The American Journal of Surgery 217 (2019) 568–571

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

The role of academic achievements and psychometric measures in the ranking process

Saad Shebrain*, Mohamed Arafeh, Gitonga Munene, Cynthia Shattuck, Lisa Miller, Steven Lu, David Schutter

Western Michigan University Homer Stryker MD School of Medicine, USA

ARTICLE INFO

Article history:
Received 23 July 2018
Received in revised form
4 September 2018
Accepted 7 September 2018

Keywords:
ERAS® application
Residents ranking
Surgery residency application

ABSTRACT

Introduction: Ranking candidates for residency positions is challenging. We hypothesize that applicant academic achievements and performance during the interview are equally important in the ranking process.

Methods: This is a retrospective study. Of 53 candidates interviewed during 2016–2017 cycle, 44 (83%) were ranked for 3 PGY1 positions. Each candidate was interviewed and scored in each of the following: USMLE Step 1 score, USMLE Step 2 score, research (RS), letters of recommendation (LOR), personal statement (PS), the way the candidate represented him/herself (RP), interest in the area (IN), answers to standardized questions (SQ), and degree of connection between the candidate and the interviewer (CN).

Results: Correlation and multiple regression analyses indicated an inverse relationship between ranking the candidates and USMLE2 ($r = -0.14$, $p = -0.364$), LOR ($r = -0.513$, $p < 0.001$), PS ($r = -0.414$, $p = 0.006$), RP ($r = -0.485$, $p = 0.001$), CN ($r = -0.605$, $p < 0.001$), IN ($r = -0.349$, $p = 0.022$), and SQ ($r = -0.480$, $p = 0.001$), USMLE1 ($r = -0.036$, $p = 0.838$) and RS ($r = -0.008$, $p = 0.96$). After controlling for the other variables, only CN reached statistical significance ($p = 0.033$).

Conclusion: Candidate non-cognitive measures during the interview weigh higher than academic performance in the ranking process.

Introduction

Every year a three-phase recruitment process of applicants (i.e. screening, interviewing, and ranking) for new Post Graduate Year 1 (PGY1) positions spans over a period of almost six months, September through March. This a daunting, time-consuming, and financially-demanding task that affects both the applicants and residency programs. While the screening process is facilitated by Electronic Residency Application Service (ERAS)®, the interview is consistently challenging and considered the most important determining factor for selecting the future residents.[1]

The in-person residency interview provides the best opportunity to evaluate long-term "fit" for both the applicant and the program for the minimum five years of post-graduate training, along with allowing the program to gauge the potential for successful completion of the surgery program. Most programs conduct the selection process using a non-standardized approach containing both subjective and objective measures. While objectively evaluating the applicant is feasible, it is challenging to do so subjectively without possible bias.

Surveys of General Surgery Program Directors in the United States indicate that candidate selection for interview is weighted toward academic (cognitive) achievements as demonstrated by United States Medical Licensing Examination (USMLE) scores, clinical clerkships grades, strength of letters of recommendation (LOR), research experience and publications, and rank in medical school.[2,3] While the screening process of potential candidates has been significantly facilitated through ERAS, the in-person interview process affords an opportunity to evaluate the second part of the equation including the non-cognitive personal characteristics of candidates such as interpersonal and communication skills. When compiling the rank list, Program Directors tend to give the most weight to the quality of the interview, the LORs, and the perception of fit within the team.[2,3] Another recent survey of general surgery residency programs in the United States shows that programs consider USMLE Step 1 as the single most important factor for

* Corresponding author.
E-mail address: saad.shebrain@med.wmich.edu (S. Shebrain).

https://doi.org/10.1016/j.amjsurg.2018.09.015
0002-9610/

preliminary screening, and the interview as the most important factor in determining the final selection. The final selection is relatively subjective and based on a combination of interview, USMLE scores, research experience, and personal judgment.[1]

The goal of this study is to identify which factors, whether cognitive, non-cognitive, or both, impact the ranking process.

