Adv in Health Sci Educ (2016) 21:389–399
DOI 10.1007/s10459-015-9637-6

Relationships between preclinical course grades and standardized exam performance

Yinin Hu[1] · James R. Martindale[2] · Robin D. LeGallo[3] · Casey B. White[2] · Eugene D. McGahren[1] · Anneke T. Schroen[1]

Received: 7 February 2015 / Accepted: 7 September 2015 / Published online: 12 September 2015

Abstract Success in residency matching is largely contingent upon standardized exam scores. Identifying predictors of standardized exam performance could promote primary intervention and lead to design insights for preclinical courses. We hypothesized that clinically relevant courses with an emphasis on higher-order cognitive understanding are most strongly associated with performance on United States Medical Licensing Examination Step exams and National Board of Medical Examiners clinical subject exams. Academic data from students between 2007 and 2012 were collected. Preclinical course scores and standardized exam scores were used for statistical modeling with multiple linear regression. Preclinical courses were categorized as having either a basic science or a clinical knowledge focus. Medical College Admissions Test scores were included as an additional predictive variable. The study sample comprised 795 graduating medical students. Median score on Step 1 was 234 (interquartile range 219–245.5), and 10.2 % (81/795) scored lower than one standard deviation below the national average (205). Pathology course score was the strongest predictor of performance on all clinical subject exams and Step exams, outperforming the Medical College Admissions Test in strength of association. Using Pathology score <75 as a screening metric for Step 1 score <205 results in sensitivity and specificity of 37 and 97 %, respectively, and a likelihood ratio of 11.9. Performance in Pathology, a clinically relevant course with case-based learning, is significantly related to subsequent performance on standardized exams. Multiple linear regression is useful for identifying courses that have potential as risk stratifiers.

✉ Yinin Hu
yh9b@virginia.edu

1 Department of Surgery, University of Virginia School of Medicine, P.O. Box 800679, Charlottesville, VA 22908-0709, USA

2 Undergraduate Medical Education, University of Virginia School of Medicine, Charlottesville, VA, USA

3 Department of Pathology, University of Virginia School of Medicine, Charlottesville, VA, USA

Keywords Medical student education · Pathology · Standardized testing · Risk-stratification

Introduction

Medical schools and residency programs are universally challenged with evaluating medical students' fund of knowledge through impartial examination (Raymond et al. 2011). Due to variability in scoring schemes for preclinical coursework and clinical clerkships, the most consistent metrics for academic achievement are the United States Medical Licensing Examination (USMLE) Step exams and the National Board of Medical Examiners (NBME) subject examinations ("shelf exams"). While USMLE Step 1 assesses basic sciences with some clinical context, Step 2 and NBME clinical subject exams focus on clinical understanding. Immense efforts by educators and students are devoted toward optimizing performance on these exams, as they comprise some of the primary academic criteria for residency selection (Torre et al. 2009; Briscoe et al. 2009; Green et al. 2009).

Numerous studies have scrutinized early predictors of standardized exam performance. Moderate correlations have been identified between USMLE Step 1 and the Medical College Admissions Test (MCAT), demographic factors, and undergraduate grade point average (Basco et al. 2002; Julian 2005; Koenig et al. 1998; Hojat et al. 2000; Donnon et al. 2007; Wiley and Koenig 1996). Performance on the Step 2 Clinical Knowledge (CK) exam has been associated with clerkship curricular format, USMLE Step 1, the MCAT, and shelf exams (Case et al. 1996; Ripkey et al. 1997, 1999). The consistent associations between the MCAT and high-stakes medical exams are intriguing, as the MCAT has traditionally comprised basic science content (physics, chemistry, biology) that is preliminary to clinical knowledge. While there are proponents who endorse that basic science knowledge and clinical knowledge comprise two largely separate entities (Patel and Groen 1986), Schmidt's encapsulation theory suggests that biomedical knowledge may be integral to clinical reasoning even if its contributions are at a sub-conscious level (Rikers et al. 2004). A second explanation may be that the question styles within the MCAT and higher-level medical standardized exams are similar. High-stakes medical standardized exams often emphasize application of knowledge, a middle-tier cognitive function within the context of Bloom's taxonomy of cognitive domains (USMLE 2013). In contrast, medical school assessments have been criticized for over-reliance on lower-level cognitive skills such as knowledge recollection (Cooke et al. 2006). Zheng and colleagues demonstrated that questions targeting higher-level cognitive functions are more common in the MCAT than in exams at the undergraduate and first-year medical school level (Zheng et al. 2008). Logically, medical school courses with clinical relevancy that are evaluated using higher-order question formats should have the strongest predictive capacity for USMLE and NBME exam performance.

