A transparent and defensible process for applicant selection within a Canadian emergency medicine residency program

Quinten S. Paterson, MD*; Riley Hartmann, MD, MSc*; Rob Woods ⓘ, MD, MMEd*; Lynsey J. Martin, MD, MHPE*; Brent Thoma ⓘ, MD, MA, MSc*

CLINICIAN'S CAPSULE

What is known about the topic?
Best practices are published to guide the applicant selection process to Canadian emergency medicine residency programs, but uptake is variable.
What did this study ask?
Can novel approaches and best practice guidelines be feasibly used to guide the resident selection process?
What did this study find?
We outline a defensible and transparent process using best practices and novel techniques that can be replicated by others.
Why does this study matter to clinicians?
The residency selection process is under scrutiny in Canada so best practices should be implemented in a transparent way.

ABSTRACT

Objectives: The Canadian Resident Matching Service (CaRMS) selection process has come under scrutiny due to the increasing number of unmatched medical graduates. In response, we outline our residency program's selection process including how we have incorporated best practices and novel techniques. **Methods**: We selected file reviewers and interviewers to mitigate gender bias and increase diversity. Four residents and two attending physicians rated each file using a standardized, cloud-based file review template to allow simultaneous rating. We interviewed applicants using four standardized stations with two or three interviewers per station. We used heat maps to review rating discrepancies and eliminated rating variance using Z-scores. The number of person-hours that we required to conduct our selection process was quantified and the process outcomes were described statistically and graphically. **Results**: We received between 75 and 90 CaRMS applications during each application cycle between 2017 and 2019. Our

overall process required 320 person-hours annually, excluding attendance at the social events and administrative assistant duties. Our preliminary interview and rank lists were developed using weighted Z-scores and modified through an organized discussion informed by heat mapped data. The difference between the Z-scores of applicants surrounding the interview invitation threshold was 0.18-0.3 standard deviations. Interview performance significantly impacted the final rank list. **Conclusions**: We describe a rigorous resident selection process for our emergency medicine training program which incorporated simultaneous cloud-based rating, Z-scores, and heat maps. This standardized approach could inform other programs looking to adopt a rigorous selection process while providing applicants guidance and reassurance of a fair assessment.

RÉSUMÉ

Objectif: Le processus de sélection des résidents au Canada a été soumis à un examen en raison du nombre croissant de diplômés non jumelés en médecine. Aussi avons-nous décidé d'exposer, dans les grandes lignes, un processus de sélection des candidats à un programme de résidence, ainsi que la manière dont ont été intégrées des pratiques exemplaires et des techniques novatrices. **Méthode**: Des réviseurs de dossiers et des intervieweurs ont été choisis afin d'atténuer les préjugés sexistes et d'accroître la diversité de l'équipe d'évaluation. Quatre résidents et deux médecins traitants ont évalué chacun des dossiers à l'aide d'un modèle de revue commun, sauvegardé dans le nuage afin de permettre des évaluations simultanées. De leur côté, les candidats ont passé un entretien dans quatre salles de rencontre montées pareillement et comptant deux ou trois intervieweurs chacune. Nous avons utilisé des cartes de couleurs pour examiner les différences d'évaluation et éliminer les variations à l'aide de la technique des écarts réduits Z. Le nombre d'heures-personnes nécessaires pour mener à terme le processus de sélection a été quantifié, et les résultats sont présentés sous forme de statistiques et de graphiques.

From *Department of Emergency Medicine, College of Medicine, University of Saskatchewan, Saskatoon, SK.

Correspondence to: Dr. Quinten Paterson, PGY-4 FRCPC Emergency Medicine, Department of Emergency Medicine, College of Medicine, University of Saskatchewan, Room 2646, 103 Hospital Drive, Saskatoon, SK S7N 0W8; Email: quintenpaterson@gmail.com

Résultats: Nous avons reçu de 75 à 90 demandes de jumelage du CaRMS au cours de chacun des cycles, entre 2017 et 2019. Le processus global de sélection a nécessité 320 heures-personnes sur une base annuelle, outre la participation à des activités sociales et les responsabilités d'adjoint administratif. Les listes préliminaires d'entrevue et de classement ont été élaborées à l'aide des écarts Z pondérés, puis modifiées dans le cadre d'une discussion structurée et guidée par les cartes de couleurs. La différence entre les écarts réduits Z des candidats relativement au seuil d'invitation à une entrevue était de 0,18-0,3 en écart type. La performance des candidats à l'entrevue a eu une forte incidence sur la liste finale de classement.

