ORIGINAL RESEARCH

Laryngoscope Investigative Otolaryngology

Sex bias in letters of recommendation and personal statements for otolaryngology residency

Garrett A. Berk BS | Tiffany D. Ho BS | Taylor J. Stack-Pyle BS | Abdullah Zeatoun MD | Keonho A. Kong MD | Mark B. Chaskes MD, MBA | Brian D. Thorp MD | Charles S. Ebert MD, MPH | Brent A. Senior MD | Adam J. Kimple MD, PhD | Christine E. DeMason MD

Department of Otolaryngology—Head & Neck Surgery, University of North Carolina, Chapel Hill, North Carolina, USA

Correspondence

Christine E. DeMason, 170 Manning Drive, CB #7070, Physician's Office Building, Room G-190, Chapel Hill, NC 27599, USA. Email: christine_demason@med.unc.edu

Funding information Center for Scientific Review, Grant/Award Number: KL2TR002490

Abstract

Objective: Application for otolaryngology residency is highly competitive, with letters of recommendation (LORs) and applicant personal statements (PSs) representing important components of the application process. However, their inherently subjective nature predisposes them to potential implicit bias. Otolaryngology has historically been predominated by male physicians and while implicit sex bias has been demonstrated in LORs for application to residency of multiple specialties, data is limited for otolaryngology.

Methods: LORs and PSs for all otolaryngology applicants to an academic medical center during the 2019–20 and 2020–21 cycles were abstracted. Quantitative analysis was performed using Linguistic Inquiry and Word Count 2015 (LIWC2015), a validated software application designed to analyze various emotional, cognitive, and structural components of written text.

Results: LORs written for females were found to be written from a perspective of higher expertise and confidence while LORs written for males were associated with a more honest, personal, and disclosing tone. Moreover, LORs written for female applicants were found to reference achievement and "grindstone" terminology more than those written for men. No differences were observed in any word category between PSs written by male and female applicants.

Conclusion: Minor linguistic differences exist in multiple domains between LORs written for male and female applicants for otolaryngology residency. These tended to favor female applicants, with their letters demonstrating higher clout, achievement, and grindstone scores. This trend was unexpected in this historically predominantly male specialty. While differences were statistically significant, the overall difference in an entire letter of recommendation is likely subtle.

Level of Evidence: 3.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made. © 2022 The Authors. *Laryngoscope Investigative Otolaryngology* published by Wiley Periodicals LLC on behalf of The Triological Society.

KEYWORDS

implicit sex bias, letters of recommendation, linguistic analysis, otolaryngology residency application, personal statements

1 | INTRODUCTION

Letters of recommendation (LORs) and personal statements (PSs) represent critical components of medical student applications for residency. They provide programs with insight into a candidate's personality, character, work ethic, communication skills, and medical aptitude that is not otherwise apparent from the other objective components of the application (including grades, board scores, and research experience). Results from the 2018 National Resident Matching Program (NRMP) Residency Program Director Survey highlight the importance of LORs, with only Step 1 of the United States Medical Licensing Examination (USMLE) being cited by a greater number of programs for use in selecting applicants to interview.¹ Letters of recommendation are even more significant, as beginning in 2022, USMLE Step 1 scores will be reported as Pass/Fail, making it more difficult for programs to screen and select applicants.

Although LORs are a crucial component of residency applications, their inherently subjective nature predispose them to potential bias, both explicit and implicit.² Implicit bias, by definition, is unconscious and can skew one's judgment in either a positive or negative direction.³ Previous research has demonstrated sex bias in academia, with LORs for men tending to be longer and containing more standout adjectives (such as "excellent" or "exceptional") and descriptors pertaining to research.⁴ LORs for residency applications in emergency medicine, ophthalmology, general surgery, and transplant surgery also demonstrate differences in language between letters or recommendation for men and women applicants.^{5–8}

