The Associations Between United States Medical Licensing Examination Performance and Outcomes of Patient Care

John Norcini, PhD, Irina Grabovsky, PhD, Michael A. Barone, MD, MPH,
M. Brownell Anderson, MEd, Ravi S. Pandian, MA, and Alex J. Mechaber, MD

Abstract

Purpose

The United States Medical Licensing Examination (USMLE) comprises a series of assessments required for the licensure of U.S. MD-trained graduates as well as those who are trained internationally. Demonstration of a relationship between these examinations and outcomes of care is desirable for a process seeking to provide patients with safe and effective health care.

Method

This was a retrospective cohort study of 196,881 hospitalizations in Pennsylvania over a 3-year period (January 1, 2017 to December 31, 2019) for 5 primary diagnoses: heart failure, acute myocardial infarction, stroke, pneumonia, or chronic obstructive pulmonary disease. The 1,765 attending physicians for these hospitalizations self-identified as family physicians or general internists. A converted score based on USMLE Step 1, Step 2 Clinical Knowledge, and Step 3 scores was available, and the outcome measures were in-hospital mortality and log length of stay (LOS). The research team controlled for characteristics of patients, hospitals, and physicians.

Results

For in-hospital mortality, the adjusted odds ratio was 0.94 (95% confidence interval [CI] = 0.90, 0.99; $P < .02$). Each standard deviation increase in the converted score was associated with a 5. 51% reduction in the odds of in-hospital mortality. For log LOS, the adjusted estimate was 0.99 (95% CI = 0.98, 0.99; $P < .001$). Each standard deviation increase in the converted score was associated with a 1.34% reduction in log LOS.

Conclusions

Better provider USMLE performance was associated with lower in-hospital mortality and shorter log LOS for patients, although the magnitude of the latter is unlikely to be of practical significance. These findings add to the body of evidence that examines the validity of the USMLE licensure program.

Medical licensing boards in the United States and its territories aim to protect public health by establishing and maintaining standards of training and competence for physicians. Regulating the physician workforce starts with medical licensure, which requires that candidates successfully complete educational requirements and a medical licensing examination series. The United States Medical Licensing Examination (USMLE) comprises a series of assessments required for the licensure of U.S. MD-trained graduates as well as those who complete their training internationally. Several prior studies have documented the relationship between licensing examinations and various indicators of competence, yet, to our knowledge, few have established a relationship with the outcomes of care.[1–9] Demonstration of such a link is desirable for a licensing process that seeks to protect the health of the public. Thus, the purpose of this study was to investigate whether physician USMLE performance was related to adjusted patient in-hospital mortality and length of stay (LOS) for 5 predetermined common medical conditions.

The USMLE is composed of 3 examinations. Performance above a minimum passing score is required for an MD-trained physician to be eligible for an unrestricted license to practice medicine in the United States and its territories. The Step 1 examination is an assessment of the candidate's ability to apply foundational science knowledge to medical practice. The Step 2 Clinical Knowledge (CK) examination targets an individual's ability to apply the medical knowledge and clinical science essential for providing supervised patient care. The Step 3 examination assesses a candidate's ability to provide unsupervised care, especially in ambulatory settings. These 3 examinations have been part of the licensure sequence since 1992, with 56% of 1,018,776 actively licensed physicians in the United States having taken all or part of the USMLE sequence (the remainder being those who are licensed by presenting prior regulatory exams for licensure such as the National Board of Medical Examiners [NBME] "Parts" or FLEX [Federation Licensing Examination] exams, or osteopathic physicians who did not take any USMLE Step examinations).[10]

Prior studies have documented the relationship between the number of test attempts and scores on licensing examinations with other markers of physician competence.[2–6] These include

Please see the end of this article for information about the authors.

Correspondence should be addressed to John Norcini, 275 South Bryn Mawr Avenue, Bryn Mawr, PA 19010; email: john.norcini@gmail.com.

