

**Diversity and Disparities in Orthopaedic Surgery (Guest Editors Alice Chu MD,
Selina Poon MD, MPH)**

Are There Gender-based Differences in Language in Letters of Recommendation to an Orthopaedic Surgery Residency Program?

Audrey N. Kobayashi PhD, Robert S. Sterling MD, Sean A. Tackett MD MPH, Brant W. Chee PhD,
Dawn M. Laporte MD, Casey Jo Humbyrd MD

Received: 3 June 2019 / Accepted: 28 October 2019 / Published online: 13 November 2019
Copyright © 2019 by the Association of Bone and Joint Surgeons

Abstract

Background Letters of recommendation are considered one of the most important factors for whether an applicant is selected for an interview for orthopaedic surgery

Each author certifies that neither he or she, nor any member of his or her immediate family, has funding or commercial associations (consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

Each author certifies that his or her institution approved the reporting of this investigation and that all investigations were conducted in conformity with ethical principles of research. This work was performed at Johns Hopkins University School of Medicine, Baltimore, MD, USA.

A. N. Kobayashi, Johns Hopkins University School of Medicine, Baltimore, MD, USA

R. S. Sterling, D. M. Laporte, C. J. Humbyrd, Department of Orthopaedic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, USA

S. A. Tackett, Department of Medicine, Johns Hopkins University School of Medicine, Biostatistics, Epidemiology and Data Management (BEAD) Core, Baltimore, MD, USA

B. W. Chee, Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA

A. N. Kobayashi (✉), Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD 21287 USA, Email: ankobayashi@jhmi.edu

All ICMJE Conflict of Interest Forms for authors and *Clinical Orthopaedics and Related Research*® editors and board members are on file with the publication and can be viewed on request.

residency programs. Language differences in letters describing men versus women candidates may create differential perceptions by gender. Given the gender imbalance in orthopaedic surgery, we sought to determine whether there are differences in the language of letters of recommendation by applicant gender.

Questions/purposes (1) Are there differences in word count and word categories in letters of recommendation describing women and men applicants, regardless of author gender? (2) Is author gender associated with word category differences in letters of recommendation? (3) Do authors of different academic rank use different words to describe women versus men applicants?

Methods Using a linguistic analysis in a retrospective study, we analyzed all letters of recommendation (2834 letters) written for all 738 applicants with completed Electronic Residency Application Service applications submitted to the Johns Hopkins Orthopaedic Surgery Residency program during the 2018 to 2019 cycle to determine differences in word category use among applicants by gender, authors by gender, and authors by academic rank. Thirty nine validated word categories from the Linguistic Inquiry and Word Count dictionary along with seven additional word categories from previous publications were used in this analysis. The occurrence of words in each word category was divided by the number of words in the letter to obtain a word frequency for each letter. We calculated the mean word category frequency across all letters and analyzed means using non-parametric tests. For comparison of two groups, a p value threshold of 0.05 was used. For comparison of multiple groups, the Bonferroni correction was used to calculate an adjusted p value (p = 0.00058).

Results Letters of recommendation for women applicants were slightly longer compared with those for men applicants (366 ± 188 versus 339 ± 199 words; $p = 0.003$). When comparing word category differences by applicant gender, letters for women applicants had slightly more “achieve” words (0.036 ± 0.015 versus 0.035 ± 0.018 ; $p < 0.0001$). Letters for men had more use of their first name (0.016 ± 0.013 versus 0.014 ± 0.009 ; $p < 0.0001$), and more “young” words (0.001 ± 0.003 versus 0.000 ± 0.001 ; $p < 0.0001$) than letters for women applicants. These differences were very small as each 0.001 difference in mean word frequency was equivalent to one more additional word from the word category appearing when comparing three letters for women to three letters for men. For differences in letters by author gender, there were no word category differences between men and women authors. Finally, when looking at author academic rank, letters for men applicants written by professors had slightly more “research” terms (0.011 ± 0.010) than letters written by associate professors (0.010 ± 0.010) or faculty of other rank (0.009 ± 0.011 ; $p < 0.0001$), a finding not observed in letters written for women.