Due to variability in preclinical curricula across medical schools, relatively few studies have assessed the relationships between preclinical coursework and standardized exam performance. However, identifying such relationships can identify students at early risk for low exam scores and also assess the quality of preclinical teaching. Cumulative performance during preclinical years has shown promise as a risk-stratifier for Step 1 failure (Coumarbatch et al. 2010). By identifying specific preclinical courses that predict standardized exam performance, educators can offer early, targeted remediation tailored to their curriculum. Furthermore, poorly predictive courses can be revised to improve relevancy. The goal of this study is to demonstrate the utility of regression modeling in identifying preclinical coursework that predicts performance on NBME and USMLE exams. We hypothesized that courses that are clinically relevant and emphasize higher-order cognitive questioning would have the highest predictive capacities.

Methods

A retrospective review was conducted of medical students at the University of Virginia who had completed USMLE Step 1 and Step 2 records between 2007 and 2012. Students who did not complete the full spectrum of preclinical courses were excluded. A prospectively collected academic database was de-identified and queried for USMLE Step 1, USMLE Step 2, and NBME shelf scores in the following subject areas: Internal Medicine, Pediatrics, Neurology, Obstetrics and Gynecology, Surgery, Psychiatry, and Family Medicine. Institutional records for preclinical course grades—each on a 100-point scoring system—were similarly assessed through the same database. During this time period, the preclinical curriculum comprised didactic and group-based courses distributed over the first 2 years of medical school training. Students' MCAT scores were made available through the Office of Admissions, and were linked to the remaining academic records. The MCAT writing sample sub-score was excluded, as it has been shown to have no correlation with medical school performance or USMLE Step exam results (Hojat et al. 2000). To preserve confidentiality, demographic data were not available for each student, but summary statistics were available for each academic class.

Linear multiple regression analyses were used to identify relationships between preclinical course scores and standardized exam scores. Prior research has shown that MCAT performance is associated with performance on standardized clinical exams (Julian 2005; Ripkey et al. 1997). Therefore, to control for this known predictive variable, the array of independent variables included the preclinical course scores and individual composite MCAT scores. The preclinical curriculum was divided by subject matter into basic science courses and clinical knowledge courses. Basic science courses included gross anatomy, cell biology, molecular genetics, biochemistry, neuroscience, physiology, and microbiology. Clinical knowledge courses included epidemiology, human behavior, psychiatry, pathology, pharmacology, and Practice of Medicine (POM) parts 1 and 2. The POM course series addressed the approach to clinical scenarios through case studies and physical exam practical sessions, and was taught in a small-group format. The dependent variables were scores on the NBME subject exams and the USMLE Step exams. Linear regression models were created to assess for significant predictors of each dependent variable using an all-subsets model selection protocol. Through this protocol, all possible combinations of independent variables are used to predict a given dependent variable. The highest-performing model is identified, and irrelevant independent variables are discarded from the model. Among variables that remain in each model, positive regression coefficients indicate that a higher preclinical course grade is associated with a higher standardized exam score, while a negative coefficient indicates the opposite association.

Students whose USMLE Step 1 score was 205 or less were considered poor performers. Similar thresholds of performance were established for USMLE Step 2 at 212, and for NBME shelf exams at 70. These thresholds were determined based on approximations of one standard deviation below the national average score for each test (United States

Medical Licensing Examination 2014). To determine the optimal screening metric for poor standardized exam performance, univariate logistic regression was performed to compare the predictive capacity of individual course scores versus a composite score consisting of an average of all preclinical courses. A receiver operating characteristic (ROC) curve was then generated using the superior screening metric. All data were analyzed using SAS statistical software (version 9.3; SAS Institute, Inc; Cary, NC). This study was approved by the University of Virginia Institutional Review Board (IRB protocol: 2012-0225-00).