Acad Med. 2024;99:325–330.
First published online October 9, 2023
doi: 10.1097/ACM.0000000000005480

demonstrated associations between USMLE performance and specialty board examination performance, clinical performance evaluations, and subsequent disciplinary actions in practice. Moreover, past studies have demonstrated a relationship between licensing examination scores and process of care measures including proper prescribing practices and adherence to preventive health screening guidelines.[7–9] Specific to the USMLE, Norcini and colleagues demonstrated that, among patients with congestive heart failure (HF) or acute myocardial infarction (AMI) who were cared for by international medical graduate (IMG) attending physicians, higher provider USMLE Step 2 CK scores were associated with lower patient mortality. Researchers determined that an increase of 1 standard deviation on the USMLE Step 2 CK score scale correlated with a 4% decrease in relative risk for patient mortality.[10]

Studies of this type have the potential to add to the overall validity argument for the use of high-stakes examinations in medical regulation. Furthermore, as part of the recent exploration of potential changes to USMLE Step 1 score reporting, a deficit has been acknowledged in the body of work seeking associations between licensing examination performance and residency performance and clinical practice outcomes. A recommendation stemming from the related 2019 Invitational Conference on USMLE Scoring (InCUS) was to "accelerate research on the correlation of USMLE performance to measures of residency performance and clinical practice."[11]

In response to the InCUS recommendation, the purpose of this study was to investigate whether physicians' USMLE performance was related to adjusted in-hospital mortality and LOS for patients hospitalized in Pennsylvania with a primary diagnosis of 1 of 5 predetermined conditions: HF, AMI, stroke, pneumonia, or chronic obstructive pulmonary disease (COPD). Our investigation seeks to add to the body of evidence examining the validity of the USMLE program.

Method

This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines.[12] The Institutional Review Board at the American Institute for Research (number EX00533) approved it as exempt.

Sources of data

Data for this study came from the Pennsylvania Health Care Cost Containment Council (PHC4), the American Medical Association (AMA) Physician Masterfile, and the examination score records of the USMLE program. These 3 data sources were linked through a series of matches based on the National Provider Identifier (NPI) number, name, self-reported sex, and birth year.

All hospitals in Pennsylvania must send data to the PHC4 each time a patient is discharged. Patient demographic characteristics, primary and secondary diagnoses, discharge status, LOS, and hospital where care was provided are included. The attending physician, defined by PHC4 as the individual who has overall responsibility for the medical care and treatment of the patient, is typically also identified by the hospital. PHC4 applies validation and editing procedures to patient data, which can be corrected by the hospitals. In this study, we analyzed hospitalizations for AMI, HF, pneumonia, COPD, and stroke from January 1, 2017 to December 31, 2019. We chose these conditions because they have been used previously as a means for judging the quality of patient care.[13]

The AMA Physician Masterfile was the primary source of information on practicing physicians. For this study, we limited analysis to self-identified family medicine and general internal medicine physicians (excluding self-identified hospitalists because their numbers were small and their training and certification backgrounds were heterogeneous) who were the attending physicians for patients with 1 or more of the 5 conditions in the PHC4 database. We did so because these physicians provide considerable care for these conditions and it prevented the results from being confounded by subspecialization. Providers' initial board certification was also available, and we excluded subspecialists and physicians who held certification in anything other than internal medicine or family medicine.

We analyzed USMLE scores on physicians' first attempt at Step 1, Step 2 CK, and Step 3 of the examination sequence (we excluded the Step 2 Clinical Skills examination since it is no longer required), but only for those physicians who had taken and passed all 3 steps. During the study period, USMLE Step 3 eligibility required passing Step 1 and Step 2 CK. In addition, physician sex and country of medical school were accessed.

Data elements

For patients, we extracted age, sex, race/ethnicity, principal and secondary diagnoses, discharge status, LOS, home county, and hospital. A modified version of the Charlson Comorbidity Index was developed as a measure of the sickness of patients.[14] In addition, we combined race and ethnicity data to a single variable with White, non-Hispanic patients in 1 group and all remaining participants in another. This was done because some race or ethnicity groups were too small for analysis. For each hospitalization, we created an index of whether the patient's home county was rural by reference to a list maintained by the state of Pennsylvania.[15]

We used discharge status to determine whether the patient died while in the hospital. For LOS analyses, we removed patients who died in hospital (n = 13,198), were transferred to another short-term inpatient facility, had missing LOS data, or had extreme values for LOS (0 days or more than 40 days). To reduce the effects of outliers, we calculated the logarithm of the LOS data.

