIM, PDs typically described an applicant's pre-residency story, scholarly contributions, clinical performance, and overall assessment. PDs often included a fifth section: special attributes, such as personal characteristics, contributions to residency, passion for education, or unique background. In LOR-AAIM, PDs wrote letters with 5 sections consistent with the guidelines. When discussing a resident's achievement in the core competencies section, PDs used different strategies: (1) providing only numerical ratings

for each core competency; (2) separate narrative description for each core competency; (3) separate description for each core competency mixing quotations and narrative; (4) combined description of all core competencies in a single narrative with a separate section for quotations; and (5) combined narrative description of all core competencies with no quotations. All 6 core competencies were rarely addressed when descriptions were combined in narrative form or when quotations were used to describe competencies. The scholarly contributions, personal characteristics/skills, and performance-related extensions in training sections were completed inconsistently.

We identified 3 themes from these LOR: what and where agentic and communal language were used, doubt raising, and career trajectory bias. Each theme will be described as follows.

### Agentic and Communal Language: What and Where

*What:* We identified different patterns of agentic and communal language use based on presence (whether terms were used), mechanism of delivery (whether language was used in narrative descriptions or evaluative quotations), and location (where the language was used in LOR). Regarding presence, both agentic and communal language were used to

**TABLE 1**  
Examples of Agentic and Communal Language

	Agentic Language	Communal Language
LOR-NonAAIM	<i>X pairs his passion for research with an equal passion for clinical medicine. He is a leader on the wards and a role model at the bedside.</i> 21: White man, narrative from clinical performance	<i>Faculty members commented on the kindness and compassion she displayed, as well as a strong sense of teamwork and collaboration.</i> 35: Latinx woman, narrative from clinical performance
LOR-AAIM	<i>Fund of knowledge and clinical judgment are outstanding. In addition, he is an outstanding leader of the team—he really functioned at the level of a resident.</i> 38: Asian man, quote from patient care section about internship	<i>She is calm and kind in the stressful CCU environment. Very friendly and always there to help and teach.</i> 12: White woman, 2 separate quotes from patient care and interpersonal skills

Abbreviations: LOR, letters of recommendation; AAIM, Alliance for Academic Internal Medicine; CCU, cardiac care unit.

describe all applicants in both letter formats; all letters had at least one instance of both types of language, though typically multiple instances. However, URC applicants were described more frequently using communal language whereas non-URC applicants were described more frequently using agentic language. This pattern remained similar for both LOR-NonAAIM and LOR-AAIM (TABLE 1). The mechanism of delivery for communal language occurred through both PD narrative description and selected attending quotations from residents' evaluations. Examples of these 2 formats are in TABLE 1.

**Where:** The location of communal language varied in LOR-NonAAIM and LOR-AAIM.

**LOR-NonAAIM:** Communal language was used in the clinical performance, special attributes, and overall assessment sections. Throughout the clinical performance section, PDs relied on communal language to describe URC applicants because they focused on these applicants' interpersonal skills. In the special attributes section, PDs discussed personality traits for URC applicants, especially communal characteristics, compared to non-URC applicants. In some cases, the entire paragraph only described communal characteristics, focusing the reader on these attributes.

*On a personal level, X has a calm demeanor that places patients at ease. His friendly smile conveys his desire to help the patient. . . He has an unending enthusiasm for medicine and a positive attitude that resonated with his peers.* (4: Black man, special attributes)

For non-URC applicants, these narrative paragraphs were often about a passion for education or unique background. *Beyond being an excellent researcher, leader, and clinician, X is a well-known, outstanding teacher, having received excellent reviews*

*for teaching medical students and residents.* (16: Asian man, special attributes)

The overall assessment paragraph ending most NonAAIM letters included a description of the most notable aspects of each applicant. In URC applicants, these sentences described and focused attention on communal characteristics as opposed to agentic characteristics.

