An Analysis of Obstetrics-Gynecology Residency Interview Methods in a Single Institution

Benjamin Beran, MD, Kristina Kaljo, PhD, Raj Narayan, MD, and Paul Lemen, MD

Department of Obstetrics & Gynecology, Medical College of Wisconsin, Milwaukee, Wisconsin

OBJECTIVE: Interviews for residency are important for resident selection, yet how to best perform the interviews remains uncertain. Multiple approaches have been described with a variety of results. Our purpose of this study was to evaluate our Obstetrics-Gynecology residency program's interview structure to determine which interview components have association with performance evaluations completed by physicians and nurses during the residents' years of training at our program.

DESIGN: Using only data from our matched residents, existing interview scores from our standard interview process were compared to performance evaluation scores completed by faculty and nursing throughout the training years. Our standard interview process consisted of 4 interviews with individual faculty, 1 structured behavioral-based interview by a blinded faculty member, and 1 interview with a pair of current residents.

SETTING: A single, multisite, Obstetrics-Gynecology residency program.

PARTICIPANTS: Interviewees of the Obstetrics-Gynecology residency program that ultimately matched with our program.

RESULTS: Interview scores from 44 residents were compared to their performance evaluation scores. Positive associations were seen between performance evaluation scores and both resident teams' interviews and unstructured faculty interviews, with the resident teams' score showing a stronger degree of association. The behavioral-based interview total score did not have association with the performance evaluation scores, nor did any of the individual questions.

CONCLUSIONS: Resident teams' interview scores of applicants show the strongest association with eventual performance evaluations completed by faculty and nursing during residency. This demonstrates that current residents should have a role in the resident selection process during interviews. This does not provide data to abandon behavior-based interview techniques, but rather encourage each program to carefully apply their use during the resident selection process. (J Surg Ed 76:414–419. © 2018 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: medical residency, interview, behavioral interviewing, Obstetrics-Gynecology, Graduate Medical Education

ABBREVIATION: SE, standard error

COMPETENCIES: Interpersonal and Communication Skills

INTRODUCTION

The resident interview process employs a variety of methods in an attempt to match applicants who will perform best in the program.[1,2] During numerous and often brief encounters, applicants engage with department chairpersons, residency directors, academic faculty, clinical faculty, as well as other residents. Typically, these interview sessions are structured with mandatory or

Funding: Supported, in part, by grant 1UL1RR031973 from the Clinical and Translational Science Award program of the National Center for Research Resources, National Institutes of Health (NIH). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.
Authors' contributions: BB, RN, and PL made substantial contributions each to the study conception and design, acquisition of data, and analysis and interpretation of data. KK made substantial contributions to the analysis and interpretation of the data. BB drafted the manuscript, while KK, RN, and PL provided additional critical revisions. All authors approved the final version submitted here for publication and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The authors declare no conflict of interest.
The contents of this research, but with a different focus, were presented as an oral presentation at the 2015 CREOG & APGO Annual Meeting, San Antonio, Texas. It was also then included as a published abstract (Beran B, Lemen P, Narayan R, Eastwood D. Association of behavioral interviewing with resident performance. *Obstet Gynecol*. 2015;126:38S).
Correspondence: Inquiries to Benjamin Beran, MD, Department of Obstetrics & Gynecology, Medical College of Wisconsin, 9200 West Wisconsin Avenue, Milwaukee, WI 53226.; e-mail: bberan@mcw.edu

  1931-7204/$30.00
https://doi.org/10.1016/j.jsurg.2018.09.001

recommended questions, or unstructured, allowing the interviewer to ask any appropriate question. In an attempt to observe desired resident attributes, some programs even incorporate simulation or role playing during their interview day; however, applicants have reported mixed results regarding acceptability.[3,4] In testament to the lack of a perfect interviewing technique, residency program directors have reported regretting nearly 1 of every 12 resident selection decisions.[5]

