Resident Selection Process and Prediction of Clinical Performance in an Obstetrics and Gynecology Program

Alexander Olawaiye
John Yeh
Matthew Withiam-Leitch

Department of Gynecology and Obstetrics
University at Buffalo School of Medicine and Biomedical Sciences
Buffalo, New York, USA

Background: To date, there is no published data showing a correlation between the National Residency Matching Program Rank list position and performance in residency.

Purpose: The background was to assess whether a residency applicant selection process is able to predict the subsequent performance of candidates during residency.

Methods: Over a period of 3 academic years, resident candidates were invited to a structured interview process. This interview process was used to generate a rank list for the National Residency Matching Program (NRMP). We evaluated the clinical performance of the residents that matched to our program at the end of the 1st postgraduate year. We then examined the correlation between the NRMP rank position and the residents’ clinical performance score.

Results: For the 3 academic years, the residents mean total performance score was similar for the residents for each of the years. There was a positive correlation between the resident candidate NRMP rank list percentile and the subsequent 1st year clinical performance evaluation score (r = .60, p < .001).

Conclusions: A structured selection process of residency applicants can predict their subsequent 1st year clinical performance.

Teaching and Learning in Medicine, 18(4), 310–315 Copyright © 2006 by Lawrence Erlbaum Associates, Inc.
M. Withiam-Leitch initially screened all applicants (> 200/year) received over the electronic residency application system (ERAS). These ERAS applications use a standardized format that included board scores, dean’s letter, medical student transcript, personal statements, letters of recommendation, and curriculum vitae. From these ERAS applications, we then invited the screened candidates to interview.

During each academic year, there were a total of three interview dates. There were eight evaluators for each interview date. We made a conscious effort to have interviewers that represented the full spectrum of career opportunities available in obstetrics and gynecology. Thus, on each interview day, we included private practice obstetrician-gynecologists, academic generalist obstetrician-gynecologists, maternal fetal medicine specialists, gynecologic oncologists, and reproductive endocrinologists. All invited candidates for all three dates were interviewed by the program director and the two associate program directors. Each candidate was interviewed by a total of six evaluators.

Prior to the beginning of each interview day, there was a meeting of the evaluators to discuss the evaluation process and to maintain uniformity in the process. Evaluators were asked to provide a score of 1 to 5 points for each of five categories (communication skills, insight into the specialty, motivation, compassion, and “fit” into the program). Each evaluator then provided a total composite score of 5 to 25 for each candidate. We then instructed the evaluators to use this composite score to rank the candidates they interviewed from best to worst. Thus, the composite score was used to generate each individual interviewer’s own rank list. Because each candidate was interviewed by six attending physicians, each candidate received six rank scores. From the rank scores, we generated an average rank score for each candidate. Finally, using the average rank scores, we ranked all the candidates from best to worst for that particular interview date.

At the end of the interview season, a meeting involving all evaluators was held. We generated a final rank list integrating all three interview dates. We facilitated the integration of the three rank lists from the three interview dates by two mechanisms: (a) the program director and the two associate program directors had interviewed all candidates on all interview dates, and (b) several other evaluators had participated on more than one interview date. We subsequently submitted this final rank list to the NRMP.

In each academic year, there were different numbers of candidates included on the final NRMP rank list. Therefore, for this study, we generated a “rank percentile” per candidate per year to depict how high the candidate was ranked in relation to the total number of candidates in that year. This rank percentile controlled for the differences in the number of candidates interviewed per academic year.

Resident Clinical Evaluation Process and Total Performance Scores

At the end of the first postgraduate year, the recruited residents were evaluated by attending physicians. Each resident’s performance was evaluated using a Global Resident Evaluation Form that we developed. We developed the Global Resident Evaluat-
tion Form to evaluate residents according to the standards recently implemented by the Accreditation Council on Graduate Medical Education (ACGME). The ACGME requires that residents be exposed to and trained in six core competencies—patient care, medical knowledge, practice-based learning and improvement, interpersonal communication skills, professionalism, and system-based learning. Full-time and volunteer attending faculty scored residents' performance on a scale ranging from 1 (poor) to 9 (exceptional) in each of the six ACGME core competency areas. For each resident, we then calculated a total performance score by taking an average of the scores from these six core competencies. We used this total performance score (potential range of 1.0–9.0) in the analysis performed in this study.

