Factors Related to Attrition in Surgery Residency Based on Application Data

Rebekah A. Naylor, MD; Joan S. Reisch, PhD; R. James Valentine, MD

Objective: To determine whether variables in the surgery resident selection process will predict attrition or performance during residency training.

Design: Retrospective study.

Setting: A university residency program.


Main Outcome Measures: Satisfactory outcome included successful completion of training and the American Board of Surgery examinations on the first attempt. Participants with a satisfactory outcome were stratified into good or marginal performance based on adverse actions during residency.

Results: Of 111 residents studied, 28 (25.2%) had an unsatisfactory outcome; attrition occurred in 25 (22.5%). Univariate analysis identified the following variables as predictors of unsatisfactory outcome: age at entry older than 29 years ($P = .005$), female sex ($P = .02$), courses repeated ($P = .01$), “C” grades on transcript ($P = .01$), no participation in team sports ($P = .02$), and lack of superlative comments in the dean’s letter ($P = .03$). The following variables were retained in the multivariate model: age older than 29 years (odds ratio [OR], 0.11; 95% confidence interval [CI], 0.02-0.47; $P = .003$), summary comments in the dean’s letter (OR, 4.57; 95% CI, 2.00-10.43; $P < .001$), participation in team sports (OR, 4.96; 95% CI, 1.36-18.05; $P = .02$), and merit scholarship in medical school (OR, 0.25; 95% CI, 0.08-0.78; $P = .02$).

Conclusions: Attrition can be predicted from factors identified on residency applications, with nonacademic factors being more important. Among residents who completed the program, no predictors of performance were identified.

Arch Surg. 2008;143(7):647-652

Author Affiliations:
Department of Surgery (Drs Naylor and Valentine) and Division of Biostatistics, Department of Clinical Sciences (Dr Reisch), University of Texas Southwestern Medical Center at Dallas.
all candidates who matched into categorical surgical residency positions at the UTSMC between 1991 and 2000 were reviewed. Files included the original application and residency performance data, including yearly summary performance evaluations and American Board of Surgery In-Training Examination (ABSITE) scores. The basic selection and evaluation processes remained constant during the study period.

SELECTION PROCESS

All applications were prescreened by the program director. To be offered an interview, candidates had to meet the following minimum criteria: USMLE step 1 score greater than 210, no grades below “C” or “P” on their transcript, and rank in the upper half of the class. All applicants from the UTSMC were invited for an interview without regard for the preselection criteria.

RESIDENT EVALUATION PROCESS

Residents received monthly written evaluations from supervising faculty, with scores on a Likert scale ranging from unsatisfactory to outstanding. Annual written evaluations were prepared for each resident by the program director summarizing overall performance, ABSITE score, and any concerns, such as inadequate cognitive knowledge, unprofessional behavior, interpersonal difficulties, or deficient technical skills. Adverse actions, such as academic probation, remedial years of training, and dismissal from the program, were documented in a final summary statement.

STUDY DESIGN

The independent variables examined in this study included sex, age at entry, marital status at entry, class rank, surgery clerkship grade, merit scholarships in medical school, membership in Alpha Omega Alpha, research experience, publications in peer-reviewed journals, extracurricular activities, leadership positions, dean’s letter summary statement, interview score, and final rank list position. Class rank was assigned by quartile and was often reported in the dean’s letter. In cases in which an objective class rank was not stated, rank was estimated from graphic performance summaries provided by each medical school. Grades were taken directly from official transcripts. In cases in which letter grades were not assigned for the surgery clerkship, a grade of “honors” was considered equivalent to an “A,” “high pass” was equivalent to a “B+,” and “pass” was equivalent to a “B.” Summary statements from the dean’s recommendation letter were recorded for each resident. This variable was ranked on a 3-point Likert scale, with 3 assigned to the best score. Residents were stratified according to the definition of keywords provided with the dean’s letter. In-Training Examination (ABSITE) scores. The basic selection criteria.

STATISTICAL ANALYSES

Analyses for this research project were performed using SAS statistical software, version 9.1 (SAS Institute Inc, Cary, North Carolina). Categorical data items were summarized using frequency counts, ranges, and percentages, and means and standard deviations were calculated for age. Age was dichotomized as 29 years or younger and older than 29 years, which was based on 1 SD above the mean age at entry for the group. To gain insight for the multivariate analyses, the individual measurements were each compared, grouped according to the outcome measure. Analyses or Fisher exact probability tests were used for group comparisons of each of the categorical measurements, and t tests for 2 independent groups were used for the numerical variables. Stepwise multiple logistic regression was used to determine which demographic and other factors were statistically related to successful outcome. The model entry criteria were relaxed to 10% so as not to miss a possible predictive factor. Factors considered after the univariate analyses included categorized age, sex, whether medical school courses had to be repeated, merit scholarship, sports participation, class rank, and categorized summary statements from the dean’s recommendation letter. The Hosmer-Lemeshow technique was used to assess the model fit.