For example, in LORs for ophthalmology, letters written for male applicants were determined to use more "authentic" words than those written for female applicants. Additionally, letters written for male applicants contained more "leisure" words and fewer "feel" words.⁶ In transplant surgery, LORs for male applicants were significantly more likely to contain terms such as "superb, intelligent, and exceptional" and were more likely to contain the phrase "future leader." Letter writers were more likely to describe female applicants using terms like "compassionate, calm, and delightful."⁸

Prior research investigating LORs for otolaryngology residency applications has largely focused on the development and use of a standardized LOR, rather than directly investigating bias.⁹⁻¹⁴ Two studies have previously investigated the impact of applicant sex on the content of their LORs for otolaryngology residency. Messner and Shimahara (2008) found that men and women applicants were similarly described, although men and women evaluators often describe applicants differently, regardless of applicant sex.¹⁵ In contrast, Friedman et al. (2017) found that while standardized LORs reduce sex bias, significant differences exist between the description of men and women applicants.¹⁶ Neither study utilized the linguistic analysis software used in this investigation that has become standard for similar analyses in recent years. Literature analyzing linguistic differences in applicant personal statements (PSs) is markedly limited; however, sex-based differences have been noted in applications to urology, internal medicine, pediatrics, and general surgery that mirror gender stereotypes found in social psychology research.¹⁷⁻²⁰ To our knowledge, no publications to date have investigated sex-based differences in PSs of applicants to otolaryngology residency.

According to the AAMC 2020 Physician Specialty Data Report, only 18.3% of active otolaryngologists were women.²¹ A better understanding of potential sex bias in residency applications is therefore particularly relevant for otolaryngology, given this extreme sex imbalance in the field. In this study, we sought to analyze the linguistic differences of LORs and PSs from otolaryngology residency applications on the basis of applicant sex.

2 | METHODS

2.1 | Study design, setting, and population

This study was a retrospective analysis of all LORs and PSs for all otolaryngology residency applications submitted to an academic medical center during the 2019–20 and 2020–21 cycles (n = 928). Two applicants from each cycle did not have LORs saved with their file and were excluded from analysis. The study was approved by the organization's institutional review board. Electronic residency application system (ERAS) files were pulled from internal department archives by an administrative coordinator and applicants were assigned a unique study identification number. Demographic information (including selfreported sex, race, and ethnicity) as well as board scores and quantitative research data were abstracted. Applicant characteristics can be found in Table 1.

2.2 | Text analysis

Text for LORs (n = 3505) and PSs (n = 928) were converted from PDF to Microsoft Word using Adobe Acrobat Pro 2020 (Adobe Systems, San Jose, California, USA). For those letters using a standardized template, the personal comments section at the end was copied for use as the narrative LOR. In order to standardize word counts and linguistic analysis across LORs, only the text included between the salutation and signature was exported. Quantitative analysis was performed using LIWC 2015.²²

Six LIWC2015 word categories were chosen for use in this analysis based on their relevance to applicant evaluation. These

	2019-20 cycle		2020-21 cycle		2019-21 cycles combined	
Parameter	Female applicants	Male applicants	Female applicants	Male applicants	Female applicants	Male applicants
Number of Applicants; N (%)	159 (35%)	293 (65%)	187 (39%)	289 (61%)	346	582
Number of Letters; N (%)	605 (35%)	1107 (65%)	700 (39%)	1093 (61%)	1305	2200
Letters per Applicant; mean (SD)	3.81 (0.43)	3.80 (0.41)	3.74 (0.45)	3.78 (0.42)	3.77 (0.44)	3.79 (0.41)
Race/ethnicity						
White	88 (55%)	165 (56%)	91 (49%)	169 (58%)	179 (52%)	334 (57%)
Asian	41 (26%)	71 (24%)	59 (32%)	55 (19%)	100 (29%)	126 (22%)
Black	14 (9%)	12 (4%)	12 (6%)	10 (3%)	26 (8%)	22 (4%)
Hispanic	8 (5%)	16(6%)	13 (7%)	28 (10%)	21 (6%)	44 (8%)
Other	3 (2%)	7 (2%)	2 (1%)	13 (5%)	5 (1%)	20 (3%)
Not Reported	5 (3%)	22 (8%)	10 (5%)	14 (5%)	15 (4%)	36 (6%)
Step 1 Score; mean (SD)	244.94 (11.18)	248.35 (25.44)	242.58 (12.69)	245.00 (13.63)	243.67 (12.06)	245.97 (13.10)
Step 2 Score ^a ; mean (SD)	252.08 (12.70)	252.45 (13.91)	253.75 (11.06)	253.57 (11.58)	252.94 (11.89)	253.07 (12.66)
^a 14.01% of applicants only had Step 1 score	es and no Step 2 scores include	ed in their application.				

