Can We Predict "Problem Residents"?

Adam M. Brenner, MD, Samuel Mathai, MD, Satyam Jain, MD, and Paul C. Mohl, MD

Abstract

Purpose
This study investigates whether data available at the time of residency application can be used to predict more accurately future problems of performance, both during and after residency.

Method
The authors identified all residents with reported problematic behavior across 20 years (1987–2007) at a single residency program and created a set of matched controls. Problems were further divided into "major" (leading to significant disruptions of performance and disciplinary action) and "minor" (remediable and resolved). Application materials were then reviewed for United States Medical Licensing Examination (USMLE) scores, evidence of academic failures, interviewer ratings, negative interviewer comments, negative comments in the dean's letter, and negative comments in letters of recommendation.

Results
The presence of any negative comments in the dean's letter yielded significant correlations with future problems. Further, those applicants with future major problems had significantly more negative comments in the dean's letter than did those with future minor problems. Other factors such as USMLE scores, failed courses, letters of recommendation, and interviewer ratings and comments did not predict future problems.

Conclusions
Most of the factors the authors assessed in prospective applicants did not predict future problems, with the exception of negative (even mildly so) comments in the dean's letter. The authors suggest that more attention should be paid to the use of the dean's letter to assess risk among applicants, and prospective study of this assessment should be performed.

Acad Med. 2010; 85:1147–1151.

Each year, residency programs expend significant resources in the recruitment and assessment of applicants. Most applicants will become successful residents who progress without interruption toward graduation, facing only the usual stumbles of normal professional development along the way. Furthermore, most residents will have few performance-related difficulties during their careers after graduation. But not all applicants will thrive. In spite of careful selection processes, residency programs regularly fail to predict which trainees will fall below minimal performance standards. The term "problem resident" has been defined as a generic term meant to describe "any resident whose behavior interferes with the resident's functioning."[1(p197)]

The presence of even just one of these problem residents can significantly affect the entire program. When problem residents are unable to continue with their duties, the program directors must take steps to protect patients and ensure appropriate care. This may tax the resources of other residents and faculty. Disciplinary action and remediation plans may require legal consultation and result in additional expenditure of time by faculty in monitoring, counseling, and evaluating the problem resident. Problem residents may also damage the morale and reputation of the residency program in less concrete but nonetheless significant ways. Further, graduates who encounter problems during training deliver a blow to the reputation of the residency.

A review of the literature assessing our capacity to predict future performance of residents yields mixed results. Several studies find that neither scores given during the selection process nor specific academic variables from the medical school record correlate with the later clinical performance of the residents as assessed by faculty.[2–6] In some of these investigations, however, United States Medical Licensing Examination (USMLE) scores or other objective measures of academic performance do predict scores on standardized in-service training exams or on specialty boards.[4–6] Other reports contradict these results and find correlations between the global assessment of the applicants by the selection committee and later clinical assessments of performance.[7,8] How should we interpret these divergent results?

The authors of a study in a surgical program that did not yield correlations between components of the residency application and clinical performance[3] suggest that different means of generating the global score or final National Resident Matching Program (NRMP) ranking may contribute to inconsistent findings. They note that some programs place more emphasis on objective, cognitively oriented data, and others on more subjective, noncognitive qualities, such as character, motivation, and affective attitudes. The authors of a report from an obstetrics–gynecology

Dr. Brenner is associate professor and director of residency training, Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas.

Dr. Mathai is assistant professor, University of Texas Southwestern Medical Center, Dallas, Texas, and medical director, OEF/OIF Program, Fort Worth VA Medical Center, Fort Worth, Texas.

Dr. Jain is assistant professor, University of Texas Southwestern Medical Center, and staff psychiatrist, Dallas VA Medical Center and North Texas State Hospital, Dallas, Texas.

Dr. Mohl is professor and vice chair for education, Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas.

