JOINING THE CONVERSATION: Predictors of Success on the United States Medical Licensing Examinations (USMLE)

Editor Note:

This empirical study builds on the following previously published TLAR article: Gandy, R.A., Herial, N.A., Khuder, S.A., & Metting, P.J. (2008). Use of curricular and extracurricular assessments to predict performance on the United States Medical Licensing Examination (USMLE) Step 1: A multi-year study. *Learning Assistance Review (TLAR), 13*(2), 27-35.

Sabry Gohara, Joseph I. Shapiro, Adam N. Jacob, Sadik A. Khuder, Robyn A. Gandy, Patricia J. Metting, Jeffrey Gold, and James Kleshinski
The University of Toledo

Author Note: Robyn Gandy is now at Department of Psychology, Lourdes College

Abstract

The purpose of this study was to evaluate whether models based on pre-admission testing, including performance on the Medical College Admission Test (MCAT), performance on required courses in the medical school curriculum, or a combination of both could accurately predict performance of medical students on the United States Medical Licensing Examination (USMLE) Steps 1 and 2. Models were produced using stepwise linear regression and feed forward neural networks. Notable accuracy in predicting Step 1 and Step 2 scores were achieved from models integrating pre-admission variables with medical school coursework grades. Of interest, the coursework grades contributed far greater to these models than the pre-admission variables except the MCAT.

Key Words: medical school admissions, neural network, regression analysis, USMLE, medical education.

For more information contact: Sabry Gohara, M.D. | Department of Surgery | The University of Toledo Health Science Campus, Mail Stop 1095 | 3000 Arlington Ave. | Toledo, Ohio 43614 | Phone: (419) 383-6021 | Fax: (419) 383-6636 | Email: Sabry.Gohara@utoledo.edu

It is critical for medical schools to graduate students who successfully obtain licensure to practice, and in most states, licensure requires the passing of all three steps of the United States Medical Licensing Examination (USMLE). USMLE Step 1 has traditionally focused on pre-clinical studies. Most medical schools require students to successfully pass the USMLE Step 1 before starting their clinical clerkships. USMLE Step 2 is typically taken in the fourth year of medical school. Passing the USMLE Step 2 is often mandatory for graduation. Because of its clinical content, the USMLE Step 2 provides a measure of the students' clinical competence and their ability to participate safely in patient care under supervision. Identifying students at risk for failing the USMLE, therefore, becomes, to a certain extent, a matter of patient safety. As such, the performance of medical students on the USMLE is not only important for the individual student but, in aggregate, is an important metric for medical schools to track.

Peterson and Tucker reported that performance during gross anatomy was a good predictor of performance on the USMLE Step 1 (Peterson & Tucker, 2005). Others have also investigated the relationship between pre-admission variables and coursework performance and its correlation with students' performance on the medical licensing exams (Roth, Riley, Brandt, & Seibel, 1997; Julian, 2005; DeChamplain, Sample, Dillon, & Boulet, 2006; Donnon, Paolucci, & Violato, 2007). Two previous publications from The University of Toledo College of Medicine, in particular, examined relationships between student performance and USMLE scores (Gandy, Herial, Khuder, & Metting, 2008; Kleshinski, Khuder, Shapiro, & Gold, 2009). One of these studies evaluated the ability of three metrics, two of them being curricular and the third being extracurricular, in predicting student performance on the USMLE Step 1 exam (Gandy et al., 2008). The curricular metrics evaluated in this previous study included the final scores each student received for two of the College of Medicine's preclinical courses, namely Human Structure and Development and Organ Systems. The extracurricular metric was each student's score on the Comprehensive Basic Science Examination (CBSE). The study concluded that the two preclinical courses were good predictors of student performance on USMLE, and their value as identifiers of students at risk was promising. The second previous study used many pre-admission variables to predict USMLE Step 1 and Step 2 performance (Kleshinski et al., 2009). Dependent variables included gender, race, age, selectivity of the undergraduate institution attended, undergraduate major, total GPA, science GPA, post-baccalaureate degrees earned, Medical College Admissions Test (MCAT) scores, parents' occupation, and scores on USMLE Step 1 and Step 2 (Clinical Knowledge). This study found that statistically significant predictors for step1 and step 2 included age of the applicant, race, college selectivity, science grade point average, and the biologic science section of the MCAT. It was also noted in this report that a feed forward neural network could improve on a linear model in terms of predicting USMLE scores based on pre-admission variables. Neural networks are mathematical models that combine information in a nonlinear way and are particularly good at identifying patterns within large datasets.