Statistics

We presented the data as mean ± standard deviation. We did data analysis with a simple linear regression model, analysis of variance tests, t tests, and a variance ratio test using SAS Version 9 (SAS Institute, Inc., Cary, NC, USA). We considered p < .05 statistically significant.

Results

During the study period, we interviewed 107 candidates—29 applicants for the academic year 2002 to 2003, 38 for 2003 to 2004, and 40 for 2004 to 2005. The NRMP rank list position and total performance score for each resident for each of the three academic years is shown in Table 1. The average total performance score did not differ significantly among the three academic years included in the study (p = .78). We therefore pooled data from these three academic years for our subsequent analysis.

Figure 2 shows the relation between the rank positions of all residents and their total performance score as evaluated by attending faculty at the end of the 1st year in residency. There was a correlation between the rank percentile of the candidates and their total performance score. The higher the rank percentile on our NRMP rank list, the better the subsequent clinical performance of the resident. The correlation coefficient was statistically significant (r = .60; p < .001).

We further subdivided the residents into two groups: those residents that were ranked in the top half (rank percentile 50% to 99%) and those ranked in the bottom half (0% to 49%). For the top ranked residents, 12 of 12 had a total performance score above 6.0 (Figure 3). On the other hand, only 1 of 14 residents in the bottom half had a total performance score above 6.0. The total performance score of the upper half was significantly higher than the lower half (6.4 ± 0.21 vs. 5.3 ± 0.98; p < .001). Additionally, use of a variance ratio test showed that there was significantly less variance in the total performance score for residents ranked in the top half as compared to residents ranked in the bottom half (Figure 3).

Conclusions

Our study indicates that the applicant selection process can predict subsequent performance of candidates in residency. The NRMP rank list generated from our selection process significantly correlated with the strength of the candidates’ clinical performance in residency as determined by attending physician evaluations. To our knowledge, this study is the first report that shows a correlation between a program’s NRMP rank list and subsequent residents’ clinical performance.

Table 1. NRMP Rank List Position and Total Performance Score for Each of 3 Academic Years

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank %&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Total PS</td>
<td>Rank %&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>1</td>
<td>93.1</td>
<td>6.4</td>
<td>94.7</td>
</tr>
<tr>
<td>2</td>
<td>82.8</td>
<td>6.0</td>
<td>81.6</td>
</tr>
<tr>
<td>3</td>
<td>79.3</td>
<td>6.4</td>
<td>78.9</td>
</tr>
<tr>
<td>4</td>
<td>75.9</td>
<td>6.4</td>
<td>76.4</td>
</tr>
<tr>
<td>5</td>
<td>48.3</td>
<td>4.8</td>
<td>31.6</td>
</tr>
<tr>
<td>6</td>
<td>34.5</td>
<td>5.8</td>
<td>28.9</td>
</tr>
<tr>
<td>7</td>
<td>31.1</td>
<td>7.2</td>
<td>21.1</td>
</tr>
<tr>
<td>8</td>
<td>27.6</td>
<td>4.7</td>
<td>18.4</td>
</tr>
<tr>
<td>9&lt;sup&gt;b&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>10.6</td>
</tr>
<tr>
<td>M Total PS (± SD&lt;sup&gt;*&lt;/sup&gt;)</td>
<td>6.0 (0.85)</td>
<td>5.9 (0.69)</td>
<td>5.7 (1.21)</td>
</tr>
</tbody>
</table>

Note: NRMP = National Residency Matching Program. There was no statistically significant difference in the mean total performance score (PS) between the 3 academic years (analysis of variance test; p = .78).

<sup>a</sup>Rank % = rank percentile. <sup>b</sup>Ninth resident transferred out of program for 2002-2003.
Our selection process is heavily weighted toward the interview. The initial screening of candidates from the ERAS system was simply to identify those candidates that presumably had the academic credentials to successfully complete a residency program in obstetrics and gynecology. These traditional credentials—medical school transcripts, board scores, and letters of recommendations—have been referred to as cognitive attributes. These cognitive attributes have been shown to predict future success on standardized exams but have not correlated with success in clinical performance.

