Matching into an Orthopedic Residency: Which Application Components Correlate with Final Rank List Order?

Joseph M. Legato, MD,* David A. Fuller, MD,* Catherine Kirbos†, Mark Pollard, MD,* Douglas Tase, MD,* Tae Won Kim, MD,* and Lawrence S. Miller, MD*

*Cooper University Hospital, Camden, New Jersey; and †Cooper Medical School of Rowan University, Camden, New Jersey

OBJECTIVE: To determine if any of the unique elements of the applications of medical students who were granted interviews to an orthopedic residency program correlate to the program's final rank order list of candidate residents.

DESIGN: Eight domains of the standard residency application were considered as independent variables for 36 applicants. Personal, identifying information was removed from the application material within each domain, thus blinding the application domains for 5 core faculty members from the program to review and rank independently. These 8 domain rank lists were then compared to the program's final rank list order to determine the correlation of each domain with the final rank list order.

SETTING: Academic medical center.

PARTICIPANTS: Applicants to a university-based orthopedic surgery residency program who were granted interviews in the 2016 academic year.

RESULTS: Two domains of the application correlated with the final rank list order: interview and personal statement. None of the other domains had a significant correlation with the final rank list order. Interobserver variability among the faculty members was high for the rankings of the different domains.

CONCLUSIONS: Interview and personal statement were the only domains within the application that had correlation with the final rank list for an orthopedic residency program. A better understanding of how these 2 components affect the rank list may provide opportunity for process improvements. (J Surg Ed 76:585−590. © 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: Residency, Application, Rank list, Personal statement, Orthopedic, Interview

COMPETENCIES: Interpersonal and Communication Skills, Professionalism, Practice-Based Learning and Improvement

INTRODUCTION

There has been a substantial growth in medical school graduates applying to orthopedic residency positions without a concomitant increase in available orthopedic training positions. In 2015, there were 1062 total applicants to orthopedic surgery residencies with an orthopedic match rate of only 66%.[1]

Applications for orthopedic residency positions contain information in multiple domains. These applications must be screened and narrowed down to a much smaller group to which interviews are extended. In most programs, at the conclusion of interviews all application information is reviewed together as a whole and utilized to create the program's final rank list. Numerous studies have reviewed screening methods for granting interviews.[2,3] There is less literature, though, that examines how the final rank list is determined after review of all variables from interviewed candidates, including the subjective component of the interview itself.

Program directors have been queried as to the importance of individual application variables to the overall orthopedic residency application. In one study, program directors stated the most important components were class rank, formality/politeness at interview, personal appearance, United States Medical Licensing Exam

Prior abstract/poster presentation: The abstract for this manuscript was presented as an ePoster at the 2017 AOA Annual Leadership Meeting in Charlotte, NC.

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

Correspondence: Inquiries to Joseph M. Legato, MD 3 Cooper Plaza, Suite 410 Camden, NJ 08103.; e-mail: jml229@gmail.com

  1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2018.08.018

(USMLE) Step 1 score, and participation in an externship.[4] Faculty surveyed in another study said the most important elements to an application were externship performance, class rank, and interview performance.[5] While these studies show some agreement in certain elements and provide valuable information, they are often done through surveys and are based solely on perceived importance of various domains within an application.

Furthermore, there is no clear consensus among applicants as to which item of their application they perceive as most important to successfully obtaining an orthopedic residency position. One study surveyed orthopedic applicants and showed their perceived importance was externship performance, letters of recommendation (LOR), and USMLE Step 1 score.[5] A similar study by Huntington et al. showed agreement in perceived importance of board scores and performance at that institution's away rotation, but their survey showed that applicants also believe the interview to be a top ranking criterion.[6] These studies along with others indicate that there are no universally agreed-upon application domains among applicants or reviewers that are more important than other domains to the overall application.

The purpose of this study is to determine if any of the unique elements of medical students whom were granted interviews to an orthopedic residency program correlate to the program's final rank order list of candidate residents. We hypothesize that there will be a difference between the individual contributions of each element in the application to the final program rank list. A better understanding could, in turn, lead to opportunities for process improvement.