Conclusions: A été exposé, dans l'article, un processus rigoureux de sélection des résidents à un programme de formation en médecine d'urgence, intégrant à la fois un modèle d'évaluation infonuagique, des formules de calcul d'écarts réduits Z et des cartes de couleurs. Cette approche structurée pourrait à la fois servir de guide à des responsables d'autres programmes à la recherche d'un processus rigoureux de sélection, et donner aux candidats une direction à suivre et l'assurance d'une évaluation juste.

INTRODUCTION

Medical students, residents, and the media have scrutinized the Canadian Resident Matching Service (CaRMS) selection process because of the rising rates of unmatched Canadian medical students.[1-5] Specifically, critics have flagged the process as being opaque and plagued by bias and subjectivity.[1,6,7] Furthermore, the CaRMS selection processes are not standardized among programs that has led to uncertainty and frustration among applicants.[8,9] While Best Practices in Applications and Selection have been published,[10] implementation rates are variable.[8] Despite recommendations for the scholarly dissemination of innovations in the selection process,[10] we are unaware of any resident selection processes that have been described comprehensively in traditional or grey literature.

In response to these concerns and calls for greater transparency,[1,8,11] we describe the selection process used by our Canadian emergency medicine (EM) residency program, its innovations, and its outcomes. We believe that our standardized approach is rigorous and defensible and hope that its dissemination will encourage other programs to share and enhance their own processes while complying with best practices and spreading innovations.

METHODS

The University of Saskatchewan Behavioral Ethics Board deemed the data presented within this manuscript exempt from ethical review. Figure 1 outlines our overall CaRMS process in a flow diagram of a de-identified sample year. Appendix 1 contains a sample spreadsheet demonstrating how application and interview data are recorded and analyzed.

Application requirements

Our application requirements and the personal characteristics assessed as part of the selection process are outlined within our program description available at www.carms.ca. Our requirements included a personal letter, three letters of reference, curriculum vitae (CV), Medical School Performance Record (MSPR), Computer-Based Assessment for Sampling Personal Characteristics (CASPer) score, medical school transcript, proof of Canadian citizenship, English language proficiency (if applicable), and a personal photo (used as a memory aid following interviews). Appendix 2 outlines our instructions for the personal letter and letters of reference and describes the characteristics we assessed throughout the selection process. We chose these characteristics as we felt that they best mapped to our program's goals and mission statement. Prior to the publication of this manuscript, applicants were not aware of our specific application review process or scoring distributions. Applicants submitted all required application materials to www.carms.ca.

Application review

Our program reviewed only complete applications. Two staff physicians and two to four residents reviewed each application. The number of file reviewers changed annually based on the total number of trainees in our program at the time and their relative availability. Reviewer training occurred by distribution of standardized instructions
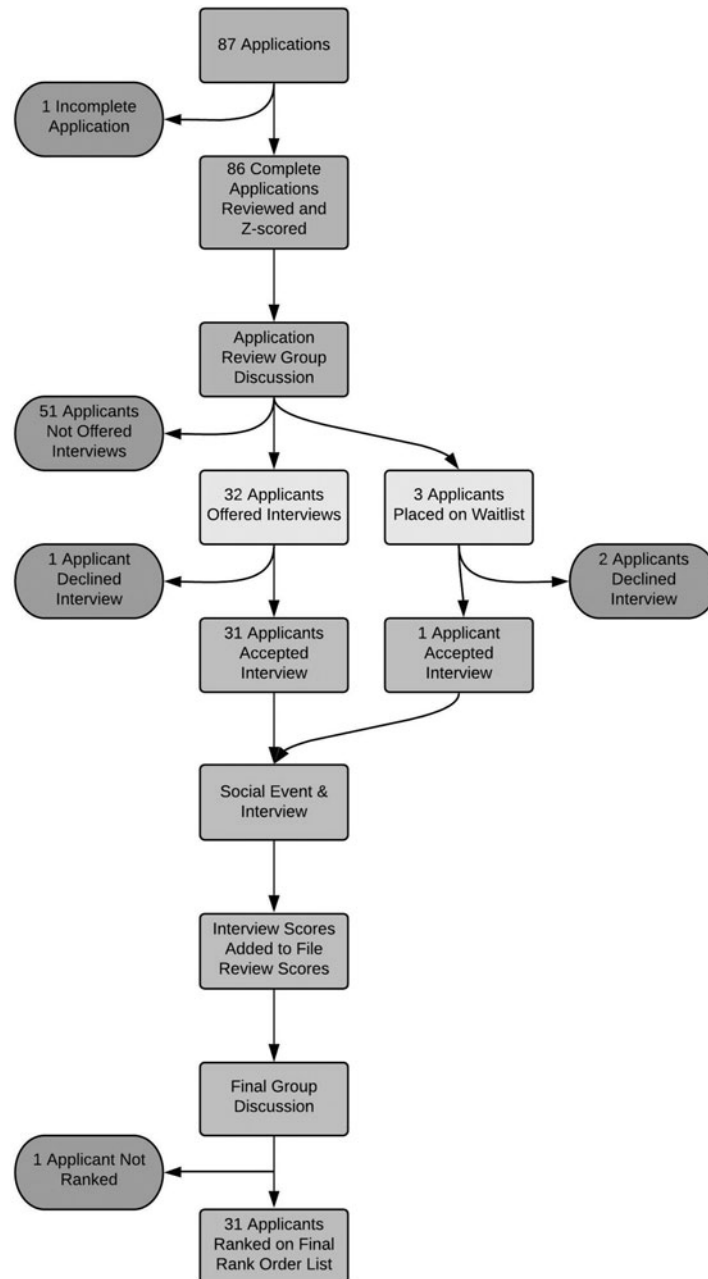
Figure 1. Flowchart outlining the application review and interview process in a de-identified year.

