



Gender and Letters of Recommendation: A Linguistic Comparison of the Impact of Gender on General Surgery Residency Applicants[☆]

Judith C French, PhD,^{*} Samuel J Zolin, MD,^{*} Erika Lampert, BS,[†] Alexandra Aiello, MSPH,[§] Kalman P Bencsath, MD,^{*} Kaitlin A Ritter, MD,^{*} Andrew T Strong, MD,^{*,†} Jeremy M Lipman, MD,^{*,‡} Michael A Valente, DO,[‡] and Ajita S Prabhu, MD^{*}

^{*}Department of General Surgery, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, Ohio; [†]Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, Ohio; [‡]Department of Colorectal Surgery, Digestive Disease and Surgery Institute, Cleveland Clinic, Cleveland, Ohio; and [§]Department of Quantitative Health Sciences, Cleveland Clinic Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio

OBJECTIVE: We investigated letters of recommendation for general surgery residency applicants to determine if any gender-based disparities exist.

DESIGN: A dictionary of over 400 terms describing applicants and 24 unique categories into which these terms were classified was created. Word count and language comparisons were performed using linguistic analysis software to assess for differences in applicant characterization, letter length, and writing style between male and female applicants and letter writers.

SETTING: A large, Midwest, academic general surgery residency program.

PARTICIPANTS: Five hundred and fifty-nine letters of recommendation received during the 2015 and 2016 interview cycles were selected for analysis.

RESULTS: Average word count was approximately equal for male and female applicants (503 vs 508, respectively). Female writers wrote longer letters (mean word count 545.5 vs 497.1, $p = 0.028$). “Standout” terms were more likely to be used to describe female applicants. Otherwise no statistically significant differences in applicant characterization were discovered.

[☆]*Funding:* This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Correspondence: Inquiries to Judith French, PhD, Department of General Surgery, Cleveland Clinic, 9500 Euclid Ave. H18, Cleveland, OH 44195; e-mail: frenchj2@ccf.org

CONCLUSIONS: Letters of recommendation for general surgery are written using similar descriptive terms and lengths for male and female applicants. This suggests that there is no specific gender disadvantage with regard to letters of recommendation when applying for general surgery residency. (J Surg Ed 76:899–905. © 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: recommendation letter, gender, general surgery, residency, graduate medical education

COMPETENCIES: Interpersonal and Communication Skills, Professionalism

INTRODUCTION

The importance of letters of recommendation (LOR) in the selection of applicants for residency is well established, both at large and within the field of general surgery. Surveys have demonstrated the importance of LOR in selecting candidates to interview as well as in ranking applicants within the National Resident Matching Program.^{1,2} Although the utility of LOR for the selection of general surgery residents has been challenged based on inconsistent presence of specialty-relevant content,³ it is clear that LOR in their current form continue to have a significant impact on interview offerings and ranking for general surgery residency applicants.

Prior analyses of LOR for positions in academia have demonstrated significant differences in language describing male and female applicants, as well as the impact of this language on hiring decisions. A linguistic analysis of LOR for junior psychology faculty positions demonstrated that LOR for men are more likely to emphasize characteristics such as agency, assertiveness, and confidence (*agentive* terms) while LOR for women are more likely to emphasize characteristics such as kindness, sensitivity, and cooperation (*communal* terms). Furthermore, usage of communal descriptors in LOR is negatively related to hiring decisions, despite blinding reviewers to applicant gender.⁴ Other studies have demonstrated that female applicants for postdoctoral positions were less than half as likely to receive an excellent letter of recommendation than their male counterparts.⁵ Female applicants for positions as medical faculty were significantly more likely to receive very short LOR compared to male applicants, and were twice as likely to receive LOR containing phrases described as “doubt raisers.” “Doubt raisers” consist of negative or potentially negative language, hedging, unexplained comments, faint praise, and irrelevancies. In this study, male applicants were also more likely to have their research accomplishments emphasized than were their female counterparts.⁶

