Background: This study examined the success of an admission screen in identifying applicants who will fail Part I or Part II of the Podiatric Boards on the first attempt.

Methods: A logistic regression model was used to predict failure on the Podiatric Boards.

Results: The biology Medical College Admission Test (MCAT) score predicted failure on Part I and Part II. Adding undergraduate grade point average and the other MCAT subscores did not significantly improve the prediction. Although a screen based on the biology MCAT score can identify nearly 90% of those who fail, there are six false-positives for every true-positive.

Conclusions: First-time Board failure can be predicted, but Board failure must be primarily addressed by interventions after matriculation rather than by an admission screen. (J Am Podiatr Med Assoc 102(6): 485-490, 2012)

The primary purpose of this study was to evaluate how well an admission screen can predict first-time failure on the Podiatric Boards Part I or Part II. Just as a medical provider should not use a medical screen without understanding its predictive success, so a medical educator should not use an admission screen without understanding its predictive success. As much as Board failure can be predicted, schools could improve their Board pass rates by setting appropriate admission screens. A secondary purpose of this study was to identify which admission variables are predictors of Board failure. This study makes an original contribution by focusing on evaluating the success of prediction rather than on identifying statistically significant predictors. Also, it makes a contribution by looking at the Podiatric Boards. The literature is almost entirely on the United States Medical Licensing Examination (USMLE). As is appropriate for the Boards, the metric of performance is pass/fail rather than a numerical score.

The National Board of Podiatric Medical Examiners is the governing agency offering the American Podiatric Medical Licensing Examination. Comparable with the USMLE, the podiatric medical examinations are subdivided into three parts assessing knowledge and skill throughout the podiatric medical education process. Parts I and II of the examination are written to evaluate whether a candidate has the knowledge base required “to practice as a minimally competent entry-level podiatric physician.” Unlike the USMLE, passing students do not receive their scores unless requested. Only a notification of passing is reported because the Boards are not designed as an achievement test but rather as an instrument to identify students who are not minimally competent, that is, those who fail.

Traditional admission criteria for acceptance into a podiatric medical program are much like those for an allopathic medical program. Students must present undergraduate grade point averages (GPAs), particular premedical courses, and scores on the Medical College Admission Test (MCAT). The similarity in both admission criteria and licensing examinations makes the literature on predictors of performance on the USMLE pertinent to the present study.

The identification of particular predictors of
Board performance has fueled a debate. The predictors investigated include traditional admission criteria and socioeconomic factors. In a study of 818 medical school students, Kleshinski et al\(^2\) found that the biology MCAT score and the undergraduate science GPA were the strongest predictors of performance on the USMLE. Peterson and Tucker\(^3\) showed that the biology MCAT score is the strongest traditional predictor of passing the Boards. However, Julian\(^4\) found that the total MCAT score, not the separate subscores, best predicts Board performance. Julian\(^4\) further found that the MCAT score alone can “essentially replace the need for undergraduate GPA’s in their impressive predication of USMLE scores.”\(^{p910}\)

The ability to identify and reject, at the admission stage, students who would not become minimally competent physicians as assessed by Board failure would be a valuable asset for medical educators. The success of diagnostic tests in predicting the presence of disease can be evaluated by using sensitivity and specificity. Admission screens have been similarly evaluated. Albanese et al\(^5\) discovered sensitivity of 57% and specificity of 90% for a screen using the total MCAT score at a cutoff point of 24 to predict USMLE Step 1 failure. Coumerbatch et al\(^6\) discovered sensitivity of 88% and specificity of 80% for a screen that included the physical science and biology MCAT scores with a measure of academic performance in the second year to predict USMLE Step 1 failure.

Although there are many available studies focusing on the predictive abilities of admission criteria for performance on the USMLE, there is only one article addressing podiatric medical education. Although the study by Smith and Geletta\(^7\) focused on admission predictors, it was limited to their use in predicting first-year podiatric medical school performance and not on their success in predicting American Podiatric Medical Licensing Examination failure. In addition, they urged the need for further studies determining admission variables applicable to clinical competence.