For attending physicians, we knew their self-designated specialty, whether they were U.S. medical graduates (USMGs) or IMGs, their sex, and whether they had initial board certification in family medicine or internal medicine. In addition, we calculated each individual physician's patient volume by counting the number of patients with any of the 5 conditions the physician treated during the study timeframe. In previous work, these characteristics have been found to be associated with patient outcomes, which we wanted to control for in this study.[16–18]

Scores across Steps 1, 2, and 3 are very highly correlated. To avoid the problems of multicollinearity, we developed a composite measure of USMLE performance for use in the analyses. First-time scores on each of the 3 steps of USMLE were on a scale that had a mean of 200 (standard deviation [SD] = 20) in an archival reference group. We converted these to *z* scores (mean = 0; SD = 1) and averaged them to develop a composite measure of USMLE performance for each physician.[19]

We also calculated the number of patients with any of the 5 conditions treated at the hospital by the study physicians (i.e., facility patient volume).

Analysis

We began by calculating descriptive statistics for hospitalizations of the total patient population and for those patients with each condition. To determine the adjusted relationship between average USMLE score and patient mortality, we applied a multivariate logistic regression model adjusted for the following:

- comorbidity index;
- condition (reference: AMI);
- patient age;

Table 1

Descriptive Information for Hospitalizations for Each of 5 Studied Conditions, From a Study of the Association Between USMLE Performance and Patient Care Outcomes at Pennsylvania Hospitals, 2017–2019

Characteristic	AMI	HF	Pneumonia	COPD	Stroke	Total
Patient age, no. (%)						
≤ 49	1,734 (8.7)	2,793 (4.4)	5,708 (11.8)	1,779 (4.1)	1,338 (6.0)	13,352 (6.8)
50–64	5,649 (28.4)	11,326 (17.9)	10,232 (21.2)	14,810 (34.4)	4,966 (22.2)	46,983 (23.9)
65–74	4,966 (25.0)	14,083 (22.3)	10,012 (20.8)	13,298 (30.9)	5,397 (24.1)	47,756 (24.3)
75–84	4,386 (22.0)	17,570 (28.8)	11,075 (23)	9,455 (22.0)	5,696 (25.4)	48,182 (24.5)
> 85	3,163 (15.9)	17,513 (27.7)	11,176 (23.2)	3,738 (8.7)	5,018 (22.4)	40,608 (20.6)
Total	19,898 (10.1)	63,285 (32.1)	48,203 (24.5)	43,080 (22.4)	22,415 (11.4)	196,881 (100)
Patient sex, no. (%)						
Female	8,281 (41.6)	31,332 (49.5)	24,350 (50.5)	24,623 (57.2)	11,489 (51.3)	100,075 (50.8)
Male	11,616 (58.4)	31,950 (50.5)	23,851 (49.5)	18,457 (42.8)	10,926 (48.7)	96,800 (49.2)
Unknown	1 (0.0)	3 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	6 (0.0)
Total	19,898 (10.1)	63,285 (32.1)	48,203 (24.5)	43,080 (21.9)	22,415 (11.4)	196,881 (100)
Patient race/ethnicity, no. (%)						
White, non-Hispanic	16,817 (84.5)	49,274 (77.9)	40,256 (83.4)	34,439 (80.0)	18,150 (81.0)	158,936 (80.7)
Other	3,081 (15.5)	14,011 (22.1)	7,947 (16.5)	8,641 (20.1)	4,265 (19.0)	37,945 (19.3)
Total	19,898 (10.1)	63,285 (32.1)	48,203 (24.5)	43,080 (21.9)	22,415 (11.4)	196,881 (100)
Patient location, no. (%)[a]						
Nonrural	12,780 (65.0)	45,627 (72.7)	32,988 (69.2)	30,211 (70.7)	16,103 (72.7)	137,709 (70.6)
Rural	6,896 (35.1)	17,167 (27.3)	14,696 (30.8)	12,519 (29.3)	6,046 (27.3)	57,324 (29.4)
Total	19,676 (10.1)	62,794 (32.2)	47,684 (24.5)	42,730 (21.9)	22,149 (11.4)	195,033 (100)
Patient length of stay in days, mean (SD)	3.8 (3.9)	5.2 (4.6)	5.3 (6.6)	4.7 (4.9)	4.2 (4.7)	4.86 (5.16)
Patient comorbid conditions, mean (SD)	1.7 (1.3)	2.8 (1.2)	1.8 (1.3)	1.7 (1.2)	1.9 (1.3)	1.7 (1.3)
Patient mortality, no. (%)	577 (2.9)	1,176 (1.9)	1,348 (2.8)	600 (1.4)	441 (2.0)	4,142 (2.1)
Facility volume/1,000 patients, mean (SD)	4.6 (2.4)	4.8 (2.6)	4.6 (2.6)	4.4 (2.6)	5.0 (2.7)	4.7 (2.6)
Physician specialty, no. (%)						
Internal medicine	16,808 (84.5)	53,917 (85.2)	40,202 (83.4)	36,221 (84.1)	19,209 (85.7)	166,357 (84.5)
Family medicine	3,090 (15.5)	9,368 (14.8)	8,001 (16.6)	6,859 (15.9)	3,206 (14.3)	30,524 (15.5)
Physician board certification, no. (%)						
Internal medicine	15,963 (80.2)	51,233 (81.0)	38,019 (78.9)	34,307 (79.6)	18,183 (81.1)	157,705 (80.1)
Family medicine	3,027 (15.2)	9,073 (14.3)	7,770 (16.1)	6,651 (15.4)	3,110 (13.9)	29,631 (15.1)
None	908 (4.6)	2,979 (4.7)	2,414 (5.0)	2,122 (4.9)	1,122 (5.0)	9,545 (4.9)
Physician sex, no. (%)						
Female	5,953 (29.9)	21,380 (33.8)	16,138 (33.5)	13,687 (31.8)	7,636 (34.1)	64,794 (32.9)
Male	13,945 (70.1)	41,905 (66.2)	32,065 (66.5)	29,393 (68.2)	14,779 (65.9)	132,087 (67.1)
Physician medical school location, no. (%)						
USMG	14,421 (72.5)	43,293 (68.4)	33,330 (69.2)	29,905 (69.4)	15,671 (69.9)	136,620 (69.4)
IMG	5,477 (27.5)	19,992 (31.6)	14,873 (30.9)	13,175 (30.6)	6,744 (30.1)	60,261 (30.6)