Surprisingly, there is a paucity of literature addressing potential gender-based differences in LOR written for general surgery residency applicants. Analyses of LOR in other surgical specialties have demonstrated differences based on applicant and letter writer gender.^{7,8} Given the importance of LOR to applicant selection for residency interview offers and ranking and the impact of gender-biased language on perceptions of employability elsewhere in academia, it is important to consider whether gender-based differences exist for LOR within the field of general surgery. If such differences exist, they may systematically disadvantage 1 gender. Raising awareness within the surgical educational community would be the first step toward lessening potential bias. In this study, we hypothesized that significant differences would exist in the language used to characterize female applicants vs their male counterparts, as well as the language and writing style utilized by male and female letter writers.

METHODS

Identification of Sample LOR for Study

Prior to initiation the Institutional Review Board approved the protocol under exempt status. A sample size calculation using estimates of percent agentive adjectives as the outcome and accounting for the possibility of high within-applicant correlation demonstrated that a total of 291 letters were needed to achieve 80% power. A binary approach to gender was utilized in this study, though the authors recognize that many individuals do not identify according to a binary system. The Electronic Residency Application Service, which was used to identify gender, only provides applicants with 2 gender choices. LOR for candidates applying to a large, Midwest, academic general surgery residency program in the 2015 and 2016 interview cycles were evaluated for language patterns. To control for applicant competitiveness, the applicant pool consisted of individuals deemed competitive enough to warrant an interview invitation and who were subsequently ranked within the National Resident Matching Program. One hundred and fifty-seven male and 91 female applicants who were invited for interviews were chosen at random for inclusion. To reach the targeted number of evaluations based on the power calculation, random application packets were chosen to contribute either 2 or 3 letters for evaluation. If an applicant had more than 3 letters, only the first 3 letters in the application packet were included. This resulted in 559 letters, 331 for male, and 228 for female applicants with a median word count (WC) of 470 (range: 182-1341; Table 1). The sample included 448 unique authors (excluding coauthored letters) with a maximum of 8 letters written by 1 author (377 authors wrote only 1 letter).

Development of Linguistic Assessment Parameters

A customized dictionary was developed in a collaborative, iterative process to define categories and content for analysis. Sample LOR from the 2014 interview cycle were reviewed by 2 female and 2 male researchers experienced in reading and writing LOR to identify key terms that might influence a reader. These were combined with

TABLE 1. Breakdown of Applicant and Writer Genders

| Applicant Gender | Total (N = 559) | Female Writer (N = 89) | Male Writer (N = 470) |
|------------------|-----------------|------------------------|-----------------------|
| Female | 228(40.8) | 42(47.2) | 186(39.6) |
| Male | 331 (59.2) | 47(52.8) | 284(60.4) |

Statistics presented as N (column %).

terms adapted from prior analyses detailed below to develop an initial dictionary. The group met repeatedly to add and categorize search terms. Online resources were utilized to identify synonyms and antonyms of descriptive words. This process continued until the dictionary reached saturation, meaning that no new search terms or word categories were added in the final meeting.

The finalized dictionary contained 24 unique categories and over 400 words. Many categories were based on previous research involving LOR (e.g., *communal*, *agentic*, and *social*). Other categories were created to capture specific wording found in surgical residency application letters (*desire*—indicating desire to match/keep candidate; *technical skill*—indicating psychomotor characteristics). The “personality” domain was further divided into 5 categories based on the Five-Factor Model of Personality,⁹ consisting of openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism. These 5 categories were further divided based on their contrasting traits (e.g., conscientiousness = *efficient/organized* vs *easy-going/careless*). **Table 2** demonstrates a complete list of search term categories and example terms.

The dictionary was built for use with the Linguistic Inquiry and Word Count program (LIWC2015; Pennebaker Conglomerates, Inc., Austin, Texas).¹⁰ This program analyzes text files and calculates the percentage of words or short word phrases used in the file based on

the dictionary uploaded to the system. The program also calculates the WC, average number of words per sentence, average number of words containing 6 or more letters, and the average number of words found in the uploaded dictionary. The effectiveness of this dictionary at capturing words used to describe the candidates was confirmed using letters from the 2014 interview cycle.