(Appendix 3).[10] We encouraged reviewers to review no more than 10 files in a single day to mitigate respondent fatigue.[12] We intentionally selected staff reviewers to ensure a diverse range of years of experience, areas of interest, gender, and certification route (e.g., Royal College of Physicians and Surgeons of Canada [Royal College] or Certification in the College of Family Physicians [CCFP-EM]). We selected resident reviewers to ensure diversity in years of training. We assigned reviewers such that each applicant's file was reviewed by both male and female reviewers. Reviewers self-disclosed conflicts of interest and were only assigned to review the files of applicants with whom they did not have a conflict. No reviewer was assigned more than 30 files, and no two reviewers reviewed the same 30 applicants. Reviews were completed within a three-week period. We suggested that reviewers spend 15–20 minutes on each file (total of 7.5–10 hours).

We created and pre-populated a cloud-based spreadsheet (Google Sheets, California, USA), with applicant names and identification codes, reviewer assignments, data entry cells, and formulas. Each reviewer was given a spreadsheet tab to rate their assigned applicants. File reviewers used a standardized, criteria-based scoring rubric for each section of the application (Appendix 4). Free-text cells on the spreadsheet allowed comments to be recorded to serve as memory aids. To help stratify which applicants near the interview cut-off threshold should be granted an interview, we asked reviewers, "Should we interview this applicant - yes or no?". The CASPer score was visible to file reviewers on our data entry spreadsheet but was not assessed by file reviewers.

This cloud-based format allowed the simultaneous collection of data from multiple file reviewers while also performing real-time ranking calculations. The pre-populated formulas converted each reviewer's scores into a Z-score for each applicant using the formulas outlined in Appendix 5. The Z-scores were then amalgamated into a pre-interview Z-score using a weighted average incorporating resident file review Z-scores (5/12ths), staff file review Z-scores (5/12ths) and the CASPer Canadian medical graduate (CMG) Z-score (2/12ths) (Figure 2; Appendix 5). Applicants were ranked by their amalgamated pre-interview Z-score.

File review group discussion

File reviewers and administrative staff met to finalize the interview and wait list after all files were reviewed. Discussion was facilitated using a spreadsheet detailing applicants' scores, ranking, and reviewers' free-text comments. We heat-mapped the ratings (red lowest score/yellow median score/green highest score colour gradient) to visually highlight discrepancies among CASPer CMG Z-scores, the file review Z-scores of each reviewer, and the overall file review Z-score of each applicant.

The files of each applicant ranked in a position that would result in an interview (top 32), as well as at least eight positions below this threshold, were discussed. This discussion followed a loose structure that included professionalism concerns and the review of discordant Z-scores. Changes to the interview list were generally slight, but substantial changes occurred if egregious professionalism concerns existed. Changes were generally made by consensus. If consensus could not be achieved, the program director made a final determination. For each application cycle, we sent 32 interview invitations

and three wait list notifications by email within four weeks of the file review opening. Applicants had 10 days to accept or decline their interview offer.