Methods and material

Our academic general surgery residency program receives between 700 and 800 applications annually for 5 PGY1 positions (3 categorical and 2 preliminary). In the 2016–2017 application cycle, we received 758 applications. 67 candidates (9%) were invited for a face-to-face interview based on review of their applications. A total of 53 confirmed candidates were interviewed over 5 interview days in November and December 2016, with approximately 10–11 candidates scheduled per interview day. After institutional review board (IRB) approval, a retrospective study of the data of the ranked candidates (44 out of 53 interviewed) was reviewed and analyzed.

The night before the interview, the program hosts a residents-only reception dinner at a current surgery resident's home. This allows the candidates to meet with the residents in an informal, stress-free environment and gives them the opportunity to engage in a friendly discussion about the program and opportunities available to residents. The next day, interview day, a formal presentation regarding the program details is given to all applicants by the Program Director followed by the interview process. During the interview, each candidate sits for five 15-min interviews with the same interviewers (2–3 surgical faculty members, 2–3 senior residents, as well as with the Chairman of Surgery, and the Program Director). Each interviewer is given a packet with the candidate's complete application and an interview evaluation scoring form (IESF). This evaluation form consists of 9 categories: USMLE Step1 & USMLE Step 2 scores, research activity (RS), letters of recommendation (LOR), personal statements (PS), the way the candidate represented him/herself (RP), degree of connection between the candidate and the interviewer (CN), interest in the area and program (IN), and answer to a standardized question (SQ). The first three categories represent the academic (Cognitive) characteristics of the candidate, while the second six categories represent the personal (Non-Cognitive) characteristics. Each category is scored on a Likert Scale from 5 (strong) to 1 (weak). The IESF is completed for each candidate by each interviewer. Therefore, each candidate will have between 6 and 7 evaluation forms completed. The evaluation form also includes a question about any red-flag behaviors observed during interview, as well as a space for additional comments about the candidate.

At the end of the interview day, the applicants have an informal lunch with junior residents followed by a tour at one of our area hospitals, where the candidates meet with current surgical residents of all levels. While the applicant lunch and tour are taking place, the interviewers meet to discuss their impressions of the candidates and report their individual rank lists for that day's applicants. The individual interviewer rank lists are then combined and averaged in real-time, resulting in a preliminary rank list for the day.

In our program, we do not assign specific interview days for our non-designated preliminary positions. However, during the interview, we do discuss with some candidates the possibility of ranking them for both categorical and preliminary positions. As a result of those discussions, a few candidates were ranked for preliminary positions as well. The non-designated preliminary rank list is separate and was not used in analysis of this study.

During the ranking process, all surgical faculty members and senior residents are given equal input with individual members providing insight into the applicants they interviewed. We hypothesized that including many faculty and senior residents would reduce subjectivity bias when evaluating the candidates' personal (non-cognitive) characteristics. If two candidates achieve the same overall score during the interview process, then the interviewers vote on who should be ranked higher for the day. After the fifth interview day, the five preliminary rank lists are combined into a single list ordered by overall scores.

In January, the program director, faculty, and senior residents meet to review the complete preliminary rank list. During this time, the final rank list is created, taking into account final feedback from the faculty and residents on each candidate's application and impressions. This final list is submitted to the National Resident Matching Program (NRMP) before the deadline in late February.

Statistical analysis

A database with no personally identifiable information was created, and statistical analysis was performed using SPSS 24 statistical software package (SPSS, Inc, Chicago, Illinois). A statistical significance was considered for $P < 0.05$. Correlation analyses and stepwise multivariate regression were carried out to assess the relationships of candidate ranking with other study variables. Categorical variables are presented as frequencies and percentages, and continuous variables were reported as means and standard deviation (SD). Intraclass correlation coefficient was used to assess the extent of agreement between the final ranking and the preliminary ranking list.

Results

During the 2016–2017 application cycle, 53 (7%) of 758 screened applicants were interviewed for 5 PGY1 surgery positions (3 categorical and 2 non-designated preliminary). A total of 44 (83%) were included in the final rank list for categorical positions. 82% were allopathic candidates, and 18% osteopathic. 32% were female candidates. The final rank for preliminary positions was extracted from the categorical list.

