How Prominent Are Gender Bias, Racial Bias, and Score Inflation in Orthopaedic Surgery Residency Recommendation Letters? A Systematic Review

Robert John Burkhart, Monish Sai Lavu, Christian Joseph Hecht, Jason Ina, Robert J Gillespie, Raymond W Liu

INTRODUCTION:

Letters of recommendation are considered an important factor in the holistic ranking of orthopaedic residency applications. The standardized letter of recommendation (SLOR) was introduced in 2017 in response to the inherent bias and limited comparative value of traditional LORs. It includes standardized questions about an applicant's orthopaedic qualifications, designed to enable fair comparisons among candidates. However, disparate and inconsistent findings have made it difficult to draw meaningful conclusions from individual studies on the SLOR and narrative letter of recommendation.

In this systematic review we asked: (1) What is the distribution of applicant SLOR rating among nine domains and summative scores? (2) Are there applicant characteristics associated with SLOR ratings? (3) Is there gender and racial bias in the SLOR and narrative letter of recommendation?

METHODS:

PubMed, EBSCO, and Google Scholar electronic databases were queried on March 20, 2023, to identify all studies evaluating SLOR and narrative letter of recommendations between January 1, 2017, and March 20, 2023. Articles that evaluated orthopaedic SLOR or narrative LORs were included. Systematic reviews, case reports, duplicate studies between databases, and grey literature such as abstracts and articles on preprint servers were excluded. Eight studies including 6179 applicants and 18,987 letters of recommendation were included in the final analysis. The applicant classes ranged from years 2014 to 2020. Two reviewers independently evaluated the quality of the included studies using the Joanna Briggs Institute (JBI) tool for cross-sectional studies. The mean JBI score of included studies was 7.4 out of a maximum of 8, with higher scores indicating better quality. The primary outcome was to determine the distribution of applicant SLOR rankings in their summative score. Summative scores were separated into ranked to match, top one-third, middle one-third, lower one-third, and not a fit. Four studies reported the summative SLOR scores of applicants. Our secondary study goal was to assess applicant characteristics associated with SLOR summative scores and assess any bias present in the SLOR or narrative recommendation. Five studies compared SLOR ratings across applicant characteristics including gender, race, USMLE Step 1 score, USMLE Step 2 score, Alpha Omega Alpha (AΩA) membership, clerkship grades, and publications. Gender and racial bias were also assessed across five studies. Four studies utilized a linguistic analysis software to search letters of recommendation for differences in word category use by gender and race.

RESULTS:

Studies consistently found that a higher percentage of candidates were identified in the top percentile blocks than is mathematically possible. For example, the two studies that tallied the proportion of candidates ranking in the top 10% of applicants found that 36% and 42% were rated as being in the top 10%. Similarly, articles found more than 87% of applicants scored in the top one-third. Studies had mixed results, but in general, they suggested that AΩA status, higher Step 1 scores, and more research publications have a slightly positive association with increased SLOR scores. We found no evidence of gender bias against women, and in fact, most studies evaluating word choices found word choices for women candidates generally were stronger. Similarly, no consistent disparities were identified in word categories utilized in SLORs based on applicant race.

DISCUSSION AND CONCLUSION:

Most notably, a mathematically impossible proportion of applicants were counted in the top percentiles in letters of recommendation. This compromises readers' abilities to differentiate candidates. Factors like AOA status and research publications displayed a modest positive correlation with SLOR scores. Gender bias against women or candidates from racial minority groups was not evident.

We suggest that group SLOR authorship, with a consensus assessment by clerkship or residency directors, should be considered. We also recommend the incorporation of mean and SD scores for each letter writer (or group) alongside their letters. Furthermore, studies indicate that submitting all four SLOR letters can substantially improve an applicant's chances of success. We advise students to choose strategically and submit the maximum allowable number of SLORs.

Table 1. Characteristics of articles included in the final analysis

Article	Journal	Study design	Data source	Applicant classes	Study aim	Sample size (number)	JBI score[a]
Girgis et al. [8]	*Journal of Surgical Education*	Retrospective	Single institution, Temple University	2016-2018	Race and gender bias	8032 LORs (2420 applicants)	6
Inclan et al. [14]	*Journal of Bone and Joint Surgery*	Retrospective	Single institution, Washington University	2018-2019	SLOR inflation	497 SLORs (179 applicants)	8
Kang et al. [16]	*Journal of the American Academy of Orthopaedic Surgeons*	Retrospective	Single institution, University of Southern California Los Angeles	2018-2019	SLOR inflation	1137 SLORs (513 applicants)	8
Kobayashi et al. [18]	*Clinical Orthopaedics and Related Research*	Retrospective	Single institution, Johns Hopkins University	2018-2019	Gender bias	2834 LORs (738 applicants)	8
Lipa et al. [21]	*Journal of Bone and Joint Surgery*	Retrospective	Single institution, Harvard combined orthopaedic surgery residency	2019-2020	Gender bias	650 LORs (182 applicants)	8
Pacana et al. [29]	*Clinical Orthopaedics and Related Research*	Retrospective	Single institution, Prisma Health-Midlands University of South Carolina	2017-2018	SLOR inflation	1376 SLORs (559 applicants)	7
Powers et al. [31]	*Journal of Bone and Joint Surgery*	Retrospective	Single institution, Washington University	2017-2018	Race and gender bias	2625 LORs (730 applicants)[b]	6
Samade et al. [33]	*Journal of Bone and Joint Surgery*	Retrospective	Single institution, The Ohio State University	2017-2018	SLOR inflation	1836 SLORs (858 applicants)	8

[a]This quality assessment tool has eight evaluation categories, each of which is awarded a score of 0 (not reported or inadequate) or 1 (reported and adequate), with a maximum score of 8; higher scores indicate better study quality.
[b]This involved SLOR and narrative; JBI = Joanna Briggs Institute; LOR = letter of recommendation; SLOR = standardized letter of recommendation.

Table 2. Key findings from all studies evaluating the distribution of applicant SLOR ratings

Article	Key findings
Inclan et al. [14]	Across all SLOR domains, approximately 75% (134 of 179) of applicants received ratings over the 85th percentile.
	Only 0.3% (16 of 4124) of domain ratings were below the 50th percentile.
Kang et al. [16]	All nine SLOR domains had at least 88% (451 of 513)[a] of applicants appraised in the top two deciles, with 26% (296 of 1137) of SLORs having all domains reported at the 100th percentile.
	More than 90% of applicants were rated as either ranked to match or in the top third of their rank list.
Pacana et al. [29]	36% (437 of 1207) of summative scores for SLORs rated applicants in the top 10% of candidates evaluated.
	An additional 51% (619 of 1207) of summative scores were ranked in the top third.
Samade et al. [33]	Approximately 42% (359 of 858) of applicants received a summative SLOR score as in the top 10% of candidates.
	An additional 46% (395 of 858) of applicants were ranked in the top third.

[a]Value derived from a histogram illustrating distribution of applicant scores across nine SLOR domains.