

Can Medical School Performance Predict Residency Performance? Resident Selection and Predictors of Successful Performance in Obstetrics and Gynecology

HINDI E. STOHL, MD
NANCY A. HUEPPCHEN, MD, MSc
JESSICA L. BIENSTOCK, MD, MPH

Abstract

Background During the evaluation process, Residency Admissions Committees typically gather data on objective and subjective measures of a medical student's performance through the Electronic Residency Application Service, including medical school grades, standardized test scores, research achievements, nonacademic accomplishments, letters of recommendation, the dean's letter, and personal statements. Using these data to identify which medical students are likely to become successful residents in an academic residency program in obstetrics and gynecology is difficult and to date, not well studied.

Objective To determine whether objective information in medical students' applications can help predict resident success.

Method We performed a retrospective cohort study of all residents who matched into the Johns Hopkins University residency program in obstetrics and gynecology between 1994 and 2004 and entered the

program through the National Resident Matching Program as a postgraduate year-1 resident. Residents were independently evaluated by faculty and ranked in 4 groups according to perceived level of success. Applications from residents in the highest and lowest group were abstracted. Groups were compared using the Fisher exact test and the Student *t* test.

Results Seventy-five residents met inclusion criteria and 29 residents were ranked in the highest and lowest quartiles (15 in highest, 14 in lowest). Univariate analysis identified no variables as consistent predictors of resident success.

Conclusion In a program designed to train academic obstetrician-gynecologists, objective data from medical students' applications did not correlate with successful resident performance in our obstetrics-gynecology residency program. We need to continue our search for evaluation criteria that can accurately and reliably select the medical students that are best fit for our specialty.

Background and Purpose

Academic residency programs train residents to conduct scientific research, prepare them to be effective clinical

educators, and develop their clinical and surgical skills to become competent physicians. Residency program directors and Residency Admissions Committees spend considerable time and effort identifying medical students who will perform well in each of these areas during residency. During the evaluation process, admissions committees typically gather data on both objective and subjective measures of a medical student's performance through the Electronic Residency Application Service, including medical school grades, standardized test scores, research achievements, nonacademic accomplishments, letters of recommendation, the dean's letter, and personal statements. The ability of the current selection process to predict success has been questioned by some, because it presumes that performance during medical school is a good predictor of performance during residency.

Several studies examining the correlation between medical school and residency performance have demonstrated a positive relationship, while other studies note no consistent correlation for any independent or

All authors are at the Johns Hopkins Medical Institutions, Department of Gynecology and Obstetrics. **Hindi E. Stohl, MD**, is Administrative Chief Resident; **Nancy A. Hueppchen, MD, MSc**, is Assistant Professor and Director of Medical Student Education; and **Jessica L. Bienstock, MD, MPH**, is Associate Professor and Residency Program Director.

Results of this study were presented at the 2009 Council on Resident Education in Obstetrics and Gynecology (CREOG) and Association of Professors of Gynecology and Obstetrics (APGO) Annual Meeting on March 12, 2009, San Diego, California.

No financial support has been provided for this study.

Corresponding author: Hindi E. Stohl, MD, Johns Hopkins Medical Institutions, 600 North Wolfe Street, Phipps No. 279, Baltimore, MD 21287, 410.955.6710, hindistohl@gmail.com

Received December 8, 2009; revision received February 22, 2010; accepted April 22, 2010.

DOI: 10.4300/JGME-D-09-00101.1

pooled variable(s).¹⁻⁶ While medical student performance may be associated with a high ranking on a program's match list and success in obtaining a residency position,^{7,8} it has not consistently been shown to predict resident overall performance in medical or surgical residencies.^{9,10}

The evaluations of the obstetrics and gynecology resident selection process in the literature yield somewhat conflicting results. Bell et al⁶ demonstrated that while an individual's performance on standardized tests as a medical student was predictive of the same individual's performance on the Council on Resident Education in Obstetrics and Gynecology examination as a resident, noncognitive skills, such as surgical dexterity, clinical judgment, patient rapport, and work ethic, could not be predicted from medical school performance or residency applications. In contrast, Olawaiye et al¹¹ concluded that a candidate's clinical performance as a postgraduate year-1 (PGY-1) resident correlated with the candidate's ranking on the program's National Residency Matching Program (NRMP) rank list. Gonnella and Hojat¹² reported a similar finding, with a positive relationship between medical student grades and performance as a PGY-1 resident. Whether these 2 correlations would have persisted throughout residency is unknown, as neither study evaluated resident performance during PGY-2 to PGY-4. Blechman and Gussman¹³ investigated the pertinence of letters of recommendation and whether they included valuable information regarding a medical student's competence in the 6 Accreditation Council for Graduate Medical Education (ACGME) core competencies. This study did not comment on cognitive metrics or on the predictive ability of the letters of recommendation to ascertain ultimate resident success.