*In summary, we are delighted to present X to you for consideration for your rigorous fellowship in cardiology. X is an exceptional young physician who has excelled in every stage of her medical career. She is energetic, compassionate, and committed. . . Her enthusiasm, dedication, and warm personality have been valued assets to our department.* (6: White woman, overall assessment)

*In summary, X is a compassionate and conscientious physician who has shown aptitude and research throughout her career. She is an energetic dedicated clinician who is an outstanding communicator and a pleasure to interact with due to her enthusiasm for all she does.* (5: Black woman, overall assessment)

PDs tended to describe non-URC applicants in the overall assessment with agentic characteristics.

*In summary, X is a highly motivated and extremely bright outstanding young physician. His engineering background, commitment to academic pursuits, and superior clinical acumen make him well poised to become a leader in cardiac electrophysiology.* (15: Asian man, overall assessment)

**LOR-AAIM:** In LOR-AAIM, communal language appeared in the core competencies, personal characteristics, and overall assessment sections. Communal language was used less for URC applicants in these structured LOR as compared to LOR-NonAAIM.

For the core competencies, PDs often used communal language in the patient care and interpersonal sections, but rarely in the medical knowledge, systems-based practice, practice-based learning and improvement, and professionalism sections. This confined use contrasted with PDs who used communal language throughout LOR-NonAAIM. Non-URC applicants continued to be described primarily with agentic language.

*X involves all members of the clinical care team effectively. He communicates well with consultants, nurses, primary care providers, patients, and families. He has a unique ability to connect with patients on a personal level when they are at their most vulnerable.* (34: Asian man, narrative from interpersonal and communication skills)

Not all PDs included the personal characteristics/skills portion of the LOR-AAIM. Similar to the special attributes paragraphs from LOR-NonAAIM, these paragraphs tended to focus on communal characteristics of URC applicants as compared to non-URC applicants.

*X has a warm, welcoming demeanor that helps him connect with patients. . . his peers consider him a great role model of compassionate care and repeatedly comment about his kindness towards team members, patients, and everyone around him.* (24: Black man, personal characteristics)

However, in contrast to LOR-NonAAIM, when PDs did include a description of personal characteristics, they typically also included a description of skills mastered beyond residency requirements.

In the overall assessment section, PDs focused on communal characteristics when describing URC applicants. The final 2 sentences of the LOR from the following excerpt focus on 2 communal characteristics, humility and integrity, by calling attention to them as the applicant's "strongest characteristics."

*She will impress you with her compassion and kindness, as well as with her powerful intellect and reasoning skills. Humility and integrity are her strongest characteristics; she is highly receptive to feedback and never needs to be told anything twice.* (29: White woman, overall assessment)

## Doubt Raising

The 3 kinds of doubt raising found in both letter formats were hedging, faint praise, and negative

language. Doubt raising was less common than the ubiquitous use of agentic and communal language.

The few examples of faint praise and hedging *only* occurred in letters for URC applicants. In the excerpt below from a LOR-NonAAIM, 2 sentences raise doubt. First, the discussion of the applicant's difficulty with the electronic record and lack of interest in general medicine is seemingly resolved by the next sentence which describes his improvement with feedback. Second, the sentence discussing his new-found insight into how individual patients differ from those in trials is an example of faint praise, as these are insights most applicants glean in medical school.

*Early in internship, X was challenged by the extensive amount of clinical data presented in the electronic record and the necessity to focus his management plans in areas outside his interest in cardiology. He improved with feedback from our academic hospitalist team, and he developed excellent work habits to help him prioritize and streamline his problem list. . . Over time, he gained understanding about how the individual patient may differ from patients in research trials, especially from the psychosocial or socioeconomic aspect.* (4: Black man, narrative from clinical skills)

In the excerpt below from a LOR-AAIM, the bolded phrase is an example of faint praise that suggests the applicant does not complete all required tasks in a timely manner.