These two prior studies formed the framework for the current study. In order to improve on the ability to predict USMLE performance, relationships were examined further on a larger set of students, exploring the relation between the students' performance in the College of Medicine's five pre-

clinical basic science curricular blocks (year 1 and 2) and seven required clinical rotations and their performance on the USMLE Step 2. (Parent occupation was not examined in this study as no significant relationship was found in the prior study.) One aim was to try to identify which students were likely to have poor performance on the USMLE.

Methods

Institutional Review Board (IRB) approval was obtained to review data from students of entering years 1998 through 2005 at The University of Toledo College of Medicine. Based on previous work, the dependent pre-admission variables of gender, race, age, selectivity of the undergraduate institution attended, undergraduate major, total grade point average (TGPA), science grade point average (SGPA), highest degree earned, MCAT scores, and scores on USMLE Step 1 and Step 2 (Clinical Knowledge) were chosen.

The curricular measures evaluated in this study include the final grade achieved in the five pre-clinical basic science curricular blocks of Cell and Molecular Biology, Human Structure and Development, and Neuroscience/ Behavioral Science in the first year curriculum, and Immunity and Infection and Organ Systems in the second year curriculum. The final grades of the student in the six required third year clinical clerkships of Family Medicine, Internal Medicine, Pediatrics, Obstetrics/Gynecology, Surgery, and Psychiatry along with the fourth year Neurology clerkship (required for graduation) were also considered. All of the clinical departments used three criteria to determine final grades. These were the National Board of Medical Examiners (NBME) "shelf exam" (which contributed 40% of the grade), the Student Clinical Competency Evaluation (40% of the grade) was provided by residents and attending physicians who had adequate exposure to the student during the rotation period, and an additional group of measures collectively called the Departmental Educational Program (e.g., case and procedure progress logs, objective standardized clinical examinations [OSCE], oral examinations, clinical vignettes, assigned readings, ethics consultation essays, quizzes before and after didactic lectures, computer assisted learning assignments, and attendance) which constituted 20% of the final grade. The final grades were determined as "Honors," "High Pass," "Pass," and "Fail." On average, about 15% of the class receives "Honors" and about 35% receives "High Pass" with the vast majority of the remaining receiving "Pass" in each of the clinical clerkships.

The outcome variables for this study were the scores on USMLE Step 1 and Step 2 (Clinical Knowledge). Only those students with complete records available in the spring of 2008, including Step 1 and Step 2 scores, were used in this analysis (816 total records). Differences in Step 1 or Step 2 scores by demographic variable were compared using a t-test or ANOVA. Multiple regression models were used to identify predictors of Step 1 and Step 2 scores. A stepwise selection procedure was used to identify the best score predictors. All analyses were carried out using Matlab™. In addition, examination was done to see if there was improvement on the linear model with a feed forward neural network. This was done either with the entire student cohort chosen for the training of the neural network or using a portion chosen by random to form the training set with the residual of students used for the testing set. For students who took the USMLE more

than once, only the first-time USMLE scores were used. For the neural network models, the architectural design was varied (e.g. number of hidden neurons ranged from 1-10), number of layers of hidden neurons (between 1-3 layers), training algorithms (Levenberg-Marquardt [LM], gradient descent [GDX] and Bayesian regularization [BR] methods) as well as the following transfer functions: two logistic and one linear transfer functions (Kleshinski et al., 2009; Hagan & Manaj, 1994; Foresee & Hagan, 1997). The quality of predictions on the testing set of data was compared by the fraction of total variance which was fit by the model. The neural network training was performed on the entirety of the dataset, replicating this process 10 times with each number of neurons, each training algorithm, and the different transfer functions. The best fit obtained with these 10 replicants for the different number of neurons is reported. The accuracy of predictions from these models were then compared to those obtained using stepwise regression models. P values are reported with $p<0.001$ substituted for zero.

Results

The records from a total of 816 students were included. Of these, 64% were male and 79% were white, 1.8% African-American, 2.3% Hispanic, 9.3% Indian-Pakistani, and 7.4% Asian. Sixty-four percent were 22 years old or younger upon matriculation, and 16.6% were from the most selective undergraduate institutions using Peterson's Four-Year Colleges 2008 as the reference (Oram, 2007). For undergraduate major, 71.3% were from wet science majors (including biology, zoology, chemistry, and pre-medicine), 10.8% psychology majors, 7.2% from dry science majors (including engineering, math, computer science, and physics), 6.9% liberal arts majors, and 3.8% business majors.