Analysis

Statistical comparisons of WC and language characteristics were conducted using mixed models. Linear mixed models and negative binomial mixed models were fit to compare language characteristics between female and male applicants’ LOR. Letters were modeled as clusters by applicant using a compound symmetry covariance structure. Negative binomial mixed models were used for most characteristics due to skewing of the factors’ underlying distributions. For these, rate was modeled using the number of words in a given category as the outcome offset by the WC per letter.

Additional models were fit in the same manner, but instead compared male and female writers. Lastly, stratified models were fit to compare applicant gender within writer gender groups. Least-squares means were calculated for each model and used to quantify group differences. All analyses were conducted using SAS (version 9.4, The SAS Institute, Cary, North Carolina), and p values

TABLE 2. Dictionary Categories With Example Search Terms.

| Gendered Terms | Grindstone | Standout | Activities | Inventive/ Curious | Consistent/ Cautious |
|--|--|----------------------------------|----------------------------------|--|--|
| Lady Man Woman | Deliberate Hardworking Tenacious | Best Leader Prodigy | Athlete Research Volunteer | Artistic Imaginative Open minded | Conventional Data driven Pragmatic |
| Efficient/ Organized | Easygoing/ Careless | Outgoing/ Energetic | Solitary/ Reserved | Friendly/ Compassionate | Challenging/ Detached |
| Common sense Dependable Fastidious | Care free Flexible Laid back | Bubbly Chatty Popular | Independent Quiet Reserved | Affable Likable Team player | Aggressive Intense Type A |
| Sensitive/ Nervous | Secure/ Confident | Physical | Age | Social | Agentic Personality |
| Anxiety Depression Vulnerable | Balanced Composed Poised | Attractive Beautiful Smile | Kid Mature Nontraditional | Children Couple Family | Able Competitive Proactive |
| Communal | First Person | Second Person | Agentic Orientation | Desire | Technical Skill |
| Agreeable Caring Nice | I Me We | You Your | Do Earn Gain | Keep Recruit Stay | Dexterity Finesse Skillful |

TABLE 3. Overall and Stratified Linear Mixed Models for Word Count

| Effect | Subgroup | Gender | Mean Words per Letter | SE | p Value |
|------------------|----------------|------------------|-----------------------|------|--------------|
| Writer gender | Overall | Female (N = 89) | 545.5 | 20.0 | 0.028 |
| | | Male (N = 470) | 497.1 | 21.6 | |
| Applicant gender | Overall | Female (N = 228) | 507.9 | 13.1 | 0.76 |
| | | Male (N = 331) | 502.7 | 16.9 | |
| | Male writers | Female (N = 186) | 491.5 | 14.2 | 0.60 |
| | | Male (N = 284) | 501.0 | 18.2 | |
| | Female writers | Female (N = 42) | 582.9 | 30.4 | 0.10 |
| | | Male (N = 47) | 513.3 | 42.1 | |

CI, confidence interval.

Female-Male = difference of means.

The p value is calculated using a *t* test.

Bold value = To highlight statistical significance.

<0.05 were considered statistically significant. All categories with statistically significant differences underwent manual evaluation by the authors to ensure the words found in the search were used in the anticipated context.

RESULTS

Descriptive statistics of the mean WC across groups are given in [Table 3](#). No significant difference was found between letters written for female and male applicants with regard to average WC (508 vs 503 words, respectively, *p* = NS); however, female letter writers wrote significantly longer letters compared to male writers (546 vs 497 words, *p* = 0.028). No statistically significant differences were found across any groups with regards to words per sentence, utilization of words consisting of 6 or more letters, or percentage of words found in the uploaded dictionary (average 10% words per letter).

