Prediction of student performance on the Comprehensive Osteopathic Medical Licensing Examination Level 1 based on admission data and course performance

MICHAEL K. COPE, PhD; HELEN H. BAKER, PhD, MBA; ROBERT FISK, PhD; JOHN N. GORBY; ROBERT W. FOSTER, DO

To predict student performance on the Comprehensive Osteopathic Medical Licensing Examination (COMLEX–USA) Level 1 examination based on academic performance during the first 2 years, stepwise regression analysis of COMLEX–USA Level 1 performance with preadmission grade point averages, Medical College Admission Test scores, and academic performance was performed on the class of 2000 to develop three formulae that were then used to predict performance on COMLEX–USA Level 1 for the class of 2001. Models ranged in accuracy of predicting the pass/fail status from 95.2% (all available data) to 96.8% (first-year grades and admissions data). A predictive model for student performance on COMLEX–USA Level 1 can be developed and has a high degree of accuracy. The model with the most variables available to choose from predicts the most failures.

(Key words: Comprehensive Osteopathic Medical Licensing Examination, medical education)

Previously, we reported the correlation between admission data and course performance with performance on the Comprehensive Osteopathic Medical Licensing Examination (COMLEX–USA) Level 1 for the Class of 2000 at the West Virginia School of Osteopathic Medicine.1 COMLEX–USA Level 1 is a norm-referenced, multiple-choice examination administered over 2 days. COMLEX–USA Level 1 is the first part of an examination series described by the National Board of Osteopathic Medical Examiners in the following way: “The COMLEX–USA program is designed to assess the osteopathic medical knowledge considered essential for osteopathic generalist physicians to practice medicine without supervision. COMLEX is constructed in the context of medical problem-solving which involves clinical presentations and physician tasks. Candidates are expected to utilize the philosophy and principles of osteopathic medicine to solve medical problems.”2

While the previous study is supportive of the use of COMLEX–USA as a measure of knowledge and licensure requirements, this type of information is of more use to the institution when applied to a process of identifying students with potential difficulty in passing COMLEX–USA Level 1.

Passage of COMLEX–USA Level 1 is important to students because it is an institutional requirement that must be met before they can continue with their clinical training programs. Also, passing COMLEX–USA Level 1 is a requirement for taking COMLEX–USA Level 2, which in turn is a requirement for graduation at all American osteopathic Association-accredited schools. If the institution develops a system that allows identification of students at high risk of failing the COMLEX–USA Level 1 early in the preclinical program, it may be possible to alert these students of the potential failure based on current performance. Accordingly, these students could then be directed to sources of assistance in preparing for the examination. Therefore, development of such predictive models would benefit the institution and its students in their ultimate goal: successful completion of requirements leading to graduation with a doctor of osteopathy degree and eventual licensure for practice.

Related research
Osteopathic medical education
In part because COMLEX–USA Level 1 was administered for the first time in June 1998, no studies could be identified describing predictive models for passage of COMLEX–USA Level 1.

Allopathic medical education
Elam and Johnson3 reported a comparison between regression equations for the old National Board of Medical Examiners (NBME) Part I and the new United States Medical Licensing Examination (USMLE) Step 1 performance. Both equations contained only two significant predictors: biology Medical College Admission Test (MCAT) subscale and second grade point average (GPA) for USMLE Step 1, and first- and second-year GPAs for NBME Part I. A nine-variable model that included all MCAT subscores, undergraduate GPAs, and first- and second-year GPAs were determined for both examinations. The models accounted for 59% of the variation in USMLE Step 1 performance.
scores and 66% of the variation in NBM E Part 1 scores.

Koenig and others4 used M CAT mean scores and adjusted undergraduate GPAs in a discriminate analysis model to predict the pass/fail status of students on USMLE Step 1, separated by ethnicity and sex. Their model overpredicted USMLE Step 1 performance, especially for African Americans, that is, students were projected to get higher scores than they actually received.

