

Developing a Predictive Model to Assess Applicants to an Internal Medicine Residency

DAVID NEELY, MD
JOSEPH FEINGLASS, PhD
WARREN H. WALLACE, MD

Abstract

Background Residency programs strive to accurately assess applicants' qualifications and predict future performance. However, there is little evidence-based guidance on how to do this. The aim of this study was to design an algorithm for ranking applicants to an internal medicine residency program.

Methods Ratings of overall performance in residency were compared to application characteristics of 230 graduating residents from 2000–2005. We analyzed 5 characteristics of the application: medical school, overall medical school performance, performance in junior medicine clerkship, United States Medical Licensing Examination (USMLE) Step 1 score, and interview ratings. Using bivariate correlations

and multiple regression analysis, we calculated the association of each characteristic with mean performance ratings during residency.

Results In multiple regression analysis, the most significant application factors ($r^2 = 0.22$) were the quality of the medical school and the applicant's overall performance in medical school ($P < .001$).

Conclusion This data has allowed the creation of a weighted algorithm to rank applicants that uses 4 application factors—school quality, overall medical school performance, medicine performance, and USMLE Step 1 score.

Introduction

Residency programs spend time and energy during the selection process in an attempt to predict future performance of applicants. Several studies from a variety of disciplines (radiology, obstetrics-gynecology, anesthesia, psychiatry, and pediatrics)^{1–5} have demonstrated poor correlations between ranking by the Intern Selection Committee (ISC) and performance as a resident.

In contrast, 2 studies have presented evidence of selection criteria that successfully predict future resident performance. A study from the University of Michigan's internal medicine program⁶ demonstrated a significant correlation ($r = 0.48$) between ISC ranking and subsequent resident performance. This study examined 123 residents who completed training from 1989 through 1992. The application characteristic with the strongest correlation to residency performance was junior medicine clerkship performance. Another study from the Department of

Obstetrics and Gynecology at State University of New York–Buffalo⁷ suggested that an intensive interview day was the single best predictor of resident performance. Although this was a small study, the correlation between ISC rank and subsequent resident performance was remarkably high ($r = 0.60$).

Due to the lack of consensus in the medical literature on how to accurately use applicant characteristics to predict residency performance, we collected application and residency performance data in our program for 6 years. The aim of this study was to develop a method to weigh application information and rank future applicants based on our past experience.

Methods

The McGaw Medical Center of Northwestern University's Chicago campus internal medicine residency program had 230 graduates from 2000–2005. The 230 individuals graduated from 66 United States medical schools and 1 international medical school. Of the 230 applicants, 50 (22%) graduated from Northwestern University Feinberg School of Medicine. The average United States Medical Licensing Examination (USMLE) Step 1 score was 228, with a range of 185 to 266. Fifteen percent of applicants were interviewed.

We analyzed the correlation of performance ratings of 230 senior residents to 5 parts of their residency application: quality of medical school, overall medical school

All authors are at Northwestern University Feinberg School of Medicine. **David Neely, MD**, is Director of Undergraduate Education at the Department of Medicine; **Joseph Feinglass, PhD**, is Faculty at the Department of Medicine; and **Warren H. Wallace, MD**, is Associate Program Director at the Department of Medicine.

Corresponding author: David Neely, MD, Director, Northwestern University, The Feinberg School of Medicine, Department of Medicine, 251 E Huron St, Galter 3-150, Chicago, IL 60611, 312.926.7252, dneely@nmh.org

Received August 15, 2009; revisions received October 5 and December 21, 2009, accepted December 23, 2009.

DOI: 10.4300/JGME-D-09-00044.1

performance, performance in the junior medicine clerkship, USMLE Step 1 score, and interview ratings. The study was approved by the Institutional Review Board of Northwestern University.

Resident Performance Ratings

In the fall of each year, 4 chief medical residents and 10 key clinical faculty members rate the overall 3-year performance of the program's senior residents on a scale of 1 to 100, where 1 represents the best resident, 50 is an average resident, and 100 is the program's worst resident. These evaluators do not have access to the residents' medical school performance information or residency application. The performance assessment includes clinical reasoning and knowledge, leadership, professionalism, patient care, and teaching. Evaluators base their ratings on their personal experience and review of the resident's summative evaluations from faculty members, peers, students, and patients.