Within structured interview questions, some programs elect to use behavioral-based questions, as traditional application factors such as United States Medical Licensing Examination Step 1 score, Alpha Omega Alpha Honor Medical Society election, medical school grades, letters of recommendation, rank of the medical school, and extracurricular activities may not be predictive of resident clinical performance.[6] Behavioral-based questions focus on noncognitive factors such as conscientiousness, recognizing limits, confidence, professionalism, and team communication. More specifically, behavioral-based questions seek evidence of past experiences to predict future operation. Use of these questions has shown positive correlation to future clinical performance,[7] or they may simply identify applicants with more past experiences based on age or prior employment.[8] Utilization of blinded interviews in which the interviewer is given minimal data regarding the interviewee, coupled with structured questions, is considered best practice outside of medical education,[9] but residency programs have not shown widespread adoption.[10] As such, it is advantageous to determine which principles are applicable within medical education.

The purpose of this study was to evaluate our Obstetrics-Gynecology residency program's interview structure to determine which interview components have association with performance evaluations completed by physicians and nurses during the residents' years of training at our program.

METHODS

The Medical College of Wisconsin Institutional Review Board approved this retrospective study and waived the requirement to obtain informed consent. The Medical College of Wisconsin Affiliated Hospitals Obstetrics-Gynecology residency program is a large multisite program in Milwaukee, Wisconsin. The program accepts 8 residents per postgraduate year, though 7 were accepted yearly prior to 2011.

Residency Interview Process

Since 2006, the resident interview day begins with a general presentation by the residency director regarding details of the program. This is followed by a series of interviews for each applicant. Each applicant interviews with the department chair (new chair appointed in 2009), residency director, each of the 3 residency associate directors (1 for each clinical site), and a team of 2 residents (members rotated each interview date). Beyond the department chair change in 2009, and the rotating team of residents at each session, the interviewers were the same each year. This amounts to 6 interviews during the interview day. Five interviews are unstructured, meaning the interviewers could ask any appropriate question and focus on any domain they felt necessary. Each of these interviewers gave a score from 0 to 500 for each applicant after the interview. For purposes of this study, the unstructured faculty interview scores were summed and averaged, while the resident team score was left as an individual component in the data analysis.

The 1 structured interview was given by the associate director from the main academic site, and consisted of 5 behavioral-based questions focusing on desirability as a future coworker, integrity, commitment to one's own education, and recognition of one's limits (Table 1). This interviewer was blinded to the applicant's file and scored each response for 2 components: clarity of thought and verbalized answer. Each component was scored as very positive (50 points), positive (40 points), equivocal (30 points), negative (20 points), and very negative (10 points), for a final score of 100 to 500.

For rank list construction, the application score and interview scores were summed, and the applicants placed in descending order. Final ranking was determined after discussion of the list by all interviewers. Through the standard methodology, the Match Process determined the final matriculants annually.

Matched Resident Evaluation Process

After each rotation in residency, individual residents were evaluated on a departmental 9-point scale (9 = highest). Faculty and nurses evaluated residents in 8 categories: patient care and clinical skills, patient care and surgical skills, medical knowledge, interpersonal and communication skills, professionalism, practice-based learning and improvement, systems-based practice, and teaching skills. An overall impression score of the resident's competence and capabilities was also given. Only the overall score was used for data analysis.

Statistical Analysis

Behavioral interview scores, averaged nonstructured faculty interview scores, and resident team interview scores were compared to the averaged overall evaluation scores from faculty, and again from nursing staff. The data were summarized and analyzed by the Medical College of Wisconsin Biostatistics Consulting Service, and the analysts were given only deidentified data compiled by a research assistant. Repeated measures analysis using mixed effect models with

TABLE 1. List of Behavioral Interviewing Questions Administered by a Consistent, Blinded Interviewer During the Study Period. Following Each Question Are Example Responses That Would Earn Top Marks Versus Low Marks