Abbreviations: USMLE, United States Medical Licensing Examination; AMI, acute myocardial infarction; HF, heart failure; COPD, chronic obstructive pulmonary disease; SD, standard deviation; USMG, U.S. medical graduate; IMG, international medical graduate.

[a] Data were missing for 1,848 records.

- patient sex (reference: male);
- patient race/ethnicity (reference: White/non-Hispanic);
- patient location (reference: nonrural);
- whether the physician was an IMG (reference: USMG);
- self-designated specialty (reference: family medicine);
- physician sex (reference: female);
- specialty board certification (with *not certified* as reference);
- physician volume; and
- number of hospitalizations in the institution for the studied conditions.

We applied a similar model to determine the adjusted relationship between USMLE average score and log LOS with a reduced number of hospitalizations as described above. To adjust for the clustering of patients within physicians and physicians within hospitals, we used generalized estimating equations. Statistical analyses were conducted in SAS software, version 7.15 (SAS Institute, Cary, NC).

We undertook an analysis to explore the possibility that physicians with higher scores on USMLE also worked at hospitals with better patient outcomes. If these 2 variables were confounded, we might mistakenly attribute patient outcomes to USMLE performance rather than to hospitals. To address this issue, we correlated the mean USMLE scores for hospitals with their mortality rates. The correlation was −0.05 (P = .54) and not statistically significant.

We were also concerned that there might be interactions between patients' age and race/ethnicity, as well as between physicians' specialization and board certification. Running our models with these interactions included did not alter our findings.

Results

Table 1 presents data for the 196,881 hospitalizations in the study, stratified by primary diagnosis. Hospitalizations are further broken down by patient and physician characteristics as well as facility volume. Most of the hospitalizations were for females (n = 100,075, 50.8%), White, non-Hispanics (n = 158,936, 80.7%), and those living in nonrural locations (n = 137,709, 70.6%). Patients had a mean LOS of 4.86 days (SD = 5.16 days) and 1.7 (SD = 1.3) comorbid conditions; and 4,142 (2.1%) hospitalizations resulted in death. Of all hospitalizations, 166,357 (84.5%) were managed by internists, 187,336 (95.2%) were managed by board-certified doctors, 132,087 (67.1%) were managed by male physicians, and 136,620 (69.4%) were managed by USMGs.