Conclusions Although there were some minor differences favoring women, language in letters of recommendation to an academic orthopaedic surgery residency program were overall similar between men and women applicants.

Clinical Relevance Given the similarity in language between men and women applicants, increasing women applicants may be a more important factor in addressing the gender gap in orthopaedics.

Introduction

Letters of recommendation are a critical component for evaluating orthopaedic surgery residency candidates. These letters offer valuable insight into a candidate’s personality, work ethic, ability to function on a team, operative capability, and bedside manner, attributes that can predict a candidate’s potential for success as a resident and future attending surgeon. According to a survey of residency program directors in orthopaedic surgery, letters of recommendation are the second-most important factor used to screen candidates for interview selection, after the United States Medical Licensing Examination Step 1 score [12]. Despite the value of letters of recommendation, previous studies have shown that letters of recommendation outside orthopaedic surgery are subject to gender bias [3, 8, 9, 11, 16, 19]. These biases may exist across different formats, including the Medical Student Performance Evaluation (“Dean’s Letter”), narrative letters, and standardized letters [3, 8, 9, 11, 16]. Positive words in letters of recommendation have been correlated with other elements of a candidate’s application, including United States Medical Licensing Examination Step 1 scores [11].

Analyses of letters in non-orthopaedic surgery specialties found that women were more often described using “compassion,” “caring,” “empathic,” “teaching,” and “grindstone” words, whereas men applicants were more often described with “research,” “ability,” “standout,” and “leadership” words [3, 8, 9, 11, 19]. A study of general-surgery residency letters of recommendation found that letters for women applicants had more “standout” terms than letters for men applicants [7]. Studies also suggest the author gender affects how men and women applicants are described, as well as the letter length [7-9, 11].

According to 2016 to 2017 data from the Association of American Medical Colleges, orthopaedic surgery residency programs had the lowest percentage (15.1% in 2017) of women residents among major residency disciplines [2]. Further, orthopaedic surgery had the lowest percentage increase (27%) in female residents among surgical specialties between 2005 and 2017 [6]. Finally, when comparing major medical specialties, orthopaedic surgery has the lowest number of women who are full-time academic faculty members [6]. Strategies such as promoting women mentors, early exposure to the field, and dispelling misperceptions of the specialty have been proposed as possible ways to reverse the current trend [4, 5, 14, 15]. Understanding potential underlying biases in letters of recommendation is particularly salient for orthopaedic surgery, given the stark gender imbalance in the field.

We therefore analyzed the letters of recommendation of applicants to an orthopaedic surgery residency program to answer the following questions: (1) Are there differences in word count and word categories in letters of recommendation describing women and men applicants, regardless of author gender? (2) Is author gender associated with word category differences in letters of recommendation? (3) Do authors of different academic rank use different words to describe women versus men applicants?

Materials and Methods

Study Design and Setting

The current study was a retrospective study of all letters of recommendation submitted to the Johns Hopkins Orthopaedic Surgery Residency program in Baltimore, MD, USA, for the 2018 to 2019 application cycle. The study was approved by the institutional review board of Johns Hopkins University School of Medicine.

Material Analyzed

All completed applications submitted through the Electronic Residency Application Service were included in the

study. Descriptive data (applicant age, gender, and self-reported race) were extracted from the Electronic Residency Application Service system along with all letters of recommendation associated with the applicant. Medical Student Performance Evaluation letters were not included in the analysis because of their variability across medical schools. The author's gender and academic rank were recorded for each letter of recommendation. If a letter was written by several authors, the author with the highest academic rank was recorded as the letter author (114 letters with multiple authors or 4% of all letters). The letters were converted from PDF to rich text format for analysis using PDF Studio Pro (Qoppa Software, Atlanta, GA, USA). For letters that used the American Orthopedic Association Standardized Letter of Recommendation template, the personal comments section was considered the narrative letter of recommendation. There were 738 applicants to the Johns Hopkins Orthopaedic Surgery Residency Program during the 2018 to 2019 application cycle. We included 2834 letters of recommendation in the study.