Social events

Our program holds social events each year to welcome the applicants to the city, encourage socializing in an informal setting, and provide additional program information. Specifically, a faculty member hosted an informal gathering for the applicants and the members of the program at their home, residents led tours of the city and the main hospital, and the program hosted a lunch that incorporated a presentation describing our program. While the social events were not formally evaluated, professionalism was expected. Professionalism concerns raised by faculty, residents, and administrative staff were considered in the final rank list discussion.

Interviews

Once again, we created and pre-populated a cloud-based spreadsheet (Google Sheets, California, USA), with applicant names and identification codes, interview room assignments, data entry cells, and formulas. Interviewers rated each applicant on a station-specific tab of the spreadsheet. Each applicant completed four 12-minute multiple-mini interview (MMI) stations with three-minute breaks between stations (one hour total). Four applicants completed the circuit each hour, with all 32 interviews being conducted on a single interview date. Each station consisted of an informal ice-breaker question, followed by a behavioural-style interview question (Appendix 6). We designed the main interview questions to assess the desired qualities of applicants listed as part of our CaRMS description. If time permitted, predetermined standardized follow-up questions or the applicants' own questions could be asked.

We provided interviewers with standardized instructions prior to the interview date (Appendix 6). Each station had two or three interviewers and consisted of at least one male and one female interviewer. One station solely comprised current residents while the remaining three stations comprised CCFP-EM and Royal College EM staff, along with an occasional resident. In 2019, an EM social worker also participated as an interviewer. Interviewers determined each applicant's score individually within each room. Each station was scored on a 10-point scale based on a predefined description of the ideal answer (Appendix 6). The pre-

Pre-Interview Scoring

Post-Interview Scoring

Figure 2. Pre- and post-interview scoring distributions.

populated formulas calculated the average interview ratings for each applicant and converted them into station-specific Z-scores using the formulas outlined in Appendix 5.

Final statistical analysis

Following the interviews, we calculated each applicant's final score using predetermined weightings: residents' file review (1/4th), staff's file review (1/4th), CASPer score (1/10th), and each interview station (1/10th per station) (Figure 2; Appendix 5). The weighting of each element was predetermined by the program director through consultation with faculty and residents. The goal of the weighting was to map the desired characteristics of applicants to the available tools in the applicants' CaRMS files.

Final group discussion and rank list entry

We invited all residents, staff physicians, interviewers, and administrative staff to participate in a two-hour

group discussion immediately following the interviews. We used a heat-mapped spreadsheet outlining the final statistical analysis to highlighted discordant Z-scores and displayed the 32 applicants' photos and names on a whiteboard in order of their final scores.

The final group discussion followed a loose structure organized by the preliminary rank list and key topics. Applicants were discussed from highest to lowest rank. Key topics that we discussed for each resident included professionalism concerns and discordant Z-scores. In some cases, we identified professionalism concerns during social events or the interview that warranted removal from or modifications to the rank list. Our discussion of discordant Z-scores clarified the reason for these discrepancies and whether they warranted minor rank list adjustments.

For applicants with nearly equivalent final scores, we further considered the diversity of our resident group and the applicants' fit within our program culture. All adjustments to the rank order list were made by consensus, with the program director making a

final determination when consensus could not be achieved.

Administrative staff submitted the final rank list of up to 32 applicants to CaRMS for a program admission quota of three residents per year. Upon completion of the interviews, we securely stored all file review data, with access restricted to the program director and spreadsheet creator. We re-evaluate our overall process annually.

RESULTS

We received between 75 and 90 CaRMS applications during each application cycle between 2017 and 2019. Our file review process required approximately 12 hours per file reviewer, 10 hours per interviewer, and 4 hours for spreadsheet generation and calculations by an experienced spreadsheet user. The cumulative time required for the entire process was approximately 320 person-hours annually, not including attendance at the social event or administrative assistant duties.

Figure 3 demonstrates the ranking of individual applicants who applied to our program in a de-identified sample year. Some applicants who ranked within the top 32 by application score were not offered an interview following the post-file review discussion. Conversely, some applicants who were ranked lower than the interview invitation threshold were moved up the rank list and offered an interview. For example, applicant 22 declined an interview, and applicants 33 and 34 declined waitlist offers that resulted in applicant 35 being interviewed. The interview scores altered the rank list based on individual applicant performance. Post-interview discussion resulted in minor rank list alterations. In one case, an applicant was interviewed but not ranked because of professionalism concerns.