Correspondence should be addressed to Dr. Brenner, UT Southwestern Medical Center, Department of Psychiatry, 5323 Harry Hines Blvd., Dallas, TX 75390-9070; e-mail: adam.brenner@utsouthwestern.edu.

program support this view: Their NRMP rank list, which did predict outcome of performance, was heavily weighted toward personal qualities.[7] One study found that both the USMLE score and the interviewer's subjective impression predicted success during residency, but in that academically oriented program, success was measured in part by publications and presentations.[9]

Some studies have focused more specifically on predicting poor clinical performance, problems related to professionalism, or attrition. A recent examination of a surgical residency found that attrition was associated with some demographic factors but also with relatively less enthusiastic summary comments in the dean's letter.[10] Another study noted that deficits of professionalism may be more closely linked to personal and motivational issues than to cognitive attributes and that medical school records largely fail to predict such difficulties, though clerkship performances may be relevant.[11] Clerkship grades also predicted professionalism problems in a study of internal medicine first-year residents.[12] This was not, however, the case in an examination of surgical residents with either professionalism problems or academic difficulties; in this study, only poor scores on personal qualities by interviewers were predictive of professional failings, whereas academic grades in medical school were actually negatively correlated.[13]

Related studies specific to psychiatry residencies are discouraging about our ability to make predictions. One study showed no correlation between interviewers' ratings of accepted applicants and later assessments of residents' performance.[14] Another group of investigators examined the relationship between the selection committee's global assessment and later evaluations of residents by the program faculty.[15] In fact, they found no significant associations between their match lists and later performance in the program. A third study identified all graduates of a residency, across a span of 27 years, who had been referred to an impaired physicians program.[16] Using a case–control format, the authors found no difference between the two groups on either scores by admission interviewers or residency evaluations. More recently, the

same group examined the relationship between admission interviews, letters of reference, and residency performance in 544 residents across 41 years.[17] They found some modest correlations but concluded that the differences were not great enough to be of practical utility.

Many different causes, including medical illness, acute psychiatric illness, character pathology, overwhelming situational stressors, and others, may lead to problematic performance. In the current study we did not attempt to distinguish among etiologies. We instead defined problematic outcomes operationally, as any significant interference with the resident's capacity to function adequately in his or her professional role. We hypothesized that examination of the applications of those residents who later developed problems would result in identification of risk factors. Our aim in this study was to determine whether any of the parameters we regularly consider in our own application process (board scores, academic transcripts, interviewer ratings and comments, the dean's letter, and letters of recommendation) could help us more accurately identify the applicants most at risk for future difficulty.

We also wanted to see if we could selectively predict whether problems could be resolved spontaneously or be remediated, or whether problems would be more intransigent, leading to major disciplinary actions, failure to successfully complete training, or performance problems subsequent to graduation. Some residency programs may wish to avoid accepting any applicants whose course will ultimately be problematic. However, in our experience, some applicants are best thought of as "high risk–high reward"—they may be likely to develop problems, but these problems do not inevitably lead to catastrophic outcomes. The very characteristics that place the resident at risk (e.g., history of an illness) may also contribute to the development of unique strengths in the future. Markers that predict an increased risk of future problems could be helpful in preparing support and proactive interventions that increase the chance of success.

Method

This study was reviewed and deemed exempt by the University of Texas

Southwestern Medical Center institutional review board, as not carrying any significant risk to the subjects studied.

We retrospectively identified all problem residents from the past 20 years (1987–2007) in our adult psychiatry program. One of the authors (P.C.M.) reviewed the rosters of each class of residents during his tenure as program director and identified those who had developed significant problems (n = 40). These designations were then reviewed and confirmed by another of the authors (A.B.). We defined "problems" as any difficulty that directly affected performance such that the resident's performance fell beneath the minimum standards of the program. These difficulties included acute psychiatric illness, character pathology, boundary violations with patients, recurrent conflicts with peers and faculty, and situational losses and stresses. They did not include acute medical illnesses or surgery. If a resident had been functioning at an exemplary level and then, in the face of some difficulty, his or her performance declined but continued to meet our standards, we did not consider this a "problem" for the purposes of our study. On the other hand, if a resident took a leave of absence in response to a stressor, we understood this to indicate that work performance either had already been affected or that the resident anticipated that he or she could not continue to maintain an adequate level of function, so these residents were classified as problem residents for our purposes. Problem residents also included some who did not meet the threshold for being so identified during training but who had significant postresidency problems. Ten of the 40 problem residents fit this category. Examples include two former chief residents who committed serious boundary violations, another who was fired after developing bipolar disorder, and another who was fired for failure to complete dictations over several months. Some, but not all, of these problem graduates had difficulty during residency, but none met our definition of problem residents.