Results

Between 2007 and 2012, 845 medical students completed USMLE Step 1 and Step 2. Of these, 50 did not have complete preclinical course scores available at the time of study. Therefore, 795 medical students across six medical school classes were included for data analysis. Females comprised 45.8 % (364/795) of the study population, and class sizes ranged from 127 to 140 students. Median scores for USMLE Step 1, Step 2, and the MCAT

Table 1 Score distributions for preclinical courses and standardized exams

	Median	IQR
MCAT	33	31–35
Basic science courses		
Anatomy	88	83–91
Cell biology	89	85–93
Molecular genetics	91	88–94
Biochemistry	88	83–92
Neuroscience	88	83–92
Physiology	85	80–90
Microbiology	86	81–90
Clinical knowledge courses		
Human behavior	93	90–95
Epidemiology	95	93–98
Psychiatry	86	82–90
Pathology	85	81–89.5
POM-1	97	95–99
POM-2	87	83–90
Pharmacology	86	80–90
NBME subject exams		
Surgery	77	71–82
Internal Medicine	79	74–85
Obstetrics and Gynecology	75	70–80
Pediatrics	79	73–85
Neurology	74	70–80
Psychiatry	82	75–87
Family Medicine	77	72–83
USMLE Step 1	234	219–245.5
USMLE Step 2	247	234–260

were 234 [interquartile range (IQR) 219–245.5], 247 (IQR 234–260), and 33 (IQR 31–35), respectively. The scoring distributions of preclinical courses and NBME shelf exams are presented in Table 1.

Linear multiple regression analyses were performed for each NBME shelf exam and USMLE Step exam, using the MCAT and preclinical course scores as the independent variables. Three exemplary models—for NBME Surgery, USMLE Step 1 and USMLE Step 2—derived through the all-subsets selection process are shown in Table 2. This process removes independent variables that do not contribute to predictive capacity within each model. Overall, preclinical course scores were significantly predictive of subsequent performance on all standardized exams—USMLE Step 1, Step 2, and NBME shelf exams ($p < 0.001$ for all exams). Adjusted R^2 statistics for the NBME clinical subject exams ranged from 0.37 (Family Medicine) to 0.46 (Neurology), indicating that performance in preclinical courses explains a moderate amount of the variability in standardized exam scores. Among preclinical courses, Pathology was universally the strongest predictor of performance for all NBME and USMLE exams, with a p value of <0.001 and the highest regression coefficient within all models (Fig. 1). With the exception of NBME subject exams for Obstetrics and Gynecology and Family Medicine, MCAT score was the second strongest predictor of performance for all standardized exams. For these two exceptions, the Practice of Medicine courses were stronger predictors of shelf exam performance than

Table 2 Preclinical course predictors of standardized exam performance

	NBME surgery		USMLE step 1		USMLE step 2	
	$F = 38.4$ Coefficient	$R^2 = 0.41$[a] p value	$F = 152.0$ Coefficient	$R^2 = 0.70$ p value	$F = 59.2$ Coefficient	$R^2 = 0.56$ p value
Basic science						
Anatomy			0.132	0.18		
Cell biology			0.145	0.082	−0.157	0.229
Molecular genetics	−0.168	0.132	−0.371	0.002		
Biochemistry			0.17	0.101		
Neuroscience						
Physiology					−0.623	<0.001
Microbiology						
Clinical knowledge						
Human behavior			−0.233	0.045		
Epidemiology	0.058	0.448	0.458	<0.001	0.176	0.273
Psychiatry	−0.276	0.001	−0.124	0.234		
Pathology	0.779	<0.001	1.735	<0.001	1.896	<0.001
POM-1	0.256	0.040			0.831	0.002
POM-2	0.321	0.003	−0.185	0.109	0.731	0.002
Pharmacology			0.601	<0.001	0.283	0.165
MCAT	0.724	<0.001	1.29	<0.001	1.496	<0.001

NBME National Board of Medical Examiners subject exam, *USMLE* United States Medical Licensing Examination, *POM* practice of medicine