Laryngoscope Investigative Otolaryngology 1747

include "analytical thinking," "clout," "authenticity," "emotional tone," "achievement," and "power." Five custom categories that were previously defined in the literature were also used: "grindstone," "ability," "standout," "research," and "teaching."23 The details of the data dictionaries used for this analysis are included in Supplemental Table S1. LIWC2015 analyzes bodies of text by counting the occurrence of words from a specified word category list and then dividing by the number of total words in that text to calculate the word category frequency (expressed as a percentage). This frequency was averaged over all letters to calculate the mean word frequencies for each category. The only exception to these calculations are the four summary categories-analytical thinking, clout, authenticity, emotional tone-which are scored based on previously published algorithms.²⁴ LIWC2015 rescales the output for these variables so that they reflect a 100-point scale ranging from 0 to 100. These variables are defined as follows:

- Analytical thinking–a high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking.
- Clout-a high number suggests that the author is speaking from the perspective of high expertise and is confident; low clout numbers suggest a more tentative, humble, even anxious style.
- Authentic-a high number is associated with a more honest, personal, and disclosing text; lower numbers suggest a more guarded, distanced form of discourse.
- Emotional tone-a high number is associated with a more positive, upbeat style; a low number reveals greater anxiety, sadness, or hostility. A number around 50 suggests either a lack of emotionality or different levels of ambivalence.

2.3 | Statistical analysis

A total of 11 linguistic word categories were evaluated in this study. Categories were chosen based on the authors' determination of relevancy to LOR/PS content and usefulness in evaluation of residency applicants. Three subsets of data were analyzed: ERAS cycle 2019–20, cycle 2020–21, and the two cycles combined. Within each sub-dataset, the mean differences between male and female applicants of each quality/characterization/variable were statistically assessed using t-test, when all the test's assumptions were met, or the Mann-Witney U test. Multiple testing was corrected using the Benjamini-Hochberg procedure to minimize false discovery rate (FDR). All analyses were conducted using R 4.1.2 (R Core Team; Vienna, Austria), and p values <.05 were considered statistically significant.

3 | RESULTS

From the application cycles studied, 3505 LORs and 928 PSs from 928 applicants were included for analysis. The total number of LORs

Applicant characteristics

TABLE 1

 TABLE 2
 Differences in word categories between LORs written for male and female applicants to otolaryngology residency,

 2019–2021 combined

	LORs for females		LORs for males		
Word category	Mean	SD	Mean	SD	p value
Analytical thinking	83.11	6.57	83.63	5.70	.54
Clout	82.86	4.78	81.94	5.11	.04
Authentic	5.32	2.86	6.05	4.09	.04
Emotional tone	95.19	4.00	94.82	4.36	.41
Achievement	3.74%	0.69%	3.61%	0.67%	.04
Power	3.14%	0.72%	3.10%	0.69%	.46
Standout	0.68%	0.26%	0.69%	0.29%	.74
Ability	0.76%	0.32%	0.73%	0.29%	.46
Grindstone	1.17%	0.35%	1.12%	0.35%	.04
Teaching	1.62%	0.52%	1.58%	0.49%	.46
Research	1.57%	0.72%	1.51%	0.75%	.26

Bold values significes p < 0.05.