RESULTS

Between 1991 and 2000, 115 applicants matched into categorical surgery residency positions at UTSMC. Four persons were excluded from the study: 2 who were still in training, 1 who never attempted the board examinations, and 1 whose file was lost. Therefore, 111 residents composed the study group. Group demographics and characteristics are given in Table 1. The proportion of women in the residency program did not change significantly throughout the study period.

FACTORS ASSOCIATED WITH UNSATISFACTORY OUTCOME

Unsatisfactory outcome occurred in 28 participants (25.2%), including 25 who did not complete the training program (22.5%) and 3 who did not pass both the qualifying and certifying examinations on the first attempt. Unsatisfactory outcome occurred for 11 of the 52 study participants (21.2%) who matched between 1991 and 1995, which was not significantly different compared with 14 of 59 study participants (23.7%) who matched between 1996 and 2000. Primary reasons for attrition were program initiated in 10 instances and resident initiated in 15 instances (Table 2). The attrition rate in women remained constant throughout the study period.

Six examined variables were significantly associated with outcome by means of χ² contingency analyses (Table 3). Unsatisfactory outcome occurred in 7 of 12
residents (58.3%) who were older than 29 years at entry into the program compared with 21 of 99 (21.2%) who were 29 years or younger (P = .005). Fourteen men (18.4%) and 14 women (40.0%) had an unsatisfactory outcome (P = .02); this was because of attrition in 12 men (15.8%) and 13 women (37.1%) and board failure in 2 men (2.6%) and 1 woman (2.9%). Two residents had repeated a course in medical school, and both had an unsatisfactory outcome. This was significantly more than 26 of 109 residents (23.9%) who had not repeated a course (P = .01). Eight residents had 1 or more grades of C on their medical school transcript. Of these residents, 5 (62.5%) had an unsatisfactory outcome compared with 23 of 103 (22.3%) who did not have any C grades (P = .01). Of 36 residents who played team sports in college or medical school, 4 (11.1%) had an unsatisfactory outcome, which was significantly less than 23 of 75 residents (30.7%) who had not played sports (P = .02). Seventy-five residents had comments ranked as 3 in their dean’s letters, 28 had comments ranked as 2, and 8 had comments ranked as 1. Unsatisfactory outcomes occurred in 15 (20.0%) of the first group, 8 (28.6%) of the second group, and 5 (62.5%) of the third group (P = .03).

Three variables that were not statistically significant showed interesting trends (Table 3). Unsatisfactory outcome occurred in 19 of 84 residents (22.6%) who were in the top quartile of their medical school class, in 6 of 23 (26.1%) in the second quartile, and in 3 of 4 (75.0%) in the third quartile (P = .06). Eleven of 30 residents (36.7%) who had received merit scholarships in medical school had an unsatisfactory outcome compared with 17 of 81 (21.0%) who did not receive merit scholarships (P = .09). The importance of the applicant’s position in the rank list was analyzed by using the unpaired t test. For matched applicants who had an unsatisfactory outcome, the mean rank position was 26.4 compared with 17.9 for residents with a successful outcome (P = .05). The range of ranks in the 2 outcome groups was the same (1-75 vs 1-74), which limits the predictability of this factor. All the other variables examined, including other academic factors, were not statistically significantly associated with outcome.

Four variables were retained in the multivariate model: age older than 29 years (odds ratio [OR], 0.11; 95% con-
22% of 88 categorical surgery residents left voluntarily. Aufses et al1 reported that 3% of categorical residents left their programs.2 In a previous publication from the UTSMC, Bergen1998. However, the database included residents in undesignated preliminary positions, and the authors estimated overall attrition of 12% to 16% in categorical residents.1

FACTORS SIGNIFICANTLY ASSOCIATED WITH PERFORMANCE

The overall performance of the 83 residents with a satisfactory outcome was then studied. Twenty residents were marginal in their performance: 4 were placed on academic probation, 1 required remedial months of training, and 20 had concerns noted in annual reviews (12, knowledge deficit; 10, unprofessional behavior; 3, deficient technical skills; and 3, patient care or service organization difficulties) (residents could be in ≥1 category). On stepwise multiple logistic regression analysis, there were no factors that were significantly associated with a marginal outcome.

This study was designed to evaluate factors significantly associated with unsatisfactory outcome (attrition or inability to pass the board examinations on the first attempt) and with performance in residency. The results show that application data can be used to identify applicants at risk for unsatisfactory outcome, but they do not show that application data are associated with performance.