Figure 4 demonstrates the tight spread of scores surrounding the interview invitation threshold (i.e., 32nd position) after file review. The range of Z-scores among the applicants rated 27th to 37th was small in 2017 (0.12–0.42), 2018 (0.12–0.42), and 2019 (0.18–0.36).
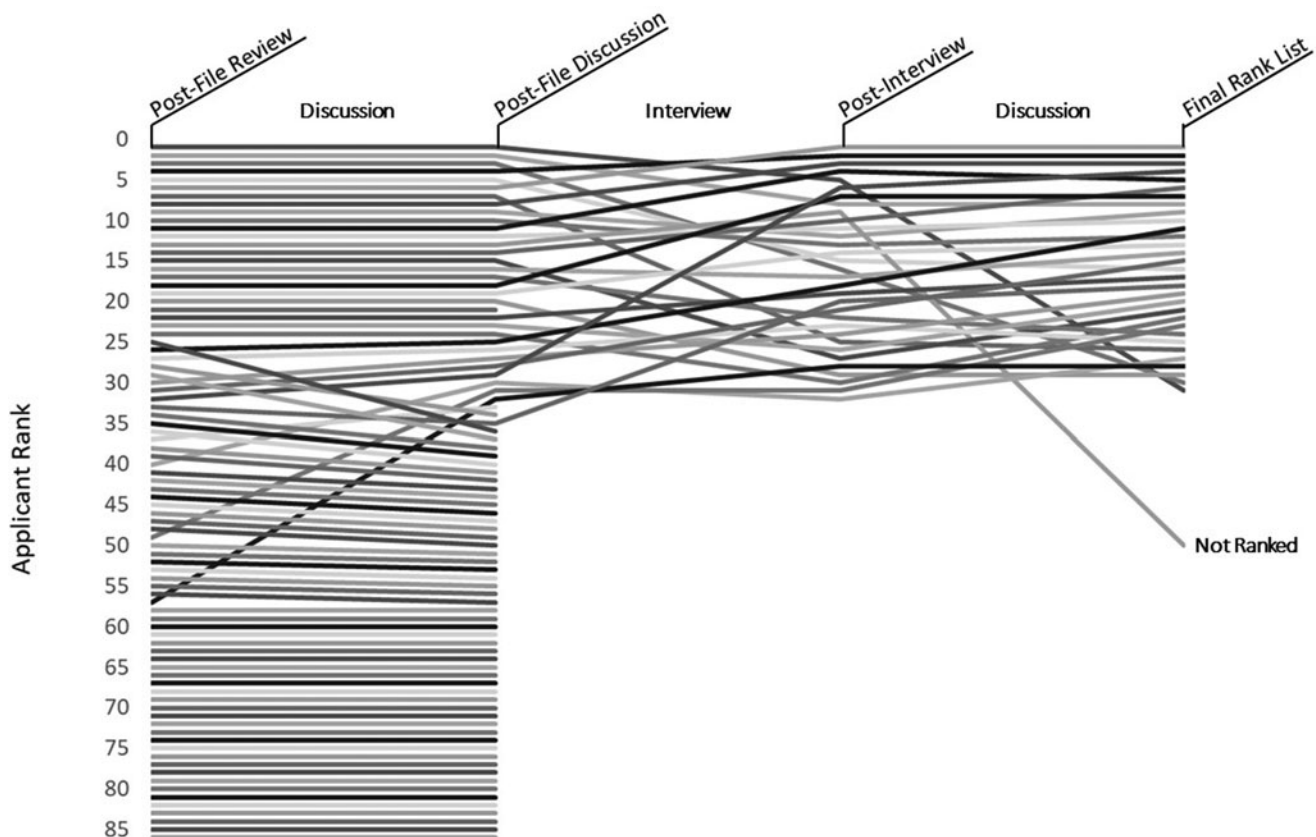