METHODS

This study was conducted at a university-based orthopedic residency program. For the 2016 academic year, 395 applications were received and reviewed from the Electronic Residency Application Service (ERAS) by the residency selection committee. Interviews were offered to 49 (12.4%) of those applicants, of whom 36 (9.1%) ultimately interviewed. The applications for these 36 applicants were deidentified, separated into their individual domains, and randomized utilizing a computerized random number generator. These were then distributed to 5 core attending orthopedic surgeons who were part of the residency selection committee. Exemption from Institutional Review Board was determined by the authors based on institutional policy.

Eight elements of the residency application were considered as independent variables (domains). These included: (1) LOR, (2) medical school dean's letter, (3) personal statement, (4) USMLE Step 1 and 2 scores, (5) educational profile (i.e., schools attended), (6) scholarly activity, (7) personal photograph, and (8) interview performance. Each application contained 3 to 4 LOR from attending orthopedic surgeons, each often reflecting a student's performance on an externship performed at the writer's institution. The dean's letter more often discussed the student's performance in medical school as a whole and personal attributes of the applicant. The educational profile domain listed all undergraduate, graduate, and medical schools attended with their associated degrees obtained and fields of study as listed under the "Medical Education" and "Education" sections of ERAS. The scholarly activity domain included publications and research-related endeavors as listed under the "Publications" section. The personal photograph domain consisted of the photograph submitted to the ERAS service for the student's application.

Applicant rankings were created by each of the 5 reviewers after reviewing a single domain in a blinded, independent fashion. The 5 independent rank lists that were generated for each of the independent domains were then averaged into one by determining a mean ranking among the 5 attending for each applicant within the domain, thus creating one rank order list for each of the 8 domains for each applicant. The rank list for the USMLE Step 1 and 2 scores domain was the same for all raters; it was created by adding the scores of the 2 exams and sorting the combined score from greatest to least. USMLE Step 1 scores ranged from 216 to 271 (mean = 243), and USMLE Step 2 scores ranged from 235 to 276 (mean = 251). There were 2 students who had not taken USMLE Step 2 before the interview, and for those students USMLE Step 1 score was doubled and factored in with the others. For the interview domain, each applicant was interviewed blindly without any application data available to the interviewer. Interviews occurred in our standard fashion, where 2 faculty members interviewed each applicant for 12-15 minutes. Because an odd number of faculty members participated in this study (5), one of the interview rooms had one faculty that was blinded while the second faculty member in the room was not blinded as to the applicant. The other 4 faculty involved with the study were divided into 2 rooms. Other interviews occurred with other faculty not related to the study in accordance with our usual interview processes. A rank list for the interview domain was created immediately following the final interview. Two pairs of interviewers submitted their interview rank lists as one combined, averaged list.

At the conclusion of the creation of the rank lists for each domain, the 5 members of the residency selection committee participating in the study met and were provided with the complete applications for each applicant. A final rank list was then created in the standard fashion

for the program utilizing a modified Delphi process. The Delphi process determines a consensus on a topic by surveying a panel of experts on the topic through multiple iterations, thereby ultimately focalizing toward a common agreement.[7] The modified process conducted by the committee is led by one core member moderator who facilitates the ranking of the candidates through multiple surveying and ultimate consensus of all five core faculty members based on discussion and open availability of all information contained within the applications. This final rank list was then compared to the lists generated for each domain to determine the relative impact of each variable upon the final rank list.

Reviewer rank lists of each component were compared to the final rank list using nonparametric, multiple regression of ranks to determine the correlation of each domain with the final rank list order. A p value of <0.01 was considered statistically significant. Additionally, inter-relator reliability was assessed comparing the rank lists when simplified into subsections of High (ranks 1-12), Medium (ranks 13-24), and Low (ranks 25-36). For a reviewer, each variable was re-reviewed blindly and similarly simplified into subsections so that intrarater reliability could also be determined.

