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

Received May 28, 2007; revised August 12, 2007; accepted August 22, 2007. Dr. S. Dubovsky is affiliated with the Department of Psychiatry at the University of Buffalo in New York; Dr. Gendel is affiliated with the Department of Psychiatry at the University of Colorado; Dr. A. Dubovsky is affiliated with the Department of Psychiatry at Harvard University and Massachusetts General Hospital; Drs. Levin and Rosse are affiliated with the Department of Psychology at the University of Colorado; Dr. House is affiliated with the Department of Psychiatry at Denver Health Medical Center. Address correspondence to Steven L. Dubovsky, M.D., University at Buffalo, Psychiatry, 462 Grider St., Rm. 1182, Buffalo, NY 14215; dubovsky@buffalo.edu (e-mail).

Copyright © 2008 Academic Psychiatry

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 psychiatry 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-

order) correlations and then with multiple regression to determine both overall predictive power and the relative contributions of each predictor. Logistic regression was used to determine if the predictors were able to distinguish among high, mid-range, and low levels of performance. We supplemented this by analysis of variance (ANOVA) with post-hoc comparisons to better identify the pattern detected by the logistic regressions.

Results

For the entire group, ratings of performance (mean ± SD) did not change significantly through the residency (PGY-2: 4.49 ± 0.74; PGY-3: 4.48 ± 0.84; PGY-4: 4.88 ± 0.75). Overall, 89% of residents completed the program (n = 484), while 4% were asked to leave the program (n = 22) and 7% left of their own volition (n = 38).

As shown in Table 1, ratings of several factors considered during admission evaluations were moderately correlated with one another. In view of the multi-co-linearity of the variables, ordinary least squares standardized regression was used to determine the relative contribution of each predictor, holding its relationship to the other predictors constant. Table 2 illustrates that using ordinary least squares standardized beta weights, the resulting regression equations were significant for each year of residency, but explained only 13% of the variance in mean performance ratings in PGY-2 and just 5% by PGY-4. Mean interview ratings had a significant but declining association with performance during PGY-2 and PGY-3 and failed to predict performance during PGY-4. Evaluations of applicants' transcripts and deans' letters failed to predict subsequent performance after controlling for the other predictors. Ratings of letters of recommendation continued to have a significant association with performance ratings across PGY-2–4, but the strength of the association decreased by about half from PGY-2 to PGY-3.

To explore the possibility that interviews are better correlated with performances of very successful and very unsuccessful residents than residents in the mid-ranges of performance—which would reduce correlation coefficients but underestimate the practical utility of the admission interview—average PGY-2 performance ratings were categorized into quartiles, and logistic regression was used to assess the ability of the set of predictors to distinguish among high (top quartile), low (bottom quartile), and middle two quartile performers. Interviews and ratings of letters of recommendation were the only factors that significantly differentiated high and low performers. Analysis of

TABLE 1. Relationships among Admission Variables

Variable	Mean	SD	1	2	3	4	5
Medical school quality	4.34	0.87					
Dean's letter	4.35	0.74	0.66*				
Transcript evaluation	4.55	0.69	0.78*	0.83*			
Recommendations	4.73	0.65	0.49*	0.57*	0.55*		
Averaged interview	4.51	0.54	0.12	0.37*	0.20*	0.36*	0.26

*$p<0.05$
1 = unacceptable; 2 = marginal, with possible mitigating factors; 3-4 = acceptable; 5 = exceptional, important contributions expected; 6 = outstanding academic potential

TABLE 2. Relationships among Predictor Variables and Performance Evaluations Across 3 Years of Psychiatry Residency

Variable	PGY-2 Rating n = 160	PGY-3 Rating n = 160	PGY-4 Rating n = 120
Interview	0.20*	0.11*	0.04
Recommendations	0.21*	0.12*	0.14*
Transcript	0.05	0.05	0.11
Dean's letter	0.10	0.01	−0.08
R	0.38*	0.34*	0.23*
Adjusted R^2	0.13	0.10	0.05

PGY = postgraduate year
*$p<0.05$

variance was used to further explore differences between mean interview scores and ratings of letters of recommendation in each of the performance quartiles (Table 3). Post hoc Scheffe tests showed that interview scores distinguished between top performers and other groups but not between mid-range and poor performers. Ratings of letters of reference seemed able to distinguish low and high performers from each other and from the two mid-range quartiles.

A series of ANOVAs (data not shown) revealed no significant relationship between any of the predictors and final status (graduated, dropped out, or asked to leave). Similarly, logistic regression showed that none of the predictors could be used to differentiate between those who completed the residency and those who voluntarily or involuntarily left before completing the residency.

Discussion

In this study of 446 physicians admitted to psychiatric residency programs over a 32-year period, an optimally weighted combination of information typically used for admission decisions explained 13% of the variance in performance ratings during PGY-2, and 5% by PGY-4. Interview scores were somewhat predictive of performance ratings during PGY-2 and PGY-3 but lost their predictive power by PGY-4. In PGY-2, interviews differentiated high performers from those in the middle and low ranges of performance, but the absolute magnitude of the differences in ratings was not great enough to be useful in actual practice. No data obtained during admission evaluations identified residents who left residency for any reason before completing the program, although the number of residents in this category may have been too low to detect a small but significant association.

