

Undergraduate Institutional MCAT Scores as Predictors of USMLE Step 1 Performance

WILLIAM T. BASCO, JR., DAVID P. WAY, GREGORY E. GILBERT, and ANDY HUDSON

Many medical schools consider the academic reputation or “selectivity” of applicants’ undergraduate schools during the admission screening process.^{1,2} Mitchell et al. found in a 1993 survey that medical school admission officers rated “quality of degree-granting undergraduate [institution]” as having “moderate” importance in their selection processes.¹ Medical schools using formal undergraduate selectivity measures do so to compensate for the psychometric inadequacies of college grade-point averages, believing that more meaning can be derived from the GPA if it is attached to a measure of institutional performance (academic rigor) or selectivity (stringent admission standards).

Researchers have reported mixed results on whether formal measures of undergraduate institution selectivity are useful contributors to predicting medical student performance.^{1–8} Adding a measure of selectivity improves prediction over GPA alone.^{4–8} However, when undergraduate GPA and scores on the Medical College Admission Test (MCAT) are used together, the addition of undergraduate selectivity adds little or nothing to prediction.^{1,4–8} Examples of formal measures of selectivity that have been studied include: average SAT scores of all freshmen at an undergraduate school^{1,4–8}; the Barron’s Selector Rating of undergraduate schools^{7,9}; and a classification developed by the Carnegie Foundation for the Advancement of Teaching, based upon the number and type of degrees available and level of federal funding.^{7,10} Use of these measures of undergraduate selectivity did improve prediction over GPA alone, but none improved prediction over GPA combined with MCAT scores.^{1,4,7} Few published data exist on the predictive validity of using average MCAT scores of an institution’s graduates to improve prediction of performance in medical school.

The purpose of this study was to explore the use of institutional MCAT scores (or MCAT scores averaged across all students from an undergraduate institution) as a measure of selectivity in predicting medical school performance.¹¹ While this measure is no doubt a *de facto* measure of undergraduate institutional selectivity, it might also be logically considered a measure of institutional academic rigor or quality. Use of institutional MCAT scores as a measure of undergraduate institution selectivity has been discouraged, but a 1986 survey of medical school admission officers revealed that 31% of responding medical schools employed institution-specific MCAT score averages to adjust the GPAs of applicants.¹² Using data from two medical schools, this study tested the hypothesis that employing MCAT scores aggregated by undergraduate institution as a measure of selectivity improves the prediction of individual students’ performances on the first sitting of the United States Medical Licensing Examination Step 1 (USMLE Step 1).

Method

Subjects consisted of the 1996–1998 matriculants of two publicly funded medical schools, one from the Southeast region of the United States and the other from the Midwest. Data were drawn from the longitudinal applicant and matriculant data sets maintained by both schools. Data were verified and entered at the individual schools, then combined. Because this study concealed student identity and used historical data, the institutional review boards of both institutions granted exemption from informed consent.

Independent variables were matriculants’ undergraduate science

grade-point averages (sciGPAs), and three MCAT scores (Physical Sciences, Biological Sciences, and Verbal Reasoning). The Writing Sample is not employed for admission selection at either medical school and so was not included. It is not clear how many medical schools use sciGPA only, but sciGPA may be a more robust predictor of USMLE Step 1 performance than cumulative undergraduate GPA.⁸

The investigational variables were the average MCAT scores attained by all students from a particular undergraduate institution that sat for the exam between April 1996 and August 1999.¹¹ The fact that the data are based on four-year rolling averages ensures adequate sample size for each undergraduate institution and mitigates temporal trends. The data include the undergraduate institution averages for the MCAT Physical Sciences Scale (UGI-PS), the Biological Sciences Scale (UGI-BS), and the Verbal Reasoning Scale (UGI-VR).

Demographic variables that included medical school; year of medical school matriculation; gender; and minority status—underrepresented minority (URM: African American, Native American, Alaskan Native, and all Hispanic) versus non-underrepresented minority were employed as control variables because previous research has shown the importance of controlling for these factors.^{5–7} Neither matriculation year nor medical school was found to be a significant confounder of USMLE Step 1 scores. Thus, data analyses were completed on the combined data in aggregate.

The dependent variable was the matriculants’ scores from their first sittings for the USMLE Step 1. Step 1 scores were chosen as a measure of medical school performance rather than medical school GPA to improve the ability to generalize the findings of this study to other institutions.