There were 1,765 attending physicians for these hospitalizations, 1,316 (75%) were internists, 731 (41%) were female, 1,009 (57%) were USMGs, and 1,663 (94%) were board certified in their self-designated specialty. Mean physician volume through the period of study was 11.2 (SD = 12.8) hospitalizations, and mean converted USMLE score was 0.64 (SD = 1.1). The hospitalizations took place at 171 institutions, with a mean hospital volume of 2,379 (SD = 2,393).

Table 2 presents results of the multivariate analysis with in-hospital mortality as the dependent measure and characteristics of the patients, physicians, and facilities as the covariates. Age, race/ethnicity, and rural location were among the patient demographic characteristics that had statistically significant associations with mortality, as did physician sex and patient volume. After adjustment, higher scores on USMLE were also associated with lower patient mortality. The adjusted odds ratio was 0.94 (95% confidence interval [CI] = 0.90, 0.99; P < .02). Each standard deviation increase in the converted score was associated with a 5.51% reduction in the odds of in-hospital mortality.

Table 3 presents the results of the multivariate analysis with log LOS as the dependent measure and characteristics of the patients, physicians, and facilities as the covariates. Sex, age, and race/ethnicity were among the patient characteristics that had statistically significant associations with log LOS, as did physician specialty and certification status and facility volume. After adjustment, higher USMLE scores were associated with shorter LOS. The adjusted

Table 2

Estimated Adjusted Odds Ratios and Confidence Intervals for In-Hospital Mortality, From a Study of the Association Between USMLE Performance and Patient Care Outcomes at Pennsylvania Hospitals, 2017–2019

Parameter	Odds ratio (95% CI)	P[a]
Patient characteristics		
Heart failure[b]	0.45 (0.41, 0.51)	*
Pneumonia[b]	0.88 (0.79, 0.98)	.015
COPD[b]	0.56 (0.49, 0.63)	*
Stroke[b]	0.60 (0.52, 0.69)	*
Comorbidity index	1.19 (1.16, 1.22)	*
Sex (male)	1.06 (0.99, 1.13)	.075
Age	1.05 (1.05, 1.05)	*
White, non-Hispanic	1.27 (1.14, 1.42)	*
Rural location	1.27 (1.16, 1.39)	*
Attending physician characteristics		
Board certified	0.99 (0.79, 1.25)	.960
Sex (female)	0.82 (0.73, 0.90)	*
Internist	1.09 (0.95, 1.25)	.229
Patient volume/10	0.99 (0.99, 1.00)	< .001
IMG	0.95 (0.86, 1.06)	.366
USMLE converted score[c]	0.95 (0.90, 0.99)	.016
Facility		
Patient volume/1,000	0.99 (0.98, 1.01)	.566

Abbreviations: USMLE, United States Medical Licensing Examination; CI, confidence interval; COPD, chronic obstructive pulmonary disease; IMG, international medical graduate.

[a] The symbol * indicates statistical significance at P < .0001.

[b] Acute myocardial infarction as reference.

[c] Measure of USMLE performance.

Table 3

Estimated Adjusted Odds Ratios With 95% Confidence Interval for Length of Stay, From a Study of the Association Between USMLE Performance and Patient Care Outcomes at Pennsylvania Hospitals, 2017–2019

Parameter	Odds ratio (95% CI)	P[a]
Patient characteristics		
Heart failure[b]	1.18 (1.17, 1.20)	*
Pneumonia[b]	1.30 (1.28, 1.32)	*
COPD[b]	1.17 (1.15, 1.19)	*
Stroke[b]	1.00 (0.99, 1.02)	.589
Comorbidity index	1.10 (1.10, 1.10)	*
Sex (male)	0.96 (0.96, 0.97)	*
Age	1.00 (1.00, 1.00)	*
White, non-Hispanic	1.02 (1.01, 1.03)	*
Rural location	1.02 (1.01, 1.02)	.069
Attending physician characteristics		
Certified	1.05 (1.01, 1.09)	.013
Sex (female)	1.01 (1.00, 1.03)	.113
Internist	1.05 (1.01, 1.09)	.013
Patient volume/10	1.00 (1.00, 1.00)	.923
IMG	0.99 (0.98, 1.01)	.534
USMLE converted score[c]	0.99 (0.98, 0.99)	.001
Facility		
Patient volume/1,000	1.02 (1.01, 1.02)	*

Abbreviations: USMLE, U.S. Medical Licensing Examination; CI, confidence interval; COPD, chronic obstructive pulmonary disease; IMG, international medical graduate.