*"X frequently completes most required tasks within the expected timeframe including documentation, responding to calls from teammates and patients as well as completing required documentation and paperwork for administrative purposes."* (10: Black man, quote from professionalism)

In both LOR-NonAAIM and LOR-AAIM, we found a common interaction in letters for URC applicants with the use of communal terms framed negatively, whereas for non-URC applicants, communal terms tended to be framed positively. This occurred in instances when applicants were described using both agentic and communal language within the same narrative, typically linked by a conjunction or preposition which served to frame the communal characteristic as negative (eg, "but," "despite") or positive (eg, "and"). In the following excerpt from a LOR-AAIM, the applicant is described as a person who does not call attention to herself (ie, humble, a communal characteristic) and is intelligent (agentic characteristic). The conjunction "but" subtly casts the humble descriptor as negative language and also

TABLE 2

Examples of Communalism Used as Positive and Negative Characteristics

	Communalism as Positive Characteristic	Communalism as Negative Characteristic
LOR-NonAAIM	<i>X is simultaneously a compassionate caregiver and an intellectually curious scientist.</i> 8: White man, overall assessment	<i>X has a very understated style that I appreciate, but has quite a remarkable fund of knowledge in cardiology for her stage of training.</i> 2: White woman, quote from clinical performance
LOR-AAIM	<i>X was a pleasure to work with. He is smart, eager, motivated, and diligent. He couples this with a humility, compassionate, and warm approach that make him a true healer.</i> 23: White man, quote from patient care <i>X did a fabulous job on this rotation. He is dedicated to providing exceptional patient care, informed by a strong foundation in clinical medicine complemented by a wonderful bedside manner.</i> 1: White man, quote from medical knowledge	<i>She was able to manage her team with great professionalism as well as allowing the intern and student to develop their sense of independence. However, in a quiet way she was in total control of information flow was able to participate vigorously in the discussion of the options available for her patients.</i> 36: Asian woman, quote from patient care

Abbreviations: LOR, letters of recommendation; AAIM, Alliance for Academic Internal Medicine.

serves to broadly undervalue this communal characteristic:

*“X is not the type of resident who calls attention to herself, but her medical knowledge, commitment to patients, and work ethic are readily apparent.”*  
(17: Asian woman, narrative paragraph about core competencies)

In contrast, a non-URC applicant is described as both intellectual (an agentic term) and compassionate (a communal term): “*X is simultaneously a compassionate caregiver and an intellectually curious scientist*” (8: White man, scholarly contributions). The conjunction “and” serves to elevate both characteristics as positive. TABLE 2 provides additional examples of conjunctions and prepositions as doubt raising devices.

### Bias in Career Trajectory

In both LOR-NonAAIM and LOR-AAIM, while PDs tend to describe URC applicants as earlier in their career, non-URC applicants were described as advanced in their career. One non-URC applicant is not only described as having a future career in academic medicine, but also is described with active verbs that frame him as a researcher.

*X has already demonstrated an interest in cardiology and research that shows he will be successful in a future career in academic medicine. While in medical school, he conducted research to improve the quality of care for patients with A... He designed an analysis that measured [this quality]... He identified areas for QI... (16: Asian man, scholarly contributions)*

In contrast, an URC applicant is described using passive and weak verbs.

*Over the course of her academic training, X has been involved in a significant amount of research... X has worked on several accomplished cardiology research teams including a project looking at A. She has also been working on a project in the use of B echocardiography to evaluate C. (33: Asian woman, scholarly contributions)*

TABLE 3 shows additional examples of bias in career trajectory. Notably, all applicants in TABLE 3 were rated in the top tier of research productivity. Despite this, URC applicants were framed as students or participants “working” with others, while non-URC applicants were framed as either already being scientists or having high potential to becoming scientists/researchers.