The performance ratings in a 2-year sample of residency graduates were examined to assess interrater reliability. There were 10 residents for whom we had evaluations from all 14 evaluators. The interclass correlation coefficient for these residents was 0.94. Thirty-nine residents were evaluated by the 6 evaluators who rated the most residents. The interclass correlation coefficient for these 39 residents was 0.93.

Application Characteristics

Each applicant's medical school and USMLE Step 1 score were obtained from residency program records. The 2 ratings that required interpretation—overall medical school performance and performance in the junior medicine clerkship—were reviewed by 2 authors (D.N. and W.H.W.). Data abstractors were not involved in resident performance ratings.

We assessed the quality of clinical education for each applicant by assigning each medical school a rank from 1 to 5, 1 for the strongest schools and 5 for the weakest schools. We used multiple inputs to obtain our ranking, including consensus opinions of ISC members, U.S. News and World Report rankings, and the prior performance of graduates from a particular school. We also adjusted the medical school ranking based on the quality of information provided regarding clinical performance. School rankings were updated yearly based on these inputs.

We assigned a score representing each applicant's overall medical school performance in deciles from 10 (strongest) to 90 (weakest) based on information in the applicant's Medical School Performance Evaluation (MSPE). We chose to rank students' performance by deciles rather than counting honors and high pass grades as in prior published studies.⁴⁻⁶ To maintain consistency, templates were prepared for each school based on the data provided by the MSPE. Therefore, all students from a particular

school were scored in a similar manner. We believe this ranking method is preferred because of the wide discrepancy in percentages of honors and high pass grades awarded in US medical schools.

We assigned a score from 10 to 90 to assess each resident's performance in the junior medicine clerkship. This rating was determined from the Chairman of Medicine's letter and/or junior medicine evaluation in the MSPE for each applicant.

We used the USMLE Step 1 score to assess the correlation between standardized examination scores and residency performance. Because they were not always available at the time of residency application, USMLE Step 2 scores were not used in this model.

For interview scores, we used a scale from 1 (strongest) to 4 (weakest). Interviewers were asked to comment on interpersonal and communication skills, enthusiasm, and potential for leadership. They were not given the transcript and were not expected to assess medical school performance. There were 2 interviewers for each applicant, and we averaged their scores.

Statistical Analysis

We estimated Pearson product moment correlations between mean ratings of third-year-resident performance and each of the 5 application characteristics (medical school rating, overall medical school performance, medicine performance rating, USMLE Step 1 score, and interview score). We separately assessed correlations for the 50 residents who attended Northwestern for medical school. We used multiple linear regression to analyze resident performance ratings by entering all 5 applicant characteristics simultaneously.

Results

The mean performance rating of the 230 residents on the 1 to 100 scale was 45, with a standard deviation of 22 and a range from 5 to 99. TABLE 1 displays bivariate linear correlations between third-year-resident performance ranking and each application characteristic. For all applicants there were modest but significant ($P < .001$) correlations between third-year-resident performance ratings and overall medical school performance ($r = 0.25$), medicine performance ($r = 0.21$), and USMLE Step 1 scores ($r = -0.26$). Medical school quality was only weakly correlated ($r = 0.14$, $P = .04$), and interview scores were not significantly correlated with residency performance.

For the 50 Northwestern students, the correlations between resident performance ratings and overall medical school performance, medicine performance, and USMLE Step 1 score were all much stronger ($r = 0.40$ to 0.63). Correlations with interview scores were not statistically significant.

TABLE 2 displays regression coefficients for each application characteristic. Application characteristics

TABLE 1 PEARSON CORRELATION COEFFICIENTS BETWEEN APPLICANT CHARACTERISTICS AND MEAN THIRD-YEAR-RESIDENT PERFORMANCE RATINGS (N = 230 INTERNAL MEDICINE RESIDENTS, 2000–2005)

Characteristic	Pearson Correlation Coefficient	
	All Students (N = 230)	Northwestern Students (n = 50)
School Quality	0.14 ^a	N/A
Overall Performance	0.25 ^b	0.57 ^b
Medicine Performance	0.21 ^c	0.63 ^b
USMLE Step 1 Score	−0.26 ^b	−0.40 ^c
Interview Score	0.08	0.24

Abbreviation: USMLE, United States Medical Licensing Examination.