In contrast, the interview has been viewed as a tool to assess noncognitive attributes. Noncognitive attributes such as communication skills, professionalism, ethics, and the like are important attributes of a clinician. In fact, the new ACGME core competencies are intended to assure that residents are exposed and educated in these noncognitive attributes. This is why we had chosen to focus our selection process on the interview. In fact, several studies have shown that residency program directors regard the interview as the most important component of the residency selection process. However, none of these studies have attempted to show a correlation between the interview process and subsequent resident clinical performance.

There are two possible reasons to explain how our interview process predicts future resident performance. First, we controlled for the interinterviewer variability by having the attending physicians rank their candidates from best to worse. The composite score of 5 to 25 points was only used by the individual interviewer to generate their personal rank for the candidates they interviewed. We then combined the ranks from these individual interviewers to calculate an average overall rank. This minimizes the potential...
for an individual interviewer’s composite scoring to skew the overall rank list. In other words, this technique helps to control for bias in tendencies by different interviewers for leniency or severity. Second, interviewers consisted of attending physicians from “all walks of life.” The generalist obstetrician-gynecologist may be most interested in whether a candidate will fit in with others in the program, whereas a subspecialist may focus on the research potential of the candidate. A private practice obstetrician-gynecologist may concentrate an interview assessing whether the candidate has the communications skills to provide good care to their patients. For a candidate to be ranked high on the NRMP rank list, it would have been necessary for multiple evaluators, regardless of their primary interest, to rank the candidate high. In our opinion, it is the breadth of viewpoints of the evaluators that has led to our interview process being able to generate a rank list that predicts the performance of candidates in residency. It should be noted that our approach is in contrast to a trend toward developing a highly structured and standardized interview process to reduce interrater variability.

Other than our study, there is little published evidence that the interview process can predict future clinical performance. For example, Komives et al. were unable to show a significant correlation between the applicant interview and resident performance among internal medicine residents. In Komives et al.’s study, the resident clinical performance evaluations were done by chief residents and may have been flawed by internal inconsistencies. In our study, the clinical evaluations were done by experienced attending physicians, and we used the average scores from all of them to calculate a total performance score. This approach limited potential bias by one particular attending toward a resident. Additionally, the total performance score we used was calculated using a scale that rated the ACGME core competencies. The ACGME core competencies evaluate many of the same noncognitive attributes as the interview, for instance, communication skills and professionalism. Thus, the qualities assessed in our interview process were similar to the qualities assessed in our total performance score. Bell et al. likewise failed to show a correlation between their selection process and subsequent resident performance in an obstetric and gynecology program. Unlike our study, Bell et al.’s resident selection process was heavily weighted toward cognitive attributes such as U.S. Medical Licensing Examination scores and honors in medical school clerkships.

Given the preceding, our data suggest that our resident interview process is relatively successful in predicting subsequent resident clinical performance. However, there are limitations and weaknesses. First, our ranking process seems to suggest that residents in the top 50th percentile of the NRMP rank list perform relatively well by our assessment standards. However, we are less able to predict clinical performance in the lower 50th percentile. The variance in clinical performance was significantly greater in these candidates. Second, although our selection process was strongly weighted toward the interview and noncognitive attributes, prior knowledge of the candidate’s cognitive attributes may have influenced to some degree the evaluators ranking of a candidate. In the future, comparing the evaluators preinterview and postinterview rank list will help to delineate the potential influence of these cognitive and noncognitive attributes. Third, an attending that evaluated a resident may have also interviewed this resident as a candidate. This “unblinded” performance evaluation could introduce bias. However, as the resident performance evaluations are done 18 months after the interview season, the evaluators would have to have excellent recall for this bias to occur. Finally, this study only addresses the selection process with regard to performance early in residency. We cannot yet address whether the selection process has any predictive value on resident performance at the end of residency training.

In summary, we have described a resident selection process heavily dependent on the interview that leads to an NRMP ranking system that correlated reasonably well with 1st-year clinical performance of residents. It is unlikely for any two programs to have the same criteria for resident selection because each program has different concerns and interests. It is, however, important for each program to formulate a uniform and objective selection process, an essential component of which is an interview process that may predict future performance of the residents.

References