[a] R^2—adjusted R^2 statistic for model fit

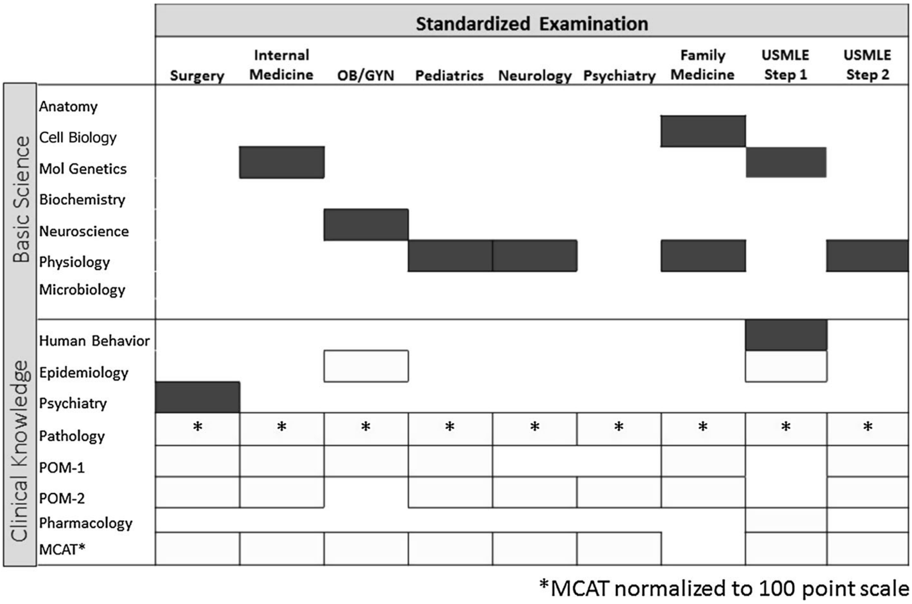

Fig. 1 Associations between preclinical course scores and standardized exam performance. Significant positive associations are shown in *white boxes*, significant negative associations shown in *dark gray*. Strongest positive predictor for each standardized exam indicated with (*). *Unmarked spaces* indicate no significant association in either direction

the MCAT. No basic science course was positively associated with performance on any standardized exam. In fact, performance on Physiology was negatively associated with scores on NBME shelf exams in Pediatrics, Neurology, and Family Medicine, as well as on USMLE Step 2.

Eighty-one students scored at or below 205 on the USMLE Step 1 exam (10.2 %). Students with higher scores on Pathology were significantly less likely to score under 205 on USMLE Step 1 ($p < 0.001$). The area under the curve (AUC) for Pathology score as a predictor of poor performance on USMLE Step 1 was 0.905, and ranged from 0.80 to 0.90 for other clinically relevant exams (Fig. 2). By comparison, the corresponding AUC for the MCAT composite score as a predictor of Step 1 performance was 0.80. Within a paradigm of identifying the highest-risk students for focused remediation, using a Pathology score of 75 as the screening threshold yielded a sensitivity of 0.37, a specificity of 0.97, and a likelihood ratio of 11.9. Conversely, using a Pathology score of 83 as the screening threshold yielded sensitivity of 0.96, specificity of 0.70, and likelihood ratio of 3.2. A cut-off score of 75 on Pathology was similarly specific for poor performance on all other NBME and USMLE standardized exams, with likelihood ratios ranging from 6.0 to 14.0 (Table 3).

Discussion

Findings from the present study suggest that select preclinical coursework may be strong predictors of standardized exam scores. These predictors can be identified specific to individual institutional curricula. Prior studies have shown that clerkship grades and