The relatively low predictive utility of admission interviews in this study is consistent with that generally found in industrial settings for unstructured employee selection interviews (10, 11). However, we had anticipated that the correlation of interviews conducted by psychiatric faculty with later performance evaluations, especially by the same group of faculty, might exceed that of typical interviewers in generic employment settings. Our results disconfirm this hypothesis and may suggest that although an unstructured interview can identify psychopathology and make a psychiatric diagnosis, it has limited ability to predict later assessments of resident performance in the same specialty.

TABLE 3. Interview and Letter of Reference Means by PGY-2 Performance Quartile

Performance	Interview Rating*	Letters of Recommendation*
Top quartile	4.73^{a}	5.03^{a}
Second quartile	4.45^{b}	4.77^{b}
Third quartile	4.40^{b}	4.70^{b}
Bottom quartile	4.29^{b}	4.47^{c}

*Means with different superscripts are significantly different from other means within the same column

Previous work in this area has involved smaller numbers of residents in specialties other than psychiatry studied over shorter periods of time. Over three academic years in an obstetrics-gynecology residency (12), there was a significant correlation between the global ranking during PGY-1 of residents who were accepted and the rank of applicants on the department's match list as determined by a structured interview, but details of performance were not studied and data about performance after PGY-1 were not obtained. Weak correlations between clinical evaluations and admission interviews were found in a study (13) comparing various admission variables with the consensus rating by 10 faculty members of the overall clinical ability of 69 pediatric house officers (agreement rate = 0.60, $p = 0.001$). In a study of 112 PGY-1 residents (14), an attempt was made to correlate internship competencies with global admission interview scores and medical school admission data. However, the only factor that predicted clinical performance during internship was having had more undergraduate humanities courses. Similarly, a prospective study (15) showed no correlation between admission interview scores and mean performance ratings at the end of PGY-1 of an internal medicine residency. In a prospective study (16), no relationship was found between data obtained on evaluation for admission to a surgery residency and final evaluation of the residents' knowledge, technical skill, maturity, judgment, and overall ability.

As was true in our study, these investigations did not use standardized or structured interviews. However, most programs, including most psychiatry residencies, use unstructured interviews to make residency admission decisions. It is possible that structured interviews designed specifically to measure those attributes critical to success in clinical settings would increase the predictive validity of admission interviews. The Accomplishment Interview is a structured interview designed to gather information about "accomplishments" in conscientiousness, empathy, confidence, recognition of limits, and interpersonal skills (1, 3). The Accomplishment Interview was tested in a prospective study of 151 applicants to a radiology residency, who were

evaluated 4 years later by their residency directors (1). There was no correlation between overall Interview scores and unstructured faculty interview scores, but the Interview domain of recognition of limits was a significant predictor of later performance.

As they are currently performed, admission interviews do not appear to be highly predictive of residency performance, and if any correlation with performance does exist, it declines substantially with increased time since the admission interview. One reason may be that interviews which do not assess abilities the faculty would like a candidate to acquire are not likely to predict later mastery of those abilities. Traits that are valued in most interviews, such as energy, confidence, charisma, verbal ability, and compatibility with the interviewer's style, which would also be considered desirable in the chief executive officer of a large business or the leader of a small country (17), do not necessarily lead to success as a resident or competence as a practitioner.

We did not have data that would allow us to be certain about whether interviews successfully screened out residents who would not do well, but many applicants who did not match at our residency and some who were dropped from the program did well elsewhere (and some returned to the area). It may be that applicants who are rejected because interviews suggest a poor general "fit" with one program may have a style that is more compatible with another program. Additionally, residencies with more highly accomplished applicants may be more likely to use interviews as a final cut in resident selection. As performance assessments become increasingly detailed and sophisticated (18), their lack of correlation with admission interview assessments is likely to increase unless admission interviews evaluate the same domains. This change would require an interview protocol that measures traits of successful practitioners, which may not be the same as traits of successful students or interviewees.

Letters of recommendation are pervasive in both the academic admission and personnel selection domains, but research has provided very limited support for their predictive utility (19) and most residency directors do not find enough variance in letters of reference to find them helpful (20). The modest sustained correlation between letters of reference and performance evaluations throughout residency in this study raises the possibility that the potential usefulness of letters of reference may arise not so much from their overt content as the ways that a skilled residency director interprets them (21). To explore this possibility, we conducted a protocol analysis interview (22) with one of the residency directors, who rated about 25% of all letters of reference during the period studied. He had observed that high-performing candidates tended to have letters from referees with whom they had worked directly, who praised specific aspects of their performance, and who would want to have them in their own residency programs. He gave high ratings to letters that followed this pattern, while others, no matter how illustrious the writer and however luminous the general praise, received lower ratings.