### Discussion

We observed that with AAIM guidelines, there appeared to be a reduced use of communal language for URC applicants, which may represent less bias. We observed that bias still existed in both types of letters. PDs described URC applicants using communal language, and non-URC applicants using agentic language, regardless of format. This pattern existed in both narrative descriptions and selected quotations. Both letter types had examples of doubt raising and bias in career trajectory. This language was readily apparent even to non-experts in the field with minimal bias training. Finally, both letter types varied widely in format despite the structure suggested by the AAIM guidelines. We will discuss our main findings illustrating the helpfulness of AAIM

**TABLE 3**  
Examples of Differences in Description of Career Trajectory

	Later in Career Trajectory	Earlier in Career Trajectory
LOR-NonAAIM	<i>X's promise as a future faculty member is clear from his prior academic experiences. . .</i> 26: White man, scholarly contributions	<i>He has shown initiative in his research pursuits and is an excellent clinician.</i> 9: Latinx man, overall assessment
LOR-AAIM	<i>He is both the first physician and the first scientist in his family. For many, the choice of a career path is a decision born out of introspection and reflection. For X, however, the decision to become a physician scientist has been a calling from the outset.</i> 11: White man, scholarly contributions	<i>During residency, X published a journal article entitled "YYY." She is presently working on 2 research projects: (1) an [approach] to YYY with Dr. A and Dr. B, and (2) working with Dr. C. . .</i> 19: Asian woman, scholarly/research contributions

Abbreviations: LOR, letters of recommendation; AAIM, Alliance for Academic Internal Medicine.

structured guidelines to reduce bias, the persistence of bias despite these guidelines, and the potential sources of this bias.

Two components of structure created by the AAIM guidelines appeared to reduce bias. First, core competencies sections forced PDs to elaborate on clinical performance areas not traditionally covered. Second, the personal characteristics and skills sections reminded PDs to discuss both aspects about an applicant. Our findings are consistent with results from a previous study of LOR-AAIM, where fellowship PDs felt that structured LOR were clearer in communicating residents' performance across 6 core competency domains than LOR-NonAAIM.<sup>5</sup>

Bias persisted within LOR-AAIM despite the AAIM guidelines. This finding aligns with previous literature for otolaryngology residency where standardized LOR reduced but did not eliminate bias, especially between men and women.<sup>21</sup> In our analysis, we saw bias persist in 3 different forms. First, it occurs when AAIM guidelines were only partially followed, as exemplified by PDs describing all clinical competencies in the same section rather than in 6 separate sections, thus incompletely addressing the competencies and straying into the pattern of biases. Second, we uniformly observed patterns of bias in the scholarly contributions and overall assessment sections. The lack of structure in the AAIM guidelines for the scholarly contributions and overall assessment sections contributed to this pattern. Third, since hedging or faint praise was only used to describe URC applicants, writers should be extra vigilant in this area.

Evaluative quotations and written narratives implicitly bring bias in both letter formats. The use of evaluative quotations in LOR is a long-standing practice requiring careful application. Selecting others' words introduces additional possibilities of bias. Furthermore, our finding that communal terms were framed as negative language for URC applicants but more positively for non-URC applicants exemplifies

the perpetuation of communal language as a negative characteristic.

Our analysis generates recommendations for PDs and for the AAIM guidelines (see TABLE 4). We recommend the creation of a new "section for growth" in LOR-AAIM. Researchers report the pervasiveness of hedging in evaluations of residents.<sup>27</sup> To rank-order residents, faculty must "read between the lines" of these evaluations, but the lack of a standard "hidden code" risks variable interpretation of evaluations.<sup>28</sup> Our experience is that a de facto system to report trainee areas for growth is in use, often communicated with doubt raising language. A required section regarding areas of strength and for growth could diminish use of doubt raising language by requiring comments for all applicants. The business world uses such a section.<sup>29</sup> We recommend further work by the Accreditation Council for Graduate Medical Education (ACGME) to create a Milestones-based system to track resident competency in research or scholarly activities. The ACGME Internal Medicine Subspecialty Milestones have a scholarship subsection (MK3) that does not exist as a Residency Milestone.<sup>30</sup> Expanding residency clinical competencies to include scholarly activities would help PDs systematically evaluate applicants. Finally, we wish to acknowledge that the bias in these letters is part of a long history of oppression against women and people of color in the United States. Despite our recommendations, as long as there remains systemic racism and sexism, bias will continue to make its way, both overtly and insidiously, into letters of recommendation.<sup>31</sup>