^a $P < .05$.

^b $P < .001$.

^c $P < .01$.

explained about 20% of the variance in mean third-year-resident performance ratings. Except for the USMLE Step 1 score, the characteristics and resident rating all use a scale in which lower numbers are better, so the coefficients are all positive. The scale for USMLE Step 1 is reversed, so its coefficients are negative. Regression results demonstrate that medical school quality and overall medical school performance are the 2 most significant factors ($P < .001$). Every 1-point increment in medical school rank correlates with an 8.6% increase in resident performance rating. An increment of 10% in overall medical school performance correlates with an increase of 2% in resident performance rating. An increment of 10% in medicine performance or a 10-point-lower USMLE Step 1 score both correlate with an additional 2% increase in resident performance rating. Interview scores did not correlate with resident performance.

Discussion

These results demonstrate that our assessment of medical school quality and overall medical school performance, taken together, were significant factors in predicting subsequent resident performance. We believe this is the first study to identify important correlations between ratings of medical school quality and residency performance. Our ability to link school quality to residency performance may result from the large number of schools represented in our study or the relatively large number of residents assessed during the 6-year study period. Our findings also confirm the accuracy of our method of assessing school quality. Other training program selection committees may wish to

TABLE 2 MULTIPLE REGRESSION RESULTS FOR APPLICANT CHARACTERISTICS AND RESIDENT PERFORMANCE RATINGS ($R^2 = 0.22$)

	β	SE	P Value
School Quality	8.6	1.7	.000
Overall Performance	0.41	0.11	.000
Medicine Performance	0.19	0.08	.02
USMLE Step 1 Score	−0.26 ^a	0.09	.01
Interview Score	0.92	2.68	.70

Abbreviation: USMLE, United States Medical Licensing Examination.

^a Higher USMLE scores correlate with better (lower) performance ratings, so the sign is negative.

consider the performance ratings of prior residents to adjust their rankings of medical schools.

We found that overall medical school performance was a stronger independent predictor than performance in junior medicine alone. Prior reports⁶ have suggested that performance in junior medicine correlated better with residency performance than overall medical school performance. However, junior medicine performance is just one assessment, and the overall clerkship performance detailed in the MSPE represents multiple assessments. Therefore, it is reasonable that 6 or 7 clerkship grades would be a more accurate assessment of clinical skills than 1 clerkship grade. In response to input from the Liaison Committee on Medical Education, many schools' MSPEs have improved significantly in the past few years. Our data suggest that MSPEs may now be more accurate measures of future clinical performance than departmental letters or junior medicine performance alone.

Consistent with other published studies,^{5–8} performance on USMLE Step 1 was a relatively minor factor in predicting resident performance. Several studies^{9–11} demonstrate that a student's score on USMLE Step 1 primarily predicts future licensing and board certification examination scores, so it is not surprising that we did not find a correlation between Step 1 score and residency performance. Similarly, interview scores did not correlate with resident performance. Perhaps this is due to the large number of faculty interviewers, or the fact that the insight gained from the 4-year medical school record overwhelms the insight gained from a 30- to 60-minute interview. Perhaps a more rigorous interview process would be more helpful. One small study⁷ has demonstrated the benefits of an intensive interview process in choosing residents. Further study is needed to determine the reproducibility of these findings.

As in the study from the University of Michigan,⁴ we found a strong local-school effect. As shown in TABLE 1, in all 230 residents, overall medical school performance and medicine clerkship performance correlated weakly

($r = 0.21$ to 0.25) with residency performance. However, when we examined the 50 Northwestern graduates in our program, overall medical school performance and medicine clerkship performance correlated much more closely ($r = 0.57$ to 0.63) with residency performance. There may be several reasons for this powerful local-school effect. The student's application and MSPE give an incomplete portrayal of a student's potential and include very little information on professionalism and other important attributes. Also, members of the ISC may know their own graduates better than students from other schools and therefore can rank them more accurately. In addition, Northwestern's student-assessment system may be particularly helpful.