RESULTS

There were 2 domains that significantly correlated with the final rank list: the interview domain (Spearman correlation coefficient, $r = 0.526$, $p < 0.001$) and the personal statement domain ($r = 0.343$, $p < 0.001$) (Table 1). The educational profile domain had a small correlation that approached significance ($r = 0.175$, $p = 0.019$). No other domains significantly correlated to the final rank list.

Multiple domains correlated amongst one another (Table 2). The board score domain had significant inverse correlations with scholastic activity ($r = -0.300$, $p < 0.001$), interview ($r = -0.227$, $p = 0.002$), and personal photograph ($r = -0.319$, $p < 0.001$). The interview domain had significant positive correlations with the personal statement domain ($r = 0.192$, $p = 0.010$) and the LOR domain ($r = 0.193$, $p = 0.009$). Finally, the educational profile domain correlated with the dean's letter domain ($r = 0.422$, $p < 0.001$).

There was substantial variability in the independent rank lists utilized to create averaged rank lists for each domain (Table 3). The interobserver reliability for all domains except for interview displayed low agreement with kappa (κ) values less than 0.25. The interview domain indicated moderate interobserver agreement with a κ value of 0.43.

Furthermore, there was substantial variability for the one reviewer that reconducted a blinded review of the domains in order to examine intraobserver reliability. The individual's intraobserver agreement was low for all domains with κ values of 0.207, 0.208, 0.291, 0.062, 0.042, and 0.005 for the personal photograph, LOR, personal statement, dean's letter, educational profile, and scholastic activity domains, respectively.

TABLE 1. Correlation of Domains with Final Rank List

Domain	Spearman Correlation Coefficient (r)	p Value
LOR	0.118	0.113
DLetter	0.088	0.239
PS	0.343	<0.001
Boards	−0.079	0.290
EduProf	0.175	0.019
SchActiv	0.133	0.074
Picture	0.053	0.484
Interview	0.526	<0.001

LOR, letters of recommendation; DLetter, dean's letter; PS, personal statement; EduProf, educational profile; SchActiv, scholastic activity.

TABLE 2. Statistically Significant Correlations of Domains Amongst Each Other

Domains	r	p Value
LOR/Interview	0.193	0.009
Dletter/EduProf	0.422	<0.001
PS/Interview	0.192	0.010
Boards/SchActiv	−0.300	<0.001
Boards/Picture	−0.319	<0.001
Boards/Interview	−0.227	0.002

r, Spearman Correlation Coefficient; LOR, letters of recommendation; DLetter, dean's letter; PS, personal statement; EduProf, educational profile; SchActiv, scholastic activity.

DISCUSSION

While there are numerous studies that report perceived importance of application components to the overall

TABLE 3. Inter-rater Reliability Among Domain Rankings

Domain	kappa (κ)	95% Confidence Interval
LOR	0.108	0.035-0.181
DLetter	0.113	0.040-0.186
PS	0.208	0.135-0.281
EduProf	0.133	0.060-0.206
SchActiv	0.241	0.168-0.314
Picture	0.188	0.115-0.261
Interview	0.433	0.360-0.506

LOR, letters of recommendation; DLetter, dean's letter; PS, personal statement; EduProf, educational profile; SchActiv, scholastic activity.

orthopedic residency application, our data indicates that few of the listed application components actually have a significant impact on the final rank list created by the program after those applicants have been selected for an interview. In fact, our study found that once an applicant is granted an interview, only the interview performance and the personal statement have any correlation to the final rank list. This contrasts what other studies have shown to be perceived paramount components of an application. In one study, 19% of faculty even stated they do not place any importance on personal statements.[4] Surveyed faculty may have acknowledged the interview as an important factor, but they have often placed the objective measures of class rank, grades, or USMLE scores to be of paramount importance.[4,5,8,9] While these factors may play a role in the initial screening process of all applications or to overall match success, our study indicates that they do not play any role in the assignment of final rank amongst those who are ultimately interviewed.

