Race- and gender-based differences in descriptions of applicants in the letters of recommendation for orthopaedic surgery residency

Alexa Powers
Katherine M. Gerull
Rachel Rothman
Sandra A. Klein
Rick W. Wright

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs
Authors
Alexa Powers, Katherine M. Gerull, Rachel Rothman, Sandra A. Klein, Rick W. Wright, and Christopher J. Dy
AOA Critical Issues in Education

Race- and Gender-Based Differences in Descriptions of Applicants in the Letters of Recommendation for Orthopaedic Surgery Residency

Alexa Powers, BA, Katherine M. Gerull, MD, Rachel Rothman, Sandra A. Klein, MD, Rick W. Wright, MD, and Christopher J. Dy, MD, MPH

Investigation performed at Washington University School of Medicine, St. Louis, Missouri

Background: Letters of recommendation (LOR) are an important component of trainee advancement and assessment. Examination of word use in LOR has demonstrated significant differences in how letter writers describe female and male applicants. Given the emphasis on increasing both gender and racial diversity among orthopaedic surgeons, we aimed to study gender and racial differences in LOR for applicants to orthopaedic surgery residencies.

Methods: All applications submitted to a single, academic orthopaedic residency program in 2018 were included. Self-identified gender and race were recorded. The LOR were analyzed via a text analysis software program using previously described categories of communal, agentic, grindstone, ability, and standout words. We examined the relative frequency of word use in letters for (1) male and female applicants and (2) white and underrepresented in orthopaedics (UiO) applicants, with the subgroup analysis based on whether standardized (using the American Orthopaedic Association template) or traditional (narrative) LOR were used.

Results: Two thousand six hundred twenty-five LOR were submitted for 730 applicants (79% men). Fifty-nine percent of applicants were self-identified as white, and 34% were self-identified as UiO. In traditional LOR, standout words (odds ratio [OR] 1.07; p = 0.01) were more likely to be used in letters for women compared with men, with no difference in any other word-use category. In standardized LOR, there were no gender-based differences in any word category. In traditional LOR, grindstone words (OR = 0.96; p = 0.02) were more likely to be used in letters for UiO than white applicants, whereas standout words (OR = 1.05; p = 0.04) were more likely to be used in letters for white candidates. In standardized LOR, there were no race-based differences in any word category use.

Conclusions: Small differences were found in the categories of words used to describe male and female candidates and white and UiO candidates. These differences were not present in the standardized LOR compared with traditional LOR. It is possible that the use of standardized LOR may reduce gender- and race-based bias in the narrative assessment of applicants.

Disclosure: The Disclosure of Potential Conflicts of Interest forms are provided with the online version of the article (http://links.lww.com/JBJSOA/A170).

Copyright © 2020 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work, provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.
Academic medical centers increasingly recognize the importance of diversity and inclusion. However, the number of women and minorities in orthopaedic surgery have been slow to increase. Orthopaedic surgery is the least gender-diverse specialty in medicine. In 2015, 38% of general surgery residents were women, compared with 15% of orthopaedic surgery residents. Women comprise 17% of orthopaedic surgery faculty at academic institutions in the United States, compared with 23% of general surgery faculty and 32% of otolaryngology faculty. Regarding racial diversity, from 2006 to 2015, there was no significant change in the representation of African American orthopaedic surgery residents over time, with a significant decrease in the overall racial diversity in the field.

Letters of recommendation (LOR) are an important aspect of promotion and advancement in all scientific disciplines. Trix and Psenka reviewed LOR for medical faculty and found that letters written for female applicants systematically differed from those written for male applicants in the increased inclusion of negative language, doubt-raising phrases, and decreased letter length. In an investigation of medical student performance evaluations, Axelson et al. found that gender bias was evident in adjective use; women were more likely than men to be described as “compassionate” and “sensitive,” whereas men were more likely to be described as “quick learners.” A study of general surgery residency applicant letters found differences in words used within men’s and women’s letters, with achievement words (such as performance, leadership, and knowledge) being used more often in men’s letters, and caring words (such as care, time, and support) being used more often in women’s letters. A study of LOR in orthopaedic surgery residency applicants found only minor differences between letters written for men and women applicants, with only subtle differences between letters written by male and female authors, and authors of various academic rank. Each of these studies focused solely on gender as the diversity variable, with no comment on the role of race in the applicant LOR. Only one study of the influence of race of those being evaluated has been conducted and focused on medical student performance evaluations, not LOR.