Our study has limitations. First, the LOR were for applicants accepted to interview at a single cardiology fellowship program. We feel that the existence of biased language in the LOR for applicants to this program shows that bias toward URC applicants is likely omnipresent. Second, our study did not consider the gender or race of the letter writers, which can impact language, letter length, and overall

TABLE 4

## Recommendations for Specific Letters of Recommendation (LOR) Sections

LOR Section	Recommendations
General recommendations	<ul style="list-style-type: none"> <li>Follow the 2017 AAIM guidelines.</li> <li>Frame communal descriptors as strengths rather than weaknesses.</li> <li>PDs and evaluators should undergo regular anti-bias training to raise awareness of use of biased language and its context.</li> <li>PD and evaluator training should include discussions of bias in a broader social context, including discussions of structural racism and sexism.</li> <li>Have LOR reviewed by a third party with anti-bias training.</li> <li>Develop guidelines for the creation of a new “section for growth.”</li> </ul>
Core competencies section	<ul style="list-style-type: none"> <li>Require a separate section to describe each competency, as opposed to a single section where all competencies are described together.</li> <li>Ensure that evaluative quotations or narrative descriptions communicate the clinical performance of applicants, including only the pertinent content (eg, descriptions of patient care should not enter the medical knowledge section). This is especially true for evaluative quotations.</li> <li>Use agentic and communal descriptors to create a fair appraisal of each applicant.</li> </ul>
Scholarly activity section	<ul style="list-style-type: none"> <li>Develop guidelines to allow better characterization of an applicant’s career trajectory in terms of skills and demonstrated competencies.</li> <li>Consider using the current scholarship subsection within the ACGME Internal Medicine Subspecialty Milestones as a guide for describing applicants’ scholarly activities.</li> <li>Use active verbs to describe applicants’ activities.</li> </ul>
Overall summary section	<ul style="list-style-type: none"> <li>Develop guidelines to standardize the overall summary section, which include a method for synthesizing the core competencies, scholarly contributions, and personal characteristics/ skills sections.</li> <li>Use agentic and communal descriptors to create a fair appraisal of each applicant.</li> </ul>

Abbreviations: AAIM, Alliance for Academic Internal Medicine; PD, program director; ACGME, Accreditation Council for Graduate Medical Education.

appraisal of the applicant being evaluated.<sup>32,33</sup> Third, coders were not blind to race/gender, introducing the possibility of confirmation bias. Fourth, given our sample, we could not comment on intersectionality of gender and racial bias, which has previously been shown to influence achievement word use in LOR.<sup>34</sup>

## Conclusions

We found that language, including communal and agentic terms, doubt raising, and bias in career trajectory, was used in a biased pattern toward URC applicants. This bias appeared reduced, though not eliminated, when PDs followed the AAIM guidelines. We have provided recommendations on how to continue to work to reduce this bias.