It is important to determine which, if any, of the metrics currently used are predictive of a candidate's success as a resident. The goal of this study is to investigate whether the objective data included in medical student applications correlate with candidates' ultimate success as residents in an obstetrics and gynecology residency program.

Methods

Participants

We performed a retrospective cohort study of all residents who matched into the Johns Hopkins University residency program in obstetrics and gynecology between 1994 and 2004 and entered the program through the NRMP as a PGY-1 resident. This study was approved by the Institutional Review Board of Johns Hopkins Medical Institutions.

Evaluation

Resident rankings were based on the evaluator's response to the question: "Knowing what you know now about this resident, would you select him or her again for admission into the residency program?" The residency program director and the department chair independently ranked

each of the residents meeting inclusion criteria by using a 4-point Likert scale (1 = minimal or no desire to "reselect", 2 = some interest, 3 = much interest, 4 = strong interest). The desire to reselect a resident was based on the evaluator's assessment of the resident's overall performance during the 4 years of training, with specific regard to the resident's achievement of the ACGME 6 core competencies. Residents who did not complete the 4-year training program owing to premature elective or mandated termination were automatically assigned a numerical value of 1.

Both of the evaluators had worked closely with every resident in the program during the time period examined. Each evaluator was blinded to the other evaluator's ratings as well as to the resident's original application to the residency program as a medical student. The evaluators were permitted to review each resident's performance folder, which included the semiannual faculty and peer evaluations of the resident during the training period.

Statistical Analysis

A mean score was calculated on the basis of these 2 independent observations, and residents were grouped according to this score. A resident with a score of 1.5 or lower was assigned to the lowest group; a score higher than 1.5 but lower than or equal to 2.5, to the second lowest group; a score between 2.5 and 3.5, to the second highest group; and a score of 3.5 or higher, to the highest group. Once ranked, all resident identities were masked to preserve resident confidentiality.

The residents' applications in the highest and the lowest groups were abstracted. Data collected from these residents' files included (1) age; (2) sex; (3) medical school attended; (4) academic degrees obtained; (5) score on the United States Medical Licensing Examination (USMLE) Step 1; (6) grade during 5 core clinical rotations (internal medicine, pediatrics, obstetrics and gynecology, surgery, and psychiatry); (7) membership in the Alpha Omega Alpha Medical Honor Society; (8) presentation at a national meeting in the medical field; (9) publication in a peer-reviewed medical journal; (10) research experience; (11) self-reported "distinctive talent," such as being a championship athlete or musician; and (12) leadership position(s) in medical school, such as being a student government officer or director of a community service program.

Each of these measures of medical student performance was compared between residents in the highest group (successful residents) and residents in the lowest group (unsuccessful residents). Data were analyzed using the Student *t* test for continuous variables and χ^2 test for categorical variables. *P* < .05 was considered statistically significant.

Results

During the study period examined, our program offered and filled 7 to 8 residency positions per academic year through

TABLE	EVALUATIVE DATA FROM RESIDENT APPLICATIONS		
	Lowest Group, % (n = 15)	Highest Group, % (n = 14)	P Value
USMLE score ≥ 220	33	50	.46
Alpha Omega Alpha membership	27	14	.65
Research experience > 1 year	47	50	1.0
Peer-reviewed publications	47	43	1.0
Other advanced degree	13	14	1.0
Leadership positions	47	64	.46
Distinctive talent	33	14	.39
Basic science honors ≥ 4	47	64	.45
Clinical core honors ≥ 4	33	39	1.0

Abbreviations: USMLE, United States Medical Licensing Examination.

the NRMP “match.” A total of 75 residents met inclusion criteria and application data were available for all study participants. All study participants matriculated into the residency program directly from medical school. Some had pursued other careers, degrees, or interests before or during medical school. Resident ages ranged between 24 and 35 years on entry into the residency program. Sixty-five percent of the residents were women. The proportion of women in the residency program did not change significantly throughout the study period; however, a statistical difference was noted between the lowest and highest groups with regard to sex. Eighty-six percent of the residents in the highest group were women, as compared to 47% in the lowest group. Of the 3 international medical graduates included in our analysis, 2 were in the lowest group and 1 in the highest group. Evaluation data collected from the residency applications are shown in the TABLE.

There was substantial agreement between the 2 evaluators regarding overall resident performance. Approximately 50% of the residents in the quartiles analyzed received the same score from both evaluators. The remainder of the residents were assigned scores that differed by a value of 1 (ie, 1 score of 3 and 1 score of 4, providing a mean of 3.5; or, 1 score of 2 and 1 score of 3, providing a mean of 2.5). No resident received scores from the 2 evaluators that differed by 2 or more. Furthermore, the number of residents assigned to the highest and lowest group did not significantly differ between the 10 years included in the study, suggesting that no one year had a particularly “bad” group or “good” group of residents. Most importantly, none of the examined objective variables from medical student applications were significantly associated with successful performance as a resident (TABLE).