LOR play an important role in the selection and ranking of orthopaedic surgery residency applicants. According to the 2018 National Resident Matching Program Director Survey, “letters of recommendation in the specialty” were second only to the United States Medical Licensing Examination (USMLE) Step 1 score in the most commonly cited factors in selecting applicants to interview among orthopaedic surgery program directors. Recently, orthopaedic surgery has been increasingly adopting the American Orthopaedic Association (AOA) standardized LOR in place of traditional LOR. The standardized LOR are forms comprising numerical ranking scales mapped to the 6 Accreditation Council for Graduate Medical Education core competencies, with a space for comments.

Despite the critical importance of LOR in residency selection and the demonstrated gender disparities in other fields, little is known about the influence of applicant race on how residency LOR are written. Given the emphasis on increasing both gender and racial diversity among orthopaedic surgeons, we aimed to study the gender and racial differences in LOR for applicants to orthopaedic surgery residencies.

### Methods

All Electronic Residency Application Service applications submitted to a single, academic orthopaedic surgery residency program for the 2018 match class were eligible for inclusion in this study. Applicants were first assigned a study ID number. Self-identified race, gender, Alpha Omega Alpha Honor Medical Society membership, USMLE Step 1 score, and medical school region were recorded for each applicant. Applicant names and all gender-specific pronouns (i.e. his, him, her, he, and she) were electronically redacted from the LOR. This study was deemed non-human subject research by the Washington University School of Medicine institutional review board. No funding was provided for this study.

### Word Use Analysis

LOR were analyzed using a text analysis software program, Linguistic Inquiry and Word Count 2015, which has been used in previous analysis of LOR. Word lists were created using previously described categories of communal, agentic, grindstone, ability, and standout words (Table I). These lists of words were then uploaded into the text analysis software, tagged by word category. The software reported the total number of words in each LOR, the total number of category words in each LOR, and the percentage of category words (# category words/total # words) in each LOR. Applicant gender was either male or female, based on self-identification.

### TABLE I | Sample List of Terms in Each Dictionary Category

<table>
<thead>
<tr>
<th>Agentic</th>
<th>Communal</th>
<th>Grindstone</th>
<th>Ability</th>
<th>Standout</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assertive</td>
<td>Agreeable*</td>
<td>Dedicate*</td>
<td>Ability</td>
<td>Amazing</td>
</tr>
<tr>
<td>Compet*</td>
<td>Caring</td>
<td>Diligen*</td>
<td>Adept*</td>
<td>Exceptional</td>
</tr>
<tr>
<td>Confident</td>
<td>Considerate</td>
<td>Effort*</td>
<td>Brilliant*</td>
<td>Outstanding</td>
</tr>
<tr>
<td>Independent</td>
<td>Helpful</td>
<td>Hardworking</td>
<td>Capable</td>
<td>Remarkable</td>
</tr>
<tr>
<td>Outspoken</td>
<td>Interpersonal*</td>
<td>Organiz*</td>
<td>Intell*</td>
<td>Superb</td>
</tr>
<tr>
<td>Strength</td>
<td>Warm</td>
<td>Persist*</td>
<td>Proficient*</td>
<td>Unique</td>
</tr>
</tbody>
</table>

*The asterisk denotes the acceptance of all letters after its appearance. For example, the word stem support* can include both supportive and supported.
identification. Applicant race (also self-identified) was dichotomized as white vs. those historically underrepresented in orthopaedics (UiO) with UiO encompassing black or African American, Hispanic, Latino or of Spanish origin, Native American or Hawaiian, Asian, Indian, and Middle Eastern. The letter types were either traditional narrative LOR or standardized LOR using the format suggested by the AOA Council of Residency Directors. The type was denoted for each letter and the “personal comments” section of standardized LOR was included in the analysis.

A descriptive statistical analysis was performed for demographic data. Univariate analyses using student t tests were used to examine both race- and gender-based differences in word use for each category. We conducted an a priori sample size calculation using pilot with 60 LOR from 20 applications in a previous application cycle. Assuming 3 LOR per applicant, we determined that the analysis of 122 male and 122 female applicants would be necessary to detect a significant difference in the use of agentic category words, based on mean 95% word use in LOR for male applicants and 77% word use in female applicants (from pilot data), with α = 0.05 and β = 0.80.