The present study was conducted at Barry University, a private institution in Miami Shores, Florida. Founded in 1940, it is a federally designated Hispanic-serving institution where nearly half of its 9,000 students are underserved minorities. For 25 years, the Barry University School of Podiatric Medicine has been providing medical education for doctors of podiatric medicine. With many more applicants than there are positions available in each first-year class, it is crucial that the best candidates are selected. In this article, the notion of “best” is made operational as “passing the Boards Part I and Part II on the first attempt.” The admission committee faces the daunting task of implementing evidence-based selection criteria as they attempt to identify candidates with high potential for passing the Boards and completing the rigorous podiatric medical program.

**Methods**

The sample includes students who entered podiatric medical school between 1995 and 2009. For students entering before 2005, reporting of Board results to the school was voluntary, and 12% of the students chose not to report. Furthermore, students were included in the sample only if they reported MCAT scores and undergraduate GPA. The sample for Part I (n = 491) constitutes 66% of first-time test takers in this period, and the sample for Part II (n = 419) constitutes 65% of first-time test takers. Although there may be some self-selection bias in the sample, for Part I, the overall pass rate in the sample was only 2% higher than the overall pass rate in the program, and for Part II, the pass rate in the sample was only 1% higher. Consequently, there does not seem to be significant self-selection bias.

The sample included 57% men and had a race/ethnic distribution of 48% white, 17% Hispanic, 14% black, 11% Asian, and 10% other race/ethnicity or unreported. Fourteen percent of the students were not US citizens, coming from 35 different countries of origin. The mean ± SD age of the sample was 25.9 ± 4.8 years. Hispanic and black students were overrepresented, composing 31% of the sample but only 14% of the population of podiatric medical students since 2007.

The conceptual framework for the statistical analysis was classical test theory.\(^8\) This framework forecasts a criterion variable (failure on Podiatric Boards) based on predictors. Prediction is never expected to precisely match the outcome owing to three sources of error: measurement errors, missing predictors, and random variation. The disease detection model was used to evaluate the predictive success of the model because measures of predictive success from this model, for example, sensitivity and specificity, are more familiar to podiatric medical faculty than are statistical measures of success, for example, log likelihood.

Logistic regression was the statistical method used to estimate the probability of failing the Boards based on measures of educational background (MCAT score and undergraduate GPA). Because demographics cannot be used as part of
an admission screen, demographic predictors (sex, race/ethnicity, age, and citizenship status) were not evaluated in this study. Logistic regression is better suited than linear regression because the dependent variable is dichotomous.

Recall that the primary purpose of this research was to estimate how well Board failure can be predicted by an admission screen. Identifying which predictors are statistically significant is of secondary interest. The identification of individual statistically significant predictors is further complicated by the high correlations (multicollinearity) among MCAT scores and undergraduate GPA (Table 1).

**Results**

**Boards Part I**

Initially, all of the predictors were entered into the model. Only biology MCAT score was statistically significant owing to multicollinearity. When the predictors were entered individually, all except the essay MCAT score were statistically significant. Because the multiple correlation between the biology score and the other predictors (undergraduate GPA and the other MCAT subscores) is 0.71, adding the other predictors did not add significantly to the predictive power of the model ($\chi^2 = 1.47, P = .83$). Proceeding with only the biology MCAT score as a predictor, logistic regression results indicated that the overall model was statistically reliable in predicting success on Part I. Model fit statistics revealed a good-fitting model. (2 log likelihood = 291.377 and Hosmer-Lemeshow goodness of fit = 12.614 [$P > .05$]. A $P$ value greater than .05 indicates no statistically significant difference between predicted and observed values. The generated model was significantly different from the constant-only model [$\chi^2 = 18.525, P < .001$].

Regression coefficients are presented in Table 2. The Wald statistic indicated that the biology MCAT score was a statistically significant predictor. The odds ratio indicated that increasing the biology MCAT score by 1 decreased the odds of failing Part I by 28% (95% confidence interval, 16%–39%). The fairly large confidence interval for the odds ratio reflects the imprecision of the prediction.

**Boards Part II**

Again, all of the predictors were initially entered into the model, and none of the predictors was statistically significant due to multicollinearity. When predictors were entered individually, all except the essay MCAT score were statistically significant. As previously noted, owing to the high multiple correlation between the biology score and the other predictors, adding the other predictors did not add significantly to the predictive power of the model ($\chi^2 = 4.967, P = .29$). Proceeding with only the biology MCAT score as a predictor, logistic regression results indicated that the overall model was statistically reliable in distinguishing between passing and failing Part II. Model fit statistics revealed a good-fitting model. (2 log likelihood = 270.438 and Hosmer-Lemeshow goodness of fit = 5.711 [$P = .33$]). A $P$ value greater than 0.05 indicates no statistically significant difference between predicted and observed values. The generated model was significantly different from the constant-only model [$\chi^2 = 11.095, P < .001$]. Regression coefficients are presented in Table 3. The odds ratio indicated that increasing the biology MCAT score by 1 decreased the odds of failing Part II by 23% (95% confidence interval, 10%–35%). Again, the fairly large confidence interval for the odds ratio reflects the imprecision of the prediction.