From this larger group, 29 resident profiles were analyzed (14 residents in the highest ranking group and 15

in the lowest ranking group). Of the 15 residents in the lowest ranking group, 8 were assigned this ranking based on their 4-year overall performance and 7, owing to premature termination from the program. Six of the residents who were terminated transferred into programs in other medical specialties (1 into anesthesiology; 2 into emergency medicine; 2 into psychiatry; 1 into pathology), and 1 left the field of medicine. No resident left our program to join another residency training program in obstetrics and gynecology or to pursue training in a different surgical specialty.

Discussion

Evaluating the residency selection process is difficult because there is no uniformly accepted or objective means of measuring a resident’s overall performance or “success” in residency training program. Scores on in-service training examinations provide 1 objective indicator of a resident’s cognitive ability, but global assessments of residents’ performance, specifically in the noncognitive competencies, are often found to be subjective. Faculty ranking of residents is the most commonly used method for assessing “success.” Other indicators, such as fellowship matching, continuation in academic medicine, and passing specialty-specific licensing board examinations, have also been used. Typically, these modalities are not used in isolation but in combination with some form of faculty assessment.^{1,3,4,10,11} In our study, we defined success on the basis of the independent rankings of 2 experienced resident educators. Although more subjective than some of the other indicators, this method allowed us to detect the greatest possible difference within our study cohort and allowed us to broadly define “success.” Residents who are clinically proficient and have achieved competence in each of the 6 ACGME competencies may be highly “successful,” yet they may not elect to pursue subspecialty training or academic

positions. Using solely markers such as fellowship placement or remaining within academic medicine as measures of success may skew the data to favor academic prowess over clinical and professional excellence. Thus, we believe that our methodology allowed us to investigate our primary question and assess whether we can reliably identify which applicants will go on to become successful residency graduates.

In our study, none of the objective measures of medical student performance predicted performance during residency. Sex was the only demographic characteristic to reach statistical significance, with a greater percentage of women in the highest group compared to the lowest group. The significance of this finding is questionable, as 80% of the residents during the study period examined were women.

Our finding that no predictable correlation exists between USMLE scores and performance during residency is in agreement with prior studies.^{5,6,9,14–16} This finding is not surprising because the USMLE only assesses a student's cognitive ability and does not measure the varied noncognitive skills that are required for resident success. Our study was not designed to investigate a relationship between performance on standardized tests as a medical student and performance on standardized tests and/or board certification examinations as a resident; however, previous studies have almost universally demonstrated a positive correlation in both medical and surgical specialties.^{6,17–20} A high score on one measure of cognitive aptitude predicts a high score on a second, similar measure.

While the purpose of our study was to specifically examine the objective components of medical student applications to assess their predictive value, even the few elements of a medical student's performance that had a subjective component (ie, grades in clinical rotations) did not correlate with resident success. This finding was a bit surprising, as we hypothesized that a partly subjective evaluation of a medical student would be associated with a partly subjective evaluation of a resident. Our study is one of a few studies^{1,2} to examine the potential significance of a “distinctive talent” (such as championship athlete or musician) or leadership position(s) during medical school. These metrics are objective yet “softer” measures of a candidate's noncognitive abilities and could be expected to indicate performance success. However, even these attributes were not predictive of a high ranking by the faculty evaluations in our study.

Although these residents may be considered a separate cohort, we included in our analysis the 7 residents who ranked in the lowest quartile because of premature termination, because they make up an important element of a “success failure.” By definition in our study, any resident who did not complete the 4-year residency training program in obstetrics-gynecology at our institution was considered unsuccessful. Their “success failures” highlight the

importance of predicting resident success from the start of residency. A resident who does not complete the training program not only loses personal time (owing to the need to start over in a new specialty) but also creates additional work for the program director in replacing them, uncertainty for their fellow residents, and disruption within the educational flow of the residency program. Predicting which residents would not complete the residency training program would save resources and time as well as enhance learning. Premature termination from our program was suggestive of a poor career choice as a medical student, with individuals finding themselves unsatisfied with a more surgically intensive specialty.

The primary limitations of our study include assessing a small study group at a single institution. It is possible that the type of medical student applying to our residency program differs from applicants to academic programs in other geographic areas and/or programs without university affiliations. Although these limitations may impact the generalizability of our data, they are commonly noted in other similar studies,^{1,6,13} especially in a relatively small specialty such as ours. Unlike programs in internal medicine or general surgery, residency programs in obstetrics and gynecology match an average of 5 residents per year. These numbers make a large study difficult to perform without compromising the currency of the data.