## References

- Dirschl DR, Adams GL. Reliability in evaluating letters of recommendation. *Acad Med.* 2000;75(10):1029. doi:10.1097/00001888-200010000-00022
- Wright SM, Ziegelstein RC. Writing more informative letters of reference. *J Gen Intern Med.* 2004;19(5 Pt 2):588–593. doi:10.1111/j.1525-1497.2004.30142.x
- Love JN, Ronan-Bentle SE, Lane DR, Hegarty CB. The standardized letter of evaluation for postgraduate training: a concept whose time has come? *Acad Med.* 2016;91(11):1480–1482. doi:10.1097/ACM.0000000000001352
- Prager JD, Perkins JN, McFann K, Myer III CM, Pensak ML, Chan KH. Standardized letter of recommendation for pediatric fellowship selection. *Laryngoscope.* 2012;122(2):415–424. doi:10.1002/lary.22394
- O’Connor A, Williams C, Dalal B, et al. Internal medicine fellowship directors’ perspectives on the quality and utility of letters conforming to residency program director letter of recommendation guidelines. *J Community Hosp Intern Med Perspect.* 2018;8(4):173–176. doi:10.1080/20009666.2018.1500424
- Grabowski G, Walker JW. Orthopaedic fellowship selection criteria: a survey of fellowship directors. *J Bone Joint Surg Am.* 2013;95(20):e154. doi:10.2106/JBJS.L.00954
- Poirier MP, Pruitt CW. Factors used by pediatric emergency medicine program directors to select their fellows. *Pediatr Emerg Care.* 2003;19(3):157–161. doi:10.1097/01.pec.0000081236.98249.ed
- Alweis R, Collichio F, Milne CK, et al. Guidelines for a standardized fellowship letter of recommendation. *Am J*



- Med.* 2017;130(5):606–611. doi:10.1016/j.amjmed.2017.01.017
9. Aggarwal S, Grob S, Banerjee D, Putzel PJ, Tao J. Key word use in letters of recommendation for ophthalmology residency applicants according to race, gender, and achievements. *J Acad Ophthalmol.* 2018;10(01):163–171. doi:10.1055/s-0038-1675842
  10. Turrentine FE, Dreisbach CN, St Ivany AR, Hanks JB, Schroen AT. Influence of gender on surgical residency applicants' recommendation letters. *J Am Coll Surg.* 2019;228(4):356–365.e3. doi:10.1016/j.jamcollsurg.2018.12.020
  11. Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in medical student performance evaluations. *PLoS One.* 2017;12(8):e0181659. doi:10.1371/journal.pone.0181659
  12. Hoffman A, Grant W, McCormick M, Jezewski E, Matemavi P, Langnas A. Gendered differences in letters of recommendation for transplant surgery fellowship applicants. *J Surg Educ.* 2019;76(2):427–432. doi:10.1016/j.jsurg.2018.08.021
  13. Trix F, Psenka C. Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse & Soc.* 2003;14(2):191–220. <https://doi.org/10.1177/0957926503014002277>
  14. Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. *J Appl Psychol.* 2009;94(6):1591–1599. doi:10.1037/a0016539
  15. Grimm LJ, Redmond RA, Campbell JC, Rosette AS. Gender and racial bias in radiology residency letters of recommendation. *J Am Coll Radiol.* 2020;17(1 Pt A):64–71. doi:10.1016/j.jacr.2019.08.008
  16. Livingston RW, Rosette AS, Washington EF. Can an agentic black woman get ahead? The impact of race and interpersonal dominance on perceptions of female leaders. *Psychol Sci.* 2012;23(4):354–358. doi:10.1177/0956797611428079
  17. Weaver JL, Garrett SD. Sexism and racism in the American health care industry: a comparative analysis. *Int J Health Serv.* 1978;8(4):677–703. doi:10.2190/AK0C-M9JF-1TR1-5UYF
  18. Coombs AAT, King RK. Workplace discrimination: experiences of practicing physicians. *J Natl Med Assoc.* 2005;97(4):467–477.
  19. Capers QI, Clinchot D, McDougle L, Greenwald AG. Implicit racial bias in medical school admissions. *Acad Med.* 2017;92(3):365–369. doi:10.1097/ACM.0000000000001388
  20. Love JN, Smith J, Weizberg M, et al. Council of Emergency Medicine Residency Directors' standardized letter of recommendation: the program director's perspective. *Acad Emerg Med.* 2014;21(6):680–687. doi:10.1111/acem.12384
  21. Friedman R, Fang CH, Hasbun J, et al. Use of standardized letters of recommendation for otolaryngology head and neck surgery residency and the impact of gender. *Laryngoscope.* 2017;127(12):2738–2745. doi:10.1002/lary.26619
  22. Keim SM, Rein JA, Chisholm C, et al. A standardized letter of recommendation for residency application. *Acad Emerg Med.* 1999;6(11):1141–1146. doi:10.1111/j.1553-2712.1999.tb00117.x
  23. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005;15(9):1277–1288. doi:10.1177/1049732305276687
  24. Santhosh L, Babik JM. Trends in racial and ethnic diversity in internal medicine subspecialty fellowships from 2006 to 2018. *JAMA Netw Open.* 2020;3(2):e1920482. doi:10.1001/jamanetworkopen.2019.20482
  25. American Board of Internal Medicine. Percentage of First-Year Fellows by Gender and Type of Medical School Attended. <https://www.abim.org/about/statistics-data/resident-fellow-workforce-data/first-year-fellows-by-gender-type-of-medical-school-attended.aspx>. Accessed March 10, 2021.
  26. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant.* 2018;52(4):1893–1907. doi:10.1007/s11135-017-0574-8
  27. Ginsburg S, van der Vleuten C, Eva KW, Lingard L. Hedging to save face: a linguistic analysis of written comments on in-training evaluation reports. *Adv Health Sci Educ Theory Pract.* 2016;21(1):175–188. doi:10.1007/s10459-015-9622-0
  28. Ginsburg S, Regehr G, Lingard L, Eva KW. Reading between the lines: faculty interpretations of narrative evaluation comments. *Med Educ.* 2015;49(3):296–306. doi:10.1111/medu.12637
  29. Hedricks CA, Robie C, Oswald FL. Web-based multisource reference checking: an investigation of psychometric integrity and applied benefits. *Int J Select Assess.* 2013;21(1):99–110. doi:10.1111/ijsa.12020
  30. Accreditation Council for Graduate Medical Education and The American Board of Internal Medicine. The Internal Medicine Subspecialty Milestones Project. <http://www.acgme.org/portals/0/pdfs/milestones/internalmedicinesubspecialtymilestones.pdf>. Accessed March 10, 2021.
  31. Hemmer PA, Karani R. Let's face it: we are biased, and it should not be that way. *J Gen Intern Med.* 2019;34(5):649–651. doi:10.1007/s11606-019-04923-w