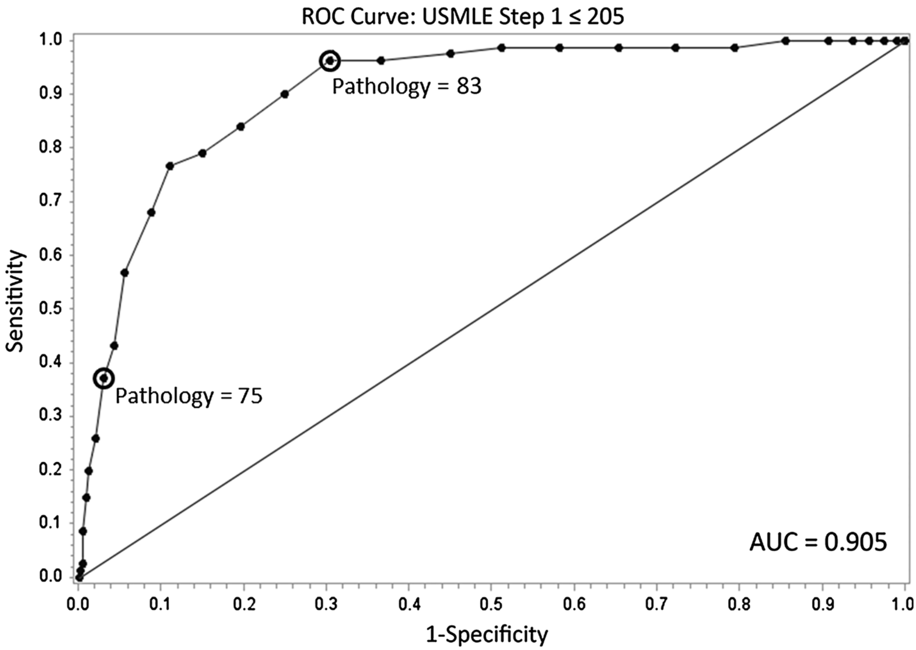

Fig. 2 Receiver operating characteristic (ROC) curve for Pathology grade as a predictor of poor performance on United States Medical Licensing Examination (USMLE) Step 1. Area under the ROC curve (AUC) is 0.905

Table 3 Pathology course grade <75 as a predictor of poor performance on standardized exams

Exam	Event	Sensitivity	Specificity	Likelihood	AUC
USMLE Step 1	<205	0.37	0.97	11.9	0.91
USMLE Step 2	<212	0.42	0.97	14.0	0.90
Surgery	<70	0.18	0.97	6.0	0.80
Neurology	<70	0.26	0.98	13.0	0.83
Ob/Gyn	<70	0.18	0.97	6.0	0.80
Pediatrics	<70	0.25	0.96	6.2	0.84
Internal Medicine	<70	0.34	0.96	8.5	0.85
Family Medicine	<70	0.25	0.97	8.3	0.81
Psychiatry	<70	0.27	0.97	9.0	0.85

AUC area under the receiver operating curve, *USMLE* United States Medical Licensing Examination

USMLE Step 1 score are the two most common selection criteria for residency (Green et al. 2009). Although correlations between NBME shelf scores and clinical rotation performance are varied (Senecal et al. 2010), most clerkship directors adopt these exams as part of the clerkship assessment process (Hemmer et al. 2002; Kassebaum and Eaglen 1999; Levine et al. 2005). Data presented here show that Pathology and Practice of Medicine are predictive of subsequent standardized exam performance. Conversely, scores from most basic science courses in this study were consistently unassociated—or negatively associated—with standardized exams. This is remarkable given the traditional perception that the USMLE Step 1 is a basic science exam (Simon et al. 2007).

Predicting failure on USMLE Step 1 using multivariate methods has historically been challenging due to a relatively low rate of occurrence, particularly at more selective medical schools (Kleshinski et al. 2009). Although exam failure is a clearly impactful outcome, performance at or lower than one standard deviation below the mean can similarly have irreparable implications. Using this as the threshold for poor performance increases the observation frequency and, by extension, the validity of multivariate analyses aimed at identifying consistently predictive disciplines. Such analyses can define performance thresholds in order to screen for students at high-risk for doing poorly on future exams. This threshold can be adjusted to cast a broad net and target sensitivity, or to focus on the highest-risk students and target likelihood ratio. In this study, multiple linear regression isolated Pathology as the single preclinical predictor that outperformed all other variables—including MCAT—in forecasting USMLE and NBME exam performance at our institution. This finding yields diverse opportunities for risk-stratification and early remediation. Because Pathology was associated with performance on both clinical and basic science standardized exams, one potential format for remediation may involve individual tutoring to address basic science deficiencies before Step 1 followed by clinical review sessions during clerkship rotations.

Isolating highly relevant, pre-clerkship courses can also allow identification of characteristics that lend superior relevancy to high-stakes exams. These factors may include the composition and development of teaching faculty and a more clinically relevant approach to assessment of pre-clerkship student performance. For example, an unexpected result of this study was the finding that no basic science course was significantly associated with performance on USMLE Step 1. These data do not necessarily contradict prior work that associated basic science ratings with Step 1 performance (Swanson et al. 1996). Instead, the relationships between basic science aptitude and Step 1 performance are likely confounded by correlations between performances on basic science courses and clinical courses such as Pathology and Practice of Medicine. It is therefore not surprising that pass/fail grading systems for basic science curricula have historically had little impact on USMLE Step 1 performance (Rohe et al. 2006; Spring et al. 2011).

