

Trust Me, I Know Them: Assessing Interpersonal Bias in Surgery Residency Interviews

Chelsea Towaij¹, MD
 Nada Gawad, MD, MAEd
 Kameela Alibhai, BSc
 Danielle Doan, MAEd
 Isabelle Raïche, MD, MAEd

ABSTRACT

Background Residency selection integrates objective and subjective data sources. Interviews help assess characteristics like insight and communication but have the potential for bias. Structured multiple mini-interviews may mitigate some elements of bias; however, a halo effect is described in assessments of medical trainees, and degree of familiarity with applicants may remain a source of bias in interviews.

Objective To investigate the extent of interviewer bias that results from pre-interview knowledge of the applicant by comparing file review and interview scores for known versus unknown applicants.

Methods File review and interview scores of applicants to the University of Ottawa General Surgery Residency Training Program from 2019 to 2021 were gathered retrospectively. Applicants were categorized as “home” if from the institution, “known” if they completed an elective at the institution, or “unknown.” The Kruskal-Wallis H test was used to compare median interview scores between groups and Spearman’s rank-order correlation (r_s) to determine the correlation between file review and interview scores.

Results Over a 3-year period, 169 applicants were interviewed; 62% were unknown, 31% were known, and 6% were home applicants. There was a statistically significant difference ($P=.01$) between the median interview scores of home, known, and unknown applicants. Comparison of groups demonstrated higher positive correlations between file review and interview scores ($r_s=0.15$ vs 0.36 vs 0.55 in unknown, known, and home applicants) with increasing applicant familiarity.

Conclusions There is an increased positive correlation between file review and interview scores with applicant familiarity. The interview process may carry inherent bias insufficiently mitigated by the current structure.

Introduction

The Canadian residency selection process, overseen by the Canadian Resident Matching Service (CaRMS), follows common guidelines; however, the scoring systems and methodology used for selection are variable across programs. The process is comprised of at least 2 common phases—the file review and the interview. Each applicant’s file includes their curriculum vitae (CV), personal statement, academic records of rotations completed, and letters of reference. In Canada, applications do not include standardized test results, and medical schools only disclose grades in pass/fail form. Files are reviewed and scored using criteria that is unique to each program. Based on the generated application scores, only those above a determined threshold are invited to an interview. The style of interview and scoring is at the program’s discretion. The results of the interview, file review, and any other tool used in the

selection process are then collated and a rank list is generated.

Several studies have demonstrated a strong correlation between file review, interview scores, and final rank.¹ That said, studies have also demonstrated large variations in rank list depending on the stage at which interview scores are incorporated in the process.² The effect of blinding the interviewer to applicant data, such as CV, academic records, and choice of electives, has consistently been shown to decrease correlation between interview and file review scores. Up to 30% of the variance in interview scores is influenced by grades and standardized test results in US studies,³ suggesting that factors external to the interview may bias the process.

Applicants and programs place importance on the interview to aid in residency selection.^{4,5} The interview is an opportunity to assess factors, including communication, insight, motivation, and compassion, among other personal factors that are thought to be predictors of success. Programs that emphasize the importance of these subjective criteria have reported higher degrees of satisfaction with selection processes⁶; however, evidence to support interview

DOI: <http://dx.doi.org/10.4300/JGME-D-21-00882.1>

Editor’s Note: The online version of this article contains a sample data set programs can use to perform their own analysis.

performance as a predictor of residency performance is limited.⁷ There are data to support that familiarity with applicants in residency interview settings and oral examination settings may contribute to inflated assessments,^{8,9} although to our knowledge these studies have not been replicated in the setting of a structure multiple mini-interview (MMI).