Correlational and multiple regression analyses were conducted to examine the relationship between the final adjusted rank list of candidates and potential predictors in the ranking process. Results of this analysis are shown in Table 1 besides the average means of scores on the Likert 5-point scale for both cognitive (USMLE 1, USMLE 2 and RS) and non-cognitive (LOR, PS, RP, CN, and IN) measures. There is a strong inverse correlation between ranking the candidates and USMLE 2 ($r = -0.14$, $p = 0.364$), LOR ($r = -0.513$, $p < 0.001$), PS ($r = -0.414$, $p = 0.006$), RP ($r = -0.485$, $p = 0.001$), CN ($r = -0.591$, $p < 0.001$), IN ($r = -0.349$, $p = 0.022$), and SQ ($r = -0.493$, $p = 0.001$). Additionally, there is a slightly weak inverse relationship between candidate ranking and USMLE 1 ($r = -0.036$, $p = 0.838$) and RS ($r = -0.008$, $p = 0.96$). The multiple regression model with all predictors produced $R2 = 0.52$, $p = 0.002$. However, after controlling for the other variables in the mode, the only predictor with statistical significance was the degree of connection between the candidate and the interviewer ($p = 0.033$). The preliminary unadjusted rank list was compared to the final rank list and there was no difference. Interclass reliability was high, 91%, with 95% CI (0.84–0.95), and $p < 0.001$. When compared using an independent-samples *t*-test (Fig. 1), both combined cognitive and combined non-cognitive measures showed significant difference in the means (mean difference = −0.6, 95% CI (−0.79 to −0.40), $p < 0.001$.

Table 1
Academic (cognitive) and personal (non-cognitive) measures: The average means of the scores on the likert 5-point scale, and the correlations with ranking.

Measure		Mean (SD)	Correlation (r)	95% CI, p-value
Cognitive	USMLE 1	3.05 (1.08)	−0.036	(-0.33 to 0.27), 0.838
	USMLE 2	3.98 (1.08)	−0.142	(-0.42 to 0.17), 0.364
	RS	3.75 (0.84)	−0.008	(-0.30 to −0.29), 0.948
Non-cognitive	LOR	4.03 (0.39)	−0.513	(-0.70 to −0.25), <0.001
	PS	4.05 (0.39)	−0.414	(-0.64 to −0.13), 0.006
	RP	4.30 (0.41)	−0.486	(-0.69 to −0.22), 0.001
	CN	4.25 (0.39)	−0.605	(-0.77 to −0.37), <0.001
	IN	4.07 (0.40)	−0.349	(-0.59 to −0.06), 0.022
	SQ	4.23 (0.38)	−0.480	(-0.68 to −0.21), 0.001

CI = Confidence Interval.
SD = Standard Deviation.
USMLE 1 = The United States Medical Licensing Examination (USMLE) Step 1.
USMLE 2 = The United States Medical Licensing Examination (USMLE) Step 2.
RS = Research activity and publications.
LOR = Letter of Recommendation.
PS = Personal Statement.
RP = The way candidate represents him/herself.
CN = Degree of connection between candidate and Interviewer.
IN = Candidate's Interest in the area and the Program.
SQ = Candidate's answer to standardized questions.

Discussion

Selecting the best fit among candidates applying for a residency position is one of the most important educational responsibilities for Program Directors and faculty. The first phase of the candidate recruitment process hinges on careful and thorough evaluation of candidate applications. Applications are hosted by the Electronic Residency Application Service® (ERAS®),[4] which provides a centralized and flexible solution to the residency application and document distribution process, thereby facilitating the interview selection process for both the applicant and the residency program.

The second phase, interviewing the candidates, is the most crucial phase, and most heavily influences the final phase, ranking of the candidates. In February, the rank order list (ROL) is submitted to the National Resident Matching Program (NRMP),[5] which provides a uniform timeline for both applicants and programs to make their training selections. Through the Main Residency Match, applicants are "matched" to programs using the certified rank order lists (ROL) of the applicants and Program Directors. If an applicant does not match successfully, they may choose to apply for available unfilled positions during the Match Week Supplemental Offer and Acceptance Program (SOAP)®. Successfully filling all available residency positions on Match Day is one of a program director's most important goals. The result of the Match is dependent on the ranking strategy by both the applicants and the programs.