The multivariable logistic regression was used to simultaneously assess the relationships of multiple predictor variables with the dependent variable. The number of times each word category was used in the LOR was the dependent variable. Separate models were used to for each word category (agentic, communal, grindstone, ability, and standout), with predictor variables dichotomized to applicant gender, applicant race, and type of LOR (traditional or standardized). A backward-elimination strategy with variable-retention p-value of 0.05 was used. All statistical tests were performed in collaboration with a biostatistician using SAS Base software version 9.4.

Source of Funding
C.J.D. was supported by K23AR073928-01 from the National Institute for Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

Results

Participant Characteristics
Two thousand six hundred twenty-five LOR were submitted for 730 applicants in this study (Table II). Five hundred seventy-six (79%) of the applicants were self-identified as men and 154 (21%) as women. Fifty-nine percent of applicants were self-identified as white, 34% self-identified as a race/ethnicity other than white (including black or African American, Hispanic, Latino or of Spanish origin, Native American or Hawaiian, Asian, Indian, and the Middle East, categorized as UiO applicants), and 7% did not self-identify their race/ethnicity and were excluded from that analysis.

Gender
Two thousand sixty-six (79%) letters were written for male applicants and 559 (21%) were written for female candidates. The average word count for men was 274 words, and the average word count for women was 305 words (p < 0.001). Standout words (odds ratio [OR] = 1.07, 95% confidence interval [CI]: 1.02-1.12) were significantly more likely to be used in female than in male applicants (Table III). There was no significant difference in the use of agentic words, communal, grindstone words, or ability words between male and female applicants.

Race
One thousand five hundred fifty-three (59%) letters were written for white applicants and 884 (34%) were written for UiO candidates. The average word counts for white and UiO applicants were 273 and 294 words, respectively (p = 0.003). Grindstone words (OR = 0.90, 95% CI: 0.87-0.93; reference category white) were significantly less likely to be used in letters for UiO applicants, whereas standout words were significantly more likely to be used for UiO applicants (OR = 1.05, 95% CI: 1.001-1.09). There was no significant difference in use of agentic, communal, or ability words between white and UiO applicants (Table IV).

Gender and Type of LOR
The length of traditional LOR was significantly longer for women as compared to men (p = 0.05). For traditional LOR, standout words (OR = 1.07, 95% CI: 1.02-1.12) were significantly more likely to be used for women compared with men (Table V). Within traditional LOR, there was no significant difference in the use of agentic, communal, grindstone, or ability words between men and women.

In standardized LOR, there was no significant difference in the use of agentic, communal, grindstone, ability, or standout words between men and women (Table V).

Race and Type of LOR
There was no difference in the length of LOR between white and UiO applicants in either standardized or traditional LOR. In traditional LOR, grindstone words (OR = 0.96, 95% CI:

TABLE II Demographic Data for 730 Applicants

<table>
<thead>
<tr>
<th>Applicant Characteristics</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>576</td>
<td>154</td>
</tr>
<tr>
<td>Male</td>
<td>576</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>394</td>
<td>166</td>
</tr>
<tr>
<td>White</td>
<td>338</td>
<td>94</td>
</tr>
<tr>
<td>Black</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>Asian</td>
<td>99</td>
<td>26</td>
</tr>
<tr>
<td>Hispanic</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>Multiracial</td>
<td>42</td>
<td>12</td>
</tr>
<tr>
<td>Native American or Hawaiian</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Did not identify</td>
<td>42</td>
<td>8</td>
</tr>
<tr>
<td>Academic performance</td>
<td>246</td>
<td>243</td>
</tr>
<tr>
<td>USMLE Step 1 (Mean)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Member of Alpha Omega Alpha</td>
<td>173</td>
<td>49</td>
</tr>
</tbody>
</table>

3
0.93-0.99) were significantly more likely to be used in letters for UiO applicants compared with white applicants (Table V). Standout words (OR = 1.05, 95% CI: 1.002-1.10) were significantly more likely to be used in letters for white candidates (Table V). There was no significant difference in the use of agentic, communal, or ability words between white and UiO applicants in traditional LOR. In standardized LOR, there was no significant difference in the use of agentic, communal, grindstone, ability, or standout words between UiO and white applicants.