A review of interview formats in medical school and residency selection demonstrates great heterogeneity, but certain factors, including structured questions, multiple observers, rating scales, rater training, and blinding to cognitive application data, are thought to improve reliability.¹⁰ The MMI is one such interview format with good feasibility, reliability, and predictive value because it employs multiple raters, which is thought to decrease the effect of any one interviewer's personal biases.¹¹ However, concerns have been raised about the MMI process and its perceived tendency to favor certain personality traits (eg, extroversion) and interviewees with an understanding of the local cultural norm.¹²

As such, the objectives of this study were to investigate whether, and if so to what extent, interviewer bias results from knowing applicants, and to provide a framework that would allow other programs to assess for bias within their own processes. To do so, file review and interview scores for known and unknown applicants applying to a Canadian general surgery residency program were compared. Given that file review is best at assessing more objective achievements such as productivity in research and academic awards, while interviews can more easily assess skills such as communication and collaboration, we hypothesized that the correlation would likely be weak. Therefore, if any pre-existing positive or negative sentiment toward an applicant biased the rating, we would expect to see higher levels of correlation between these scores. Determining the effect of bias on interviewers will inform our institution's residency selection process and identify strategies to mitigate these biases for future CaRMS cycles.

Methods

Setting and Participants

This retrospective cohort study included all Canadian medical graduates (CMGs) and international medical graduates (IMGs) in the 2019-2021 CaRMS cycles applying to the University of Ottawa General Surgery Residency Training Program—an urban, university-based program with 32 residents. There were 6 available residency spots each year over the 3 years, which consisted of 14 CMG and 4 IMG positions. Applicants were categorized as “home” applicants if

Objectives

To identify the extent to which prior knowledge of applicants influences interviewers during residency selection in a multiple mini-interview (MMI) format.

Findings

Interview scores of applicants better known to the residency program correlated more strongly with file review scores than applicants who had no previous interactions with the program.

Limitations

Small sample size, difficulty quantifying the degree to which applicants were truly familiar to their raters, and variability in scoring systems across the country limited generalizability.

Bottom Line

Residency interview processes may suffer from biases that may result in discrepancies between how known and unknown applicants are scored and ultimately selected.

they were enrolled in the University of Ottawa's undergraduate medical education (UME) program, “known” applicants if they were enrolled in another institution's UME program but had completed an elective in general surgery at our institution, or “unknown” if they had not.

The data for each applicant in the 2019-2021 CaRMS cycles was gathered retrospectively from the CaRMS database at our institution. Data included the candidate's name, their home school, whether they had completed an elective in Ottawa, whether they were a CMG or IMG applicant, file review score, and interview scores. All data was de-identified prior to analysis to ensure the anonymity of applicants who may be known to some of the study authors. The de-identification was performed by one author (D.D.), who is the program administrator and who already had access to the data given her role within the program. She does not have a role in the scoring of the file review or interview processes, nor does she contribute to resident rankings, acceptance, or assessments. One study author (C.T.) extracted the following information from the de-identified CaRMS database for each applicant: (1) CMG or IMG status; (2) enrolled at our institution (yes/no); (3) completed elective at our institution (yes/no); (4) file review score; and (5) interview score.

Selection Process

To generate file review scores, teams of resident and staff surgeons evaluate each applicant's personal statement, CV, letters of reference, and elective experience using a rubric. The file review process also considers residents' and staff surgeons' feedback about their experience working with any of the applicants on elective. This informal feedback was not gathered in 2021 so as not to disadvantage

students whose electives were cancelled due to the COVID-19 pandemic. The applicants who score the highest on the file review are offered an interview, and then the final rank list is calculated using a weighted proportion of the interview and file review scores.

Interview scores are generated through an MMI, where applicants rotate through structured stations. There are 8 to 10 stations per year which have been designed to highlight attributes identified by the program as essential to success as a surgical resident. Raters (2 per station: one senior resident and one staff surgeon) are asked to score applicants using a standardized rubric. Rater training is performed on the day of the interview to promote reliability of scoring. Use of the full scale and the importance of assessing only the applicant's interview performance are reviewed. The MMI score is calculated by averaging both raters' scores to generate a station score and adding up the values for each station. Since the total number of MMI stations differed each year, in order to compare the interviews scores across all 3 cycles for the purposes of this study, the sum of each station's score was then converted to a score out of 100. Interviewers do not have access to the applicant files or aggregate file review score during the interview; however, some interviewers may have participated in the file review process. During the 2021 cycle, interviews were conducted on a virtual video platform; otherwise, the format was unchanged.