A meta-analysis by Donnon and colleagues noted correlations between MCAT and USMLE Steps 1–3 ranging from 0.38 to 0.60 (Donnon et al. 2007). Substantial heterogeneity across studies implied that institution-specific analyses might reveal predictors more relevant to individual students. At our institution, Pathology test questions were unique in that they were primarily case-based, emphasizing knowledge application and information synthesis. Not only do these questions condition students for common standardized exam question stems, they also require higher-order cognitive processes within Bloom's taxonomy (Anderson et al. 2000; Zoller and Pushkinb 2007). Within this taxonomy, knowledge recall is classified as the lowest level among cognitive processes, and knowledge application is considered mid-level. While it is difficult for multiple-choice style questions to address true top-level cognitive processes such as analysis, evaluation, and synthesis, question stems and response options can be designed to engage mid-level processes. Optimizing basic science instruction such that it consistently recruits these higher-order functions and aligns with subsequent standardized exams remains an opportunity for applying improved pedagogical methods. Work by Palmer, Miller and others have provided guidelines that improve cognitive content of multiple-choice questions in the basic sciences (Miller et al. 1991; Palmer and Devitt 2007).

Relatively few studies have focused on preclinical courses as predictors for standardized exam scores. Kozar and colleagues identified the preclinical Pathology NBME subject exam score as a sensitive, though not specific, predictor of poor performance on the Surgery shelf exam (Kozar et al. 2007). Similar conclusions were noted in earlier work by Holtman and colleagues, in which national Pathology NBME subject exam scores explained 58 % of within-school USMLE Step 1 variance (Holtman et al. 2001). That same study noted that medical school courses in the second year explained a greater percentage of Step 1 variance than first-year courses. Coumarbatch and colleagues later corroborated these findings, reporting an AUC of 0.92 for the predictive capacity of cumulative second-year performance toward Step 1 scores (Coumarbatch et al. 2010). Despite different study approaches, these authors' results lend credence to the present study's findings. Like most traditional curricula, the first year of training at the University of Virginia focused on basic sciences while the second year contained more clinically relevant courses such as Pathology.

Although this study revealed an absence of association between basic science course performance and clinical standardized testing outcomes, these results do not necessarily disparage the importance of basic science knowledge. Within the knowledge encapsulation paradigm put forth by Schmidt and colleagues, basic science concepts are encapsulated within clinical facts and experience (Rikers et al. 2005). In the process of tackling a clinical problem, experienced physicians continue to refer to basic science knowledge. Because this connection often occurs on a subconscious level, it is not entirely surprising that direct testing of basic science knowledge is not strongly associated with subsequent clinical testing performance. The challenge for educators is to integrate basic science teaching within clinical contexts in order to enhance this encapsulation process. Toward this goal, the University of Virginia, like many other medical schools, has recently converted to an integrated curriculum that employs more clinician instructors and assimilates basic science concepts into systems-based and disease-based didactics. As systems-based preclinical courses become more pervasive, Pathology is often dispersed across multiple course blocks. However, these curricular changes do not diminish the importance of this study's findings. We believe that electronically tracking test performance on Pathology-related questions throughout the preclinical curriculum will continue to hold value in predicting subsequent standardized test performance.

This study has several limitations. Specifically, its single-institutional, retrospective design weakens the generalizability of its findings. This is particularly true as grading scales for preclinical courses are not standardized across institutions. However, rather than suggesting that any one subject is preeminent, the data simply endorse the identification of early markers for future success or failure. Toward this end, the use of linear multiple regression to identify early predictors of performance followed by logistic regression to risk-stratify students is an approach that is applicable to any institution that tracks preclinical course grades. Moving forward, it will be important to investigate whether pathology content and assessment methods are associated with standardized testing outcomes at other institutions. Although there is a possibility of type 1 error inflation in the setting of multiple linear regression models, the relationships between Pathology and each standardized exam remain meaningful with p-value thresholds adjusted to 0.001. Additionally, personalized demographic information was not available for statistical analysis. While undergraduate major may be unrelated to USMLE Step 1 performance, other factors such as age, gender, and race have previously shown evidence of association and may contribute to multivariate modeling (Haist et al. 2000; Huff and Fang 1999; Smith 1998).