These results are limited by the retrospective design. Although the large number of residents studied over multiple generations of residents and faculties increases confidence in the findings, we did not study the outcomes of applicants whose interview scores fell below the threshold for acceptance to the program under investigation or who chose to go elsewhere. Generalizability is reduced by the number of international medical graduates (IMGs) that was too small to analyze separately (about 10 in the entire sample), perhaps as a result of the large number of applicants and the location of the residency. Many residency programs now contain larger percentages of IMGs; however, no IMGs in this study were dismissed from the program and none did more poorly than U.S. graduates on the measures discussed below. The percentage of women in most residencies now is higher than that in our overall sample, but like most programs, the number of women has increased exponentially in recent years, and the current proportion of women is similar to the national ratio. The very low turnover of residency directors, while increasing reliability of assessments, may also limit generalizability to the many residencies with more frequent changes in leadership.

The addition of other outcome measures such as Psychiatry Resident-In-Training Examination scores might produce greater variability and perhaps more objective results, but these scores were only available for later cohorts and there was no greater variability of scores of residents who took the examination than there was in faculty evaluations. The overall good performance of this large group could reflect easy grading, although this tendency would have persisted for generations without change. It is also possible that screening for medical school was so effective, especially in highly qualified applicants, that the variance in outcome in residency is low enough to make it impossible to identify the very small number of high-risk residents, although interview scores were not correlated with performance in the lowest quartile while letters of recommendation were. An alternative explanation is that if qualities of high-performance residents could be better

quantified, those qualities could be more specifically measured in admission interviews, and the likelihood of disappointing educational outcomes—especially of residents being dismissed or dropping out—could be reduced. In the meantime, current data suggest that admission interviews might focus as much on helping the candidate to evaluate the program as helping the program to evaluate the candidate.

References

1. Altmaier EM, Smith Wl, O'Halloran CM, et al: The predictive utility of behavior-based interviewing compared with traditional interviewing in the selection of radiology residents. Invest Radiol 1992; 27:385–389
2. Balentine J, Gaeta T, Spevack T: Evaluating applicants to emergency medicine residency programs. J Emerg Med 1999; 17:131–134
3. Wood PS, Smith Wl, Altmaier EM, et al: A prospective study of cognitive and noncognitive selection criteria as predictors of resident performance. Invest Radiol 1990; 25:855–859
4. McDonald JS: Performance of residents in anesthesiology as related to measures of personality and interests. Psychol Rep 1991; 68:979–994
5. Smilen SW, Funai EF, Bianco AT: Residency selection: should interviewers be given applicants' board scores? Am J Obstet Gynecol 2001; 184:508–513
6. Takayama H, Grinsell R, Brock D, et al: Is it appropriate to use core clerkship grades in the selection of residents? Curr Surgery 2006; 63:391–396
7. McDonald JS, Lingam RP, Gupta B, et al: Psychologic testing as an aid to selection of residents in anesthesiology. Anesthesia & Analgesia 1994; 78:542–547
8. Hojat M, Borenstein B, Veloski JJ: Cognitive and noncognitive factors in predicting the clinical performance of medical school graduates. J Med Educ 1988; 63:323–325
9. Oggins J, Inglehart M, Brown DR, et al: Gender differences in the prediction of medical students' clinical performance. JAMWA 1988; 43:171–175
10. Arvey RD, Campion JE: The employment interview: summary and review of recent literature. Personnel Psychol 1982; 35:281–322
11. Levin RA, Zickar MJ: Investigating self-presentation, lies and bullshit: understanding faking and its effects on hiring decisions using theory, field research and simulation, in The Psychology of Work: Theoretically Based Empirical Research. Edited by Drasgow F, Brett J. Mahwah, NJ, Erlbaum, 2002, pp 253–276
12. 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
13. Borowitz SM, Saulsbury FT, Wilson WG: Information collected during the residency match process does not predict clinical performance. Arch Pediatr Adolesc Med 2000; 15: 256–260
14. Rolfe IE, Pearson S, Powis DA, et al: Time for a review of admission to medical school? Lancet 1995; 346:1329–1333
15. Komives E, Weiss ST, Rosa RM: The applicant interview as a predictor of resident performance. J Med Educ 1984; 59:425–426
16. Papp KK, Polk HC, Richardson JD: The relationship between criteria used to select residents and performance during residency. Am J Surgery 1997; 173:326–329
17. Laskin DM, Lesny RJ, Best AM: The residents' viewpoint of the matching process, factors influencing their program selection, and satisfaction with the results. J Oral Maxillofac Surg 2003; 61:228–233
18. Benjamin S, Robbins LI, Kung S: Online resources for assessment and evaluation. Acad Psychiatry 2006; 30:498–504
19. Reilley RR, Chao GT: Validity and fairness of some alternative employee selection procedures. Personnel Psychol 1982; 35:1–62
20. Wright SM, Ziegelstein RC: Writing more informative letters of reference. Gen Intern Med 2004; 19:588–593
21. Ericsson KA, Charness N: Expert performance: its structure and acquisition. Am Psychol 1994; 49:725–747
22. Ericsson KA, Simon HA: Protocol Analysis: Verbal Reports as Data (revised). Cambridge, Mass, Bradford Books/MIT Press, 1993