**Gender of Letter Writer and Applicant**

When male letter writers wrote LOR for male applicants, they were more likely to use communal words (OR 1.07 [95% CI: 1.02-1.12]) and standout words (OR 1.08 [95% CI: 1.03-1.13]) than when male letter writers wrote LOR for female applicants. There were no differences in the use of agentic, grindstone, and ability words. When female letter writers wrote LOR for male applicants, there were no differences compared with when female letter writers wrote LOR for female applicants.

**Discussion**

LOR are an important component of the residency application to orthopaedic surgery. The analysis of LOR content may indicate unconscious biases in how applicants are viewed by their recommenders. In this study, we have demonstrated both gender- and race-based differences in how applicants for orthopaedic surgery residencies are described. Women were more likely than men to be described using communal and standout words in LOR for orthopaedic surgery residency. UiO applicants were more likely than white applicants to be described using grindstone words, whereas white applicants were more likely to be described using standout words.

Women being described using communal words is well supported by previous literature, whereas women being described using standout words differs significantly from previous studies in other fields. Although we detected a statistically significant difference in standout words, the actual differences are modest (with OR very close to 1), suggesting that men and women are described quite similarly in LOR for orthopaedic surgery residency. These results support those found in a recent study on gender differences in LOR for orthopaedic surgery, which showed an overall similarity in the language used to describe men and women applicants to orthopaedic surgery. A potential reason for this difference seen in orthopaedic surgery could be attributed to letter writers unconsciously using more stereotypically masculine words to describe female trainees to explicitly demonstrate their fit within

### TABLE II. Gender Differences*

<table>
<thead>
<tr>
<th>Word Category</th>
<th>No. of Times Used in LOR (Unadjusted Mean Values)</th>
<th>OR from Multivariable Regression Model (Reference: Male)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agentic</td>
<td>2.8</td>
<td>1.02 (0.97-1.07)</td>
<td>0.44</td>
</tr>
<tr>
<td>Communal</td>
<td>2.4</td>
<td>1.06 (0.99-1.12)</td>
<td>0.06</td>
</tr>
<tr>
<td>Grindstone</td>
<td>4.2</td>
<td>0.99 (0.95-1.03)</td>
<td>0.48</td>
</tr>
<tr>
<td>Ability</td>
<td>1.7</td>
<td>1.01 (0.95-1.08)</td>
<td>0.71</td>
</tr>
<tr>
<td>Standout</td>
<td>2.1</td>
<td>1.07 (1.02-1.12)</td>
<td>0.01</td>
</tr>
</tbody>
</table>

*OR = odds ratio and LOR = letters of recommendation. † Adjusted for applicant race and type of LOR (traditional or standardized).
this stereotypically male-dominated field. A more likely explanation is that letter writers may use a basic LOR template and add modifications to fit different applicants. In addition, this study looked at the overall word categories. Previous studies, particularly those in otolaryngology, focused on differences in single word usage. We chose to use the overall word categories, as opposed to individual words, to comment on overarching trends in word usage. Although modest, our analysis shows statistically significant differences in the way applicants are described between traditional and standardized LOR. To highlight this, all analyses were included in the final analysis and discussion. Further research is required to comment on possible differences in the word context.

Although studies on gender differences in LOR to orthopaedic surgery have been performed previously, there are no previous studies evaluating racial differences in LOR in orthopaedic surgery. Day et al. found that Asian American, African American, Hispanic, and Latino residents constituted only 19.8% of orthopaedic surgery residents in 2006, which was significantly lower than that found in general surgery. Racial differences in LOR in orthopaedic surgery may be one barrier in the goal of diversifying the field of orthopaedics.

To combat bias and variability in interpretation of traditional LOR, the AOA Council of Orthopaedic Residency Directors introduced the standardized LOR to provide a more objective assessment of an applicant. Several studies have been performed assessing the usefulness in using the Council of Residency Directors Standardized LOR rather than a traditional LOR; however, these studies have focused on assessing the value of the summative rank statement in stratifying applicants. There have been no previous studies performed to determine whether the Standardized Letter of Recommendation (SLOR) reduces the gender or racial differences that previous studies have found during evaluation of the traditional narrative LOR. Because LOR are an important component of the residency application, reducing potential biases may contribute to the further diversification of this field. When stratifying letters by type (either traditional or standardized LOR), we found gender- and race-based differences with traditional LOR that were not present in the standardized LOR. These findings corroborate previous work in emergency medicine and otolaryngology which demonstrated that standardized LOR reduced subjectivity and gender bias compared with traditional LOR. Although statistically significant, the effect sizes in our study are modest, with OR very close to 1, suggesting that the actual effect of SLORs may be quite small.