Interview performance and personal statement are highly subjective components of the application. They can be considered part of the candidate's "affective domain," and as such can be notoriously difficult to measure.[3,10,11] Even if utilizing solely for a "rule-out" purpose, any psychopathology that may be indicative of future troublesome behavior may be masked in a carefully constructed essay or 12-minute interview. Furthermore, the components of these domains that may be contributing to the overall ranking may be far more arbitrary; indeed one study showed that interviewers do take personality into consideration simply by unconsciously choosing those that mirror their own.[12]

While these factors may foster a sense of camaraderie and similar work styles, little data supports it as a predictor of resident outcome. In fact, literature supports more objective components of an application as predictors of outcome, such as USMLE Step 1 and medical school grades.[2,13,14,15] This indicates that possibly some of the most time-consuming and subjective components for programs to review in an overall application are indeed guiding the selection of residents, while the more straightforward and overtly ranked components are the ones that predict outcome.

Our study also indicated that certain components of an application correlate to one another. For one, board scores inversely correlated with scholastic activity, interview, and personal photograph. This may be possibly due to a complacency effect of scoring well on USMLE exams; those that score high may not be as incentivized to perform additional scholastic activity and may prepare less for their subsequent interviews due to overconfidence. Inversely, those that perform poorly on the USMLE exams may augment their application with additional scholastic activity and have more motivation in the interview. The correlation of board score with personal photograph may be more enigmatic. Kleisner et al. showed that there is positive relationship between perceived intelligence when judging photographs and actual intelligence quotient for men only, and indeed most of the applicants in the study were men.[16] The inverse correlation with photograph ranking may be due to a similar additional attention to detail and preparation for application photographs by those who have not performed as well on the USMLE exams.

Interview performance correlated to personal statement and LOR. The interview performance and personal statement both are reflective of a common affective domain, and one may expect them to correlate as such.[10] Additionally, letter writers may have been positively influenced by the same personality traits of the applicant, which in turn may have garnered stronger letters. Finally, the correlation of educational profile and dean's letter may simply reflect a common affinity by raters for certain medical schools over others, as one may often construe from which program a student is based upon cues in the dean's letter.

The lack of strong interobserver and intraobserver reliability emphasizes the subjective and arbitrary nature of how individual components of an application are reviewed. Additionally, it displays that there are no common agreed-upon criteria in quantifying many of the individual components of an application. While interview performance and personal statement may have ultimately correlated to the final rank list, there was substantial variability in what raters considered to be a strong interview or statement. A common groundwork in which to rate each domain could provide for a more consistent ranking of domains. If the domains could be reproducibly ranked, then one could reliably measure them against outcomes such as resident performance in order to make an educated decision as to what components should be weighed differently to make a final rank list.

One limitation of this study is that it only considers applicants that were granted interviews. While this study indicates that only interview performance and personal statement factor into the final rank list of interviewed applicants, one cannot determine which of the other components may have been involved in the initial interview selection. While other components likely played a role in interview selection, this study does support the idea that once an interview is granted, objective measures such as board scores no longer play a significant role in final ranking.

Another limitation is the fact that while the application domains were de-identified, reviewers may not have been completely blind to certain domains. For one,

while all of the applications were reviewed for the granting of interviews by only one faculty member, that faculty member was one of the core who participated in the study so he may have still been able to identify some candidates based on his knowledge of all applications to the program. However, given the volume of applications and the time lag between interview offerings and the completed interviews, we believe this information had minimal carry-over from the interview selection process to confound the study. The study, though, does involve some applicants of whom the faculty members had prior knowledge, such as those that had performed externships at the program's institution, so the interview domain may not have been completely blinded for those applicants. Finally, certain domains of the application could potentially have overlapping information. For example, the dean's letter may mention scholarly activity pursuits or discuss the applicant's educational profile.