Only 1 of the 24 dictionary categories had a statistically significant difference when comparing the gender of the applicants. More *standout* terms (e.g., amazing, outstanding, etc.) were used to describe females vs males ([Tables 4](#) and [5](#)). Several significant differences were found when analyzing the gender of the letter writer; however, these differences lack statistical power due to the paucity of female letter writers randomly selected (89 female authored letters vs 470 male

authored letters). Female writers used more *grindstone*, *outgoing/energetic*, *agentic*, and *communal* terms compared to male writers. Female writers also used more *efficient/organized* terms to describe female vs male applicants (mean rate % 0.193 vs 0.110, *p* = 0.020).

Two additional categories (*sensitive/nervous* and *challenging/detached*) were initially found to have statistically significant differences with regard to the gender of the letter writer. Once these categories were assessed for context, most terms were found to be describing things other than the candidates, e.g., “sensitive exam” and “intense workload.” All letters that contained these categories were reevaluated to exclude inapt phrases. Terms from these categories that were used to describe the candidates were rare and not statistically significant.

DISCUSSION

Our study is the first to examine gender-based differences in LOR for general surgery residency applicants. The results of our study indicate more similarities than differences in the terms used in LOR to describe female and male candidates. Female writers tend to use a different style compared to their male counterparts, but the overall descriptions of candidates are relatively the same. There was no difference between genders with regard to the percentage of dictionary words found in letters.

TABLE 4. Linear Mixed Model for Standout Terms by Applicant Gender

| Applicant Gender | Mean Percent of Words Identified as Standout Terms per Letter (%) | SE | p Value |
|------------------|---|------|--------------|
| Female | 1.1 | 0.04 | 0.050 |
| Male | 1.00 | 0.06 | |

Female-Male = difference of means.

The p value is calculated using a *t* test.

TABLE 5. Negative Binomial Mixed Models by Applicant Gender

| Category | Applicant Gender | Mean Rate of Category Terms per Letter (%) | SE | p Value |
|---------------------|------------------|--|------|---------|
| Grindstone | Female | 0.858 | 0.04 | 0.87 |
| | Male | 0.851 | 0.03 | |
| Agentic personality | Female | 0.536 | 0.04 | 0.44 |
| | Male | 0.517 | 0.03 | |
| Communal | Female | 0.261 | 0.08 | 0.38 |
| | Male | 0.239 | 0.06 | |
| Social | Female | 2.282 | 0.02 | 0.054 |
| | Male | 2.533 | 0.02 | |
| Desire | Female | 0.079 | 0.12 | 0.49 |
| | Male | 0.087 | 0.10 | |
| Technical skill | Female | 0.197 | 0.09 | 0.26 |
| | Male | 0.222 | 0.06 | |

Female:Male = difference of means.

The p value is calculated using the Wald test.

Bold value = To highlight statistical significance.

The importance of this finding is that seemingly, neither male nor female applicants are placed at a disadvantage regarding LOR for surgery residency applications.

Only 1 of the 24 categories in the dictionary, *standout*, showed a statistically significant difference between applicant genders, with more *standout* terms being used to describe female applicants when compared to male applicants. Trix⁶ found roughly the same percent of *standout* terms used between genders for their population; however, upon further review they determined that letters written for men had more occurrences of *standout* terms used in a given letter when compared to women. The findings in our study are based on percentages per letter. While the reason behind this result goes beyond the scope of our study, this may either accurately represent applicants or perhaps reflect evolving LOR writing styles with more attention being paid to equality for female candidates.

Even though over 550 letters were randomly selected from the 2 year pool, only 89 were written by females, which did not provide enough power for some of the analyses. The Association of American Medical Colleges indicates that about 22% of full time surgery faculty at United States medical schools are women.¹¹ The non-purposeful sample collected for this study with regard to writer gender only had 16% female writer representation. One explanation for the low percentage of female writers is that candidates are requested to submit a letter from their department Chair. As of July 2018, only 21 women are Chairs of Surgery departments across the United States and Canada.¹² Candidates may also be more likely to obtain letters from faculty of higher academic rank, and men are more likely than women to be full professors at academic medical centers.¹³ These disparities obviously contribute to a gender gap in LOR

authorship. Importantly, however, we found that male writers still write the same caliber of letter regardless of applicant gender.