We then divided the problem cases into "major" and "minor." Two of the authors (A.B. and P.C.M.) reviewed the files, discussed the cases together, and arrived

at a consensus. "Minor problems" resulted in resolution of the difficulty and did not require significant disciplinary action. A "major problem," in contrast, involved severity such that significant disciplinary action was required by the program or by an external governing body. Major problems all involved substantial deficiency in professional behavior, including in some cases profound transgressions of boundaries with patients. Major problems generally did not result in full remission of the difficulties, regardless of whether the underlying causes were characterologic, situational, or medical.

We then created a set of control residents (n = 42), matching for as many of the following variables as possible, in the following order: year of graduation, gender, age, medical school, and status as an international medical graduate (IMG). Names and other obvious identifiers were then deleted from the files of both cases and controls.

The files were reviewed for the following variables: total rating score by the admissions interviewers, number of negative interviewer comments, USMLE scores, failing grades in academic courses in medical school, negative comments in letters of recommendation, and negative comments in the dean's letter. To establish reliability, the four authors together first reviewed and scored six application files that were not included in our study. We were able to rapidly achieve consensus and consistency in our ratings. Two of the authors then served as independent, blinded raters (S.J. and S.M.) and reviewed the application files of our cases and controls. Neither of the two raters was associated with the residency training administration, and both were new to the department, reducing any risk that the content of the files might lead to unintended identification of a resident by the rater.

A low threshold was set for deeming a comment in the letters of recommendation and dean's letters negative. Such letters are generally overwhelmingly positive in content, and we hypothesized that even subtle negativity could be meaningful. For example, a dean's letter comment that on a particular rotation the student could have read more and been better prepared would qualify as negative (see List 1 for

List 1
Sample Negative Comments From Dean's Letters

- "Very nervous, timid initially."
- "Displayed little curiosity."
- "Had difficulty applying knowledge clinically."
- "He came across as confrontational."
- "Maybe somewhat overconfident for his level of training."
- "Lack of enthusiasm and problems in organization."
- "Needs to read more on her own."
- "Lots of effort, uneven outcome."

more examples). We found that the files of IMGs were qualitatively different in terms of the content of the dean's letter, and so we decided to also analyze the U.S. medical graduates (USMGs) and the IMGs as separate groups.

We analyzed the data to assess whether the application file variables correlated with problematic outcome, and then whether this correlation was specific for major or minor problems. Statistical measures included chi-square analysis and t tests.

Results

We found negative comments in 21 of the dean's letters of USMGs and in none of the dean's letters of IMGs. The presence of these negative comments showed a strong correlation with problem outcomes among USMGs ($\chi^2 = 7.5241$, $P < .01$) (Table 1). This correlation remained significant even when we looked at the entire pool of residents, which were attenuated by the

addition of the IMGs ($\chi^2 = 8.5255$, $P < .01$). The amount of negative comments in the letter was also predictive. Our problem cases had significantly more (1.22 ± 1.72) negative dean's letter comments than controls (0.53 ± 1.10, $P < .04$, two-tailed t test). In addition, the dean's letters of those with major problems had significantly more negative comments than did those of residents with minor problems (0.62 ± 0.96, 1.54 ± 1.10, $P < .03$, two-tailed t test).