Two outcome measures were chosen to represent unsatisfactory outcome because we believed that they were penultimate markers of failure of the educational process. Graduates of this program have been highly successful on American Board of Surgery examinations, making this an uncommon reason for unsatisfactory outcome. Most participants (25 of 28 participants) had unsatisfactory outcome because of attrition; therefore, the study results may be largely interpreted as factors associated with attrition more than outcome per se.

The magnitude of attrition in surgery residencies has been difficult to define nationally. Estimates range from 3% to 26%, depending on the definition of attrition, the number of years studied, and the scope of the database. Morris et al1 surveyed 206 surgery program directors and found that 3% of categorical residents left their programs voluntarily during a 1-year period; nonvoluntary dismissals were not included. Aufses et al1 reported that 22% of 88 categorical surgery residents left voluntarily during a 15-year period at a single institution. Kwakwa and Jonasson2 accessed data from the American Medical Association Medical Education Research Information Database, the American College of Surgeons Resident Masterfile, and the Association of American Medical Colleges’ Graduate Medical Education Tracking Census Database. The combined data showed that general surgery residents had an attrition rate of 26% from 1993 to 1998. However, the database included residents in undesignated preliminary positions, and the authors estimated overall attrition of 12% to 16% in categorical residents.2 In a previous publication from the UTSMC, Bergen et al3 reported voluntary attrition of 11.5% and involuntary attrition of 2%. The present data show that more UTSMC residents have been dismissed involuntarily in recent years (13.5% and 9.0%, respectively), but the present data are generally in keeping with the recently reported national attrition data. The overall attrition rate of categorical residents who leave programs for any reason remains unknown.

The present study suggests that unsatisfactory outcome (ie, attrition) cannot be predicted on the basis of traditional measures of academic achievement in medical school. Of the academic, demographic, and social variables studied, age was the most important factor. We cannot explain why older residents had a higher risk of unsatisfactory outcome, other than to note that family and lifestyle issues tend to become more important with increasing age. Most of the voluntary resident attrition during the study was attributed to family and lifestyle issues, especially for women (Table 2). Although there was a positive association between female sex and unsatisfactory outcome on univariate analysis in the present study, it was not retained in the multivariate model. Overall, the present results suggest that age is a more important factor in outcome.

Of the multiple academic variables evaluated in this study, the dean’s recommendation letter seems to give the most accurate picture of an applicant’s suitability for residency training. Although the letters vary from institution to institution, most include common phrases or have a list of key terms that help put an applicant in context. The present study supports the validity and accuracy of the dean’s summary evaluation: the highest level of recommendation was associated with the best chance of satisfactory outcome.

The most surprising result in this study was the negative association of a merit scholarship with outcome. We cannot explain this finding other than to note that merit scholarships are usually awarded to students on the basis of exceptional academic performance during the undergraduate years. This adds further support to the notion that academic factors are not significantly associated with outcome in residency.

Of the extracurricular activities examined, only participation in team sports was positively associated with satisfactory outcome. It is somewhat surprising that playing a musical instrument was not significant. We acknowledge that playing a musical instrument may have occurred in the context of a group or orchestra (ie, “team”), but we were unable to determine this information from the application data. Other investigators10 found no correlation between extracurricular activities, such as playing a musical instrument, and overall performance during residency.

The second purpose of this study was to examine whether application data were associated with the performance of residents who had a satisfactory outcome. The data show that marginal performance in residency was not significantly associated with any of the application data evaluated. This should not be surprising because previous studies5,8 have been unable to identify consistent predictors of residency performance, including postgraduate year 1 performance or ABSITE scores.
As in the previous studies, no application factors were significantly associated with performance.

Other researchers have shown that performance during the first year of residency is predictive of overall performance. A post hoc analysis was performed on the present data to determine whether unsatisfactory outcome could be anticipated early in residency. Indeed, either unsatisfactory faculty evaluations or concerns noted by the program director during the first year of training were statistically significantly associated with unsatisfactory outcome. We caution that these findings are based on subjective data that may be prone to individual variation.

Several limitations must be acknowledged. First, this study is from a single large university program; the findings may not be applicable to smaller programs or to community-based programs. Second, the study is retrospective. The files, particularly evaluations, required some degree of interpretation, which may not be accurate. Third, the preselection criteria used to invite applicants to interview may have created bias that prevented a true analysis of the examined factors. It is possible that academic factors were not significant because only top academic performers were interviewed; the lack of low-performing individuals in the study group risks a type II error. We also acknowledge that a variety of academic variables not evaluated in this study, such as grade point average and USMLE scores, may have been significant. However, the wide variation in reported grades and the change in national board examinations during the study period made this analysis beyond the scope of the present study. A final limitation is that a significant change in resident expectations was introduced during the study period: the 80-hour workweek. It is possible that changes in expectations have led to changes in faculty evaluations, but this is beyond the scope of the present study. The possible effect of this change on the results obtained is unknown.