Desirability as a future coworker
1. You are an intern in our program. You introduce yourself to a patient who is waiting in the preoperative area. She looks at you and says, "I go by first impressions and I don't want you participating in my surgery." What do you do?
2. Describe a situation you have encountered with a difficult team member and how you handled it.
Top marks: The candidate looks at the situation from the other person's point of view first and tries to allay any misconceptions and find an agreement.
Low marks: The candidate refuses to respect the other's point of view or examine his or her personal beliefs about the situation.
Integrity
3. Tell us about a time when you made a major mistake. How did you handle it?
Top marks: The candidate describes a situation, why it happened, and what she or he learned from it.
Low marks: The candidate describes the situation and places blame on someone else.
Commitment to own education
4. For what kind of supervisor do you work best? Provide an example.
Top marks: The candidate prefers a supervisor that gives clear expectations, allowing appropriate autonomy and oversight balance.
Low marks: The candidate prefers complete autonomy without criticism of mistakes.
Recognition of own limitations
5. What is the most difficult experience you have had in medical school?
Top marks: Observing an undesired outcome of a patient, such as the death of a child.
Low marks: Being disappointed in evaluations, such as complaining about not getting a grade they felt they deserved.

a random effect for subject (resident) was used to evaluate and compare the associations of the 2 interview methods with resident performance evaluation scores. Results of the association analyses were multiplied by 100 to give the "slope100," which is the expected increase in evaluation scores for every additional 100 points on interview scores (rather than "slope," which would be increase per 1 point on interview scores). The limited number of residents available for this study limited the statistical power available for some of the evaluation scores. A 2-sided Fisher's z-test of the null hypothesis that the Pearson correlation coefficient $r = 0.0$ and type I error = 5% gave 80% power to detect an r of 0.42 when the sample size is 44. Repeated measures analysis allowed up to 4 evaluations per resident (1 per year of completed residency) to be considered, or about 98 resident-years, potentially allowing weaker associations to be detected. The data were analyzed using SAS version 9.3 (The SAS Institute, Cary, NC).

RESULTS

Data were submitted for all 44 residents (100%) who participated in the above interview process from 2006 to 2011, matriculated to our residency program, and have at least 1 year of performance evaluations. Distribution of interview scores is shown in Table 2. A total of 251 performance evaluation scores were submitted (129 from faculty and 122 from nursing).

Behavioral-based questions did not have association with faculty or nursing evaluations. This lack of association was seen even after normalizing interviewing scores for each interview year. Subanalysis of each individual behavioral-based interview question did not show significant association with faculty or nursing evaluations. Unstructured faculty interviews had association with faculty evaluations, but not nursing evaluations. The resident team interview scores had association with both faculty and nursing evaluations (Fig. 1).

As unstructured faculty interview and resident team interviews had significant associations with evaluations, the strength of associations was compared in a multivariate model. In this model, only the resident team scores had significant association with faculty (slope100 $\pm$ standard error [SE] = 0.3588 ± 0.1176, $p < 0.0001$) and nursing evaluations (slope100 $\pm$ SE = 0.2893 ± 0.1905, $p = 0.0068$), demonstrating a stronger association with performance evaluation scores than the unstructured

TABLE 2. Distribution of Interview Scores by Type of Interview

Type of Interview	Minimum	Maximum	Mean	Standard Deviation
Behavioral based	340	500	436.4	46.64
Unstructured faculty	325	492.5	420	42.74
Resident team	300	500	430.6	55.20

The minimum possible score was 0, while the maximum possible score was 500.