In contrast, there was no significant correlation between interviewers' ratings of applicants and whether those applicants developed problems (either minor or major) for all the residents ($P < .65$, two-tailed t test); this was also true for the U.S. graduates ($P < .8$, two-tailed t test) and the IMGs examined separately. Further, the interviewer ratings did not correlate differently with major and minor problems for the residents as a whole ($P < .90$, two-tailed t test) or for the subgroup of IMGs ($P < .14$, two-tailed t test). However, the difference in interviewer ratings for residents who exhibited major and minor problems in USMGs did achieve statistical significance ($P < .02$, two-tailed t test). There were no significant differences in the number of negative comments recorded by the interviewers for residents with problems versus those without problems whether we looked at all residents ($P < .38$, two-tailed t test), just USMGs ($P < .41$, two-tailed t test), or just IMGs ($P < .46$, two-tailed t test). The number of negative comments from interviewers also did not correlate differently with major versus minor problems ($P < .67$, two-tailed t test).

Table 1
Presence of Negative Comments in Dean's Letters of Cases (Eventually Reported for Problems) and Controls (Never Reported for Problems) in a Psychiatry Residency, University of Texas Southwestern Medical Center, 1987–2007

Cohort	Negative comments in dean's letter	No negative comments in dean's letter
U.S. medical graduates (n = 66)*		
Cases (n = 34)	21	13
Controls (n =32)	9	23
All residents (n = 82)†		
Cases (n = 40)	21	19
Controls (n = 42)	9	33

* $\chi^2 = 7.5241$, $P < 0.01$, $df = 1$.
† $\chi^2 = 8.5255$, $P < 0.01$, $df = 1$.

There were no significant correlations between a history of academic failures and problem outcomes ($\chi^2 = .6679$, $P < .25$, $df = 1$). No significant findings were uncovered regarding the presence of negative or lukewarm comments in recommendation letters ($\chi^2 = .0038$, $P < .8$, $df = 1$). USMLE scores could not be analyzed because too many were missing from the records. This was at least partly related to Texas being on the Federal Licensing Examination through much of the period under study.

Because the most robust discriminator between problem residents and controls was the dean's letter, which is not comparable between IMGs and USMGs, we provide some further qualitative description of our IMGs who had difficulty. Nine of the problem residents were IMGs: Two were from the Middle East, three were South Asian (two of whom were born and raised in the United States but attended medical school in India), three were from East Asia, and one was from Latin America. Five of the nine (the three East Asians, the Latin American, and one of the South Asians) developed problems that appeared to be due to language or cultural differences. Two of the South Asians and one of the Middle Easterners had personality issues that affected their clinical work, and the other Middle Eastern resident developed a major psychiatric illness after graduation that disrupted her career.

Discussion

We have found a rather robust multilevel correlation between residents who have problems, major or minor, during or after residency, and negative statements, even subtle ones, in the dean's letter. Further, to the extent that we were able to test others' findings, we were unable to replicate them. The power of the dean's letter is, in some ways, not surprising. After all, it is a summation of three years of academic data and one year of intense, multiobserver reports of actual behavior and performance in clinical settings.

There are several limitations to our study. First, we looked at data from only one U.S. psychiatry residency program. We must therefore be cautious in assuming generalizability, although we have no reason to believe that our applicants, application materials, or resident

outcomes are unusual. Another limitation was the lack of standardization of the interviewers and their ratings over the years. At some points a large pool of faculty interviewed candidates, while at other times a small group who were highly involved in residency training, conducted all interviews. We attempted to adjust for this by matching cases with controls who had been interviewed in the same year.

It is also possible that our identification of problem graduates (i.e., those whose problems occurred after residency) was incomplete. Because there was no systematic reporting method for postgraduation problems, it is entirely conceivable that some graduates had significant problems that are still unrecognized. But because of the increasing transparency of state medical boards' responses to problem physicians and the tendency of employers to call the training program when a relatively recent graduate has problems, we do not believe we have missed a significant number.