Statistical Analysis

Descriptive statistics were used to report the medians and interquartile ranges for file review and interview scores for the "home," "known," and "unknown" applicant groups. The Kruskal-Wallis H test was used to compare the median interview scores between the 3 groups of applicants. A comparison of the median file review scores was not conducted because in the CaRMS 2019 and 2020 cycles, points were afforded to applicants who had completed an elective in Ottawa unlike in 2021 due to the COVID-19 pandemic. Spearman's rank-order correlation (r_s) was calculated to determine the correlation between file review and interview scores for each group. Interpretation of Spearman's r_s was based on precedents set in psychology literature, seeing as education research is similarly focused on human factors as compared to clinical research. Accordingly, <0.3 is considered a weak association, 0.4 - 0.6 is a moderate association, and >0.7 is a strong correlation.¹³ P values $<.05$ were considered statistically significant. SPSS 25 (IBM Corp, Chicago, IL) software was used for all analyses.

TABLE 1

Characteristics of Interviewed Applicants

Cycle	Unknown Applicants, n	Known Applicants, n	Home Applicants, n
2019	27	25	3
2020	26	26	4
2021	51	2	5
Total	104	53	12

Ethics approval was waived by the Ottawa Health Science Network Research Ethics Board because the primary purpose was to identify bias within our selection process to inform and enable quality improvement.

Results

The study cohort included 169 applicants who were interviewed between 2019 and 2021. The majority were "unknown" applicants (62%, 104 of 169), while "known" (31%, 53 of 169) and "home" applicants (6%, 12 of 169) were the minority. The 2021 cycle was a clear outlier due to the COVID-19 pandemic, when 88% (51 of 58) of applicants were unknown compared to previous groups where the known to unknown ratio was close to 1:1 (TABLE 1). In 2020 and 2021, of the 9 home applicants who applied to the program, all 9 (100%) were interviewed. Of the 34 known applicants, 28 (82%) were interviewed. Of the 244 unknown applicants, 78 (32%) were interviewed. These data were not retained from 2019. File review scores of the home applicants were above the cut-off for those offered an interview.

There was a statistically significant difference ($H=8.51$, $P=.01$) between the median interview scores, with home applicants scoring highest with a median (IQR) of 76.0 (13.8) compared to known applicants at 73.0 (10.0) and unknown applicants at 68.0 (10.3) out of a total 100 (TABLE 2). Comparing interview and file review scores, unknown applicants had a weak positive correlation between scores ($r_s=0.15$, $P=.14$). The strength of the positive association was greater with increasing familiarity between applicants and program, with known applicants showing a moderate-weak correlation ($r_s=0.36$, $P=.006$) and home applicants showing a moderate correlation ($r_s=0.55$, $P=.06$) between interview and file review scores (TABLE 2).

Discussion

This study demonstrates that there is greater correlation between file review and interview scores for applicants who are known to our program and that the extent of correlation increases with increasing

TABLE 2

Median File Review and Interview Scores and Spearman's rank-order correlation (r_s)

Applicant Type	n	Median File Review Score (IQR)	Median Interview Score (IQR)	Spearman's rank-order correlation (r_s)
Home	12	73.5 (9.50)	76.0 (13.8)	0.55
Known	53	70.0 (16.0)	73.0 (10.0)	0.36
Unknown	104	67.0 (11.0)	68.0 (10.3)	0.15

Abbreviation: IQR, interquartile range.

level of familiarity. It also demonstrates that there is a statistically significant difference in interview scores between “home,” “known,” and “unknown” applicants, with “home” applicants scoring the highest. Identifying applicant familiarity as a source of bias may help optimize residency selection processes. To our knowledge this is the first study isolating the potential impact of familiarity on the residency selection process when an MMI interview is used.