TABLE 3 Differences in word categories between PSs written by male and female applicants to otolaryngology residency, 2019 to 2021 combined

	Female PSs		Male PSs		
Word Category	Mean	SD	Mean	SD	p value
Analytical thinking	88.16	7.68	88.41	7.03	.98
Clout	40.12	12.62	39.66	12.48	.84
Authentic	61.60	18.11	61.09	17.99	.84
Emotional tone	84.26	14.46	83.07	14.48	.27
Achievement	4.02%	1.26%	3.85%	1.20%	.27
Power	2.65%	0.90%	2.61%	0.89%	.84
Standout	0.33%	0.26%	0.32%	0.24%	.98
Ability	0.67%	0.37%	0.63%	0.37%	.27
Grindstone	0.75%	0.46%	0.74%	0.42%	.98
Teaching	1.0%	0.58%	1.01%	0.68%	.84
Research	0.93%	0.69%	0.97%	0.71%	.84

per applicant ranged from two to five with a median of four. Applicant characteristics can be found in Table 1. There was no statistically significant difference in number of letters per applicant or average length of letter/PS on the basis of sex.

When analyzing all applicants, LORs written for female applicants were found to have a higher mean score for clout (82.86 [4.78] > 81.94 [5.11], p = .0396) and greater mean score for achievement (3.74% [0.69%] > 3.61% [0.67%], p = .0396) and grindstone (1.17% [0.35%] > 1.12% [0.35%], p = .0437) word usage. LORs written for male applicants were found to have a higher mean authenticity score (6.05% [4.09%] > 5.32% [2.86%], p = .0396) (Table 2). No differences were observed in any word category between PSs written by male and female applicants (Table 3). Differences in word categories for LORs and PSs separated by application cycle can be found in Supplemental Tables S2, S3, S4, S5.

4 | DISCUSSION

Prior research outside of otolaryngology has demonstrated differences in the language of LORs used to describe male and female applicants in numerous specialties.⁵⁻⁸ However, this is not consistent across all investigations regarding LORs in residency applications, as an evaluation of sex-based differences in LORs to an orthopedic surgery residency program reported that language used was overall similar between men and women.²⁵

The aim of this study was to elucidate the extent to which sex bias is present in LORs of otolaryngology residency applicants. In a historically male dominated surgical subspecialty, we hypothesized that there may be an implicit bias favoring male applicants. What we found was in fact the opposite, with the majority of statistically significant differences in the language of LORs favoring female applicants. However, it is important to note that having statistical significance in mean differences does not imply practical relevance. For example, our analysis found that letter writers were statistically more likely to use language referencing achievement and grindstone traits when discussing female applicants compared to their male counterparts (mean frequency values of $3.74 \pm 0.69\%$ vs. $3.61 \pm 0.67\%$ (p = .0396) and $1.17 \pm 0.35\%$ vs. $1.12 \pm 0.35\%$ (p = .0437), respectively). Word category frequency scores outputted by LIWC2015 are expressed as a percentage of words that fall into the category. This means that a 0.1% difference in mean word frequency results from one additional word from a category in a 1000-word LOR/PS., given that the average length of letters in our sample was 388.85 words, program directors and other evaluators are highly unlikely to view an applicant more favorably solely due to a single additional descriptor per three letters of recommendation.

When comparing the four summary categories, the greater clout score in LORs written for females (82.86 vs. 81.94) would suggest that those letters were written from a perspective of greater expertise and confidence, while the greater authenticity score in LORs written for males (6.05 vs. 5.32) suggests a more honest, personal, and disclosing text. However, when these comparisons are viewed in the context of their 100-point scales, both variables are similar for both male and female applicants despite the statistically significant difference. Interestingly, similar studies investigating LORs for residency applications to ophthalmology and urology have also reported increased authenticity scores in letters for male applicants.^{6,26} The high-clout values for both groups suggests that authors are universally confident in their assessments of both male and female applicants, while the extremely lowauthenticity scores reflect a more guarded, distanced form of discourse used to describe both. Looking at the remaining summary variables for the LORs, the high-analytical thinking values indicate a more formal and logical writing style and the high-emotional tone values reflect a positive, upbeat style.

