Strategies for Identifying Students at Risk for USMLE Step 1 Failure

Jira Coumarbatch, MD; Leah Robinson, EdS; Ronald Thomas, PhD; Patrick D. Bridge, PhD

Background and Objectives: *Failing Step 1 of the US Medical Licensing Examination (USMLE) or a delay in taking the exam can negatively affect a medical student's ability to match into a residency program. Unfortunately, identifying students at risk for failing Step 1 is challenging, but it is necessary to provide proactive educational support. The purpose of this study was to develop a strategy to identify students at risk for failing Step 1.* Methods: *Using a retrospective study design, 256 students from the class of 2008 were eligible for the study. Independent variables included Medical College Admission Test (MCAT) scores and cumulative grades from years 1–2 of medical school. The dependent variable was their score on the USMLE Step 1. Variables with a significant univariate relationship were loaded into a series of binary logistic regression models. A receiver operating characteristic (ROC) curve examined the significant variables.* Results: *Both year-2 standard score and the MCAT biological sciences score were significant as predictors of failure. The ROC curve provided a range of values for establishing a cutoff value for each significant variable. Conclusion: Using internal and external predictors, it is possible to identify students at risk for failing Step 1 of the USMLE.*

(Fam Med 2010;42(2):105-10.)

Most studies that have examined the prediction of medical students' performance on Step 1 of the United States Medical Licensing Exam (USMLE) have focused primarily on the "first-time test taker." These studies suggest that Medical College Admission Test (MCAT) scores are the best predictor of passing Step 1.[1-4] In particular, the MCAT biological science score has shown to be the strongest predictor of passing.[3] It has been relatively easy to create models for predicting passage of Step 1 due to the large number of students who pass on their first attempt. As a result, clear conclusions can be drawn from the results based on the power of the studies. It has been more difficult, on the other hand, to create models to predict failure for first-time test takers because of the low number of failures that occur at most schools.[3]

The inability to identify students at risk for Step 1 failure can affect a medical school but more importantly, it can affect a student's career. At many medical schools, students failing Step 1 are often delayed from continuing course work, which affects graduation and increases costs to the student. More importantly, Step 1 failure can affect a student's ability to enter a residency program and in some instances restrict them from applying for residency in specific states (eg, Alaska).[5] In fact, 21 states have some type of limitation on the number of attempts a student can have to pass all three steps of the USMLE.[5]

Determining the characteristics of a student who will fail Step 1 is challenging. There are instances where a student can excel academically during the first 2 years of medical school and yet fail Step 1. On the other hand, there are students who have academic difficulty during the first 2 years of medical school but pass Step 1 on the first attempt. In short, while predicting Step 1 failure with a high level of confidence is the goal, the goal is

From the Department of Academic and Student Programs (Dr Bridge, Ms Robinson), Faculty of Family Medicine and Public Health Sciences (Drs Bridge and Coumarbatch), Faculty of Pediatrics (Dr Thomas), Wayne State University.

difficult to achieve. The purpose of this study was to develop a strategy for identifying students at risk for failing Step 1 on the first attempt.

Methods

Subjects

Subjects for this study were the 256 students from the graduating class of 2008 at Wayne State University, which represents students matriculating from 2002–2004. The Wayne State University Institutional Review Board approved our study methods.

Data Collection

We collected data on the test results and academic performance of the students. These data were obtained from students' records in the Office of Admissions and Records and Registration.

Independent Variables. Independent variables included the average grade-point average (GPA) during college, MCAT scores, and course scores during years 1 and 2 of medical school. The mean GPA for the 256 students was 3.58. The mean MCAT score for biological science (BS) was 10.05, for physical science (PS) was 9.70, and for verbal reasoning (VR) was 8.97. We also collected scores on students' MCAT writing sample (WS).

Course scores during medical school included course standard scores and cumulative standard scores from year 1 and year 2. Course standard scores are based on students' performance for each course examination, as well as overall performance for the course. To calculate a standard score, a student's raw score is converted to a standard score for each examination with a mean set at 500 and one standard deviation of 100. At the end of the year, standard scores are converted to a cumulative standard score. For the graduating class of 2008, the first quintile (lowest) for cumulative standard scores ranged from 415.24 and below, second quintile 415.25–473.63, third quintile 473.64–524.39, fourth quintile 524.40–571.80, and the fifth quintile (highest) 571.81–672.95.

Dependent Variable. The dependent variable was the student's score for the first attempt at taking Step 1 of the USMLE. The national pass rate for the USMLE during the time frame of the matriculating cohort was 182.