Female letter writers crafted longer letters on average compared to male writers. Previous research examining writer gender and WC has yielded inconsistent findings. Among graduate school applicants, the shortest letters were written by males for males.¹⁴ Letters written for applicants to faculty positions in biochemistry or chemistry had no difference in length based on the writer gender.¹⁵ When examining letters for otolaryngology applicants, no difference in WC was found based on writer gender.⁸ Looking at other documents in medical education, specifically Medical Student Performance Evaluations, a study analyzing these evaluations from over 100 different medical schools found male writers tended to produce shorter overall evaluations compared to female writers.¹⁶

Does WC matter? A 1994 study looking at surgical residency LOR found that letters which scored highly on their scale, on average, were twice as long as letters which scored poorly.¹⁷ Shorter, more general letters reflected poorly on the depth of the writer-applicant relationship. In a simulated laboratory study, letter length was positively correlated with the “judged favorability” of a letter.¹⁸ While WC may not be the most important factor determining perceived letter quality, it does play a role. In our study, neither female nor male applicants were disadvantaged by letter length.

Although there were insufficient female writers to perform a full analysis of the differences between the letters written by males vs females, several themes did emerge. Male and female writers described candidates differently, with female writers using more *agentive adjective*, *communal*, *grindstone*, and *outgoing/energetic* terms

than male writers. Other research has reported the opposite finding with regard to *agentive adjectives* in which male writers used more *agentive* terms than female writers.⁴ One explanation for this may lie in the differences in methodology by which the letters were analyzed. Alternatively, the use of *agentive* or *communal* terms may vary by specialty. Interestingly, male and female letter writers had roughly the same percent of dictionary words in their letters.

Female writers also used more *efficient/organized* terms to describe female applicants. This category is a subset of the Five-Factor Model of Personality, conscientiousness.⁹ Conscientious people can be described as hard working or reliable which are considered *grindstone* traits. This falls in line with prior research demonstrating that descriptions of women were more often associated with effort and men with natural ability.⁶ Unlike the findings in our study, this difference was not correlated with writer gender.

Several limitations exist with regard to the current study. We analyzed letters for applicants to a single academic residency program. It is possible that letters for applicants to nonacademic residencies might demonstrate different results as the personal characteristics and activities emphasized in the letters might differ. Our residency program selects for highly competitive candidates when interview offers are made. Therefore the content and writing style of the letters analyzed in this study is likely not representative of the full spectrum of general surgery residency applicants. While analyzing letters only for applicants selected for interview helped to control for applicant competitiveness, it is possible that this approach selected for applicants with higher quality letters and that greater gender-based differences in LOR might have been present in letters for applicants not selected for interview. Even though 377 authors in our sample wrote only 1 letter, we did not account for the number of different institutions these authors represent. Letters written by authors from the same institutions could use similar language formats. Content analysis relied on counting term usage. While context was reviewed in cases of statistically significant differences, it was not possible to verify the context of every word or phrase used. If a spelling error was present in a letter, usage of a word might not have been detected. Accommodations were, however, made for common wording differences such as 'well liked' vs the hyphenated version 'well-liked.' Finally, LOR in general surgery lack standardization, with no set criteria upon which residency applicants are guaranteed to be evaluated. The combination of personal subjective evaluation and variations in individual writing style make interwriter comparisons challenging.