This study additionally sought to determine whether male and female applicants tend to inherently describe themselves differently in their own PSs. We hypothesized that if male and female applicants used different language to describe themselves, it might influence the language used by evaluators to do the same; however, this proved not to be the case as no differences were found between PSs on the basis of sex. When looking at the four summary categories for PSs, similar trends were observed for analytical thinking and emotional tone as compared to their respective LORs, with high values in both signifying formal yet upbeat styles. PS clout scores were universally much lower than for their respective LORs and PS authenticity scores were universally much higher than their respective LORs. This suggests that applicants write about themselves in a more tentative or humble tone, but also in a more personal manner than their evaluators.

To our knowledge, this is the first study to examine both LORs and PSs for applicants to otolaryngology residency. Prior research investigating PSs in any capacity is limited across all specialties. Our study is also the first to utilize the LIWC2015 linguistic analysis software for the otolaryngology subspecialty to objectively compare these texts across multiple word categories. In recent years, LIWC2015 has become the standard tool for analyzing written language, particularly among literature examining LORs in other specialties.

This study does have limitations. Utilizing data from applicants to a single institution, does not include the entire national cohort of applicants to otolaryngology residency. NRMP archives indicate that there were 505 total otolaryngology applicants in the 2019–2020 cycle and 559 in the 2020–21 cycle.^{27,28} The sample sizes of 452 and 476 included in this study represent 89.5% and 85.1% of the total numbers for the 2019–20 and 2020–21 cycles respectively and therefore still represent the vast majority of applications. Lastly, although the word frequency calculation conducted by LIWC2015 has become widespread for linguistic analysis of LORs, its design does not necessarily capture how a human reader will respond to a LOR or PS. More advanced language analysis might be able to identify more subtle differences in linguistics in the future.

5 | CONCLUSION

Minor linguistic differences exist in multiple domains between LORs written for male and female applicants for otolaryngology residency. These tended to favor female applicants, with their letters demonstrating higher clout, achievement, and grindstone scores. This trend was unexpected in this historically male dominant specialty. While differences were statistically significant, the overall difference in an entire letter of recommendation is likely subtle.

FUNDING INFORMATION

The project described was supported by NIH grants KL2TR002490 to AJK. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. No authors have any financial conflicts of interest.

REFERENCES

- Results of the 2018 NRMP Program Director Survey. https://www. nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf.
- Greenwald AG, Banaji MR. Implicit social cognition: attitudes, selfesteem, and stereotypes. *Psychol Rev.* 1995;102:4-27. doi:10.1037/ 0033-295x.102.1.4
- Gopal DP, Chetty U, O'Donnell P, Gajria C, Blackadder-Weinstein J. Implicit bias in healthcare: clinical practice, research and decision making. *Future Healthc J.* 2021;8(1):40-48. doi:10.7861/fhj.2020-0233
- 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
- Li S, Fant AL, McCarthy DM, Miller D, Craig J, Kontrick A. Gender differences in language of standardized letter of evaluation narratives for emergency medicine residency applicants. *AEM Educ Train.* 2017; 1(4):334-339. doi:10.1002/aet2.10057
- Lin F, Oh SK, Gordon LK, Pineles SL, Rosenberg JB, Tsui I. Genderbased differences in letters of recommendation written for ophthalmology residency applicants. *BMC Med Edu.* 2019;19(1):476. doi:10. 1186/s12909-019-1910-6