Residency interviews are supposed to measure different characteristics of the applicants as compared to the academic dossier.¹⁰ The ideal interview format has not been established, but current literature suggests that the interview should contribute to the final rank list by identifying and assessing applicants on specialty-specific traits.¹⁰

The MMI has been proposed as a solution to certain rater bias found in unstructured interviews¹⁴; however, the MMI has also been found to be susceptible to bias.^{11,12,15} Recently, authors have suggested that local institutions should attempt to collect validity evidence for their own MMI.¹⁵ Our study isolates applicant familiarity as a confounding factor that contributes to higher levels of correlation between file review and interview scores, which suggests that this bias may dilute the value of the interview in residency selection. This is in keeping with the well-established halo effect phenomenon, where a rater's overall perception of an applicant impacts the assessment of their attributes, and has been demonstrated at many levels of medical student and resident assessment.^{16,17} This study provides a framework for programs to analyze their file review and interview scores and determine Spearman's r_s , as it may provide clarity on the degree of halo effect intrinsic to their process. The online supplementary data provides a sample data set programs can use to perform their own analysis.

Unblinded interviews, where interviewers access applicant files, have higher levels of correlation between file review components and interview scores as compared to blinded interviews.^{3,18,19} In our MMI, although some interviewers may have been involved in the file review process in preceding months, they do not have direct access to the applicants' files. Overall,

our interview process is reflective of a structured and blinded interview format, which according to existing literature should mitigate the risk of rater bias.

Positive bias toward familiar applicants in interviews is partially mitigated by the file review process where a more directed and analytic comparison is performed.⁹ Higher interview scores noted in home applicants is in keeping with previous studies that suggest a degree of positive bias toward better known applicants.^{16,17,20} Known applicants who are viewed favorably (as evidenced by higher file review scores) do seem to score higher as compared to strong unknown applicants, suggesting that familiarity may yield a positive bias in interview scores. Higher interview scores among known and home applicants may also be secondary to the applicants' greater understanding of the program's cultural norms.

Assessment of unknown applicants is particularly relevant for the upcoming selection cycle given that in Canada visiting medical student electives continue to be on hold. The dynamics of the applicant pool has shifted as a result of these changes, with a small in-group of students who are very well known and a very large out-group of students who will interact with the program only virtually.²¹ Performance during visiting electives has been identified as the single most important variable in the selection process by certain programs.^{22,23} Our study shows that elective rotations can give an advantage to strong applicants. However, elective rotations are costly, so equity is a concern,²⁴ and the criteria used by programs to attribute them remained poorly defined.²⁵ As programs have been asked to revisit their selection process,²⁶ they should be aware of the potential risk of bias linked to the familiarity with the known applicants and the implications on diversity, equity, inclusion, and justice. Efforts to decrease bias in the interview process have shown that rater training and raising awareness regarding implicit bias can help decrease their influence.²⁷

Limitations of our study relate to the small sample size, especially of our “home” students. This has been mitigated by pooling data from 3 years when the selection process was comparable. Given the small number of home students, demographic information

(ie, age, gender) was not collected, as it would have been identifying, which also limited the potential to conduct further statistical analyses. Analysis of correlation tests carries intrinsic limitations and specific difficulties given that even minor correlations can be statistically significant. In this study, the comparison of increasing strength of correlation between each of the 3 groups provides the most meaning to this analysis. Further limitations included an inability to quantify the degree to which applicants were truly familiar to their raters. With respect to generalizability of this framework, we do note that, even though all programs have access to the same information, internal scoring systems may vary.

Efforts to reduce bias in favor of known applicants could include de-identifying files for review and integrating anecdotal feedback regarding applicants after interview scores have been collected to decrease interviewer bias. Future studies should investigate the impact of additional mitigation measures on interviewer bias, as well as study the degree to which secondhand knowledge of an applicant may bias interviews.