In conducting a study that attempts to predict risk based on data from the application file, we must also consider the accuracy of those data. A number of studies across a range of specialties, including psychiatry, suggest that a significant minority of claims regarding publications contain substantial inaccuracies, either by being wholly fabricated or by being inflated (e.g., changing one's status in the author listing, upgrading the journal where the article was published on a CV).[18–21] In addition, and particularly relevant to our results, one study examined over 500 dean's letters and found a surprising number of discrepancies.[22] In these instances, the dean's letter omitted a significant academic difficulty that was evident either in the transcript or in some other part of the application. In our opinion, this tendency to burnish or enhance one's own application, or the applications of the institution's graduating medical students, suggests that any evidence of problems in the record is likely to represent the tip of the iceberg. Fortunately, in the last few years the Association of American Medical Colleges has made efforts to standardize the format of the medical student performance evaluation, as dean's letters are now called, and has strongly

encouraged those writing them to be neutral and objective.

We looked at the number of negative interviewer comments along with the interviewer's global rating because we hypothesized that there might be applicants whose global interview score was boosted by an impressive record of achievement but who, nevertheless, raised red flags for some of the interviewers. Yet, neither the global rating nor the number of distinctly negative comments differentiated the future problem residents from the controls. This does not necessarily imply inaccuracy in assessing the relative aptitude of applicants in general; that is a separate question not addressed by this study. It is worth noting that our rating form for interviewers asked for both strengths and weaknesses of each applicant. We believe this resulted in our interviewers having a very low threshold for including lukewarm or negative comments.

Our analyses found the dean's letter to be the only consistently significant correlate of future difficulties. Prior studies in other specialties that examined the dean's letter summary statements found conflicting results, with one group finding a correlation with attrition in a surgical residency[10] and another group finding no correlation with radiology resident performance.[6] Our study, however, did not limit itself to the brief summary statements, and it may be unique in scoring the entire dean's letter, line by line, for negative comments.

Until recently, deans were in the position of wanting to support their own students and even to effectively market the graduates of the medical school to prospective residency training programs. But, unlike letters of recommendation, the dean's letter must also fulfill the function of providing an accurate summary report of the student's performance with the imprimatur of the medical school administration. It is possible that writers of dean's letters might feel compelled to include any significantly negative information but would, nonetheless, do their best to present the most positive picture that is consistent with the facts. As a result, most dean's letters during the period covered by this study are overwhelmingly positive; therefore, when any issue rises

to the level of prominence to merit inclusion, it may indeed indicate the "tip of the iceberg." Some studies suggest that clerkship performance is predictive of problems during residency[11,12]; the dean's letter, with its qualitative evaluations from clerkships, may provide a more sensitive reading of those performances than does the final grade on the transcript. The simple presence of any negativity in the dean's letter increased the chance of difficulty down the road, and this increased further according to the number of negative comments in the dean's letter. This finding should be further investigated in prospective studies of incoming residents.

What are the implications of better prediction of problem residents? Certainly, we might endeavor to be more exacting about whom we accept into our training programs and, thereby, attempt to avoid all hints of future problems. At our own program, we plan to integrate this study's findings into our current system of grading applicants to give more negative weighting to evidence of problems in the dean's letter. It will be interesting to see whether, over time, we are more successful in our outcomes. Yet we are also cautious about becoming too restrictive and guarded in our assessment of applicants. Screening out all potential problems would come at a significant cost to our residencies and to psychiatry as a whole. For example, some residents may have a mental illness that increases the risk of problems, but that illness may also be profoundly generative of empathy, scholarship, activism, or other forms of professional productivity. In an intriguing study of internal medicine residents, Girard and Hickam[23] found that some measures of clinical performance actually correlated with higher levels of depression among interns. Being able to identify at-risk applicants might help us to be more proactive in supporting them, setting up systems of monitoring, and planning for additional supervision. Our goal, then,

would be to contain potential problems to what we defined as "minor"—recognizing that problems will inevitably occur, but trying to use available tools to ensure that we choose residents who are able to recover from performance problems and support them in this recovery, ensuring the best outcomes for our trainees as well as our patients.