Analyses were completed first with individual school data and then with combined school data. The two approaches yielded similar findings, and results of the combined analyses are reported. To remove the influence of scale, individual MCAT scores and sciGPAs were standardized to means of 100 and standard deviations of 10. Multiple regression with blockwise selection was completed to determine the incremental effects of adding the block of three average institutional MCAT predictors to models already including the block of individual predictors (sciGPA, MCAT Physical Sciences, Biological Sciences, and Verbal Reasoning) while controlling for the relevant demographic variables. Multicollinearity was assessed,¹³ and beta coefficients were adjusted using ridge regression.¹⁴ Cross-validation procedures were employed to estimate the shrinkage that results from using multiple regression for prediction and demonstrated that multiple regression results would be generalizable to a similar sample of students. Descriptive statistics were completed. Correlations with performance on the USMLE Step 1 were adjusted for restriction of range.¹⁵

Results

There were 16,954 applicants and 933 matriculants in the combined data set. Applicants’ average age was 23.9 years (SD = 4.0). A total of 41.9% of the applicants were female and 12.5% were underrepresented minorities (URMs). Means and standard deviations for independent and the dependent variables are shown in Table 1.

Bivariate analyses (Table 1) demonstrated moderate correlations

TABLE 1. Descriptive Statistics and Correlation Matrix* for Investigational Variables and United States Medical Licensing Examination (USMLE) Step 1 Scores, 1996–1998 Applicants (*n* = 16,954) and Matriculants (*n* = 933), Combined Cohort from the Ohio State University College of Medicine and Public Health and the College of Medicine of the Medical University of South Carolina

	Mean (SD)	SciGPA	Individual MCAT			UGI MCAT			USMLE Step 1
			MCAT-PS	MCAT-BS	MCAT-VR	UGI-PS	UGI-BS	UGI-VR	
SciGPA†	3.236 (.516)	1.00	.370	.380	.238	.011	.031	.041	.398
MCAT-PS‡	8.687 (2.19)		1.00	.744	.519	.445	.441	.187	.491
MCAT-BS	8.974 (2.13)			1.00	.557	.427	.440	.189	.553
MCAT-VR	8.449 (2.19)				1.00	.316	.323	.204	.397
UGI-PS§	8.250 (1.12)					1.00	.981	.445	.224
UGI-BS	8.472 (1.11)						1.00	.449	.241
UGI-VR	8.028 (1.77)							1.00	.439
USMLE Step 1¶	214.756 (21.15)								1.00

* For correlation, all variables are standardized to mean of 100 and standard deviation of 10.

† Undergraduate science grade point average.

‡ Individual Medical College Admission Test scores.

§ Average undergraduate institution MCAT scores.

¶ Mean and standard deviation are derived from all matriculants in the cohort having USMLE scores (*n* = 933). Correlations with USMLE Step 1 are adjusted for restriction in range.

between sciGPA and the individual MCAT scores. Additionally, there was moderate correlation between sciGPA and USMLE Step 1 scores. There were substantial correlations between individual MCAT scores and USMLE Step 1 scores, including the individual Verbal Reasoning scores. Correlations between individual MCAT scores and the USMLE Step 1 scores were slightly higher than institutional MCAT scores, in part due to adjustment for restriction in range.

Regression analyses are shown in Table 2. For the model without undergraduate selectivity measures, multicollinearity was observed in MCAT Physical Sciences (MCAT-PS) scores and MCAT Biological Sciences (MCAT-BS) scores. Undergraduate institutional Physical Sciences and undergraduate Biological Sciences also demonstrated multicollinearity in addition to URM status, MCAT-PS scores, and MCAT-BS scores in the model with the selectivity measures. Ridge regression allowed us to make the parameter estimates substantially more accurate, reducing the variation at times by 85%, while introducing only a small amount of bias to the parameter estimates.

The base multiple regression model containing gender, URM status, and SciGPA accounted for 13.9% of the variation in USMLE Step 1 scores ($R = .3763$, with $df = 3, 930$; $F = 51.14$; $p < .0001$).

1. USMLE Step 1 = $67.06671 - 1.80834(\text{gender}) - 4.44266(\text{URM}) + 0.34523(\text{sciGPA})$. When applicant MCAT scores were added to the model, the model explained 29.1% of the variation in USMLE Step 1 scores (R^2 change = .153, with $df = 6, 927$; $F = 64.91$; $p < .0001$).

2. USMLE Step 1 = $2.12889 - .15462(\text{gender}) + .75448(\text{URM}) + .24183(\text{sciGPA}) + .19178(\text{MCAT-PS}) + .17584(\text{MCAT-VR}) + .34235(\text{MCAT-BS})$

Finally, when institutional MCAT scores were added to the predictive model, .94% additional percentage of variation in USMLE Step 1 scores was explained (R^2 change = .0094, with $df = 9, 923$; $F = 45.52$; $p < .0001$; Table 2).