Variables, Outcome Measures, Data Sources, and Bias

For this study, we analyzed letters using the word category dictionary available in Linguistic Inquiry and Word Count software (LIWC2015, Pennebaker Conglomerates, Austin, TX, USA). Previous studies have quantified gender differences in letters of recommendation using this software [7, 9, 10, 17]. Words are grouped into predetermined word categories that allow text analysis across a range of psychological, social, and emotional categories that have been previously validated [18]. Of the 77 predefined Linguistic Inquiry and Word Count categories available in the software, 39 word categories were used in this analysis. These categories were chosen before analysis based on categories that would be relevant to letters of recommendation. "Male" and "female" word categories were included as internal validation. Five additional word categories defined by Trix and Psenka [19] ("grindstone," "ability," "standout," "research," and "teaching") were also used. Letters were edited to remove the first and last name of the applicants, and placeholder categories were substituted in the letter. Forty-six word categories were used in this analysis (Supplemental Digital Content 1, <http://links.lww.com/CORR/A263>). We analyzed letters by counting the occurrence of words within a word category and then dividing by the number of total words in a letter to calculate the word category frequency. This frequency was then averaged over all letters to calculate the mean word category frequency. Based on the average length of the letters in our sample, a 0.001 difference in mean word frequency translates into one more word from that word

category appearing when you compare three letters for one gender with three letters for the other. We compared differences in word category frequency by applicant gender, author gender, and author academic rank. In addition, we compared mean word category frequencies between men and women applicants stratified by author gender and author academic rank. Authors were separated into three categories of academic rank: "professor," "associate professor," and "other."

Statistical Analysis

When comparing mean word frequencies, we used non-parametric tests. We used the Wilcoxon rank sum test for comparing word categories between letters of recommendation for men and women applicants. This test was also used to compare word categories between letters of recommendation written by men and women authors. To compare word categories used by authors of different academic ranks, we used the Kruskal-Wallis test. To compare proportions for descriptive analyses, we used chi-square testing. When making multiple comparisons, we used the Bonferroni correction to calculate an adjusted p value ($p = 0.00058$) threshold for statistical significance. When not making multiple comparisons, a p value less than 0.05 was considered statistically significant. For comparing letter length between men and women applicants, the p value threshold of 0.05 was used (question 1). For comparing word categories by applicant gender (question 1), by author gender (question 2), and by author academic rank (question 3), the p value threshold of 0.00058 was used. Stata version 13 (Stata Corp, College Station, TX, USA) was used for data analysis.

Description of Study Population

The study sample included mostly men applicants (80%) with a mean of 4 ± 0.4 letters per candidate (Table 1). The mean age and race distributions were similar between men and women applicants. Letters for both women and men applicants had more men authors than women authors. There were nearly 10 times as many men authors as women authors (2588 men and 246 women). Women authors accounted for 9% of letter authors. Men applicants had a greater number of letters written by men authors (92%) than did women applicants (88%). The distribution of letter authors by academic rank was similar between men and women applicants. Women authors with the rank of associate professor were the most represented followed by those with the rank of other academic ranks. Men authors were most often at the rank of professor compared with men authors of other academic ranks. Men professors

Table 1. Basic characteristics of letters of recommendation for orthopaedic surgery residency applicants, 2018 to 2019

Parameter	All applicant letters	Letters for men applicants	Letters for women applicants	p value
Number of applicants	738	586 (80%)	152 (21%)	
Number of letters	2834	2244 (79%)	590 (21%)	
Letters per applicant, mean \pm SD	4 \pm 0.4	4 \pm 0.4	4 \pm 0.4	0.164
Applicant age (years), mean \pm SD	28 \pm 2.9	28 \pm 2.9	28 \pm 2.7	0.271
Applicant race				0.222
White	1701 (60%)	1343 (60%)	358 (61%)	
Black	153 (5%)	115 (5%)	38 (6%)	
Asian	485 (17%)	380 (17%)	105 (18%)	
Hispanic	98 (4%)	81 (4%)	17 (3%)	
Other	235 (8%)	199 (9%)	36 (6%)	
Not reported	162 (6%)	126 (6%)	36 (6%)	
Author gender				0.001
Men	2588 (91%)	2069 (92%)	519 (88%)	
Women	246 (9%)	175 (8%)	71 (12%)	
Author rank				0.712
Professor	1201 (42%)	943 (42%)	258 (44%)	
Associate professor	768 (27%)	609 (27%)	159 (27%)	
Other	865 (31%)	692 (31%)	173 (29%)	

Selected p values comparing men and women applicants are also reported.

comprised the largest proportion of letter writers for both men and women applicants.