Data Analysis

Initial Identification of Variables. Variables thought to have an association with failure were initially examined by a comparison of those who passed and those who failed Step 1 using the Mann Whitney U test. We chose the non-parametric Mann Whitney U because the data were not normally distributed.[6] The variables identified were MCAT scores for BS, PS, VR, and WS and standard scores for years 1 and 2 of medical school.

Regression Analysis. These variables were then examined using a series of binary logistic regression models. The final model contained only variables whose odds ratios were significantly (or close to significantly) associated with failure. Odds ratios and the 95% confidence intervals of the odds ratio are reported, along with significance values.

Additionally, we report the Nagelkerse pseudo R^2 as a measure of percent of variance (in Step 1 failure rate) explained by the model. It is called pseudo because there is no widely accepted direct analog to a linear regression's R^2. Nagelkerke's R^2 is a modification of the Cox and Snell coefficient that yields a result that varies from 0 to 1. Nagelkerke's R^2 will normally be higher than the corresponding Cox and Snell measure but will tend to run lower than the corresponding linear R^2.

Assessment of Multicollinearity. To assess for the presence of multicollinearity, we used the dependent variable from the logistic regression analysis as a dependent variable in a linear regression. The collinearity diagnostic statistics are based on the independent variables only, so the choice of the dependent variable does not matter. We examined Tolerance and the Variance Inflation Factor (VIF) for each variable. Since for each independent variable, Tolerance = $1 - R^2$, where R^2 square is the coefficient of determination for the regression of that variable on all remaining independent variables, low values indicate high multivariate correlation. The VIF is 1/Tolerance, so it is always ≥ 1, and it is the number of times the variance of the corresponding parameter estimate is increased due to multicollinearity as compared to what it would be if there was no multicollinearity. Although there is no formal cutoff value to use with VIF for determining presence of multicollinearity, values of VIF exceeding 10 are often regarded as indicating multicollinearity, but in weaker models, which is often the case in logistic regression, values above 2.5 may be a cause for concern.

Validation of Regression Model

Regression models can be validated or tested against a similar set of data to show that they explain what they seek to explain. One method of validation used is to develop the model on 75% of the data.[7] We therefore ran the technique with a random sample of 75% and then 50% to examine whether the overall model found held in these respective random samples.

Receiver Operating Characteristics. Lastly, a receiver operating characteristic (ROC) curve was examined on those variables found to be significant in the final logistic regression model. The ROC curve and the area under the curve (AUC) provide a graphical plot of the sensitivity versus 1-specificity (false positives) for the binary classifier (pass-fail) as its discrimination threshold is varied.

Results

Forty-three students failed Step 1 of the USMLE on their first attempt and 213 passed. The mean scores obtained for three parts (BS, PS, VR) of the MCAT exam, as well as standard scores for both year 1 and year 2 of medical school, were all significantly lower in the failure group than for those who passed. There was no difference between the two groups in scores on the MCAT WS (Table 1).

Initial Logistic Regression Analysis

Of the six variables entered into the original logistic regression procedure, only two retained statistical significance (year 2 standard score and BS score), and PS was marginally significant. These three variables were retained in the model for further analysis. Table 2 lists the odds ratios and 95% confidence limits for these variables as predictors of failing USMLE Step 1. Year 2 standard scores were the strongest predictor of failure.

Year 2 standard scores ranged from 239.26 to 689.07. Based on the results reported in Table 2, a one-unit change in the year 2 standard score, on the average, increases the odds of failure by 0.976. PS and BS scores range, in single digits, from 4 to 14. A one-unit change in the biological science score, on average, increases the odds of failure by 0.404. A one-unit change in the physical sciences score increases the odds of failure by 0.677.

The Nagelkerse pseudo R^2 was 0.74, indicating that 74% of the variance in failing USMLE Step 1 can be attributed to the three variables in Table 2. The overall predictive accuracy was 94.1%. The model predicted a high percentage of those who passed (97.2%) and most importantly, almost 80% of those who failed.

Table 1

Mean Score Differences on the Independent Variables of Students Passing and Students Failing Step 1 on First Attempt

	Students Passing Step 1	*Students Failing Step 1*	
Independent Variables	*Mean/SD*	*Mean/ SD*	*Mann-Whitney U*
MCAT: Verbal Reasoning	9.14±1.61	7.40±2.26	z=-5.118, *P*=.000
MCAT: Biological Sciences	10.30±1.27	7.67±1.66	z=-8.343, *P*=.000
MCAT: Physical Sciences	9.94±1.56	7.37±1.73	z=-7.727, *P*=.000
MCAT: Writing Sample*	5.32±1.92	4.80±1.78	z=-1.919, *P*=0.06
Year 1 Standard Score	511.27±87.35	398.79±65.33	z=-7.905, *P*=.000
Year 2 Standard Score	514.44±77.96	379.06±55.20	z=-8.769, *P*=.000

MCAT—Medical College Admission Test
SD—standard deviation

* Writing sample converted to numerical value with a range of J=5 and T=15.[8]

Multicollinearity

Analyses for multicollinearity revealed fairly high tolerance values (.625, .595, and .753) for the PS, BS, and year-2 standard scores, respectively, which, in turn, indicates low multicollinearity. VIF values were 1.59, 1.68, and 1.32, respectively—well below a value of 2.5 for even a weak model. We therefore made a determination that adjustments to deal with the presence of multicollinearity were not warranted.