References

1 Scheiber S, Tasman A. Special problems. In: Kay J, ed. Handbook of Psychiatry Residency Training. Washington, DC: American Psychiatric Association; 1991:185–203.

2 Metro D, Talrico J, Patel R, Wetmore A. The resident application process and its correlation to future performance as a resident. Anesth Analg. 2005;100:502–505.

3 Papp K, Polk H, Richardson J. The relationship between criteria used to select residents and performance during residency. Am J Surg. 1997;173:326–329.

4 Dirschl D, Campion E, Gilliam K. Resident selection and predictors of performance: Can we be evidence based? Clin Orthop Relat Res. 2006;449:44–49.

5 Bell J, Kanellitsas I, Shaffer L. Selection of obstetrics and gynecology residents on the basis of medical school performance. Am J Obstet Gynecol. 2002;186:1091–1094.

6 Boyse T, Patterson S, Cohan R, et al. Does medical school performance predict radiology resident performance? Acad Radiol. 2002;9:437–445.

7 Olawaiye A, Yeh J, Witham-Leitch M. Resident selection process and prediction of clinical performance in an obstetrics and gynecology program. Teach Learn Med. 2006;18:310–315.

8 Ozuah P. Predicting residents' performance: A prospective study. BMC Med Educ. 2002;2:7.

9 Daly K, Levine S, Adams G. Predictors for resident success in otolaryngology. J Am Coll Surg. 2006;202:649–654.

10 Naylor R, Reish J, Valentine J. Factors related to attrition in surgery residency based on application data. Arch Surg. 2008;143:647–652.

11 Brown E, Rosinski E, Altman D. Comparing medical school graduates who perform poorly in residency with graduates who perform well. Acad Med. 1993;68:806–808.

12 Greenburg DL, Durning SJ, Cohen DL, Cruess D, Jackson JL. Identifying medical students likely to exhibit poor professionalism and knowledge during internship. J Gen Intern Med. 2007;22:1711–1717.

13 Brothers T, Wetherold S. Importance of the faculty interview during the resident application process. J Surg Educ. 2007;64:378–385.

14 Kandler H, Plutchik R, Conte H, Siegel B. Prediction of performance of psychiatric residents: A three-year follow-up study. Am J Psychiatry. 1974;132:1286–1290.

15 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.

16 Dubovsky SL, Gendel M, Dubovsky AN, Rosse J, Levin R, House R. Do data obtained from admissions interviews and resident evaluations predict later personal and practice problems? Acad Psychiatry. 2005;29:443–447.

17 Dubovsky SL, Gendel MH, Dubovsky AN, Levin R, Rosse J, House R. Can admissions interviews predict performance in residency? Acad Psychiatry. 2008;32:498–503.

18 Boyd AS, Hoook M, King LE Jr. An evaluation of the accuracy of residency applicants' curricula vitae: Are the claims of publications erroneous? J Am Acad Dermatol. 1996;35:606–608.

19 Roellig MS, Katz ED. Inaccuracies on applications for emergency medicine residency training. Acad Emerg Med. 2004;11:992–994.

20 Yang GY, Schoenwetter MF, Wagner TD, Donohue KA, Kuettel MR. Misrepresentation of publications among radiation oncology residency applicants. J Am Coll Radiol. 2006;3:259–264.

21 Caplan JP, Borus JF, Chang G, Greenberg WE. Poor intentions or poor attention: Misrepresentation by applicants to psychiatry residency. Acad Psychiatry. 2008;32:225–229.

22 Edmond M, Roberson M, Hasan N. The dishonest dean's letter: An analysis of 532 dean's letters from 99 U.S. medical schools. Acad Med. 1999;74:1033–1035.

23 Girard D, Hickam D. Predictors of clinical performance among internal medicine residents. J Gen Intern Med. 1991;6:150–154.