The stepwise regression model based on demographics and preadmission variables could predict approximately 17.4% and 13.0% of the variation in USMLE Step 1 and Step 2 scores, respectively. Using a feed forward neural network, some improvements on this performance could be made. First, there were fairly similar results from the neural network models regardless of whether the Levenberg-Marquardt (LM), gradient descent (GDX), or Bayesian regularization (BR) methods were used; similarly, both the logsig and tansig transfer functions yielded virtually identical results (data not shown). With all of the different training algorithms applied on the pre-admission variables, a maximum R2 of about 21% could be achieved with the best results somewhere between 5-9 neurons for predicting USMLE Step 1 scores. As far as predicting USMLE Step 2 scores, again the different algorithms gave similar results with maximal predictions occurring between 5-9 neurons. Because previous experiences indicated superior generalization with the Bayesian regularization method, focus was placed on this approach for subsequent predictions. Increasing the number of hidden layers from one to three did not improve on the results obtained with the single hidden layer model.

Using a stepwise regression model, USMLE scores could be much better predicted if the grades that students obtained in the courses taken before the USMLE were added. Specifically, adding the pre-clinical course grades to the demographics and pre-admission variables improved the R2 achieved with USMLE Step 1 to 57.3% (Table 1), and adding all required clinical

courses and USMLE Step 1 scores improved the R2 achieved with USMLE Step 2 results to 64.7% (Table 2). For Tables 1 and 2, variables with a p value less than 0.05 were not included in the model. Using a neural network approach, only a trivial improvement in predictive capability was obtained for both USMLE Step 1 and Step 2 predictions. Specifically, the prediction of USMLE Step 1 scores could be increased using a neural network based on pre-admission variables and pre-clinical course work to an R2 of about 59%, whereas a neural network based on these data plus USMLE Step 1 scores plus all coursework prior to USMLE Step 2, including clinical rotations, allowed for the prediction of about 68% of the variance in USMLE Step 2 scores. The large improvement in predictive ability with the addition of coursework grades led next to examing whether coursework grades alone could predict USMLE Step 1 and Step 2 scores. Interestingly, using stepwise regression performed on the preclinical course grades alone, 53.1% of the variance in USMLE scores could be predicted. Only the Immunity and Infection curricular block scores dropped out of the model with stepwise regression. It is noted that this is only 4.5% less accurate than that seen with the model including the preadmission variables. Applying the neural network model to this group of variables, the R2 achieved could only be increased to 54.8%. These same pre-clinical course grades could also predict 40.5% of the variance in USMLE Step 2 scores with the stepwise regression model and 41.8% of the variance in USMLE scores with a neural network model.

Table 1: Stepwise Linear Regression Predicting USMLE Step 1 Score

Variable	Coefficient	Std.Err.	p-Value
Sex	-1.59	1.04	0.128
Age	-0.25	0.16	0.127
Race	-0.16	0.29	0.563
Undergraduate College Selectivity	-0.13	0.46	0.785
Undergraduate Major	-0.53	0.51	0.303
Highest Degree Earned	0.34	0.76	0.655
TGPA	-1.71	1.71	0.317
SGPA	-0.66	1.46	0.650
Verbal Reasoning	0.63	0.32	0.048
Physical Science	1.21	0.33	<0.001
Writing Sample	-0.21	0.25	0.389
Biologic Science	2.31	0.41	<0.001
Cell & Molecular Biology Block	-4.54	0.96	<0.001
Human Structure & Development Block	-4.75	0.98	<0.001
Neuroscience/Behavioral Science Block	-2.89	0.88	0.001
Immunity & Infection Block	-1.40	1.00	0.160
Organ Systems Block	-11.48	0.93	<0.001

Note: Overall R^2 = 57.3245%

Table 2:

Stepwise Linear Regression Predicting USMLE Step 2 Score

Variable	Coefficient	Std.Err.	p-Value
Sex	0.94	1.01	0.351
Age	-0.41	0.17	0.016
Race	-0.07	0.28	0.806
Undergraduate College Selectivity	-0.69	0.49	0.153
Undergraduate Major	1.11	0.49	0.025
Highest Degree Earned	2.70	0.79	0.001
TGPA	3.98	1.72	0.021
SGPA	-0.73	2.38	0.759
Verbal Reasoning	0.49	0.31	0.116
Physical Science	0.13	0.31	0.668
Writing Sample	-0.17	0.24	0.471
Biologic Science	-0.05	0.38	0.890
Cell & Molecular Biology Block	0.16	0.94	0.865
Human Structure & Development Block	2.29	0.93	0.014
Neuroscience/Behavioral Science Block	-1.85	0.86	0.032
Immunity & Infection Block	1.57	0.99	0.113
Organ Systems Block	-2.48	1.00	0.013
Family Medicine	-4.50	0.73	<0.001
Internal Medicine	-2.57	0.71	<0.001
Neurology	-3.04	0.75	<0.001
Obstetrics & Gynecology	-4.52	0.67	<0.001
Pediatrics	-2.31	0.73	0.002
Psychiatry	-1.10	0.69	0.112
Surgery	0.01	0.70	0.984
USMLE Step 1 Score	0.36	0.03	<0.001

Note: Overall R^2= 64.735%

To further test this concept, students were sorted based on the results in the biological science portion of the MCAT, the single best pre-admission predictor of USMLE Step 1 performance and the lowest 10th percentile. When USMLE Step 1 and Step 2 scores were reviewed in this subset, the averages were 204 and 215, respectively. Next, it was examined how well USMLE Step 1 and Step 2 scores in this subset could be predicted with linear and neural network models. Similarly to the case for the overall dataset, 58.5% and 43.1% of the variance could be predicted in USMLE Step 1 and Step 2 scores, respectively, with the stepwise regression model. Significant improvements on these predictions by using a neural network approach on this subset could not be made.

Discussion

A three-step process comprising standardized tests is used to grant license to practice medicine in the United States. The first step (USMLE Step 1) tests basic pre-clinical science, while the second (USMLE Step 2) addresses clinical application of medical knowledge, specifically the competency to participate in patient care under supervision in postgraduate training programs. Passing the third step (USMLE Step 3) allows for the unsupervised practice of medicine.

In a previous report from our institution, Gandy and colleagues identified the strong correlation of some coursework to USMLE Step 1 scores (Gandy et al., 2008). In this current report, it was noted that performance of students in medical school was a far better predictor of USMLE results than pre-admission variables except the MCAT. In fact, addition of the pre-admission variables added very little to the predictive capacity of the multiple logistic regression model compared with a model containing only coursework performance. It was also noted that while the neural network was still better than the linear model with preadmission variables (as reported in earlier work), once course performance was included, most of the advantages of the neural network approach over the linear approach which had been previously noted appeared to be lost. One explanation for this may be the overriding effect of the coursework on the neural network model or possibly the architectural design and transfer functions that were used. This current study has considerable implications. First, perhaps mandated visits to the UT Health Science Campus learning center, The Academic Enrichment Center (AEC), should be implemented earlier during the medical school curriculum for students who are struggling. Having board review courses starting in the first rather than second year, small group supplemental instruction, and required completion of board-like practice tests could also be of benefit. Findings from this study will be presented to the Senior Leadership Team of the College of Medicine as well as to members of the College of Medicine Executive Curriculum Committee to discuss other potential academic interventions.

Additionally, a number of medical schools around the country have Master degrees or other post-baccalaureate programs geared toward students whose ultimate goal is to gain entry into medical school. Performance in these programs may assist schools in better selecting those students likely to be successful because the curriculum in these programs often contains some of the same courses that the medical students are taking. Another possibility would be to consider examining entry criteria to medical school and instituting a pyramidal process within the pre-clinical years to identify those who will perform successfully. While this would give additional students an opportunity to matriculate into medical school, it would obviously have tremendous financial implications to those students who did not progress through completion. Suffice it to say that movement towards this pyramidal system would constitute a paradigm shift for U.S. medical schools (Barzansky & Etsel, 2009).