Figure 3. Visual display of applicant rankings through the stages of the CaRMS application and interview ranking process in a sample year.
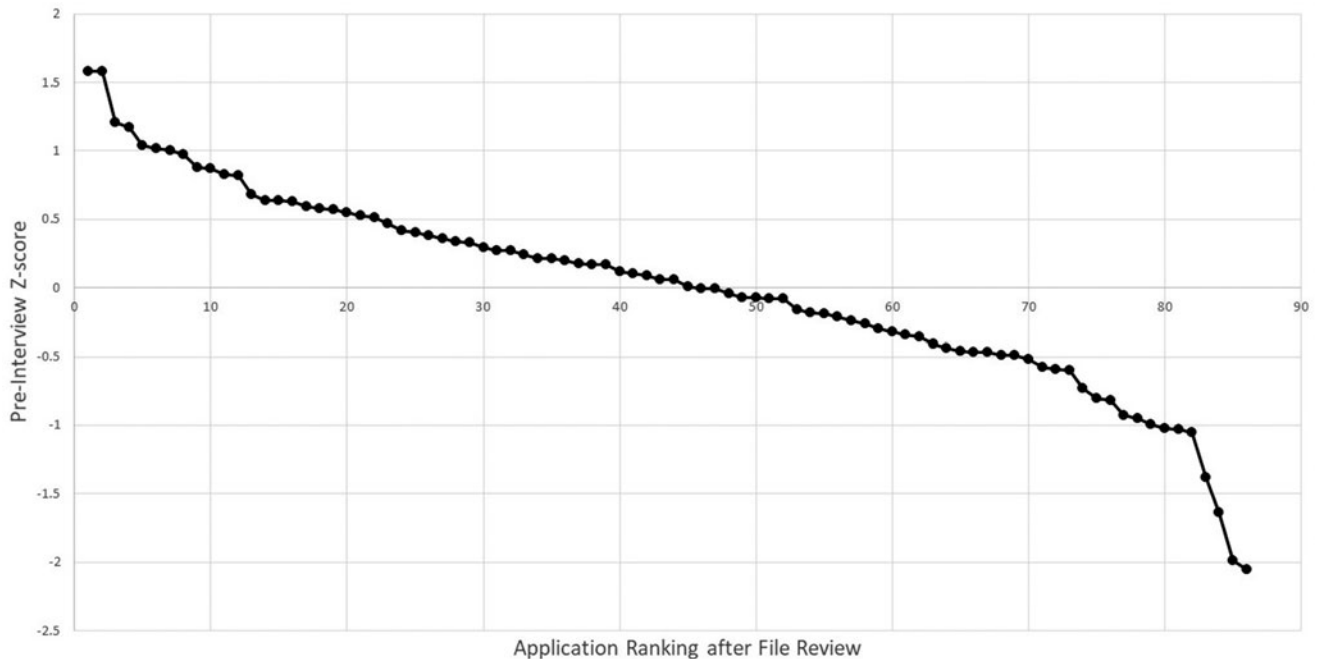
Figure 4. Pre-interview Z-score versus application ranking after file review.

DISCUSSION

While there are many robust and viable approaches to resident selection, we believe medical students and residency programs would benefit from openly published examples of best practices. We have comprehensively described our resident selection methods to address well-founded concerns with the CaRMS process. We hope that reviewing this description would allow applicants to understand how they would be assessed and give programs a blueprint on which to build. Though largely a descriptive manuscript, we have also described several innovations that have not previously been discussed in the literature.

Comparison with previous work

Research on the emergency medicine match is prominent in the United States. However, it generally focuses on evaluating specific questions.[13-20] One previous Canadian study provided a detailed description and analysis of their interview process.[21] To our knowledge, this is the first comprehensive description of an emergency medicine program's entire resident selection process in the literature.

Our process was developed in keeping with the Best Practices in Applications and Selection recommendations[10] and the practices of other residency programs about which we learned through personal communications. We incorporated strong assessment practices such as criterion-referenced rating scales[22] and multiple assessments from diverse raters.[23] Furthermore, we assessed characteristics that are associated with success in residency (reference letters and publication numbers[17]) and have demonstrated reliability (MMIs[24] consisting of structured interviews with predefined assessment tools[25]). Informal communications with other programs suggest that our use of simultaneous cloud-based rating, Z-scores, and heat maps are innovations that are not commonly used.

Moving forward, our approach to resident selection would differ dramatically from traditional processes because of its transparency. Applicants are historically unaware of the weighting of each component of the file and interview for different programs.[9, 26-29] Within this report, we have outlined this clearly (Figure 2) and remain committed to sharing this information openly with our applicants as our selection processes evolves. As applicants' knowledge of the selection process improves, we anticipate that they will have a greater understanding of how, why, and what is being assessed throughout the selection process.