For a variety of reasons, however, risk-stratifying students based on these demographic factors may not be possible or appropriate.

In summary, linear multiple regression models can identify preclinical predictors of standardized exam performance. Clinically relevant courses such as Pathology have significant and broad predictive value and may out-perform the MCAT in predictive capacity. By setting thresholds of performance, students can be risk-stratified with high specificity, facilitating focused strategies of remediation. In more modern systems-based curricula, there remain basic science "threads" and examination questions that can be tracked to inform students and faculty regarding performance. Many traditional basic science courses show little association with performance on NBME and USMLE exams, suggesting that these courses could benefit from greater integration of higher-level cognitive domains.

Funding Funding support was provided by NIH T32 CA163177 (to YH).

Compliance with ethical standards

Conflicts of interest None.

References

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., et al. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Abridged ed.). Boston, MA: Pearson.

Basco, W. T, Jr, Way, D. P., Gilbert, G. E., & Hudson, A. (2002). Undergraduate institutional MCAT scores as predictors of USMLE step 1 performance. *Academic Medicine, 77*(10 Suppl), S13–S16.

Briscoe, G. W., Fore-Arcand, L., Levine, R. E., Carlson, D. L., Spollen, J. J., Pelic, C., & Al-Mateen, C. S. (2009). Psychiatry clerkship students' preparation, reflection, and results on the NBME Psychiatry Subject Exam. *Academic Psychiatry, 33*(2), 120–124.

Case, S. M., Ripkey, D. R., & Swanson, D. B. (1996). The relationship between clinical science performance in 20 medical schools and performance on Step 2 of the USMLE licensing examination. 1994–1995 Validity Study Group for USMLE Step 1 and 2 Pass/Fail Standards. *Academic Medicine, 71*(1 Suppl), S31-3.

Cooke, M., Irby, D. M., Sullivan, W., & Ludmerer, K. M. (2006). American medical education 100 years after the Flexner report. *New England Journal of Medicine, 355*(13), 1339–1344.

Coumarbatch, J., Robinson, L., Thomas, R., & Bridge, P. D. (2010). Strategies for identifying students at risk for USMLE step 1 failure. *Family Medicine, 42*(2), 105–110.

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. *Academic Medicine, 82*(1), 100–106.

Green, M., Jones, P., & Thomas, J. X, Jr. (2009). Selection criteria for residency: Results of a national program directors survey. *Academic Medicine, 84*(3), 362–367.

Haist, S. A., Wilson, J. F., Elam, C. L., Blue, A. V., & Fosson, S. E. (2000). The effect of gender and age on medical school performance: An important interaction. *Advances in Health Sciences Education Theory and Practice, 5*(3), 197–205.

Hemmer, P. A., Szauter, K., Allbritton, T. A., & Elnicki, D. M. (2002). Internal medicine clerkship directors' use of and opinions about clerkship examinations. *Teaching and Learning in Medicine, 14*(4), 229–235.

Hojat, M., Erdmann, J. B., Veloski, J. J., Nasca, T. J., Callahan, C. A., Julian, E., & Peck, J. (2000). A validity study of the writing sample section of the medical college admission test. *Academic Medicine, 75*(10 Suppl), S25–S27.

Holtman, M. C., Swanson, D. B., Ripkey, D. R., & Case, S. M. (2001). Using basic science subject tests to identify students at risk for failing step 1. *Academic Medicine, 76*(10 Suppl), S48–S51.

Huff, K. L., & Fang, D. (1999). When are students most at risk of encountering academic difficulty? A study of the 1992 matriculants to U.S. medical schools. *Academic Medicine, 74*(4), 454–460.

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

Kassebaum, D. G., & Eaglen, R. H. (1999). Shortcomings in the evaluation of students' clinical skills and behaviors in medical school. *Academic Medicine, 74*(7), 842–849.

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

Koenig, J. A., Sireci, S. G., & Wiley, A. (1998). Evaluating the predictive validity of MCAT scores across diverse applicant groups. *Academic Medicine, 73*(10), 1095–1106.

Kozar, R. A., Kao, L. S., Miller, C. C., & Schenarts, K. D. (2007). Preclinical predictors of surgery NBME exam performance. *Journal of Surgical Research, 140*(2), 204–207.