TABLE 2. Standardized Ridge Regression Coefficients (β), *p*-values, and Adjusted R^2 Values for Predicting United States Medical Licensing Examination (USMLE) Step 1 Performance Using Standardized Undergraduate Science Grade-point Average (sciGPA) and Standardized Medical College Admissions Test (MCAT) Scores,* 1996–1998 Matriculants (*n* = 933), Combined Cohort from the Ohio State University College of Medicine and Public Health and the College of Medicine of the Medical University of South Carolina†

USMLE Step 1 Predictor Variables	Predictors‡ without Selectivity Adjustments			Predictors with Selectivity Adjustments		
	β §	<i>p</i>	Adj. R^2	β ¶	<i>p</i>	Adj. R^2
SciGPA	.181	<.001		.199	<.001	
MCAT-PS	.166	<.001		.143	<.001	
MCAT-BS	.235	<.001		.209	<.001	
MCAT-VR	.121	<.001	.291	.106	<.001	
Selectivity adjustments						
UGI-PS				.026	.5168	
UGI-BS				.001	.3167	
UGI-VR				.080	.0141	.301

* Adjusted for gender and minority status.

† All variables are standardized to mean of 100 and standard deviation of 10.

‡ Predictors are: undergraduate science GPA, MCAT Physical Sciences score, MCAT Biological Science score, and MCAT Verbal Reasoning score.

§ Ridge regression constant $c = .10$.

¶ Ridge regression constant $c = .15$.

$$3. \text{USMLE Step 1} = -6.11036 - .24005(\text{gender}) + 1.26509(\text{URM}) + .27677(\text{sciGPA}) + .16979(\text{MCAT-PS}) + .15498(\text{MCAT-VR}) + .31311(\text{MCAT-BS}) + .86329(\text{UGI-PS}) + 1.86368(\text{UGI-VR}) - 1.43293(\text{UGI-BS})$$

Overall, the model multiple correlation coefficients (with control variables) were .54 (GPA and MCAT scores) and .55 (GPA, MCAT scores, and selectivity).

Discussion

Consistent with findings from previous studies, this study demonstrated that undergraduate science GPAs and MCAT scores are strong predictors of standardized test performances during medical school. In contrast to prior studies, this study employed institutional MCAT averages and demonstrated that their inclusion in regression models, as a measure of selectivity, can produce a small improvement when used in a theoretical model in the prediction of a medical student's performance.

While all applicants to an undergraduate school may take the SAT, far fewer will take the MCAT, leading to the observation that average institutional SAT scores would be a more appropriate measure of the overall academic environment to which a medical school applicant was exposed.¹² An additional argument over the use of MCAT as a measure of institutional selectivity is that institutions differ in how they encourage students to take the MCAT (e.g., through academic advising). One could argue that the average institutional MCAT scores may be at least partially influenced by which students are encouraged to take the MCAT, a scenario that would limit the usefulness of the data for those who do take it. On the other hand, it can be argued that the average institutional MCAT scores should be a reflection of the academic rigor or instructional quality of the applicant's undergraduate education. Regardless of how the average institutional MCAT scores are interpreted as a measure of selectivity, a measure of academic rigor, or a measure of educational climate; this study shows it to be a useful addition to the traditional prediction model used for admission.

This study has the limitation of using samples from only two medical schools. However, multiple years of data were used, the schools were from different regions of the country, the analyses were cross-validated, and the overall correlations among the independent and dependent variables are of magnitudes similar to those found in studies using more nationally-representative data.^{1,4,6,8} This study also employed only one measure of medical school performance (USMLE Step 1), albeit one with wide generalizability. It remains to be seen whether adjusting for average undergraduate MCAT scores can help with prediction of performance beyond the basic science portion of medical school. Another study limitation is that because this study employed ridge regression the standardized beta coefficients are now biased estimates. However, when an estimator has only a small amount of bias and is more precise it is often the preferred estimator since it has an increased probability of being close to the true value of the parameter.¹⁴

The improvement in prediction shown in this study was limited, and schools should consider whether this gain in prediction warrants the effort involved in including undergraduate institutional measures. The intent of this study was to demonstrate empirically whether undergraduate institutional measures might contribute to predicting basic science achievement in medical school. Further studies will have to be done to determine how these results can be practically applied. We feel that the results of this study warrant further investigation with larger sample sizes, broader participation by medical schools, and exploration of other regression models. In fact, two of the current authors (DW and AH) have been adjusting for undergraduate selectivity using a similar method since 1995 and have demonstrated an improvement in prediction of roughly 10% when the analysis approach is one that is intended to maximize prediction.

