Can Admissions Interviews Predict Performance in Residency?

Steven L. Dubovsky, M.D., Michael H. Gendel, M.D.
Amelia N. Dubovsky, M.D., Robert Levin, Ph.D., Joseph Rosse, Ph.D.
Robert House, M.D.

Objective: The authors aim to determine whether admission interviews predict performance in residency.

Methods: The authors determined whether interview and other admission data were correlated with performance during postgraduate years 2–4 and with remaining in the residency in 544 residents enrolled in a single psychiatry residency program between 1963 and 2004.

Results: Considered together, admissions data predicted 13% of the variance in performance ratings in postgraduate year 2 (PGY-2) and 5% in PGY-4. Interview scores were moderately correlated with performance ratings in PGY-2, modestly correlated with performance in PGY-3, and not correlated with performance ratings in PGY-4. Letters of reference were moderately correlated with performance ratings in PGY-2 and modestly correlated with performance in PGY-3 and PGY-4. In PGY-2, interview scores differentiated between the top quartile of performance and the other three quartiles, while letters of reference differentiated performance in the top and bottom quartiles from the middle quartiles. Numerical differences among groups were not great enough to be practically useful, and no variables predicted which residents would leave the program before completing it.

Conclusion: As they are currently conducted, application interviews do not have sufficient power to predict performance during residency. Letters of reference may be useful to the extent that they reflect personal experience with the applicant, but differences in ratings of these letters are not great enough to base admission decisions on them. As it is currently performed, the interview process may be more useful as a means of interesting applicants in the program than of evaluating their potential for success in the residency.

Academic Psychiatry 2008; 32:498–503

Residency is a time- and labor-intensive phase of medical education; admitting individuals who will not perform well, and especially those who will not complete their education successfully, represents significant actual and opportunity costs to the institution, the applicant, and the other applicants who were not admitted. Moreover, because residency represents the last step in most physicians’ formal academic training, residency admission decisions represent the final selection decision before most physicians practice medicine on their own.

Postgraduate programs have traditionally based admission decisions on test scores, grades, clerkship performance evaluations, letters of recommendation, and interviews, despite limited evidence that such data reliably predict clinical performance in medical school or residency (1, 2). Scores on examinations such as the United States Medical Licensing Examination (USMLE) steps II and III have been shown to be significantly correlated with later performance in some reports, but the correlation is primarily with written tests of knowledge rather than actual clinical ability, and the magnitude of the correlation is low (3–5). A study of 121 of 122 U.S. medical schools whose graduates applied to a university surgery residency (6) found so much variability in clerkship grading methods and percentages of students receiving outstanding evaluations in different clerkships that even clerkship grades in the
specialty to which students were applying were not reliable predictors of residency performance.

Even if such data accurately reflected knowledge and skills, many medical educators believe that a candidate with excellent personal attributes but weaker academic credentials may do very well as a resident and practitioner, while a candidate with excellent academic credentials but poor personal attributes may do poorly in residency (7). A prospective study of 600 residents (8) found that noncognitive factors (e.g., interpersonal skills) were best correlated with offers of continued residency training. In one study using standardized personality inventories (7), 95 anesthesiology residents at six training centers completed a battery of psychological tests and questionnaires toward the beginning of the 3-year residency and were rated by supervisors at the end of years 1 and 2. California Psychological Inventory scores on measures of independence, empathy, socialization, responsibility, well-being, and ability to excel in situations requiring compliance with rules and structure correlated modestly with performance ratings in residency (highest r = 0.20–0.26).

Most resident selection committees use interviews to assess personality factors on which acceptance decisions are based (2, 5). Although some studies suggest that admission interview scores are correlated with ratings of relationships with patients and clinical skills (9), other experience suggests that ratings on unstructured admission interviews do not predict residency performance (1, 2, 5). It is possible that the variability in findings reflects variable skill in interviewing. Because psychiatrists use interview techniques in most aspects of their work and the interview is an essential component of evaluation for a psychiatric residency, we thought that assessments of applicants to a psychiatry residency might be more comprehensive and therefore better correlated with later performance than in other specialties. To test this hypothesis, we studied the correlation between assessments of a large number of candidates for a psychiatric residency and performance during residency.

Methods

This study was determined to be exempt by our institutional review board. Data from admission and resident dossiers were obtained on all residents (N = 544) enrolled in a psychiatry residency program between 1963 and 1995, a period that permitted sufficient follow-up of the cohort into practice for further research. Of these cases, 98 were dropped because of missing data on the outcome measures. An additional 280 cases were missing at least some values for the predictor variables. Missing data in archival records are likely to be randomly distributed, so mean values were substituted for missing predictor (but not outcome) variables in most analyses. Because mean substitution produces restriction of range, it is an inherently conservative process and resulted in an effective number of 446. The sample was predominantly male (70%), with an average age of 28.9 years old at the time of admission. The majority of candidates were Caucasian. Each resident was randomly assigned a tracking number by an investigator who had no association with the residency. After all data were collated, the link between resident names and tracking numbers was destroyed, making it impossible to identify any individual resident.

The resident selection process included three to six interviews by senior faculty members, who rated applicants on a numerical scale with the same anchor points used in evaluations of performance during the residency. The available scores were 1 (unacceptable), 2 (marginal, with possible mitigating factors), 3–4 (acceptable), 5 (exceptional, important contributions expected), and 6 (outstanding academic potential). Although the specific wording of anchor points varied somewhat across time, the substance did not. Faculty members did not receive formal training in standardized interview and evaluation formats, but correlations between numerical ratings of individual residents by different evaluators indicated a moderate to high degree of interrater reliability (on a randomly selected group of five residents from each class, r = 0.72). The residency director rated dean’s letters, medical school transcripts, letters of recommendation, and other data such as medical school attended on the same scale of 1 to 6 before the interview data were reviewed. The three residency directors during the study period followed the same protocol. As a result, consistency of evaluation might not have been perfect, but was sufficient given the large number of candidates reviewed by each individual over many years.

Performance during residency was assessed by clinical supervisors every 6 months using a scale from 1 to 6 with the same anchor points as the admission ratings. As was true of interview ratings, numerical performance ratings by each faculty member represented a single global assessment that was supported by narrative comments. Because residents used to enter a program after internship elsewhere before the consolidation of PGY-1 with years 2–4, evaluations were considered only for PGYs-2, 3, and 4. Each year’s rating represented the mean of the two 6-month evaluations.

Our hypotheses were initially tested by simple (zero-