Comparing old and new M CATs, Swanson and colleagues reported three regression models for predicting USMLE Step 1 performance for each of the M CAT forms. For each M CAT, the first model included undergraduate admission data (science GPA, nonscience GPA, undergraduate school selectivity index, the product of the selectivity index, and the science GPA) and all of the appropriate M CAT parts. The first models predicted 36% of the variability in USMLE Step 1 scores for the new and old M CATs. The second models included only the undergraduate information and predicted 17% of the variability for the old M CAT and 18% of the variability for the new M CAT. The third models included only M CAT scores and predicted 32% of the variability for the old and new M CATs.

In an article discussing relationships between nonacademic variables and USMLE Step 1 performance in two medical schools, Webb and colleagues also reported regression models. These academic factors models included undergraduate GPAs, M CAT total score, and competitiveness of undergraduate college attended. These models accounted for 38% of USMLE Step 1 score variability in one school and 42% of the variability in the other school.

Wiley and Koenig9 reported five predictive models based on students from 16 medical schools. The models included various combinations of undergraduate GPAs, M CAT scores, and selectivity index of undergraduate college. The models were able to predict between 48% and 75% of the variability in USMLE Step 1 scores. The highest predictive level was shared by a model containing all three variables and a model containing undergraduate GPAs and M CAT scores.

Additionally, eight reports were published describing models for predicting performance on the old NBM E Part 1 examination2,6-14 and one for predicting performance on NBM E subject ("shelf") examinations15 in three basic sciences (anatomy, physiology, and biochemistry). These models used combinations of M CAT scores, undergraduate GPAs, country of origin, gender, and age to develop predictive formulae. Six of the authors2,6-9,12 reported R2 values that ranged between 0.20 and 0.73. Two of the authors reported the accuracy of their models, one at 79%10 and the other at 83%.11

Regression analysis of pertinent student data has been shown to provide reasonable predictions of student performance on licensing examinations. The greater the number of variables available to select from with the regression analysis, the greater the accuracy of the models. Models developed after the student has completed all didactic coursework are the most accurate, but this method may not provide sufficient time for counseling of students with potential difficulties, nor might it allow time for students to take appropriate remedial action. In this study, we attempted to determine an appropriate time for counseling.

Methods
Subjects
Subjects were West Virginia School of Osteopathic Medicine students in the Class of 2000 who sat for the COMLEX–USA Level 1 in June 1998 and students in the Class of 2001 who sat for the Level 1 examination in June 1999.

Admissions data model
Admissions data (GPAs and M CAT scores) from the class of 2000 were entered into the stepwise regression analysis using SPSS Base 9.0. Stepwise regression selected three variables (phase 1 GPA, pharmacology course, and physical sciences M CAT) for use in predicting COMLEX–USA Level 1 performance. The R2 for this model is 0.610. This model accounts for 61% of the variability in the COMLEX–USA Level 1 scores. The three models developed in this study have similar R2 values to those in other studies.

Results
The predicted values for each of the models and the actual results are presented in the Table. All variables chosen from the Class of 2000 for use in the prediction models for the Class of 2001 were evaluated by independent samples t-test for comparison of means to ensure that the samples were from the same population.

Admissions data model
The admissions data model (ADM) did not predict any failures. The ADM predicted an average score of 496 with a standard deviation of 19.9. Comparison of actual performance with the ADM (continued on page 89)
shows that the model correctly predicted the pass/fail status of 60 (96.8%) out of 62 students. Two students failed that were predicted to pass. The model overpredicted the performance of 24 students and underpredicted the performance of 38 students. The correlation between the predicted score and the actual score for the ADM was 0.441, significant at the 0.000 level.