The cost of adding this one element to prediction equations is minimal after the initial "up-front" time of creating the variable. It appears from the beta coefficients presented that it was the UGI-VR score that added the improvement in prediction. However, since the undergraduate mean MCAT scores were entered in block-wise fashion, one can not interpret the individual UGI-score betas.

This study did show a greater correlation between individual MCAT-VR scores and USMLE Step 1 performances than would be expected, and we are not sure how to interpret this finding. In general, previous studies with other data demonstrated less correlation of the MCAT-VR score with performance on a science-laden exam such as USMLE Step 1.⁸ Our data suggest that perhaps schools should pay more attention to the Verbal Reasoning section score in the selection process, particularly if it has better correlation with performance in the later years of medical school.

There are other cautions to consider when medical schools adjust for undergraduate selectivity. Because undergraduate institutional measures are likely to be a reflection of an undergraduate institution's selectivity, any use of these measures may have unintended effects upon the mix of applicants who are offered interviews or admission. Medical schools should decide a priori how much a selectivity adjustment should affect the admission process, and medical schools should evaluate how applying selectivity adjustments would affect the relative ranking of applicants the school otherwise feels should have priority (e.g., applicants from rural or urban underserved areas).

In summary, adding a measure of an applicant's undergraduate institutional selectivity by employing the average MCAT score of applicants from that institution can produce a small improvement in prediction of performance on USMLE Step 1.

The authors gratefully thank Dr. Judith Westman and Dr. Mark Notestine for their support of this project. Also, Ms. Rebecca Hasegawa, Ms. Sandra Morris, and Ms. Wanda Taylor contributed to data collection, data entry, and logistical support. The authors also express their gratitude to Dr. Ellen Julian, Assistant Vice President and Director of the Medical College Admission Test, Association of American Medical Colleges, for her support and advice throughout this project.

Correspondence: William T. Basco, Jr., MD, Division of General Pediatrics, Medical University of South Carolina, 326 Calhoun Street, P.O. Box 250106, Charleston, SC 29425; e-mail: (bascob@musc.edu).

References

1. Mitchell K, Haynes R, Koenig J. Assessing the validity of the updated Medical College Admission Test. *Acad Med.* 1994;69:394-401.
2. Hall F. Association of American Medical Colleges, Washington, DC. Electronic communication, February 2000.
3. Zelesnik C, Hojat M, Veloski JJ. Predictive validity of the MCAT as a function of undergraduate institution. *J Med Educ.* 1987;62:163-9.
4. Wiley A, Koenig J. The validity of the Medical College Admission Test for predicting performance in the first two years of medical school. *Acad Med.* 1996;71(10 suppl):S83-S85.
5. Huff KL, Fang D. When are students most at risk of encountering academic difficulty? A study of the 1992 matriculants to U.S. medical schools. *Acad Med.* 1999;74:454-60.
6. Koenig JA, Sireci SG, Wiley A. Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Acad Med.* 1998;73:1095-106.
7. Blue AV, Gilbert GE, Elam CL, Basco WT Jr. Does institutional selectivity aid in the prediction of medical school performance? *Acad Med.* 2000;75(10 suppl):S31-S33.
8. Swanson DB, Case SM, Koenig J, Killian CD. Preliminary study of the accuracies of the old and new Medical College Admission Tests for predicting performance on USMLE Step 1. *Acad Med.* 1996;71(1 suppl):S25-S27.
9. Barron's Profiles of American Colleges. 23rd ed. Hauppauge, NY: Barron's Educational Series, Inc., 1998.
10. Boyer E. A Classification of Institutions of Higher Education. Pittsburgh, PA: The Carnegie Foundation for the Advancement of Teaching, 1994.
11. Medical College Admission Test, Summary of scores by undergraduate college

- attended, April, 1995–August, 1998. Washington, DC: Association of American Medical Colleges, 1999.
12. Mitchell KJ. Use of MCAT data in selecting students for admission to medical school. *J Med Educ.* 1987;62:871–9.
 13. Freund RJ, Littell RC. *SAS System for Regression.* Cary, NC: SAS Institute Inc., 1986.
 14. Neter J, Wasserman W, Kutner MH. *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs.* 2nd ed. Homewood, IL: R. D. Irwin, 1985.
 15. Cohen J, Cohen P. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.