From a program perspective, ranking candidates is challenging and carries a risk of wide variability and personal bias. In 2011, a survey of surgery Program Directors, Department Chairs, and Associate Program Directors demonstrated that USMLE Step 1 scores are the most important factor used in screening criteria (37%), followed by USMLE Step 2 (24%). Final selection and ranking was based on a combination of USMLE scores, research experience, personal judgment, and the interview.[1,6] Traditionally "non-cognitive" attributes, such as interpersonal and professional communication and overall fit for the program, were more highly correlated with the final rank than academic performance alone when ranking candidates.

The purpose of this study was to determine which of the applicants' academic achievements and personal characteristics predicted the final ranking. Using the Likert 5-point scale, we found that USMLE Step1 and USMLE Step2 scores, as well as research and scholarly activity, were the least important factors in the final selection of candidates for the ranking list, with scores of 3.05, 3.75, and 3.98 respectively. One explanation for this is that most

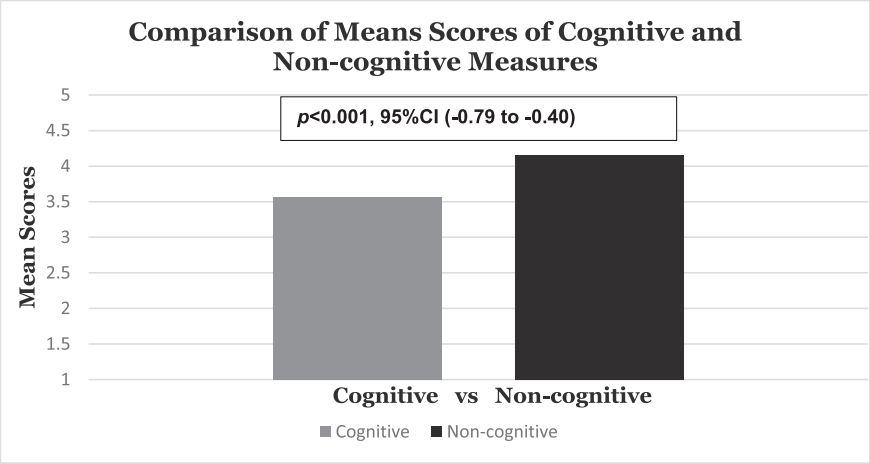

Fig. 1. Difference in mean scores between cognitive and non-cognitive measures.

programs adhere to a specific set of screening and selection criteria in ERAS prior to inviting the applicant to interview. These criteria rely heavily on academic performance demonstrated via USMLE Step 1 and 2 scores, clerkship grades, and scholarly activity. These criteria are standard within each residency, but undoubtedly vary between programs. Because screening of academic performance is completed before an invitation is made, a second re-assessment of known objective data during the interview adds no additional benefit. Therefore, the interview should be focused on assessing the "fit" and "alignment" between the applicant and the program.

In our study, we found that the personal (non-cognitive) characteristics carry more weight than the academic performance scores when ranking the candidates. In our study, the most important factors were: the way the candidate presented him/herself, the perceived degree of connection between the candidate and the interviewer, the candidate's answer to the standardized questions, the candidate's interest in the area and the program, the personal statement, and the strength of the letters of recommendation. Collectively, these subjective assessments are, in fact, the interview itself. This finding is supported by strong evidence showing that non-cognitive qualities of candidates, such as interpersonal and communication skills, professionalism, and patient care, are best evaluated through the personal interview, and that it remains a central part of the selection process.[7–9]

There is no doubt that the traditional interview process carries the risk of several weaknesses that could adversely impact the integrity of this phase of recruitment. Various studies have identified these pitfalls, including: interviewers making up their minds early in the interview, poor inter-rater and intra-rater reliability between the interviewers, content covered in unstructured interviews varying widely from applicant to applicant, interviewers tending to be influenced in their judgment more by unfavorable than by favorable information, and an unconscious tendency for interviewers to like and select people who look like and act like themselves.[6] One challenge when using subjective criteria to evaluate lies in the difficulty of establishing a standard measurement to ensure inter-rater reliability between evaluators on the faculty team. Only one recent study demonstrates the reliability of general surgery residents in evaluating ERAS applications.[9] In our study, each of the 44 candidates received between 6 and 8 evaluations, and we believe this helped reduce the variability and bias.