This paper has several limitations. First, it represents assessments of resident performance at 1 program in a single specialty. In addition, our program only looks at a small range of the entire population of US medical students. The reproducibility of our findings in other settings and programs is unknown. Second, we used subjective, global assessments in conjunction with summative evaluations to assess resident performance. Although our interrater reliability was high, there is no gold standard for clinical assessment, and the best method of assessing clinical performance remains controversial. Lastly, $r^2 = 0.22$ for our regression analysis shows that much of the variance in mean performance ratings is unexplained. This may be due to limited information in residency applications in such critical areas as leadership skills, teamwork, and professionalism.

As a result of these data, a member of our ISC now calculates an overall weighted score for each applicant to assist our committee in ranking applicants (Appendix). Although we can adjust applicant rankings based on interviews, research accomplishments, or other life experiences, the two most important factors in our scoring system are medical school quality and overall medical school performance. We believe a comprehensive review of resident performance over a 6-year period has identified several factors that predict performance in our internal medicine residency program. Further work is needed to assess this model in other programs and disciplines.

References

- Adusumilli S, Cohan HR, Marshall KW, et al. How well does applicant rank order predict subsequent performance during radiology residency? *Acad Radiol.* 2000;7:635–640.
- Dawkins K, Ekstrom RD, Maltbie A, Golden, RN. The relationship between psychiatry residency applicant evaluations and subsequent residency performance. *Acad Psychiatry.* 2005;29:69–75.
- Metro DG, Talarico JF, Patel RM, Wetmore AL. The resident application process and its correlation to future performance as a resident. *Anesth Analg.* 2005;100(2):502–505.
- Bell JG, Kanellitsas I, Shaffer L. Selection of obstetrics and gynecology residents on the basis of medical school performance. *Am J Obstet Gynecol.* 2002;186(5):1091–1094.
- 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):256–260.
- Fine PL, Hayward RA. Do the criteria of resident selection committees predict residents' performances? *Acad Med.* 1995;70(9):834–838.
- Olawaiye A, Yeh J, Withiam-Leitch M. Resident selection and prediction of clinical performance in an obstetrics and gynecology program. *Teach Learn Med.* 2006;18(4):310–315.
- Gunderman RB, Jackson VP. Are NBME examination scores useful in selecting radiology residency candidates? *Acad Radiol.* 2000;7(8):603–606.
- McCaskill QE, Kirk JJ, Barata DM, Wludyka PS, Zenni EA, Chiu TT. USMLE Step 1 scores as a significant predictor of future board passage in pediatrics. *Ambul Pediatr.* 2007;7(2):192–195.
- Swanson DB. Validity of NBME Part I and Part II scores in prediction of Part III performance. *Acad Med.* 1991;66(suppl 9):S7–S9.
- Sosenko J, Stekel KW, Soto R, Gelbard M. NBME Examination Part I as a predictor of clinical and ABIM certifying examination performances. *J Gen Intern Med.* 1993;8(2):86–88.

Appendix

The following are formulas used in this study.

In order to guide our ranking meeting, we used the standardized coefficients from our regression analysis to derive a formula:

$$4 \times \text{school rank} + 4 \times \text{overall rank} + 1 \times \text{medicine rank} + 1 \times \text{board score}$$

The method used to calculate a total weighted score was:

$$4 \times \text{school rank} (1-5) + .2 \times \text{overall rank} (10-90) + .05 \times \text{medicine rank} (10-90) + 1 \times \text{board score number} (>250 = 1, >240 = 2, >230 = 3, >220 = 4, >210 = 5, >200 = 6)$$

A student from a 2.5 school with an overall rank of 20%, a medicine grade of 40%, and a board score of 235 would have a score of:

$$10 + 4 + 2 + 3 = 19$$

A student from a 1.5 school with an overall rank of 40%, a medicine grade of 60%, and a board score of 255 would have a score of:

$$6 + 8 + 3 + 1 = 18$$

Most of our applicants' scores are between 10 and 25.