Correlation Between Trainee Candidate Selection Criteria and Subsequent Performance

Jesse C Selber, MD, MPH, FACS, Winnie Tong, MD, John Koshy, MD, Amir Ibrahim, MD, Jun Liu, PhD, Charles Butler, MD, FACS

BACKGROUND: The objective of trainee recruitment is to identify candidates likely to perform well as trainees and subsequent faculty. The effectiveness of this process has not been established. The goal of this study was to identify trainee selection criteria predictive of excellent performance.

STUDY DESIGN: Twenty-nine microsurgery fellows were enrolled from 2008 to 2012. Each candidate was interviewed and rated based on presentation, plastic surgery (PS) training experience, academic potential, personality, social skills, communication skills, and ability to be a team player. An unadjusted rank list was generated based on weighted averages, and an adjusted rank list was then generated at a faculty meeting. At the conclusion of fellowship, each fellow was rated based on the ACGME core competencies. Spearman correlation coefficients (r) were used to measure the correlations between fellow selection criteria and fellow performance.

RESULTS: Plastic surgery training and academic potential had, by far, the strongest correlation to overall performance (r: 0.678, $p < 0.001$ and r: 0.56, $p < 0.002$), and to all ACGME competencies. When reformulated to weight PS training and academic potential more heavily than subjective criteria, the scoring system was significantly more predictive of excellent performance (r: 0.49 vs 0.70). The unadjusted rank list was more predictive of excellent performance than the adjusted rank list (r: 0.45 vs 0.65).

CONCLUSIONS: Plastic surgery training experience and academic potential were better predictors of performance than any subjective information ascertained during the interview. Adjustments to the rank list based on faculty discussion resulted in lower performance candidates moving up in ranking. Ranking criteria and interview techniques must be refined to improve predictive power. It may be beneficial for semi-objective criteria to carry more weight than subjective criteria and raw scores to remain unadjusted by extraneous information. (J Am Coll Surg 2014;219:951–957. © 2014 by the American College of Surgeons)

Besides patient care, training future practitioners is our most critical task. As a result, selecting trainees represents a challenging and time consuming process for academic medical programs. Program directors and faculty spend considerable effort to identify and recruit the best candidates who will be both excellent fellows and go on to become excellent practitioners, yet our definition for what constitutes excellence in these areas is lacking. Candidates are ranked based on performance parameters that vary widely among programs, including a combination of subjective and objective criteria such as in-service scores, letters of recommendation, personal statements, and a large and diverse set of personality and character traits. Despite these efforts, there is no consensus on the best strategy to identify candidates who will be top performers in training. Most selection processes are nonstandardized and often unreliable, obfuscating comparison between candidates. The lack of consensus regarding performance further thwarts efforts to refine the selection process. Therefore, the effectiveness of the current selection process to identify the best trainees, although critical, has not been verified. There is a paucity of data on the predictive validity of current selection processes in almost every area of medicine.

Disclosure Information. Nothing to disclose.

Abstract presented at the American Association of Plastic Surgeons Annual Meeting, Miami, FL, April 2014.

Received June 24, 2014; Revised July 21, 2014; Accepted July 25, 2014.

From the Department of Plastic Surgery, The University of Texas, MD Anderson Cancer Center (Selber, Tong, Ibrahim, Liu, Butler) and the Department of Surgery, Baylor College of Medicine (Koshy), Houston, TX. Correspondence address: Jesse C Selber, MD, MPH, FACS, Department of Plastic Surgery, The University of Texas, MD Anderson Cancer Center, 1400 Pressler St, Unit 1488, Houston, TX 77030. email: jcselber@mdanderson.org

http://dx.doi.org/10.1016/j.jamcollsurg.2014.07.942
ISSN 1072-7515/14

Abbreviations and Acronyms
AP = academic potential
ARL = adjusted rank list
PS = plastic surgery
URL = unadjusted rank list

The goal of this study was to identify factors in fellow candidate selection that most strongly predict excellent performance in the fellowship, and to adjust our ranking system to reflect those factors.

METHODS

Patients and study design

This was a retrospective cohort study of 29 microsurgery fellows who matched into the MD Anderson Cancer Center (MDACC) microsurgery fellowship between 2008 and 2012.

Fellow candidate selection

All applicants were screened based on their submitted applications including a curriculum vitae, 3 letters of recommendation, and a personal statement. A subset of candidates was invited to a formal interview over 2 days. During the interview, each applicant was provided with information about the fellowship, had tours of the facility, met with current fellows, and was interviewed by all available clinical faculty in the plastic surgery department. Candidates were evaluated on the basis of their submitted application and interview, and then rated. The rating system was a custom designed, standardized, web-based online system in which all fields required completion and were submitted before faculty discussion. The home-grown system included 7 characteristics: presentation, plastic surgery training experience (PS training), academic potential (AP), personality, social skills, communication skills, and ability to be a team player. Each characteristic was scored 1 to 10 (worst to best). Each interviewer gave an overall impression score and a preliminary ranking of the candidates blinded to other evaluators. A final score was then generated using a weighted average of the 7 individual categories and overall impression, with overall impression being more heavily weighted than the other categories.

$$\text{Current Formula}: \text{Final Score} = 0.3 \times \text{overall impression} + 0.1 \times (\text{presentation, plastic surgery training, academic potential, personality, social skills, communication skills, team player})$$

Resulting scores were placed in descending order to generate an unadjusted rank list (URL). During a faculty meeting convened immediately after the interviews, candidates were discussed and additional information provided. Candidates could be moved up, down, or off the list until consensus was reached. This resulted in a final adjusted rank list (ARL) that was subsequently submitted.

Fellow performance evaluation

After completing 1 year of training, the clinical performance of each fellow at our institution was evaluated by the clinical plastic surgery faculty. Each fellow was given a score by the faculty for 37 distinct items within the 6 core competencies of the ACGME: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and system-based practice. The 37 subcategories included information geared specifically toward the evaluation of microsurgical skills, such as the harvest of common free flaps and performance of a microvascular anastomosis. The scores ranged from 1 to 5 (worst to best) To study the effectiveness of our selection process, factors used in fellow candidate selection were correlated with fellow performance evaluation to determine which of the 7 individual criteria were most strongly predictive of high performing clinical fellows. These results were used to modify our scoring system as described below. In addition, the correlations between both the URL and ARL and performance were compared to determine if group consensus improved the ability to predict high performance fellows.

Statistical analysis

Descriptive statistics were used to summarize the scores of the interview and performance evaluation. Spearman correlation coefficients (r) were used to measure the correlations between selection criteria and performance. An internal bootstrap analysis was used to determine whether the modified scoring system was better than the existing scoring system at predicting performance. The analyses were performed in SAS 9.2 (SAS Institute Inc).

Validation

After performing the analysis, factors that were mostly strongly correlated (PS training and AP) were used to recalculate the final score, using weighted averages adjusted to reflect the strength of the individual factors. In order to accomplish this, the weight given to selection criteria was altered in the formula used to calculate the URL. Then we determined if the reformulation was more predictive by using bootstrapping internal validation techniques. One hundred random samples were taken within the cohort in order to test the 2 scoring systems against one another.