Other limitations may involve the weighting of various core members' influences on the final rank list. The modified Delphi approach assumes a common consensus of all faculty members with equal input, while in actuality some faculty members may be more vocal or may have opinions that are more heavily weighted in the group discussion. There was also variability in the submission of the interview performance rankings; 2 pairs of interviewers submitted combined lists and one submitted his own list. The combined lists likely were influenced differently by each member of the pair, while individually submitted lists prior to any discussions may have provided more independent data. Therefore, the individually submitted list received slightly greater weighting in the average over the paired lists. Finally, the data gathered represents a snapshot of only one interviewing cycle of one program; findings may not necessarily be representative of other programs. Future studies may involve multiple programs or may examine comparisons of individual rater's contributions to final rank lists.

CONCLUSIONS

Overall, our study indicates that for interviewed applicants, the interview performance and personal statement are the only application domains that correlate to the applicant's final ranking. These are the 2 domains that provide the most insight into the applicant's personality and behavioral attributes. Quantifiable measurements for these behavioral attributes may help to decrease variability of component rankings and allow reliable comparisons to resident outcomes. Knowledge of how these reported application domains contribute to the final rank list may allow programs to more efficiently and better manage their application review process.

ACKNOWLEDGMENT

The authors wish to thank Dr. John P. Gaughan for his statistical assistance and expertise.

REFERENCES

1. National Resident Matching Program and Association of American Medical Colleges, Charting Outcomes in the Match: Characteristics of Applicants Who Matched to their Preferred Specialty in the 2015 Main Residency Match, 2015.
2. Dirschl DR, Campion ER, et al. "Resident selection and predictors of performance: can we be evidence based?". *Clin Orthop Relat Res*. 2006;449:44–49.
3. Evarts CM. "Resident selection: a key to the future of orthopaedics". *Clin Orthop Relat Res*. 2006;449:39–43.
4. Bernstein AD, Jazrawi LM, et al. Orthopaedic resident-selection criteria". *J Bone Joint Surg Am*. 2002;84-A:2090–2096.
5. Bajaj G, Carmichael KD. "What attributes are necessary to be selected for an orthopaedic surgery residency position: perceptions of faculty and residents". *South Med J*. 2004;97:1179–1185.
6. Huntington WP, Haines N, et al. "What factors influence applicants' rankings of orthopaedic surgery residency programs in the National Resident Matching Program?". *Clin Orthop Relat Res*. 2014;472:2859–2866.
7. Martin M, Salzberg L. Resident characteristics to evaluate during recruitment and interview: a Delphi study". *Educ Prim Care*. 2017;28:81–85.
8. Thordarson DB, Ebramzadeh E, et al. "Resident selection: how we are doing and why?". *Clin Orthop Relat Res*. 2007;459:255–259.
9. Wagoner NE, Suriano JR. "Program directors' responses to a survey on variables used to select residents in a time of change". *Acad Med*. 1999;74:51–58.
10. Egol KA, Collins J, et al. "Success in orthopaedic training: resident selection and predictors of quality performance". *J Am Acad Orthop Surg*. 2011;19:72–80.

11. Mallott D. Interview, Dean's Letter, and Affective Domain Issues". *Clin Orthop Relat Res*. 2006;449:56–61.

12. Quintero AJ, Segal LS, et al. "The personal interview: assessing the potential for personality similarity to bias the selection of orthopaedic residents". *Acad Med*. 2009;84:1364–1372.

13. Clark R, Evans EB, et al. "Characteristics of successful and unsuccessful applicants to orthopedic residency training programs". *Clin Orthop Relat Res*. 1989: 257–264.

14. Turner NS, Shaughnessy WJ, et al. "A quantitative composite scoring tool for orthopaedic residency screening and selection". *Clin Orthop Relat Res*. 2006;449:50–55.

15. Klein GR, Austin MS, et al. "Passing the Boards: can USMLE and Orthopaedic in-Training Examination scores predict passage of the ABOS Part-I examination?". *J Bone Joint Surg Am*. 2004;86-A:1092–1095.

16. Kleisner K, Chvatalova V, et al. "Perceived intelligence is associated with measured intelligence in men but not women". *PLoS One*. 2014;9:e81237.

SUPPLEMENTARY INFORMATION

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.jsurg.2018.08.018.