32. Stauffer JM, Buckley MR. The existence and nature of racial bias in supervisory ratings. *J Appl Psychol*. 2005;90(3):586–591. doi:10.1037/0021-9010.90.3.586
33. Isaac C, Chertoff J, Lee B, Carnes M. Do students' and authors' genders affect evaluations? A linguistic analysis of medical student performance evaluations. *Acad Med*. 2011;86(1):59–66. doi:10.1097/ACM.0b013e318200561d
34. Akos P, Kretchmar J. Gender and ethnic bias in letters of recommendation: considerations for school counselors. *Prof Sch Counsel*. 2016;20(1):102–113. doi:10.5330/1096-2409-20.1.102



**Neil Zhang, MD, MS**, is Clinical Instructor, Department of Medicine, University of California, San Francisco; **Sarah Blissett, MD, MHPE**, is Assistant Professor, Division of Cardiology,

Department of Medicine, and Centre for Education Research and Innovation, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; **David Anderson, MD**, is a Cardiology Fellow, Division of Cardiology, Department of Medicine, University of California, San Francisco; **Patricia O'Sullivan, EdD**, is Professor, School of Medicine, University of California, San Francisco; and **Atif Qasim, MD, MSCE**, is Cardiology Fellowship Program Director and Associate Professor, Division of Cardiology, Department of Medicine, University of California, San Francisco.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Neil Zhang, MD, MS, University of California, San Francisco, neil.zhang@ucsf.edu, Twitter @NeilSijiaZhang

Received August 14, 2020; revisions received October 25, 2020, and February 15, 2021; accepted February 17, 2021.