[a] The symbol * indicates statistical significance at $P < .0001$.

[b] Acute myocardial infarction as reference.

[c] Measure of USMLE performance.

estimate was 0.99 (95% CI = 0.98, 0.99; $P < .001$). Each standard deviation increase in the converted score was associated with a 1.34% reduction in log LOS.

Discussion

The aim of this study was to investigate whether the USMLE performance of physicians was related to the adjusted in-hospital mortality and LOS of their patients with a primary diagnosis of HF, AMI, stroke, pneumonia, or COPD. In terms of mortality, each increase of 1 standard deviation in converted USMLE scores was associated with a 5.51% decline in the odds for mortality. This is comparable to findings from earlier research that focused only on IMGs.[9] If replicated, our findings may be of clinical significance when compared to the magnitude of the effects of recommended medical treatments such as low-dose aspirin in the secondary prevention of cardiovascular and cerebrovascular events (18%).[20] In a normal distribution of scores, the worst performers are conservatively 4 standard deviations below the best performers, which would translate to a 22.04% difference in odds for the mortality of their patients (i.e., 4 × 5.51%).

Similarly, shorter lengths of stay were associated with higher USMLE scores. Each increase of 1 standard deviation in converted USMLE scores was associated with a 1.34% decrease in log LOS. This translates to a reduction of less than 2 hours for the mean LOS (4.86 days). Although it is statistically significant, even at the extremes, this is unlikely to be of practical importance.

It is noteworthy that our findings likely understate the true relationship between USMLE and practice performance. Those physicians for whom the USMLE is a licensure requirement must meet or exceed the passing scores for all 3 steps to practice (and therefore have patient outcomes); those who were ultimately unsuccessful were not included in the study. This restricts the range of scores we could analyze and attenuates the magnitude of the relationships we report.

This study has several limitations. It is important to note that in a retrospective observational study, we can establish an association, but not causality. However, we adjusted for the number of patient, physician, and hospital characteristics known to be related to outcomes, tested for interactions among several of them, and found that unmeasured hospital effects did not significantly influence our findings. Nonetheless, there may be factors that are not included in our study but that are correlated with both USMLE scores and practice performance that influenced the results. There may have also been variability across hospitals in terms of the process for identifying the attending physician. Finally, we confined our analyses to 5 common hospital conditions in 1 U.S. state. Broader sampling of conditions, sites of care, and locations are needed to increase confidence in our findings.

Despite these limitations, it is incumbent upon assessment organizations and licensing and certification bodies to generate evidence that their examinations are associated with relevant outcomes—optimally, those that relate to patients, such as process of care measures or patient health or morbidity and mortality. Doing so helps to build a broader validity argument for the information that results from a single assessment or system of assessment. In the specific case of this study, our results provide some support for the component of Kane's validity framework known as extrapolation, or the extent to which scores reflect real-world performance.[21,22] Until recently, the ability to develop such a body of evidence for the USMLE program has been limited in part by the fact that practice data were generally unavailable in the quality and quantity needed for analysis. Such data are now accessible, and guidelines for analyzing and reporting observational research studies based on them have been strengthened. Future research is needed to address the limitations of this study noted above, as well as to look more closely at the separate steps in the USMLE process to further contribute to validity arguments.

This study demonstrates the ongoing and important role of standardized assessments for higher-stakes decision-making and medical regulation and, along with other evaluations, adds to the inferences we can make about trainees and practicing physicians based on program participation and performance. However, given the many factors that ultimately influence patient outcomes, we must design overall systems to include multiple types of content and assessment to make broader, more comprehensive inferences about trainees and examinees and to more fully support validity arguments. This would include other structured assessments of performance as well as authentic workplace-based assessment. An ideal assessment system would include an educational system focused on outcomes, leveraging assessment results for feedback and performance improvement, complemented by high-stakes examinations with validity evidence related to patient outcomes. This approach would benefit learners and would be optimal for patients, particularly since unrestricted licensure is one of the last stages of physician regulation prior to providing patient care without supervision.