An additional study limitation is that resident success in our study was determined by only 2 individuals (the residency program director and department chair). By nature, a faculty evaluation is subjective and may be dependent upon the clinical setting in which the resident is evaluated.²¹ Including scores from a larger group of faculty members from all departmental divisions would have allowed for a more robust determination of a resident's success, but obtaining a faculty consensus is often quite difficult and thus a mechanism for decreasing interobserver variability is necessary. We attempted to overcome this problem in our study by selecting the program director and department chair to serve as the faculty evaluators. Both individuals had worked closely with each resident in the program in a variety of clinical and nonclinical settings, allowing them to best evaluate a resident's overall performance. Additionally, the 2 evaluators independently ranked the residents to avoid biasing each others' opinions. The 2 evaluations were given completely equal weight in the ultimate determination of a resident ranking. Interestingly, there was remarkable agreement between the evaluations, even with this blinding.

Conclusion

Medical students most likely to be ranked highest on a program's NRMP rank list are those who have a high academic standing in medical school, perform well during interviews, and are perceived by program directors to be well-rounded individuals.²² Our study questions the weight

given to objective measures of a medical student's performance in determining a candidate's ranking, as no objective measure is predictive of resident success. We specifically evaluated the extremes of performance (ie, very weak or very strong) by comparing residents ranked in the highest and lowest groups in an effort to detect the greatest possible effect. While there may be a relationship between academic achievement in medical school and resident performance at the extremes,²³ our results do not support this argument. Therefore, our specialty might be better served by using nonobjective tools as part of the resident selection process to predict successful resident performance. The "best" students do not always make the "best" residents, so we need to continue our search for evaluation criteria that can accurately and reliably select the medical students that are best fit for our specialty.

References

- Daly KA, Levine SC, Adams GL. Predictors for resident success in otolaryngology. *J Am Coll Surg*. 2006;202(4):649–654.
- 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):206–210.
- 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.
- Dirschl DR, Campion ER, Gilliam K. Resident selection and predictors of performance: can we be evidence based? *Clin Orthop Relat Res*. 2006;449:44–49.
- 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.
- 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.
- Brown E, Rosinski EF, Altman DF. Comparing medical school graduates who perform poorly in residency with graduates who perform well. *Acad Med*. 1993;68(10):806–808.
- Calhoun KH, Martinez SA, Stevens MH, Hokanson JA, Bailey BJ. The resident selection process in otolaryngology-head and neck surgery. *Arch Otolaryngol Head Neck Surg*. 1990;116(9):1041–1043.
- Dirschl DR, Dahners LE, Adams GL, Crouch JH, Wilson FC. Correlating selection criteria with subsequent performance as residents. *Clin Orthop Relat Res*. 2002(399):265–271.
- Naylor RA, Reisch JS, Valentine RJ. Factors related to attrition in surgery residents based on application data. *Arch Surg*. 2008;143(7):647–652.
- 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.
- Gonnella JS, Hojat M. Relationship between performance in medical school and postgraduate competence. *J Med Educ*. 1983;58(9):697–685.
- Blechman A, Gussman D. Letters of recommendation: an analysis for evidence of Accreditation Council for Graduate Medical Education core competencies. *J Reprod Med*. 2008;53(10):793–797.
- Gunderman RB, Jackson VP. Are NBME examination scores useful in selecting radiology residency candidates? *Acad Radiol*. 2000;7(8):603–606.
- Berner ES, Brooks CM, Erdmann JB. Use of the USMLE to select residents. *Acad Med*. 1993;68(10):753–759.
- Thordarson D, Ebramzadeh E, Sangiorgio SN, Schnall SR, Patzakis MJ. Resident selection: how we are doing and why? *Clin Orthop Relat Res*. 2007;459:255–259.
- Sosenki 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.
- Spellacy WN. The use of national board scores in selecting residents for obstetrics and gynecology. *Am J Obstet Gynecol*. 1985;153(6):605–607.
- Case SM, Swanson DB. Validity of NBME part I and part II scores for selection of residents in orthopaedic surgery, dermatology, and preventive medicine. *Acad Med*. 1993;68(2 suppl):S51–S56.
- Armstrong A, Alvero R, Nielson P, et al. Do U.S. medical licensure examination step 1 scores correlate with council on resident education in obstetrics and gynecology in-training examination scores and American board of obstetrics and gynecology written examination performance? *Mil Med*. 2007;172(6):640–643.
- Kastner L, Gore E, Novack AH. Pediatric residents' attitudes and cognitive knowledge, and faculty ratings. *J Pediatr*. 1984;104(6):814–818.
- Provan JL, Cuttress L. Preferences of program directors for evaluation of candidates for postgraduate training. *CMAJ*. 1995;153(7):919–923.
- Yindra KJ, Rosenfeld PS, Donnelly MB. Medical school achievements as predictors of residency performance. *J Med Educ*. 1988;63(5):356–363.