1750 Laryngoscope Investigative Otolaryngology–

- French JC, Zolin SJ, Lampert E, et al. Gender and letters of recommendation: a linguistic comparison of the impact of gender on general surgery residency applicants^{*}. J Surg Educ. 2019;76(4):899-905. doi: 10.1016/j.jsurg.2018.12.007
- 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
- Hu AC, Gu JT, Wong BJF. Objective measures and the standardized letter of recommendation in the otolaryngology residency match. *Laryngoscope*. 2020;130(3):603-608. doi:10.1002/ lary.28054
- Kominsky AH, Bryson PC, Benninger MS, Tierney WS. Variability of ratings in the otolaryngology standardized letter of recommendation. Otolaryngol Head Neck Surg. 2016;154(2):287-293. doi:10.1177/ 0194599815623525
- Messner A, Teng M, Shimahara E, et al. A case for the standardized letter of recommendation in otolaryngology residency selection. *Laryngoscope*. 2014;124(1):2-3. doi:10.1002/lary.24476
- Perkins JN, Liang C, McFann K, Abaza MM, Streubel SO, Prager JD. Standardized letter of recommendation for otolaryngology residency selection. *Laryngoscope*. 2013;123(1):123-133. doi:10.1002/ lary.23866
- Prager JD, Myer CM 3rd, Pensak ML. Improving the letter of recommendation. Otolaryngol Head Neck Surg. 2010;143(3):327-330. doi: 10.1016/j.otohns.2010.03.017
- Kimple AJ, McClurg SW, del Signore AG, Tomoum MO, Lin FC, Senior BA. Standardized letters of recommendation and successful match into otolaryngology. *Laryngoscope*. 2016;126(5):1071-1076. doi:10.1002/lary.25637
- 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(8):1335-1344. doi:10.1097/MLG.0b013e318175337e
- 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
- Demzik A, Filippou P, Chew C, et al. Gender-based differences in urology residency applicant personal statements. *Urology*. 2021;150: 2-8. doi:10.1016/j.urology.2020.08.066
- Osman NY, Schonhardt-Bailey C, Walling JL, Katz JT, Alexander EK. Textual analysis of internal medicine residency personal statements: themes and gender differences. *Med Edu*. 2015;49(1):93-102. doi:10. 1111/medu.12487

- Babal JC, Gower AD, Frohna JG, Moreno MA. Linguistic analysis of pediatric residency personal statements: gender differences. *BMC Med Educ*. 2019;19(1):392. doi:10.1186/s12909-019-1838-x
- Ostapenko L, Schonhardt-Bailey C, Sublette JW, Smink DS, Osman NY. Textual analysis of general surgery residency personal statements: topics and gender differences. J Surg Edu. 2018;75(3): 573-581. doi:10.1016/j.jsurg.2017.09.021
- 21. Results of the AAMC 2020 Physician Specialty Data Report. https://www.aamc.org/data-reports/workforce/interactive-data/ active-physicians-largest-specialties-2019
- 22. Tausczik YRP, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol*. 2010;29:24-54.
- 23. Trix FP, Psenka C. Exploring the color of glass: letters of recommendation for female and male medical faculty. *Discourse Soc.* 2003;14:191-220.
- Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The Development and Psychometric Properties of LIWC2015. University of Texas at Austin; 2015.
- Kobayashi AN, Sterling RS, Tackett SA, Chee BW, Laporte DM, Humbyrd CJ. Are there gender-based differences in language in letters of recommendation to an Orthopaedic surgery residency program? *Clin Orthop Relat Res.* 2020;478(7):1400-1408. doi:10. 1097/CORR.00000000001053
- Filippou P, Mahajan S, Deal A, et al. The presence of gender bias in letters of recommendations written for urology residency applicants. Urology. 2019;134:56-61. doi:10.1016/j.urology.2019.05.065
- National Resident Matching Program. Results and Data: 2020 Main Residency Match. Available at: https://www.nrmp.org/wp-content/ uploads/2021/12/MM_Results_and-Data_2020-rev.pdf
- National Resident Matching Program. 2021 Results and Data: Main Residency Match. https://www.nrmp.org/wp-content/uploads/ 2021/08/MRM-Results_and-Data_2021.pdf

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Berk GA, Ho TD, Stack-Pyle TJ, et al. Sex bias in letters of recommendation and personal statements for otolaryngology residency. *Laryngoscope Investigative Otolaryngology*. 2022;7(6):1745-1750. doi:10. 1002/lio2.932