**Classification Success**

For Part I and Part II, classification success was
evaluated using a biology MCAT score of 7 or less as the cutoff value to distinguish predicted passing from predicted failure. This cutoff score was determined based on a receiver operating characteristic curve analysis and was chosen to maximize the separation of passing and failing, that is, the cutoff score that maximizes the distance between the receiver operating characteristic curve and the diagonal line. The overall accuracy of prediction was 43% for Part I and 45% for Part II.

Table 4 shows how the observed Board outcome matched the predicted Board outcome for Part I and Part II. The columns represent the observed pass/fail outcomes, and the rows represent the predicted pass/fail outcome. For example, the third column shows that of the 47 students who failed Part I, five were incorrectly predicted to pass and 42 were correctly predicted to fail.

Sensitivity, specificity, and positive and negative predictive values were also calculated to assess the predictive success of the model (Table 5). Sensitivity is true-positives as a proportion of all students who failed. Specificity is true-negatives as a proportion of all students who passed. In the logistic regression model for Part I, the sensitivity of 89% represented the ratio of the 42 students who were correctly predicted to fail to the 47 who actually failed. For Part I, the specificity of 38% represented the ratio of the 170 students who were correctly predicted to pass to the 444 who actually passed. Sensitivity and specificity are measures of the model getting the prediction correct, and the values do not depend on the incidence of failure. Consequently, sensitivity and specificity can be used to compare the predictive success of these models with other screens. Positive predictive value is the proportion of those predicted to fail who actually failed. For Part I, the positive predictive value of 13% represents the ratio of the 42 students who were correctly predicted to fail to the 316 students who were predicted to fail. Negative predictive value is the proportion of those predicted to pass who actually passed. Positive and negative predictive values depend on the incidence of failure, so the low positive predictive value reflects the low incidence of failure.

Although high sensitivity shows that the admission screen is successful in identifying students who fail, this comes at the cost of a high rate of false-positives: 87% for Part I and 85% for Part II. Choosing a cutoff score requires a trade-off between more accurate overall prediction and predictions that are less accurate but better identify students at elevated risk for failing the Boards. Perhaps the cutoff score was set too high and a substantially lower cutoff score would give a more acceptable rate of false-positives. If the cutoff score is set at 4, 38% of those who failed Part I and 25% of those who failed Part II would have been correctly identified by the screen. However, the rate of false-positives is still high—76% for Part I and 82% for Part II. Although the rate of false-positives is not qualitatively lower when the cutoff score is reduced from 7 to 4, the sensitivity is less than half. Apparently, a high rate of false-positives is unavoidable for a podiatric medical admission screen. Comparing the admission screens with the prostate cancer medical screen, the prostate-specific antigen test, and the pulmonary embolism screen, the D-dimer assay, the admission screen is only slightly better than the prostate-specific antigen test and does not perform as well as the D-dimer assay.

**Discussion**

This study makes an important original contribution to the literature on podiatric medical education in two ways. First, it brings to bear rigorous statistical methods using a large sample to predict Board success, the most objective measure of competence

### Table 3. Logistic Regression Analysis Summary for Biology MCAT Score Predicting Failure on Part II

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biology MCAT score</td>
<td>-0.267</td>
<td>0.081</td>
<td>10.727</td>
<td>0.77</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.479</td>
<td>0.502</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note: df = 1, 2 log likelihood = 270.438, $\chi^2_1 = 11.095, P = .001 \ (n = 419)$.

Abbreviations: MCAT, Medical College Admission Test; NA, not applicable.

* $P < .001$. 