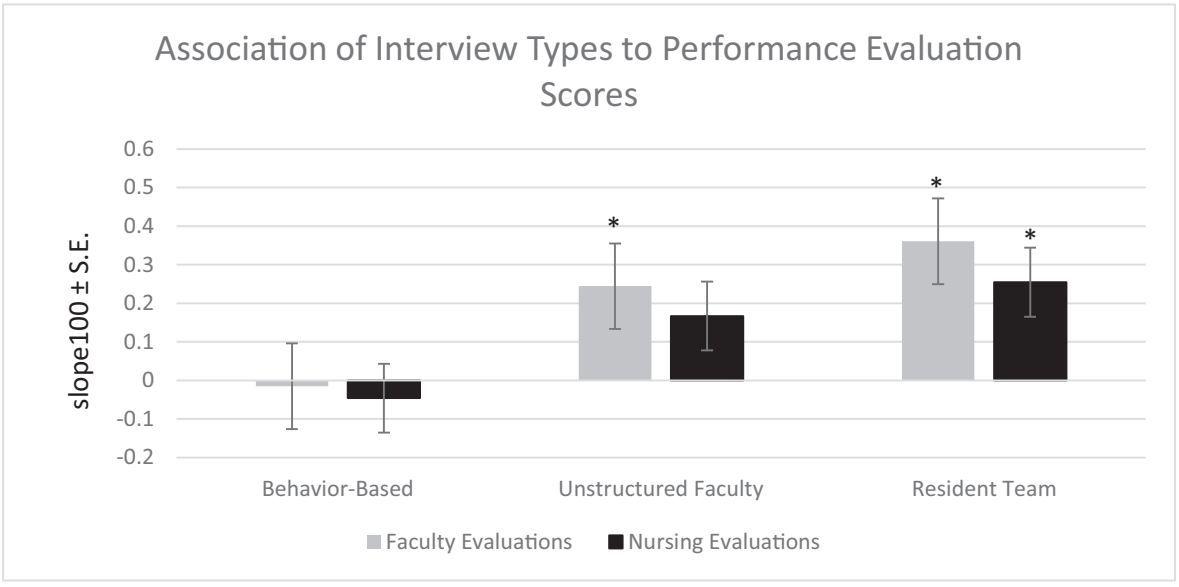

FIGURE 1. Results of association testing between interview scores and performance evaluations completed by faculty and nursing teams (*$p < 0.05$).

faculty interviews (faculty evaluations slope100 ± SE = 0.0874 ± 0.1464, p = 0.3424; nursing evaluations slope100 ± SE = 0.1856 ± 0.2374, p = 0.3043).

DISCUSSION

Our data show mixed results in the association between different resident candidate interview methods and subsequent resident performance evaluations completed by faculty and nursing teams. Much research and consensus focuses on finding applicants that will "fit" the program, as this has been considered as the best predictor for resident success in training and preparation for unsupervised practice.[3,11,12] Notably, our results reflect that the resident team is best able to predict applicants that will succeed in our residency program. This may be due to their presence at all 3 clinical sites of the residency program and understand the characteristics that make a successful resident in each unique environment. The faculty work at 1 site only with no faculty overlapping between sites, which may hinder a full understanding of assessing compatible applicants. Resident team interview scores providing the strongest association with performance evaluations coincides with other published literature showing a structured resident interview predicting compatibility with the program[12] and best practice recommendations to include peer groups of residents during the interview day.[11]

The lack of association between behavioral-based questions and performance evaluations is a complex interpretation. Mostly, positive association has been seen in prior publications,[8,13,14] but may represent publication bias. Importantly, prior studies have looked at correlation between behavioral interview of a given resident at that institution and the performance of that resident at the program into which they subsequently matriculated.[14] We postulate that correlations found in these analyses may not actually reflect how that resident would have performed at the institution where the behavioral interview took place. For instance, the applicant may have been a better (or worse) "fit" at the program in which the matched and so performed better (or worse) than the behavioral interview would have predicted for that institution. Additionally, the respondent pool of program directors represents an extremely diverse group of individuals and programs. In contrast, our study has looked at a consistent group of interviewers and evaluators gauging performance in a single program. This provides a consistency of perspective that is lacking in prior studies.

Strengths of our study include 6 years of interview data using consistent interviewers, apart from a new department chair in 2009. Faculty conducting these interviews are from the same pool of faculty that provide resident performance evaluations. By using multiple sources of performance evaluations (faculty and nursing), a more comprehensive evaluation of resident performance is achieved.