Levine, R. E., Carlson, D. L., Rosenthal, R. H., Clegg, K. A., & Crosby, R. D. (2005). Usage of the National Board of Medical Examiners Subject test in psychiatry by U.S. and Canadian clerkships. *Academic Psychiatry, 29*(1), 52–57.

Miller, D. A., Sadler, J. Z., Mohl, P. C., & Melchiode, G. A. (1991). The cognitive context of examinations in psychiatry using Bloom's taxonomy. *Medical Education, 25*(6), 480–484.

Palmer, E. J., & Devitt, P. G. (2007). Assessment of higher order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. *BMC Medical Education, 7*, 49.

Patel, V. L., & Groen, G. J. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*, 91–116.

Raymond, M. R., Mee, J., King, A., Haist, S. A., & Winward, M. L. (2011). What new residents do during their initial months of training. *Academic Medicine, 86*(10 Suppl), S59–S62.

Rikers, R. M., Loyens, S. M., & Schmidt, H. G. (2004). The role of encapsulated knowledge in clinical case representations of medical students and family doctors. *Medical Education, 38*(10), 1035–1043.

Ripkey, D. R., Case, S. M., & Swanson, D. B. (1997). Predicting performances on the NBME Surgery Subject Test and USMLE Step 2: The effects of surgery clerkship timing and length. *Academic Medicine, 72*(10 Suppl 1), S31–S33.

Ripkey, D. R., Case, S. M., & Swanson, D. B. (1999). Identifying students at risk for poor performance on the USMLE Step 2. *Academic Medicine, 74*(10 Suppl), S45–S48.

Rikers, R. M., Schmidt, H. G., & Moulaert, V. A. (2005). Biomedical knowledge: Encapsulated or two worlds apart? *Applied Cognitive Psychology, 19*(2), 223–231.

Rohe, D. E., Barrier, P. A., Clark, M. M., Cook, D. A., Vickers, K. S., & Decker, P. A. (2006). The benefits of pass-fail grading on stress, mood, and group cohesion in medical students. *Mayo Clinic Proceedings, 81*(11), 1443–1448.

Senecal, E. L., Askew, K., Gorney, B., Beeson, M. S., & Manthey, D. E. (2010). Anatomy of a clerkship test. *Academic Emergency Medicine, 17*(Suppl 2), S31–S37.

Simon, S. R., Bui, A., Day, S., Berti, D., & Volkan, K. (2007). The relationship between second-year medical students' OSCE scores and USMLE Step 2 scores. *Journal of Evaluation in Clinical Practice, 13*(6), 901–905.

Smith, S. R. (1998). Effect of undergraduate college major on performance in medical school. *Academic Medicine, 73*(9), 1006–1008.

Spring, L., Robillard, D., Gehlbach, L., & Simas, T. A. (2011). Impact of pass/fail grading on medical students' well-being and academic outcomes. *Medical Education, 45*(9), 867–877.

Swanson, D. B., Ripkey, D. R., & Case, S. M. (1996). Relationship between achievement in basic science coursework and performance on 1994 USMLE Step 1. 1994–1995 Validity Study Group for USMLE Step 1/2 Pass/Fail Standards. *Academic Medicine, 71*(1 Suppl), S28–S30.

Torre, D., Papp, K., Elnicki, M., & Durning, S. (2009). Clerkship directors' practices with respect to preparing students for and using the National Board of Medical Examiners Subject Exam in medicine: Results of a United States and Canadian Survey. *Academic Medicine, 84*(7), 867–871.

United States Medical Licensing Examination. (2014). *Scoring and score reporting*. http://www.usmle.org/bulletin/scores/ [3/25 2014].

USMLE. (2013). *Bulletin of information*. Philadelphia, PA: National Board of Medical Examiners.

Wiley, A., & Koenig, J. A. (1996). The validity of the Medical College Admission Test for predicting performance in the first 2 years of medical school. *Academic Medicine, 71*(10 Suppl), S83–S85.

Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy debunks the "MCAT myth". *Science, 319*(5862), 414–415.

Zoller, U., & Pushkinb, D. (2007). Matching Higher-Order Cognitive Skills (HOCS) promotion goals with problem-based laboratory practice in a freshman organic chemistry course. *Chemistry Education Research and Practice, 8*(2), 153–171.