End-of-first-year model
The end-of-first-year model (Y1M) predicted two failures. The Y1M predicted an average score of 492 with a standard deviation of 47.3. Comparison of actu-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Admission</th>
<th>Year 1 and admission</th>
<th>Year 2</th>
<th>COMLEX–USA level 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall GPA,* mean ± SD</td>
<td>3.38 ± 0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biology MCAT† score, mean ± SD</td>
<td>7.25 ± 1.57</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 GPA, mean ± SD</td>
<td>86.40 ± 4.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacology course GPA, mean ± SD</td>
<td>83.12 ± 6.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical sciences, MCAT score, mean ± SD</td>
<td>7.03 ± 1.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase 1 and 2 GPA, mean ± SD</td>
<td>87.58 ± 4.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathology discipline GPA, mean ± SD</td>
<td>84.98 ± 6.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical sciences MCAT score, mean ± SD</td>
<td>7.03 ± 1.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>498 ± 19.9</td>
<td>492 ± 47.3</td>
<td>484 ± 62.3</td>
<td>518 ± 81.7</td>
</tr>
<tr>
<td>Maximum</td>
<td>542</td>
<td>579</td>
<td>603</td>
<td>695</td>
</tr>
<tr>
<td>Minimum</td>
<td>455</td>
<td>378</td>
<td>320</td>
<td>281</td>
</tr>
<tr>
<td>Range</td>
<td>87</td>
<td>201</td>
<td>283</td>
<td>414</td>
</tr>
<tr>
<td>Predicted/actual failures</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Percent correct</td>
<td>96.8</td>
<td>96.8</td>
<td>95.2</td>
<td></td>
</tr>
<tr>
<td>Pass prediction</td>
<td>62</td>
<td>60</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Fail prediction</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Pass that fail</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fail that pass</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overpredictions, No.</td>
<td>24</td>
<td>23</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>Underpredictions, No.</td>
<td>38</td>
<td>39</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>Correlation with actual and P value</td>
<td>.441</td>
<td>.768</td>
<td>.824</td>
<td></td>
</tr>
</tbody>
</table>

*GPA = grade point average.
†MCAT = Medical College Admission Test.
al performance with the Y1M shows that the model correctly predicted the pass/fail status of 60 (96.8%) out of 62 students. One student predicted to fail passed and one student predicted to pass failed. The model overpredicted the performance of 23 students and underpredicted the performance of 39 students. The correlation between the predicted score and the actual score for the Y1M was 0.768, significant at the 0.000 level.

**End-of-second-year model**

The end-of-second-year model (Y2M) predicted five failures. The Y2M predicted an average score of 484 with a standard deviation of 62.3. Comparison of actual performance with the Y2M shows that the model correctly predicted the pass/fail status of 59 (95.2%) out of 62 students. Three students predicted to fail passed. The model overpredicted the performance of 16 students and underpredicted the performance of 46 students. The correlation between the predicted score and the actual score for the Y2M was 0.814, significant at the 0.000 level.

**Comments**

All three models tended to underpredict performance. The amount of this underprediction increased as the models included scores from later in the curriculum. The number of students whose scores were underpredicted rose from 38 with the admission model to 59 with the Y1M to 46 with the Y2M. The predicted range of values also increased, from 87 in the admissions model to 201 in the Y1M to 283 in the Y2M. The actual range of scores was 414 (281 to 695).

The accuracy of all three models for predicting the pass/fail status of students is high. The admissions model was 96.8% accurate, failing to predict the failure of two students. The Y1M was 96.8% accurate, failing to predict one passage and one failure. The Y2M was 95.2% accurate, predicting three failures by students who passed the test.

When used to predict scores for COMLEX–USA Level 1 performance, these models give a conservative representation of potential. In other words, they all provide more underpredictions than overpredictions of student performance. When these models are used to predict the pass/fail status of students, they get more pessimistic when more variables are available.

The ADM evaluation of potential student performance was 96.8% accurate, but did not predict any failures. This overprediction of student potential is not useful in counseling students because it does not identify any potential failures.

The Y1M was also 96.8% accurate in that it predicted one failure that was a pass and one pass that was a failure. This model is somewhat better in that it does predict some failures and is available at an appropriate time for counseling students. However, the Y1M did miss one failure.

The Y2M evaluation of student performance was 95.2% accurate and overpredicted failures. This model, being more conservative, would be most useful for counseling purposes, but the predictions are available just prior to taking COMLEX–USA and therefore would be of little or no value in alerting students to possible difficulty with the COMLEX–USA. Future model development should look to provide information for counseling at the end of the first semester of the second year.

**Acknowledgment**

The authors gratefully acknowledge assistance from colleagues too numerous to list for their many helpful comments on earlier versions of this manuscript.

**References**