While the magnitudes of association seen are small, they must be interpreted in the context that interview and evaluation scores tend to have small ranges. A slope100 = 0.3606 (resident interview score to faculty evaluation scores) can be stated more clearly as "for every additional 100 points on the interview score, a resident's performance evaluation improved by 0.36 points." Very roughly, this implies about one-third of evaluation scores are 1 point higher (on a 9-point scale) for a 100-point increase on the interview

score. No consensus exists regarding what defines a significant increase in performance evaluation scores. A major difficulty in identifying valid interviewing techniques is the lack of a "gold standard" for resident performance evaluation, which is also a limitation of this study. We used performance evaluations in place during the years of the study, but recognize the limited applicability of their results. Other limitations include the small number of residents in the study based on number of matriculated residents, but all residents during the study period were included, and our data only reflect our single program's experience with the complicated task of resident selection.

CONCLUSION

Resident teams' interview scores of applicants show the strongest association with eventual performance evaluations completed by faculty and nursing during residency. Unstructured interviews conducted by faculty showed association with faculty performance evaluations, but not nursing evaluation. Behavioral-based interview questions did not have association with resident performance evaluations. This demonstrates that current residents should have a role in the resident selection process during interviews. We do not argue that behavioral-based questions should be abandoned, but rather encourage each program using them to carefully apply their use during the resident selection process.

ACKNOWLEDGMENTS

The authors recognize Daniel Eastwood, MS, for his contributions to the statistical analysis of this project.

REFERENCES

1. 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;7:539–548.
2. Kamangar F, Davari P, Azari R, et al. The residency interview is still paramount: results of a retrospective cohort study on concordance of dermatology residency applicant evaluators and influence of the applicant interview. *Dermatol Online J*. 2017;23(5):22.
3. Ogunyemi D, Alexander C, Tangchitnob E, Kim DS. Mini surgical simulation, role play, and group and behavioral interviews in resident selection. *J Grad Med Educ*. 2016;8:410–416.
4. Cannon K, Hartsell Z, Ivanov I, et al. Perceptions of internal medicine residency program candidates on the use of simulation in the selection process. *J Grad Med Educ*. 2014;6:338–340.
5. VanOrder T, Robbins W, Zemper E. Residency program directors' interview methods and satisfaction with resident selection across multiple specialties. *J Am Osteopath Assoc*. 2017;117:226–232.
6. Chole RA, Ogden MA. Predictors of future success in otolaryngology residency applicants. *Arch Otolaryngol Head Neck Surg*. 2012;138:707–712.
7. 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;18:310–315.
8. Breitkopf DM, Vaughan LE, Hopkins MR. Correlation of behavioral interviewing performance with obstetrics and gynecology residency applicant characteristics? *J Surg Educ*. 2016;73:954–958.
9. Huffcutt A. From science to practice: seven principles for conducting employment interviews. *Appl HRM Res*. 2010;12:121–136.
10. Kim RH, Gilbert T, Suh S, Miller JK, Eggerstedt JM. General surgery residency interviews: are we following best practices? *Am J Surg*. 2016;211. 476-481.e3.
11. Black CC. Examination of the residency interview process for academic pathology departments: how to make the most of a resource-heavy process. *Acad Pathol*. 2016;3:1–6.
12. Min AA, Leetch A, Nuno T, Fiorello AB. How well will you FIT? Use of a modified MMI to assess applicants' compatibility with an emergency medicine residency program. *Med Educ Online*. 2016;21:29587.
13. Altmaier EM, Smith WL, O'Halloran CM, Franken EA Jr. The predictive utility of behavior-based interviewing compared with traditional interviewing in the selection of radiology residents. *Invest Radiol*. 1992;27:385–389.
14. Strand EA, Moore E, Laube DW. Can a structured, behavior-based interview predict future resident success? *Am J Obstet Gynecol*. 2011;204. 446.e1-446.e13.

SUPPLEMENTARY INFORMATION

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jsurg.2018.09.001.