Strengths and limitations

Some aspects of our process may be controversial. For example, the use of discussion to adjust the final

interview and rank lists could be considered subjective and, therefore, inappropriate.[10] However, previous publications have highlighted how having close colleagues and high social relatedness are important to resident wellness and success.[30,31] Further, our program finds value in these discussions. Like qualitative research, we believe that when they are conducted in a rigorous and focused fashion,[1] they can provide important insights. Though some have postulated that using vaguely defined constructs such as "fit" are likely to disadvantage non-traditional applications,[7] we use them as an opportunity to discuss inequities and underrepresented populations.[11] The minute differences in scores among applicants near the interview invitation threshold (Figure 4) suggest that the differences in the applications between applicants who received and did not receive an interview invitation was minimal.

The major limitation of our description was that we have not performed a robust evaluation of the validity or reliability of our selection process. The lack of access to the outcomes of applicants that were not interviewed or matched to our program also makes an evaluation of validity challenging. Our historic data allow us to track some measures of reliability. Our preliminary analyses of these data suggest that the reliability of our file review varies from year to year. We have considered rater training to improve their reliability,[32,33] but given the extensive volunteer commitment that we already require of reviewers, we anticipate that there would be little interest. We have also considered having raters evaluate a single component of all applications (e.g., the personal letter) but believe that this reductionist approach could miss insights that would be gleaned from a more holistic application review.[34,35] Instead, we utilized Z-scores to normalize variations in scoring. While not ideal, this technique eliminates the advantage or disadvantage received by applicants reviewed by a team of generous or stringent reviewers.

One threat to the validity of our process is that we were not able to confirm its fidelity. For example, we advised file reviewers to limit their reviewing to 10 files per day but did not take steps to ensure this was adhered to. Further, we did not ensure that interviewers rate applicants independently so they could have possibly influenced each other's ratings. As interviewers also occasionally served as file reviewers, it was also possible that the file review could have positively or negatively influenced interview rating. While we had a basic framework to guide our interview and rank list discussions, it was not always strictly enforced. These variations represent opportunities to improve in future years.

Lastly, we recognize that our process has been developed over the nine years since our program's founding and is rooted in the local culture of our site and specialty. It will need to be altered prior to adoption by other programs to account for site- and specialty-specific contexts. In particular, the recruitment of sufficient file reviewers and interviewers is unlikely to be possible in programs that have fewer faculty and residents, receive substantially more applications, or have a different culture.

CONCLUSION

We have described a reproducible, defensible, and transparent resident selection process that incorporates several innovative elements. This description will allow its adoption and modification to the context of other Canadian residency training programs. We encourage all Canadian residency programs to develop and disseminate their selection processes to decrease the opacity of the process while allowing best practices and innovations to be disseminated.

REFERENCES

1. Ryan T. Addressing bias and lack of objectivity in the Canadian resident matching process. *CMAJ* 2018;190:E1211–2.

2. Apramian T, Ramazani F, Lee D, et al. Position paper: Support for unmatched Canadian medical students. 2017. Available at: https://www.cfms.org/files/position-papers/agm_2017_support_unmatched.pdf (accessed January 13, 2019).

3. Owens B. If Canada needs more doctors, why hasn't medical school enrolment increased? *CMAJ News* 2018. Available at: https://cmajnews.com/2018/10/03/if-canada-needs-more-doctors-why-hasnt-medical-school-enrolment-increased-cmaj-109-5649/ (accessed January 13, 2019).

4. Vogel L. Record number of unmatched medical graduates. *CMAJ* 2017;189(21):E758–9.

5. AFMC Board of Directors. *Reducing the number of unmatched Canadian medical graduates: A way forward* 2018. Available at: https://afmc.ca/sites/default/files/pdf/AFMC_reportreducingunmatchedcdnmg_EN.pdf (January 15, 2019).

6. Persad A. The unmatched. *CMAJ* 2018;9(2):e89–92.

7. Persad A. The overall culture of residency selection needs fixing. *CMAJ* 2018;190(14):E443.

8. Bandiera G, Wycliffe-Jones K, Busing N, et al. Resident selection in Canada: What do program directors think about best practice recommendations? Poster presented at: Family Medicine Forum; 2015 Nov 12-14; Toronto, ON.

9. McInnes M. Residency matching woes. *CMAJ* 2015;187 (5):357.

10. Bandiera G, Abrahams C, Cipolla A, et al. Best practices in applications & selection; 2013. Available at: https://pg. postmd.utoronto.ca/wp-content/uploads/2016/06/BPAS-DraftFinalReportPGMEACMay2013.pdf (accessed December 6, 2018).