Results

Letters of recommendation for women had slightly more words and some minor word category differences compared with letters for men, though these differences were small. Letters written for women were slightly longer than letters written for men, regardless of author gender or academic rank (mean 366 \pm 188 words versus mean 339 \pm

188 words, respectively, $p = 0.003$, Table 2). The word category analysis of letters of recommendation for men and women applicants showed that letters for women applicants had more “achieve” words (mean frequency 0.036 \pm 0.015 versus 0.035 \pm 0.018, $p < 0.0001$) and had less use of their first name than letters for men applicants (mean frequency 0.014 \pm 0.013 versus 0.016 \pm 0.009, $p < 0.0001$, Table 3). Letters of recommendation for men applicants had greater use of “young” terms than those for women applicants (mean frequency 0.001 \pm 0.003 versus 0.000 \pm 0.001, $p < 0.0001$). These differences were very small as each 0.001 difference in mean word frequency was

Table 2. Differences in word counts of letters of recommendation written for applicants to an orthopaedic surgery residency program, 2018 to 2019

Parameter	All applicant letters	Letters for men	Letters for women	p value, men vs	
				women applicants	p value, author traits
Total	345 \pm 197	339 \pm 199	366 \pm 188	0.003	
Author gender					
Men	341 \pm 196	337 \pm 199	359 \pm 185	0.024	0.002
Women	382 \pm 200	366 \pm 198	422 \pm 201	0.044	
Author rank					
Professor	350 \pm 199	345 \pm 202	367 \pm 185	0.126	0.281
Associate professor	347 \pm 207	345 \pm 212	352 \pm 183	0.721	
Other	336 \pm 199	325.4 \pm 180	379.1 \pm 198	0.001	

Word counts are presented as the mean \pm SD. Selected p values comparing men and women applicants as well as comparing author traits are reported.

Table 3. Differences in word categories among letters of recommendation written for men and women applying to orthopaedic surgery residency, 2018 to 2019

Word category	Letters for men (n = 2244)		Letters for women (n = 590)		p value
	Mean	SD	Mean	SD	
Ability	0.006	0.007	0.007	0.006	0.002
Achieve	0.035	0.018	0.036	0.015	< 0.0001
Affect	0.056	0.021	0.057	0.020	0.168
Agentic terms	0.009	0.015	0.009	0.010	0.009
Communal terms	0.003	0.005	0.003	0.005	0.234
Drive	0.088	0.032	0.088	0.026	0.759
Female	0.000	0.003	0.057	0.017	< 0.0001
Male	0.056	0.017	0.001	0.003	< 0.0001
First name	0.016	0.013	0.014	0.009	< 0.0001
Grindstone	0.011	0.010	0.011	0.008	0.604
Insight	0.021	0.012	0.021	0.010	0.332
Leader	0.000	0.000	0.000	0.000	0.305
Long terms	0.307	0.052	0.313	0.044	0.002
Power	0.029	0.018	0.029	0.017	0.983
Research	0.010	0.010	0.011	0.010	0.317
Social	0.108	0.025	0.108	0.024	0.999
Standout	0.010	0.014	0.011	0.010	0.003
Teaching	0.013	0.014	0.013	0.009	0.287
Young	0.001	0.003	0.000	0.001	< 0.0001

Word categories are presented as the mean \pm SD. Mean values represent the average word frequency in a letter that are from that word category (for example, 0.006 means on average 0.6% of words in letters for men were from the Ability word category).

equivalent to one more word from the word category appearing when comparing three letters for women to three letters for men. There were no differences in the use of other word categories including “ability,” “affect,” “agentic,” “communal,” “drive,” “grindstone,” “insight,” “leader,” “power,” “research,” “social,” “standout,” and “teaching” words between letters of recommendation for men and women applicants (Supplemental Digital Content 2, <http://links.lww.com/CORR/A264>).