In conclusion, the present study shows that social and demographic factors may be more important factors related to outcome for surgery residents than traditional markers of academic achievement in medical school. Of the academic factors evaluated, summative comments in the dean’s letter of recommendation was the only significant variable associated with successful outcome. The only other positive factor was participation in team sports; negative factors included age older than 29 years and having a merit scholarship in medical school. Identification of these factors may affect selection or mandate early intervention. In residents who had a successful outcome, no variables were associated with overall performance in the residency.

Accepted for Publication: January 5, 2008.
Correspondence: Rebekah A. Naylor, MD, Department of Surgery, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75390-9156 (rebekah.naylor@utsouthwestern.edu).
Author Contributions: Dr Naylor had access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Naylor and Valentine. Acquisition of data: Naylor and Valentine. Analysis and interpretation of data: Naylor, Reisch, and Valentine. Drafting of the manuscript: Naylor, Reisch, and Valentine. Critical review of the manuscript for important intellectual content: Naylor, Reisch, and Valentine. Financial Disclosure: None reported.

Previous Presentation: This paper was presented at the 2007 Annual Meeting of the Western Surgical Association; November 5, 2007, Colorado Springs, Colorado; and is published after peer review and revision. The discussions that follow this article are based on the originally submitted manuscript and not the revised manuscript.

REFERENCES

Question number 1 for you, Dr Naylor. Will the authors share with us how they believe we are actually going to be able to use these data in resident selection? For example, I doubt that you would fail to rank somebody because they are 31 years old or if they didn’t play team sports.

Number 2, 65% of the residents of Southwestern are AOA [Alpha Omega Alpha] and have extremely low board fail rates, suggesting that maybe they are not typical surgery residents. Can the data used in this study be used to extrapolate to their more average surgical brethren across the country?

Number 3, I was concerned in reading the manuscript that all of the variables that you looked at were subjective, and you didn’t include objective variables, most notable of which were the USMLE and national boards scores. Is there not a way that those numbers could be utilized in some fashion using percentiles or something similar so that they could be included in these data? Those tests tend to be so important in our screening and the way we rank our candidates at the end of the interview season. As a follow-up to that question, have you yourself instituted any programs to decrease the attrition rate in your own residency program?

Dr Valentine: Dr Dayton suggested that our attrition rate is higher than the national average. However, our results are very much in keeping with the current reported attrition rate in the United States. The national attrition rate in general surgery programs is much higher than most people would imagine. The best estimate is 26%, based on a paper from Olga Jonasson that used data from the AMA [American Medical Association], the American College of Surgeons, and the AAMC [Association of American Medical Colleges]. However, the real attrition rate is not known, and that is why Dr Dick Bell from the American Board of Surgery and a resident from Yale University, Dr Heather Yeo, are examining this issue directly. Based on his preliminary data, Dr Bell recently noted that the attrition rate may be as high as 25% in the United States. I would encourage all program directors to examine their own 10-year attrition rate. Some of you will be very surprised.

Your first question was whether we can use these data in resident selection. We certainly will in our program. Although we used to put a premium on playing musical instruments as a marker for technical skill, we are now going to pay more attention to whether the applicant plays team sports. We are also going to pay more attention to age, considering that being older than 29 years was a high risk for attrition in our program. We will also put heavier emphasis on the dean’s letter of recommendation. It is not surprising that the dean’s letter was predictive of attrition because deans have access to more information about a candidate than most other faculty members do. The fact that deans are actually requesting feedback about their letters from program directors suggests that they are paying a lot of attention to their recommendations.

Your second question was whether our data can be extrapolated to all programs in the United States. This is clearly a limitation of our study. We have a large program based at a busy academic medical center. Our data may not apply to community-based programs or to programs with fewer residents. Because our interview selection criteria led to inclusion of high academic performers in the study group, our data may also not apply to less competitive applicants.

Your third question asked whether there was a way that we could have included more objective variables, especially USMLE scores. There are 2 reasons why we did not include these scores in the analysis. First, the tests changed halfway through our study period: the USMLE was substituted for National Board of Medical Examiners, with a profound change in grading scales. Second, normative values were not provided with the USMLE to allow us to compare applicants from one year with those from another year. Judging from recent USMLE graphs provided by the ERAS [Electronic Residency Application Service], there has been significant grade inflation. We felt that using raw scores was not appropriate because of this trend.

Financial Disclosure: None reported.