Validation of Regression Model

After repeating the binary logistic regression procedure, using first a 75% random sample and then a 50% random sample of subjects, the same pattern of results were seen. That is, year-2 standard scores were the best predictor of failure, followed by BS and then PS scores (Table 3)

Table 2

Results of Initial Logistic Regression Involving Data from All Subjects

Variable	*Odds Ratio for Failing USMLE Step 1*	*95% Confidence Limits*
Year 2 Standard Score	0.976	0.966–0.986
Physical Sciences MCAT Score	0.677	0.452–1.014
Biological Sciences MCAT Score	0.404	0.241–0.675

USMLE—United States Medical Licensing Examination
MCAT—Medical College Admissions Test

Receiver Operating Characteristics

The area under the ROC curve for the year-2 standard score was very high, with an AUC of 0.92 (95% CI=0.89 to 0.96). Additionally, the BS score had an AUC value of 0.89 (95%CI=0.84 to 0.95).

Because of the large number of cutoff values associated with the standard score, a cross-sample of values that can be used in establishing a single cutoff value are shown in Table 4. A year 2 standard score of 427, for example, had a sensitivity of .791 and a 1-specificity (false positive) of .132. Increasing the standard score cutoff to 436 increases our sensitivity to .860, but also increases our false positives to .165. Similar data are presented in Table 5 for cut-off values for MCAT BS and PS cut-off scores.

Discussion

Based on the outcomes from the logistic regression analysis, we were able to identify two variables, the MCAT BS score and the year-2 standard score, as significantly associated with failure of the USMLE step-1 examination. A third variable (MCAT PS score) had a near-significant association.

With the predictor variables identified, the next step was to define the cutoff values for each of the variables that would allow medical school administrators to benchmark poor student performance. Choosing a cutoff value and the balance between sensitivity and 1-specificity (false positive) is dependent on the available resources in a medical school. Medical schools with abundant resources to work with at-risk students may identify their cutoff scores with a higher sensitivity, thus a higher false positive. This inevitably will increase the pool of at-risk students to monitor and who potentially need interventions. Schools with fewer resources may opt to have cutoff scores with a lower sensitivity, sacrificing their ability to capture the true failures but also decreasing the number of false positives and a smaller number of at-risk students.

Once the cutoff values are identified for the three variables, how are these values integrated into practice to identify students at risk for failing Step 1? The strategy we designed tracks at-risk students from the time of matriculation until the time of the Step 1 exam, using a three-tiered approach (Table 6). Other medical schools need to identify their own institutional model to help guide the prediction and intervention processes students at risk for failing Step 1.

Table 3

Results of the Logistic Regression Using Random Sample of 75% and 50% of Subjects

Variable	*Odds Ratio (95% CI) for Failing USMLE Step 1*	
	75% Sample	*50% Sample*
Year 2 Standard Score	0.976 (0.966–0.986)	0.983 (0.972–0.994)
Physical Sciences MCAT Score	0.686 (0.445–1.058)	0.664 (0.338–1.136)
Biological Sciences MCAT Score	0.444 (0.261–0.758)	0.533 (0.298–0.952)

USMLE—United States Medical Licensing Examination
MCAT—Medical College Admissions Test
CI—confidence interval

Table 4

Cross-Sample of the Coordinates of the Receiver Operating Characteristics Curve for Year-2 Standard Cut-off Scores

Year 2 Standard Cutoff Scores	*Sensitivity (True Positives)*	*1-Specificity (False Positives)*
384	0.465	0.047
388	0.535	0.047
395	0.651	0.066
405	0.674	0.108
406	0.721	0.108
422	0.744	0.132
427	0.791	0.132
431	0.814	0.156
436	0.860	0.165
440	0.884	0.175
448	0.884	0.203
450	0.907	0.203
462	0.953	0.259
476	0.953	0.316
479	1.000	0.325

Limitations

Although our analysis found that these three variables are associated with failure, caution should be exercised in the interpretation of these results—for two important reasons. First, we used our original sample to test the accuracy of the model, so it is possible that our study overestimates the ability of the model to correctly classify a new observation.