In terms of future direction, a purposeful sampling method and increased number of letters would allow for

increased power to detect effects related to writer gender. One possibility would be to perform a multi-institutional analysis similar to this to increase sample size and decrease bias related to institutional characteristics and methods of applicant evaluation. While we examined LOR with regard to descriptor utilization and length with some implications regarding quality based on prior research, future studies might address letter quality as a function of gender with attention to characteristics such as authenticity and personal detail. Ultimately, the most important questions that remain are whether LOR are an accurate and sufficiently representative marker of applicant quality and whether the field of general surgery should transition to a more standardized evaluation similar to those used in other fields.^{19,20,21}

CONCLUSIONS

LOR for general surgery residency are written using similar descriptive terms and lengths for male and female applicants. This suggests that LOR, an important part of the interview selection process, place neither gender at a disadvantage in general surgery residency application. Further research is necessary to determine whether this applies to the greater general surgery applicant pool as well as to other specialties. As for differences based on the gender of the writer, our sampling methods did not provide enough letters from females writers to produce meaningful results. Future research with purposeful sampling methods is needed.

ACKNOWLEDGMENTS

We wish to thank Valery Vilchez Parra, Dominykas Burneikis, and Talia Burneikis for their input into the design and analysis of our work.

REFERENCES

1. National Resident Matching Program, Data Release and Research Committee: Results of the 2018 NRMP Program Director Survey. Washington, DC. Available at: <http://www.nrmp.org/wp-content/uploads/2018/07/NRMP-2018-Program-Director-Survey-for-WWW.pdf>. Accessed July 31, 2018.
2. Makdisi G, Takeuchi T, Rodriguez J, et al. How we select our residents—a survey of selection criteria in general surgery residents. *J Surg Educ*. 2011;68:67-72.
3. Fortune JB. The content and value of letters of recommendation in the resident candidate evaluative process. *Curr Surg*. 2002;59:79-83.

4. Madera JM, Hebl MR, Martin RC. Gender and letters of recommendation for academia: agentic and communal differences. *J Appl Psychol*. 2009;94:1591-1599.
5. Dutt K, Pfaff DL, Bernstein AF, et al. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nat Geosci*. 2016;9:805.
6. Trix F, Psenka C. Exploring the color of Glass: letters of recommendation for female and male medical faculty. *Discourse Soc*. 2003;14:191-220.
7. Filippou PL, Mahajan SL, Wallen EM, et al. Gender bias in letters of recommendation for urology residency applicants. *J Am Coll Surg*. 2017;225:S76.
8. 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.
9. Digman JM. Personality structure: emergence of the five-factor model. *Annu Rev Psychol*. 1990;41:417-440.
10. Pennebaker JW, Boyd RL, Jordan K, et al. *The Development and Psychometric Properties of LIWC2015*. Austin, TX: University of Texas at Austin; 2015.
11. Association of American Medical Colleges. Medical School Faculty Tables. Washington, DC. Available at: <http://aamcdiversityfactsandfigures2016.org/report-section/u-s-medical-school-faculty/>. Accessed July 31, 2018.
12. Association of Women Surgeons. Female chairs of departments of surgery. Chicago, Illinois. Available at: <https://www.womensurgeons.org/page/Female-SurgeryChairs>. Accessed July 31, 2018.
13. Jena AB, Khullar D, Ho O, Olenski AR, Blumenthal DM. Sex differences in academic rank in US medical schools in 2014. *JAMA*. 2015;314:1149-1158.
14. Watson C. Sex-linked differences in letters of recommendation. *Women Lang*. 1987;10:26-28.
15. 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.
16. 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.
17. Greenburg AG, Doyle J, McClure DK. Letters of recommendation for surgical residencies: what they say and what they mean. *J Surg Res*. 1994;56:192-198.
18. Judge T, Higgins C. Affective disposition and the letter of reference. *Organ Behav Hum Decis Process*. 1998;75:207-221.
19. Girzadas DV, Harwood RC, Dearie J, Garrett S. A comparison of standardized and narrative letters of recommendation. *Acad Emerg Med*. 1998;5:1101-1104.
20. Kaffenberger JA, Mosser J, Lee G, et al. A retrospective analysis comparing the new standardized letter of recommendation in dermatology with the classic narrative letter of recommendation. *J Clin Aesthet Dermatol*. 2016;9:36-42.
21. Perkins JN, Liang C, McFann K, et al. Standardized letter of recommendation for otolaryngology residency selection. *Laryngoscope*. 2013;123:123-133.