There are several limitations to this study. Our results are based on LOR submitted to a single institution for the 2018 application cycle. However, the applicants reviewed represent the average candidate to orthopaedic surgery for USMLE Step 1 score and represent 86% (133 individual medical schools) of all US allopathic medical schools, 24% (10 individual medical schools) of all US osteopathic medical schools, and 30 international schools of medicine. In the number of applications received, 730 applicants applied for 8 residency spots at this institution, for an average of 91.25 applicants per position. Nationally, 1,037 applicants applied for a total of 755 orthopaedic surgery positions across 175 residency programs in 2019. Applicant applied to an average of 88 orthopaedic surgery residency programs, resulting in 91,256 applications for the 755 residency spots or 120.87 applications per position. Although the number of applicants to this residency program seems large, it is less than the national average. Thus, our results can be reasonably extrapolated to the entire orthopaedic surgery applicant population and to all residency programs. The national match rate for all orthopaedic surgery applicants (including US MD, DO, and International) was 72.5% in 2019.

In addition, these study results may be a consequence of the particular terms specified in each word category. The dictionary was compiled based on the terms used in previous studies; however, these studies were performed in other fields. It is possible that the area of orthopaedics is looking for particular traits and qualities indicative of a successful applicant in this specialized field. Having a more orthopaedic-driven dictionary of terms may yield more specific results. Finally, this study focuses specifically on word text analysis; it does not take into consideration the context of word usage. It is possible that letter writers mention specific qualities equally between men and women, but the manner in which applicants are described differs linguistically between the genders.

The results of this study have implications for future analysis. Further work using a qualitative analysis approach to the LOR content is needed to comment on the context of the descriptive words and to investigate thematic differences in the

---

**TABLE V ORs from the Subgroup Analysis Stratified by the Type of LOR—Traditional or Standardized**

<table>
<thead>
<tr>
<th></th>
<th>Traditional LOR Only</th>
<th></th>
<th>Standardized LOR Only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>By Gender (Reference: Male)</td>
<td>By Race (Reference: White)</td>
<td>By Gender (Reference: Male)</td>
</tr>
<tr>
<td>Agentic</td>
<td>1.02 (0.97-1.07)</td>
<td>0.78 (0.93-1.02)</td>
<td>1.03 (0.89-1.20)</td>
</tr>
<tr>
<td>Communal</td>
<td>1.06 (0.99-1.12)</td>
<td>0.98 (0.93-1.03)</td>
<td>1.06 (0.90-1.25)</td>
</tr>
<tr>
<td>Grindstone</td>
<td>0.99 (0.95-1.03)</td>
<td><strong>0.96 (0.93-0.99)</strong></td>
<td>0.99 (0.89-1.11)</td>
</tr>
<tr>
<td>Ability</td>
<td>1.01 (0.95-1.08)</td>
<td>1.00 (0.94-1.06)</td>
<td>1.02 (0.87-1.19)</td>
</tr>
<tr>
<td>Standout</td>
<td><strong>1.07 (1.02-1.12)</strong></td>
<td><strong>1.05 (1.002-1.10)</strong></td>
<td>1.00 (0.86-1.15)</td>
</tr>
</tbody>
</table>

*CI = confidence interval, OR = odds ratio, and LOR = letters of recommendation. †Statistically significant odds ratios marked in bold.*

---

(continued...
content. Studies, such as this, have the potential to impact applicants to the field of orthopaedic surgery and the way LOR are viewed during the application process.

Alexa Powers, BA1,2
Katherine M. Gerull, MD1
Rachel Rothman1
Sandra A. Klein, MD1
Rick W. Wright, MD1
Christopher J. Dy, MD, MPH1

1Department of Orthopaediac Surgery, Washington University School of Medicine, St. Louis, Missouri
2Saint Louis University School of Medicine, St. Louis, Missouri

E-mail address for C.J. Dy: dyc@wustl.edu

ORCID iD for A. Powers: 0000-0003-2139-2202
ORCID iD for K.M. Gerull: 0000-0001-8534-2963
ORCID iD for R. Rothman: 0000-0002-7731-0176
ORCID iD for S.A. Klein: 0000-0003-2416-1186
ORCID iD for R.W. Wright: 0000-0003-4018-7132
ORCID iD for C.J. Dy: 0000-0003-1422-2483

References