There were no differences in word categories between men and women authors (Table 4). The use of other word categories, including “ability,” “achieve,” “affect,” “communal,” “drive,” “grindstone,” “leader,” “research,” “social,” “standout,” and “teaching” terms, did not differ between men and women authors (Supplemental Digital Content 3, <http://links.lww.com/CORR/A265>).

Letters written for men by professors were more likely to contain “research” terms than those written by associate professors or writers of other ranks (mean frequency 0.011 ± 0.010 versus 0.010 ± 0.010 or 0.009 ± 0.011 , respectively, $p < 0.0001$, Table 5). This pattern was not observed among letters of recommendation for women applicants. There was no difference in the use of “ability,” “achieve,” “grindstone,” “standout,” or “teaching” words

between men and women applicants when stratified by author rank (Supplemental Digital Content 4, <http://links.lww.com/CORR/A266>)

Discussion

Previous studies outside of orthopaedic surgery have demonstrated differences in language used to describe men and women applicants [3, 8, 9, 11, 16, 19]. We sought to understand whether similar biases exist in the language used in letters of recommendation for men and women applicants to an academic orthopaedic residency program. Elucidating these differences may yield insight into the current gender imbalance in the field. Although there were very minor differences in word choices that seemed to favor women applicants, overall language used to describe candidates was similar.

This study had several limitations. This was a single-institution study using letters from a single year that may not represent orthopaedic surgery residency programs in general. Data from the National Resident Matching Program for the 2018 to 2019 application year showed that 1037 medical students applied for orthopaedic surgery

Table 4. Differences in word categories in letters of recommendation between men and women applying to an orthopaedic surgery residency program, 2019 to 2019, stratified by author gender

Word category	Letters for men (n = 2244)					Letters for women (n = 590)				
	Men authors		Women authors		p value	Men authors		Women authors		p value
	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Ability	0.006	0.007	0.007	0.006	0.023	0.007	0.006	0.008	0.005	0.177
Achieve	0.035	0.018	0.038	0.019	0.008	0.036	0.015	0.040	0.013	0.006
Affect	0.056	0.022	0.056	0.017	0.198	0.057	0.021	0.053	0.015	0.081
Communal terms	0.003	0.005	0.003	0.004	0.394	0.003	0.005	0.002	0.003	0.182
Drive	0.088	0.032	0.091	0.026	0.019	0.087	0.027	0.089	0.019	0.320
Grindstone	0.011	0.011	0.012	0.008	0.175	0.010	0.008	0.012	0.006	0.010
Insight	0.021	0.012	0.022	0.011	0.145	0.020	0.010	0.022	0.009	0.317
Leader	0.000	0.000	0.000	0.000	0.200	0.000	0.000	0.000	0.000	n/a
Power	0.029	0.018	0.030	0.014	0.282	0.029	0.017	0.028	0.010	0.872
Research	0.010	0.010	0.010	0.010	0.757	0.011	0.010	0.011	0.010	0.839
Social	0.107	0.026	0.112	0.023	0.038	0.108	0.023	0.110	0.026	0.906
Standout	0.010	0.014	0.010	0.009	0.745	0.011	0.011	0.010	0.007	0.419
Teaching	0.013	0.014	0.012	0.007	0.908	0.013	0.010	0.014	0.006	0.216

Word categories are presented as the mean ± SD. Mean values represent the average word frequency in a letter that are from that word category (for example, 0.006 means on average 0.6% of words in letters for men by men authors were from the Ability word category).

residency [13]. Our sample of 738 applicants to our academic orthopaedic surgery program is 71% of this reported number and thus encompasses most applicants for the 2018 to 2019 year. The study sample had a higher representation of women (21%) compared with the overall applicant pool,

which comprised 16% women according to Association of American Medical Colleges data [1]. The study sample also had fewer women writers who were faculty members (9%) than the average representation of women faculty in orthopaedics (18%) [6]. A similar discrepancy has been

Table 5. Differences in word categories in letters of recommendation between men and women applying to an orthopaedic surgery residency program, 2018 to 2019, stratified by author academic rank