Conclusions

The degree of familiarity with applicants to a residency program corresponded to higher levels of correlation between file review scores and interview scores during the residency selection process. There were statistically significant higher interview scores seen among the “home” applicants most familiar to the residency program.

References

- Christakis PG, Christakis TJ, Dziura J, Christakis JT. Role of the interview in admissions at the University of Toronto ophthalmology program. *Can J Ophthalmol*. 2010;45(5):527-530. doi:10.3129/i10-041
- Gong H, Parker NH, Apgar FA, Shank C. Influence of the interview on ranking in the residency selection process. *Med Educ*. 1984;18(5):366-369. doi:10.1111/j.1365-2923.1984.tb01284.x
- Robin AP, Bombeck CT, Pollak R, Nyhus LM. Introduction of bias in residency-candidate interviews. *Surgery*. 1991;110(2):253-258.
- Wagoner NE, Suriano JR, Stoner JA. Factors used by program directors to select residents. *J Med Educ*. 1986;61(1):10-21. doi:10.1097/00001888-198601000-00002
- Wagoner NE, Gray GT. Report on a survey of program directors regarding selection factors in graduate medical education. *J Med Educ*. 1979;54(6):445-452. doi:10.1097/00001888-197906000-00001
- Perrault D, Arquette C, Fox P. Improving Resident Education through Directed Feedback *Plas Reconstruct Surg Glob Open*. 2021;9(25):14. doi:10.1097/01.GOX.0000735004.43197.10
- Olawaiye A, Yeh J, Withiam-Leitch M. Resident selection process and prediction of clinical performance in an obstetrics and gynecology program. *Teach Learn Med*. 2006;18(4):310-315. doi:10.1207/s15328015tlm1804_6
- Stroud L, Herold J, Tomlinson G, Cavalcanti RB. Who you know or what you know? Effect of examiner familiarity with residents on OSCE scores. *Acad Med*. 2011;86(suppl 10):8-11. doi:10.1097/acm.0b013e31822a729d
- Hauge LS, Stroessner SJ, Chowdhry S, Wool NL. Evaluating resident candidates: does closed file review impact faculty ratings? *Am J Surg*. 2007;193(6):761-765. doi:10.1016/j.amjsurg.2007.01.020
- Stephenson-Famy A, Houmard BS, Oberoi S, Manyak A, Chiang S, Kim S. Use of the interview in resident candidate selection: a review of the literature. *J Grad Med Educ*. 2015;7(4):539-548. doi:10.4300/JGME-D-14-00236.1
- Rees EL, Hawarden AW, Dent G, Hays R, Bates J, Hassell AB. Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37. *Med Teach*. 2016;38(5):443-455. doi:10.3109/0142159X.2016.1158799
- Alweis RL, Fitzpatrick C, Donato AA. Rater perceptions of bias using the multiple mini-interview format: a qualitative study. *J Educ Train Stud*. 2015;3(5):52-58. doi:10.11114/jets.v3i5.818
- Akoglu H. User's guide to correlation coefficients. *Turkish J Emerg Med*. 2018;18(3):91-93. doi:10.1016/j.tjem.2018.08.001
- Lemay JF, Lockyer JM, Collin VT, Brownell AKW. Assessment of non-cognitive traits through the admissions multiple mini-interview. *Med Educ*. 2007;41(6):573-579. doi:10.1111/j.1365-2923.2007.02767.x
- Reiter H, Eva K. Vive la différence: the freedom and inherent responsibilities when designing and implementing multiple mini-interviews. *Acad Med*. 2018;93(7):969-971. doi:10.1097/ACM.0000000000002042
- Nisbett RE, Wilson TD. The halo effect: evidence for unconscious alteration of judgments. *J Pers Soc Psychol*. 1977;35(4):250-256. doi:10.1037/0022-3514.35.4.250
- Sherbino J, Norman G. On rating angels: the halo effect and straight line scoring. *J Grad Med Educ*. 2017;9(6):721-723. doi:10.4300/JGME-D-17-00644.1
- Miles WS, Shaw V, Risucci D. The role of blinded interviews in the assessment of surgical residency