---

**Table 4. Classification Success for Predictive Models of Board Failure**

<table>
<thead>
<tr>
<th>Predicted Performance</th>
<th>Observed Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pass</td>
</tr>
<tr>
<td>Part I</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>170</td>
</tr>
<tr>
<td>Fail</td>
<td>274</td>
</tr>
<tr>
<td>Total</td>
<td>444</td>
</tr>
<tr>
<td>Part II</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>148</td>
</tr>
<tr>
<td>Fail</td>
<td>227</td>
</tr>
<tr>
<td>Total</td>
<td>375</td>
</tr>
</tbody>
</table>
for podiatric students. More important, this study focuses on how well we can predict Board failure not just on identifying the statistically significant predictors. In this way, this article contributes to a critical evidence-based discussion of admission policy in podiatric medical education.

The results of this study were largely consistent with the literature in terms of identifying the statistically significant predictor, biology MCAT score, and in terms of the predictive success of the model. Predicting Board failure is made difficult by the low prevalence of failure in the population of podiatric medical students. Intuitively, it is more difficult to predict a rare event. However, this article shows that we can predict Board failure at the admission stage for Part I and Part II. However, the prediction is much less precise than is usually imagined, even at low cutoff scores. The low precision of prediction is shown by the large confidence intervals for sensitivity and specificity in Table 5. The conclusion that our ability to predict Board failure at the admission stage is only slightly better than the prostate-specific antigen screen is faint praise of screening precision, especially given the recent recommendation that healthy men not use the prostate-specific antigen screen.

Several study limitations must be identified. It is a threat to external validity that the sample came from a single institution. Although this institution had a relatively high minority student representation, there is no reason to believe that the racial/ethnic distribution is a source of bias because race/ethnicity was not a predictor of Board failure. We believe that the results are relevant to all podiatric medical students and possibly allopathic medical students. In this study, there were no qualitative measures of affective influences, for example, motivation, on Board success. This is a threat to internal validity because these influences are uncontrolled sources of variation in Board success. This raises the question of whether interviews or other qualitative screening processes could accurately identify some of these affective influences and significantly improve predictive success at the admission stage. The underlying concern is to screen out applicants who, after training, would not “practice as a minimally competent entry-level podiatric physician.” However, it is not clear that there is a strong empirical basis supporting the Podiatric Boards as valid instruments for making operational the construct of minimal competence in entry-level practice.

The demonstrated imprecision of an admission screen, in this study and in the literature, raises the question of whether an admission screen should be used. The answer to that question depends on the value the program places on a true-positive, correctly identifying an applicant who would go on to fail the Boards, versus a false-positive, incorrectly screening out an applicant who would have passed the Boards. Identifying applicants who would fail the Boards is beneficial for the applicant and the program. The applicant is spared taking on educational debt that may be more difficult to repay if the applicant cannot enter podiatric medical practice. Removing these applicants from the enrolled class would improve academic quality and may allow the program to recruit more

<table>
<thead>
<tr>
<th>Screen and Cutoff Score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive Predictive Value</th>
<th>Negative Predictive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I biology MCAT score ≤7</td>
<td>89 (77–96)</td>
<td>38 (34–43)</td>
<td>13 (10–18)</td>
<td>97 (94–99)</td>
</tr>
<tr>
<td>Part II biology MCAT score ≤7</td>
<td>89 (75–96)</td>
<td>39 (35–45)</td>
<td>15 (11–20)</td>
<td>97 (93–99)</td>
</tr>
<tr>
<td>Part I biology MCAT score ≤4</td>
<td>38</td>
<td>87</td>
<td>24</td>
<td>93</td>
</tr>
<tr>
<td>Part II biology MCAT score ≤4</td>
<td>25</td>
<td>87</td>
<td>18</td>
<td>91</td>
</tr>
<tr>
<td>USMLE Step 1 biology MCAT score ≤8</td>
<td>48</td>
<td>93</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>PSA cutoff point 5 ng/mL</td>
<td>75</td>
<td>44</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>D-dimer assay</td>
<td>85</td>
<td>68</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

Note: Data are given as percentages (95% confidence intervals).
Abbreviations: MCAT, Medical College Admission Test; NR, not reported; PSA, prostate-specific antigen; USMLE, United States Medical Licensing Examination.
predicting Board success are statistically significant, it is important to understand that they have low precision. Many applicants who would otherwise have passed the boards would have to be unnecessarily rejected to appropriately reject a single applicant who would fail. The problem of Board failure cannot be “solved” with an admission screen. Follow-up screening and successful remediation are necessary to ensure program improvement and student success.

Financial Disclosure: None reported.  
Conflict of Interest: None reported.

References