Future Study:

Although this study did not specifically address the impact of the institution's learning assistance center, The Academic Enrichment Center (AEC), during the time of the study, the center now has the ability to collect specific data on its ability to provide academic support specifically to the College of Medicine. According to an internal AEC report (2010), between June 2008 to October 2010, the AEC provided assistance to 654 distinct medical students, resulting in 4,960 tutoring sessions, totaling 10,130 hours of support. Consequently, the College of Medicine support resulted in 61.4% of the total academic assistance provided for the entire University of Toledo Health Science Campus. Specifically, within that overall percentage, 43.4% was service to students in the first two years (in preparation for USMLE Step 1) and 17.7% for students in the second two years (in preparation for USMLE Step 2). Furthermore, between 2008 and 2010, the center provided approximately160 hours of help in study skills, 90 hours of drop-in tutoring for content questions, 1,170 hours of group tutoring, 1,240 hours of individual tutoring, and 5,390 hours of Supplemental Instruction in addition to 2,300 hours of USMLE Step 1 review, and 230 hours of USMLE Step 2 review. A future study could look at that impact of the learning center providing opportunity for supplemental instruction, study techniques, and content assistance that impacts coursework grades.

The National Board of Medical Examiners (NBME) provides a content outline for material presented on the USMLE Step 1 and Step 2 examinations (Federation of State Medical Boards and National Board of Medical Examiners, 2010). This content outline assists medical schools as they evaluate and refine their respective curriculums. Though the approach may vary among medical schools with regard to pedagogy, the five preclinical blocks and clinical clerkships examined in this study likely represent the same overall content provided at medical schools across the country in preparation for the USMLE examinations. Therefore, our results could be applicable to all medical schools given the above assumptions.

Conclusion

In summary, it was found that notable accuracy in predicting USMLE Step 1 and Step 2 scores could be achieved from models integrating pre-admission variables with coursework grades from medical students. Of interest, the coursework grades contributed far more to these models than the pre-admission variables except the MCAT.

Acknowledgements

The authors would like to thank Ms. Diane Pfaff, Educational Coordinator, Registrar's Office and Mr. John Cavins, Director of Academic Computing, for assistance in retrieving data on course grades and USMLE scores used in this study.

References

Academic Enrichment Center (2010, October 22). Statistics of students' use: Academic Enrichment Center UTHSC 6-8-2008 to 10-22-2010. Internal report AEC, UTHSC.

Barzansky, B., & Etzel, S.I. (2009). Medical schools in the United States, 2008-2009. *Journal of the American Medical Association*,*302*(12), 1349-1355.

De Champlain, A., Sample, L., Dillon, G.F., & Boulet, J.R. (2006). Modeling longitudinal performances on the United States Medical Licensing Examination and the impact of sociodemographic covariates: An application of survival data analysis. *Academy of Medicine*, *81*(10 Supl), S108-111.

Donnon, T., Paolucci, E.O., & Violato, C. (2007). The predictive validity of the MCAT for medical school performance and medical board licensing examinations: A meta-analysis of the published research. *Academy of Medicine*, *82*(1), 100-106.

Federation of State Medical Boards and National Board of Medical Examiners (2010). *United States Medical Licensure Examination content description*. Retrieved November 11, 2010, from http://www.usmle.org.

Foresee, F.D., & Hagan, M.T. (1997). Gauss-Newton approximation to Bayesian regularization. *Proceedings of the 1997 International Joint Conference on Neural Networks*, 1930-1935.

Gandy, R.A., Herial, N.A., Khuder, S.A., & Metting, P.J. (2008). Use of curricular and extracurricular assessments to predict performance on the United States Medical Licensing Examination (USMLE) Step 1: A multi-year study. *Learning Assistance Review*, *13*(2), 27-35.

Hagan, M.T., & Menhaj, M. (1994). Training feed-forward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, *5*, 989-993.

Julian, E.R. (2005). Validity of the Medical College Admission Test for predicting medical school performance. *Academy of Medicine*, *80*(10), 910-917.

Kleshinski, J., Khuder, S.A., Shapiro, J.I., & Gold, J,P. (2009). Impact of preadmission variables on USMLE step 1 and step 2 performance. *Advanced Health Science Education and Practice*, *14*(1), 69-78.

Oram, F.A. (Ed.) (2007). *Peterson's four-year colleges 2008*. Lawrenceville, NJ: Peterson's/Nelnet.

Peterson, C.A., & Tucker, R.P. (2005). Medical gross anatomy as a predictor of performance on the USMLE Step 1. *The Anatomical Record Part B: The New Anatomist, 283*(1), 5-8.

Roth, K.S., Riley, W.T., Brandt, R.B., & Seibel, H.R. (1996). Prediction of students' USMLE step 2 performances based on premedical credentials related to verbal skills. *Academic Medicine, 71*(2), 176-80.