- candidates. *Am J Surg*. 2001;182(2):143-146. doi:10.1016/S0002-9610(01)00668-7
19. Smilen SW, Funai EF, Bianco AT. Residency selection: should interviewers be given applicants' board scores? *Am J Obstet Gynecol*. 2001;184(3):508-513. doi:10.1067/mob.2001.109868
 20. McKinsty BH, Cameron HS, Elton RA, Riley SC. Leniency and halo effects in marking undergraduate short research projects. *BMC Med Educ*. 2004;4:1-5. doi:10.1186/1472-6920-4-28
 21. Bernstein SA, Gu A, Chretien KC, Gold JA. Graduate medical education virtual interviews and recruitment in the era of COVID-19. *J Grad Med Educ*. 2020;12(5):557-560. doi:10.4300/JGME-D-20-00541.1
 22. Drolet BC, Brower JP, Lifchez SD, Janis JE, Liu PY. Away rotations and matching in integrated plastic surgery residency: applicant and program director perspectives. *Plast Reconstr Surg*. 2016;137(4):1337-1343. doi:10.1097/PRS.0000000000002029
 23. Weissbart SJ, Stock JA, Wein AJ. Program directors' criteria for selection into urology residency. *Urology*. 2015;85(4):731-736. doi:10.1016/j.urology.2014.12.041
 24. Winterton M, Ahn J, Bernstein J. The prevalence and cost of medical student visiting rotations. *BMC Med Educ*. 2016;16(1):1-7. doi:10.1186/s12909-016-0805-z
 25. Huebner C, Adnan M, Kraeutler MJ, Brown S, Mulcahey MK. Use of the United States Medical Licensing Examination Step-1 Score as a screening tool for orthopaedic surgery away rotations. *J Bone Joint Surg Am*. 2019;101(20):e106. doi:10.2106/JBJS.19.00265
 26. Sternberg K, Jordan J, Haas MRC, Deiorio NM. Reimagining residency selection: part 2—a practical guide to interviewing in the post-COVID-19 era. *J Grad Med Educ*. 2020;12(5):545-549. doi:10.4300/JGME-D-20-00911.1
 27. Maxfield CM, Thorpe MP, Dessler TS, et al. Awareness of implicit bias mitigates discrimination in radiology resident selection. *Med Educ*. 2020;54(7):637-642. doi:10.1111/medu.14146



Chelsea Towaij, MD, is a Postgraduate Year 5 Resident, Division of General Surgery, Department of Surgery, Faculty of Medicine, University of Ottawa, The Ottawa Hospital, Ottawa, ON, Canada; **Nada Gawad, MD, MAEd**, is a Surgical Fellow, Division of General Surgery, Department of Surgery, Faculty of Medicine, Department of Innovation in Medical Education (DIME), University of Ottawa, The Ottawa Hospital, Ottawa, ON, Canada; **Kameela Alibhai, BSc**, is a Third-Year Medical Student, Division of General Surgery, Department of Surgery, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada; **Danielle Doan, MAEd**, is a Program Administrator, Division of General Surgery, Department of Surgery, Faculty of Medicine, Eric Poulin Office of Education, University of Ottawa, The Ottawa Hospital, Ottawa, ON, Canada; and **Isabelle Raiche, MD, MAEd**, is an Assistant Professor of Surgery, Division of General Surgery, Department of Surgery, Faculty of Medicine, DIME, University of Ottawa, The Ottawa Hospital, Ottawa, ON, Canada.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

Corresponding author: Chelsea Towaij, MD, University of Ottawa, Ottawa, Ontario, Canada, ctowaij@toh.ca, Twitter @ChelseaTowaij

Received September 8, 2021; revisions received January 24, 2022, and February 27, 2022; accepted February 28, 2022.