Word category	Letters for men (n = 2244)							Letters for women (n = 590)						
	Professor		Associate professor		Other		p value	Professor		Associate professor		Other		p value
	Mean	SD	Mean	SD	Mean	SD		Mean	SD	Mean	SD	Mean	SD	
Ability	0.006	0.007	0.007	0.008	0.007	0.007	0.137	0.007	0.007	0.007	0.006	0.007	0.006	0.293
Achieve	0.034	0.019	0.035	0.017	0.035	0.017	0.022	0.036	0.015	0.038	0.016	0.035	0.014	0.181
Affect	0.055	0.021	0.057	0.023	0.056	0.021	0.125	0.058	0.023	0.057	0.019	0.055	0.018	0.755
Communal terms	0.003	0.005	0.003	0.006	0.003	0.005	0.162	0.003	0.005	0.004	0.005	0.003	0.003	0.342
Drive	0.088	0.035	0.088	0.031	0.089	0.027	0.089	0.088	0.027	0.087	0.023	0.088	0.028	0.964
Female	0.001	0.004	0.001	0.003	0.000	0.001	0.473	0.057	0.017	0.060	0.018	0.054	0.017	0.007
Grindstone	0.011	0.009	0.011	0.010	0.011	0.013	0.203	0.010	0.009	0.012	0.008	0.010	0.007	0.039
Male	0.055	0.018	0.055	0.017	0.057	0.018	0.380	0.001	0.004	0.001	0.002	0.001	0.002	0.193
Research	0.011	0.010	0.010	0.010	0.009	0.011	< 0.0001	0.011	0.009	0.011	0.010	0.011	0.011	0.850
Social	0.106	0.026	0.108	0.025	0.109	0.025	0.491	0.109	0.023	0.108	0.024	0.107	0.024	0.683
Standout	0.011	0.018	0.010	0.011	0.010	0.009	0.511	0.012	0.010	0.012	0.010	0.011	0.011	0.053
Teaching	0.013	0.019	0.013	0.009	0.013	0.009	0.653	0.013	0.009	0.013	0.009	0.014	0.010	0.946
Tentative	0.015	0.011	0.016	0.012	0.016	0.011	0.006	0.015	0.010	0.015	0.009	0.017	0.010	0.018

Word categories are presented as means ± SDs. Mean values represent the average word frequency in a letter that are from that word category (e.g. 0.006 means on average 0.6% of words in letters for men by professors were from the Ability word category).

noted in studies analyzing letters of recommendation for surgical residency and suggests that applicants preferentially ask men faculty over women faculty for letters of recommendation [7, 8, 11]. If applicants believe that letters from writers of higher academic rank carry more weight, then the larger proportion of men at higher academic rank could be one explanation for this difference [6]. Despite some of these differences, we believe our findings are representative of general trends in the 2018 to 2019 application year given that our overall study sample size approached the total number of candidates applying to orthopaedics.

Another limitation is that word category frequencies may not be sensitive enough to capture subtle biases expressed in letters of recommendation. For example, both “the applicant has skills” and “the applicant does not have skills” were counted the same when looking for the word “skill.” In addition, the analysis depended on exact word matches and did not account for misspellings. Despite these shortcomings, word frequency analysis, particularly using LIWC categories, has been used in multiple studies of letters of recommendation in and outside of medicine [7, 9, 10, 17]. More advanced language analysis capable of detecting subtleties such as hedging or hidden language may reveal more subtle differences.

In our analysis, we attributed letters written by multiple authors to the author with the highest academic rank. It was hard to know exactly who the true author is in letters with multiple authors. Some letters may truly be an amalgam of all author’s thoughts. Others may be written by one member and signed off by the others. Some letters could have been written by the junior-most faculty member. This would have affected our results. Letters with multiple authors only represented 4% of the study sample and any error introduced would not be expected to exert a huge effect on our findings. Further, for this study, the Medical Student Performance Evaluation was excluded from the analysis since these letters are often collections of comments from numerous authors with unknown genders. Word category analyses of this type of letter would not have allowed for proper attribution of word category differences with author gender or academic rank, and their inclusion would have reduced the ability to detect differences.