Second, the strongest predictor of Step 1 failure was a poor standard score, but standard scores are based on course performance and examinations scores at our medical school. These scores are potentially unique to our medical school, and other schools may award grades and test scores differently. Thus, there may be differences in the ability of course scores to predict USMLE failure.

Conclusions

Using internal (course-related) and external (MCAT score) predictors, it is possible to identify students at risk for failing Step 1 of the USMLE. Medical schools may wish to focus resources and interventions on students at risk of Step 1 failure.

Table 5

Coordinates of the Receiver Operating Characteristics Curve for MCAT Biological and Physical Sciences Cutoff Scores

MCAT Cut-off Score Range	*MCAT Biological Science*		*MCAT Physical Science*	
	Sensitivity (True Positives)	*1-Specificity (False Positives)*	*Sensitivity (True Positives)*	*1-Specificity (False Positives)*
4.50	0.023	0.000	0.023	0.000
5.50	0.116	0.000	0.093	0.000
6.50	0.256	0.005	0.326	0.019
7.50	0.419	0.009	0.605	0.042
8.50	0.698	0.052	0.767	0.184
9.50	0.814	0.241	0.907	0.368
10.50	1.000	0.627	0.977	0.670
11.50	1.000	0.825	0.977	0.844
12.50	1.000	0.943	1.000	0.943
13.50	1.000	0.995	1.000	0.991

MCAT—Medical College Admissions Tests

Table 6

Description of At-Risk Groups and Current Associated Educational Intervention Strategies

At-risk Level	*Factors*	*Educational Intervention*
Risk Level 1	Does not meet MCAT Basic Science cut-off value OR Does not meet MCAT Physical Science cut-off value	If student does not meet MCAT Basic or Physical Science cut-off values internal tracking and monitoring of progress during year 1. If student meets the MCAT cutoff values, but remains in the lower 25th percentile of the class for the first 6 months, internal tracking and offer of academic support services (tutoring and lunchtime reviews).
Risk Level 2	Does not meet MCAT Basic or Physical Science cut-off values AND Student begins to fail Year-1 exams or course(s)	Students with no course failures but continue to be in the lower 25th percentile for two or more courses will be invited to meet with the academic counselor to discuss academic difficulties and design an educational plan. Students with two exam failures in the first two courses of Basic Science (anatomy and histology) are invited to decelerate their first year into 2 years. If a student fails two courses, they are placed on academic probation and must meet with the academic counselor on a regular basis to design an educational follow-up intervention.
Risk Level 3	Does not meet MCAT Basic or Physical Science cut-off values AND Student fails Year-1 exams or course(s) AND Does not meet the Year-2 standard score cut-off following the completion of the first course	Required to meet with the academic counselor to discuss academic difficulties and design an educational intervention, which would include mandatory academic skills counseling, tutoring, back-to-school-reviews, case studies, 3,500 Question Challenge, as well as other board reviews.

MCAT – Medical College Admissions Tests

Corresponding Author: Address correspondence to Dr Bridge, Wayne State University School of Medicine, Academic and Student Programs, 540 E. Canfield, Room 1206, Detroit, MI 48201. 313-577-5076. Fax: 313-577-1457. pbridge@med.wayne.edu.

References

1. Julian ER. Validity of the Medical College Admission Test for predicting medical school performance. Acad Med 2005;80(10):910-7.
2. Koenig JA, Sireci SG, Wiley A. Evaluating the predictive validity of MCAT scores across diverse applicant groups. Acad Med 1998;73(10):1095-106.
3. Kleshinski J, Khuder SA, Shapiro JI, Gold JP. Impact of preadmission variables on USMLE step 1 and step 2 performance. Adv Health Sci Educ Theory Pract 2007;Nov 7.
4. Donnon T, Paolucci EO, Violato C. The predictive validity of the MCAT for medical school performance and medical board licensing examinations: a meta-analysis of the published research. Acad Med 2007;82(1):100-6.
5. State Medical License Requirements. www.healthcarelicensing.com/state_requirements.htm. Revised June 2006. Accessed September 12, 2008.
6. Sawilowsky SS. Misconceptions leading to choosing the *t* test over the Wilcoxon Mann-Whitney Test for Shift in Location Parameter. Journal of Modern Applied Statistical Methods 2005;4(2):598-600.
7. Lang TA, Secic M. How to report statistics in medicine. Annotated guidelines for authors, editors, and reviewers, second edition. Philadelphia: American College of Physicians, 2006.
8. Association of American Medical Colleges. Medical School Admission Requirements (MSAR) 2009–2010. Washington, DC: Association of American Medical Colleges, 2009:222-3.