11. Bandiera G, Abrahams C, Ruetalo M, et al. Identifying and promoting best practices in residency application and selection in a complex academic health network. *Acad Med* 2015;90(12):1594–601.

12. Ben-Nun P. Respondent fatigue. In: Lavrakas PJ, editor. *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage Publications, Inc.; 2011. p. 743. Available at: http://meth-ods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n480.xml (accessed February 25, 2019).

13. DeSantis M, Marco CA. Emergency medicine residency selection: factors influencing candidate decisions. *Acad Emerg Med* 2005;12(6):559–61.

14. Love JN, Howell JM, Hegarty CB, et al. Factors that influence medical student selection of an emergency medicine residency program: implications for training programs. *Acad Emerg Med* 2012;19(4):455–60.

15. Thurman RJ, Katz E, Carter W, et al. Emergency medicine residency applicant perceptions of unethical recruiting practices and illegal questioning in the match. *Acad Emerg Med* 2009;16(6):550–7.

16. Hayden SR, Hayden M, Gamst A. What characteristics of applicants to emergency medicine residency programs predict future success as an emergency medicine resident? *Acad Emerg Med* 2005;12(3):206–10.

17. Bhat R, Takenaka K, Levine B, et al. Predictors of a top performer during emergency medicine residency. *J Emerg Med* 2015;49(4):505–12.

18. Borowitz SM, Saulsbury FT, Wilson WG. Information collected during the residency match process does not predict clinical performance. *Arch Pediatr Adolesc Med* 2000;154 (3):256–60.

19. Hopson LR, Burkhardt JC, Stansfield RB, et al. The multiple mini-interview for emergency medicine resident selection. *J Emerg Med* 2014;46(4):537–43.

20. Oyama LC, Kwon M, Fernandez JA, et al. Inaccuracy of the global assessment score in the emergency medicine standard letter of recommendation. *Acad Emerg Med* 2010;17(2 Suppl 2): S38–41.

21. Dore KL, Kreuger S, Ladhani M, et al. The reliability and acceptability of the Multiple Mini-Interview as a selection instrument for postgraduate admissions. *Acad Med* 2010;85 (10 Suppl):S60–3.

22. Boursicot K, Etheridge L, Setna Z, et al. Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach* 2011;33(5):370–83.

23. Gormley G. Summative OSCEs in undergraduate medical education. *Ulster Med J* 2011;80(3):127–32.

24. Pau A, Jeevaratnam K, Chen YS, et al. The Multiple Mini-Interview (MMI) for student selection in health professions training - a systematic review. *Med Teach* 2013;35(12):1027–41.

25. Bandiera G, Regehr G. Reliability of a structured interview scoring instrument for a Canadian postgraduate emergency medicine training program. *Acad Emerg Med* 2004;11 (1):27–32.

26. Eneh AA, Jagan L, Baxter S. Relative importance of the components of the Canadian Residency Matching Service application. *Can J Ophthalmol* 2014;49(5):407–13.

27. Gorouhi F, Alikhan A, Rezaei A, Fazel N. Dermatology residency selection criteria with an emphasis on program characteristics: A national program director survey. *Dermatol Res Pract* 2014;2014:692760.

28. Kenny S, McInnes M, Singh V. Associations between residency selection strategies and doctor performance: a meta-analysis. *Med Educ* 2013;47(8):790–800.

29. McCann SD, Nomura JT, Terzian WT, Breyer MJ, Davis BJ. Importance of the emergency medicine application components: the medical student perception. *J Emerg Med* 2016;50(3):466–70.e.1.

30. Cohen JS, Patten S. Well-being in residency training: a survey examining resident physician satisfaction both within and outside of residency training and mental health in Alberta. *BMC Med Educ* 2005;5(21):21.

31. Raj KS. Well-being in residency: A systematic review. *J Grad Med Educ* 2016;8(5):674–84.

32. Feldman M, Lazzara EH, Vanderbilt AA, DiazGranados D. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof* 2012;32(4):279–86.

33. Woehr DJ, Huffcutt AI. Rater training for performance appraisal: A quantitative review. *J Occup Organ Psychol* 1994;67(3):189–205.

34. Kreiter CD. A proposal for evaluating the validity of holistic-based admission processes. *Teach Learn Med* 2013;25(1):s–7.

35. Witzburg RA, Sondheimer HM. Holistic review—shaping the medical profession one applicant at a time. *N Engl J Med* 2013;368(17):1565–7.