Although responses to the global rating items on standardized letters were not analyzed, text written in the personal comments section was included since it could influence a program’s impression of an applicant. Since standardized letters of recommendation are not mandated for the orthopaedic residency application process and one would expect these written comments to contain similar underlying biases, we felt that treating these comments equal to letters of recommendation was the most conservative approach.

We found few differences between letters written for men and women. The few differences we observed—including slightly higher word count and slightly greater use of “achieve” words—seemed to favor women candidates, although this is hard to conclude definitively. Our study’s findings are in line with studies from surgical fields that have found relatively similar lengths in letters [7, 11]. It is hard to know whether the mean difference of 27 words between men and women letters is enough to make an impact on a reader. Previous studies on letters of recommendation outside of surgical specialties have suggested that longer letters are favorable [19, 20]. It is difficult to ascertain, though, whether a longer letter is positive in every case because authors may write longer letters for candidates they know well or could be using words to talk around a candidate’s weaknesses.

In terms of language difference by applicant gender, the results of our analysis were very similar to those done in other surgical specialties, which showed relatively little difference in language [7, 11]. In a study by French et al. [7], only the one word category of “standout” terms were more frequent for women. It is also notable that the effect size was similar with a difference in mean word frequency of 0.001, similar to our study’s small differences. We also found greater use of “young” terms in letters about men, which was also seen in the study by Messner et al. [11]. In conjunction with the greater use of first names for men applicants, we believe this represents a bias against terms like “young woman” or colloquial use of first name for women applicants, which may be seen as derogatory or stereotypical.

It is noteworthy that the lack of major differences in word categories describing surgical program residency applicants is in stark contrast to a number of studies outside of surgery that have shown more differences by applicant gender [3, 9, 19]. It may be that letters writers in surgical fields are influenced by the gender disparity in their field [6] and feel compelled to promote women applicants. Another possibility is that authors have enough knowledge of the applicant to overcome any stereotyped biases and focus on their unique attributes and accomplishments. Finally, authors may all purposefully focus on describing a prescribed set of attributes (for example, operative skill or medical knowledge), which reduces the variability between letters. Overall, this study in combination with others [7, 11] seems to support the theory that letters for men and women applicants are largely homogeneous rather than heterogeneous. It is hard to comment with a 1-year study whether there are changes over time and therefore studying temporal trends could yield important insight.

We also found that there was no difference in the language used by men and women authors in letters of recommendation. There is conflicting evidence on whether author gender influences the language of letters. Referring

to the studies from surgical residency applicants, one study found no difference by author gender [7] while others found that men authors tended to focus more on personal life, family history, and hobbies for men applicants while women authors focused on professional accomplishments and were more likely to write letters of minimal assurance [8, 11]. As noted above, authors may have a fixed idea of what aspects of an applicant to comment on in letters of application, which reduces variation by author gender. Alternatively, perhaps women authors are influenced by what they think their men colleagues want to read about in an applicant and actively tailor their language and focus.

In the analysis by author academic rank, there were very few differences in language describing men and women applicants. Professors who wrote letters for men applicants used more “research” words than did authors of other academic ranks; however, this was not observed for women applicants. Other studies on letters of recommendations in surgical residencies have not expressly analyzed differences in language by author academic rank [7, 8, 11]. They have noted that professors make up a large portion of letter writers, although this is driven by the need for a chairman letter [7, 8]. One possibility for the finding in our study is that professors may have an emphasis on research-related activities compared with authors of a lower rank. An alternative possibility is that professors may have only superficial familiarity with applicants and therefore focus on research. This could be particularly true for letters that are written as an obligatory letter from a chairperson. It seems reasonable that how much an author knows, and thus describes, an applicant is not necessarily driven by the author’s rank.

In our study of letters of recommendation to an academic orthopaedic residency program, we did not find major differences in language used to describe men and women candidates. Given the contrast to other studies that have reported a difference in the language used for applicants of different genders [3, 8, 9, 11, 19], the results of our study suggest that either there is no gender bias in these letters or that authors are trying to overcome gender biases that might otherwise be present. Although it is hard to say definitively, we believe that the latter is more likely. It would be interesting to know whether the similarity in language has evolved over time.