Although the adjustment of the rank list based on raw interview scores by the faculty using extraneous information is commonly performed, this potentially introduces significant subjective opinion bias to the largely objective process. Studies have found that unadjusted rank lists were more predictive of subsequent performance than the rank list adjusted by faculty after the interviews.[10] In our study, we made very few adjustments to the preliminary rank list when creating the final rank list. An Intraclass correlation coefficient was used to assess the extent of agreement between the final ranking and the preliminary ranking lists and was found to be high (91%, $p < 0.001$). Although this was an excellent agreement, we will, in the future, evaluate the potential factors that lead to the difference.

Our study faces several limitations in addition to those expected of retrospective studies. First, there is a small sample size, which adversely affects its ability to detect potential significant effects of many independent variables on the measured outcomes. Interviews were conducted on five different dates with five different sets of interviewers (core faculty and residents remain the same, but one community faculty rotates each date). This could lead to potential bias, especially among the community faculty. Some of the candidates had been M3 core students and others had been M4 rotating sub-interns at this institution leading to bias based on previous interactions not included in the interview date. While we stressed to interviewers that reading LORs and personal statements was an important component of the interview, subjective bias undoubtedly exists when scoring these items. This is further compounded because many LORs are becoming generic and do not add much value, especially those from writers who did not observe the applicants closely. Finally, this study was performed at a single institution, therefore limiting the generalizability of the results, and adding the potential for selection bias.

While the in-person residency interview remains the standard of practice in recruiting, there has been increased enthusiasm regarding the potential value of online interviews as an alternative path that could help ameliorate some of the financial burden of face-to-face interviews on both candidates and programs. Conflicting opinions on this option exist within the surgical community, however, a recent study showed that no difference in rate of acceptance into the program from either online or face-to-face interviews was observed, and both the groups noted that they were able to evaluate the program sufficiently.[11] Future studies are needed to support the possibility of a paradigm shift to more utilization of online interviews in the era of unparalleled advanced technology.

Conclusion

Ranking candidates for surgical residency is very challenging. Although candidates' academic achievement is an important factor for selecting and inviting candidates to interview, candidate interactions and psychometric measures during the interview are important factors in the ranking process.

Conflicts of interest

No conflict of interest.

References

1. 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(1):67–72.
2. Bernstein AD, Jazrawi LM, Elbeshbeshy B, et al. An analysis of orthopaedic residency selection criteria. *Bull Hosp Jt Dis*. 2002-2003;61(1-2):49–57.
3. Nies MS, Bollinger AJ, Cassidy C, Jebson PJ. Factors used by program directors to select hand surgery fellows. *J Hand Surg Am*. 2014;39(11):2285–2288.
4. https://www.aamc.org/services/468576/centralizedapplicationservices.html.
5. http://www.nrmp.org/intro-to-main-residency-match/.
6. Quintero AJ, Segal LS, King TS, Black KP. The personal interview: assessing the potential for personality similarity to bias the selection of orthopaedic residents. *Acad Med*. 2009;84(10):1364–1372.
7. Gilbart MK, Cusimano MD, Regehr G. Evaluating surgical resident selection procedures. *Am J Surg*. 2001;181(3):221–225.
8. Blouin D, Day AG, Pavlov A. Comparative reliability of structured versus unstructured interviews in the admission process of a residency program. *J Grad Med Educ*. 2011;3(4):517–523.
9. Dennis Y, Kim MD, Edward Gifford MD, et al. General surgery residents can Be a reliable resource in the evaluation of residency applications. *J Surg Educ*. 2015;72(6):e172–e176.
10. Selber JC, Tong W, Koshy J, Ibrahim A, Liu J, Butler C. Correlation between trainee candidate selection criteria and subsequent performance. *J Am Coll Surg*. 2014;219(5):951–957.
11. Vadi MG, Malkin MR, Lenart J, et al. Comparison of web-based and face-to- face interviews for application to an anesthesiology training program: a pilot study. *Int J Med Educ*. 2016;7:102–108.