One may assert that standardized letters of recommendation are superior to narrative letters as they force a more objective ranking of an applicant. Studies on standardized letters have suggested that these letters are more objective, as well as more efficient and comparable [8, 11]. We agree with Friedman et al. [8] that narrative comments in letters of recommendation may be subject to bias but it would be a mistake to solely rely on standardized scales to describe an applicant’s characteristics. Important information can be gleaned from letters of recommendation and perhaps efforts

should be dedicated to mandating the format of letters more rather than leaving it up to authors on whether to use a fully standardized or fully narrative format. If letters of recommendation for applicants are more similar than not, it suggests that increasing women applicants to orthopaedics may be a more fruitful way to address the gender inequality in the field.

Acknowledgments None.

References

1. Association of American Medical Colleges. Orthopaedic surgery. Available at: <https://www.aamc.org/download/359600/data/orthopaedicsurgery.pdf>. Accessed September 5, 2019.
2. Association of American Medical Colleges. Report on Residents. Table B3: Number of Active Residents, by Type of Medical School, GME Specialty, and Sex. Available at: <https://www.aamc.org/data/493922/report-on-residents-2018-b3table.html>. Accessed May 3, 2019.
3. Axelson RD, Solow CM, Ferguson KJ, Cohen MB. Assessing implicit gender bias in medical student performance evaluations. *Eval Health Prof.* 2010;33: 365-385.
4. Baldwin K, Namdari S, Bowers A, Keenan MA, Levin LS, Ahn J. Factors affecting interest in orthopedics among female medical students: a prospective analysis. *Orthopedics.* 2011;34: e919-932.
5. Blakemore LC, Hall JM, Biermann JS. Women in surgical residency training programs. *J Bone Joint Surg Am.* 2003;85: 2477-2480.
6. Chambers CC, Ilnow SB, Monroe EJ, Suleiman LI. Women in orthopaedic surgery: population trends in trainees and practicing surgeons. *J Bone Joint Surg Am.* 2018;100:e116.
7. French JC, Zolin SJ, Lampert E, Aiello A, Bencsath KP, Ritter KA, Strong AT, Lipman JM, Valente MA, Prabhu AS. Gender and letters of recommendation: a linguistic comparison of the impact of gender on general surgery residency applicants. *J Surg Educ.* 2019;76:899-905.
8. Friedman R, Fang CH, Hasbun J, Han H, Mady LJ, Eloy JA, Kalyoussef E. Use of standardized letters of recommendation for otolaryngology head and neck surgery residency and the impact of gender. *Laryngoscope.* 2017;127:2738-2745.
9. 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;86:59-66.
10. Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. *J Appl Psychol.* 2009;94:1591-1599.
11. Messner AH, Shimahara E. Letters of recommendation to an otolaryngology/head and neck surgery residency program: their function and the role of gender. *Laryngoscope.* 2008;118: 1335-1344.
12. National Resident Matching Program. Results of the 2018 NRMP Program Director Survey National Resident Matching Program. Available at: <https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf>. Accessed May 3, 2019.
13. National Resident Matching Program. Results and Data: 2019 Main Residency Match. Available at: https://mk0nrmp3oyqui6wqfm.kinstacdn.com/wp-content/uploads/2019/04/NRMP-Results-and-Data-2019_04112019_final.pdf. Accessed September 5, 2019.

14. Nguyen L, Amin NH, Vail TP, Pietrobon R, Shah A. Editorial: a paucity of women among residents, faculty, and chairpersons in orthopaedic surgery. *Clin Orthop Relat Res.* 2010;468:1746-1748.
15. Rao RD, Khatib ON, Agarwal A. Factors motivating medical students in selecting a career specialty: relevance for a robust orthopaedic pipeline. *J Am Acad Orthop Surg.* 2017;25:527-535.
16. 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:e0181659.
17. Schmader T, Whitehead J, Wysocki VH. A linguistic comparison of letters of recommendation for male and female chemistry and biochemistry job applicants. *Sex Roles.* 2007;57:509-514.
18. Tausczik YRP, Pennebaker J.W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology.* 2010;29:24-54.
19. Trix FP, Psenka C. Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse and Society.* 2003;14:191-220.
20. Watson C. Sex-linked differences in letters of recommendation. *Women and Language.* 1987;10:26-30.