Funding/Support: This study was funded by the United States Medical Licensing Examination.

Other disclosures: None reported.

Ethical approval: The Institutional Review Board at the American Institute for Research (No. EX00533) approved this study as exempt.

J. Norcini is research professor, Department of Psychiatry, SUNY Upstate Medical University, Syracuse, New York; ORCID: https://orcid.org/0000-0002-8464-4115.

I. Grabovsky is senior psychometrician, National Board of Medical Examiners, Philadelphia, Pennsylvania; ORCID: https://orcid.org/0000-0002-3695-0572.

M.A. Barone is vice president of competency-based assessment, National Board of Medical Examiners, Philadelphia, Pennsylvania, and adjunct associate professor of pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland; ORCID: https://orcid.org/0000-0002-4724-784X.

M.B. Anderson is associate professor, University of Minho School of Medicine, Braga, Portugal.

R.S. Pandian is vice president of operations management, National Board of Medical Examiners, Philadelphia, Pennsylvania.

A.J. Mechaber is vice president, United States Medical Licensing Examination, National Board of Medical Examiners, Philadelphia, Pennsylvania, and professor emeritus, University of Miami Miller School of Medicine, Miami, Florida; ORCID: https://orcid.org/0000-0002-0277-8428.

References

1 Arnhart KL, Cuddy MM, Johnson D, et al. Multiple United States Medical Licensing Examination attempts and the estimated risk of disciplinary actions among graduates of U.S. and Canadian medical schools. Acad Med. 2021;96:1319–1323. doi:10.1097/ACM.0000000000004210.

2 Hamstra SJ, Cuddy MM, Jurich D, et al. Exploring the association between USMLE scores and ACGME milestone ratings: a validity study using national data from emergency medicine. Acad Med. 2021;96:1324–1331.

3 Cuddy MM, Young A, Gelman A, et al. Exploring the relationships between USMLE performance and disciplinary action in practice: a validity study of score inferences from a licensure examination. Acad Med. 2017;92:1780–1785.

4 Sharma A, Schauer DP, Kelleher M, et al. USMLE Step 2 CK: best predictor of multimodal performance in an internal medicine residency. J Grad Med Educ. 2019; 11:412–419.

5 Cuddy MM, Liu C, Ouyang W, et al. An examination of the associations among USMLE Step 3 scores and the likelihood of disciplinary action in practice. Acad Med. 2022;97:1504–1510.

6 Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298:993–1001.

7 Tamblyn R, Abrahamowicz M, Brailovsky C, et al. Association between licensing examination scores and resource use and quality of care in primary care practice. JAMA. 1998;280:989–996.

8 Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. JAMA. 2002;288:3019–3026.

9 Norcini JJ, Boulet JR, Opalek A, et al. The relationship between licensing examination performance and the outcomes of care by international medical school graduates. Acad Med. 2014;89:1157–1162.

10 Young A, Chaudhry HJ, Pei X, et al. FSMB census of licensed physicians in the United States, 2020. J Med Regul. 2021;107: 57–64.

11 Barone MA, Filak AT, Johnson D, Skochelak S, Whelan A. Summary report and preliminary recommendations from the Invitational Conference on USMLE Scoring (InCUS), March 11–12, 2019. Accessed September 13, 2023. https://www.usmle.org/sites/default/files/2021-08/incus_summary_report.pdf.

12 von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. Ann Intern Med. 2007; 147(8):573–577.

13 Centers for Medicare & Medicaid Services. 2022 Condition-Specific Mortality Measures Updates and Specifications Report. Accessed September 29, 2023. https://www.cms.gov/files/document/2022-condition-specific-mortality-measures-updates-and-specifications-report.pdf.

14 Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43:1130–1139.

15 Center for Rural Pennsylvania. Population Density by County. Accessed September 25, 2023. https://www.rural.pa.gov/data/rural-urban-definitions.cfm.

16 Reid RO, Friedberg MW, Adams JL, et al. Associations between physician characteristics and quality of care. Arch Intern Med. 2010;170:1442–1449.

17 Tsugawa Y, Jena AB, Figueroa JF, et al. Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. JAMA Intern Med. 2017;177:206–213.

18 Norcini JJ, Kimball HR, Lipner RS. Certification and specialization: do they matter in the outcome of acute myocardial infarction? Acad Med. 2000;75:1193–1198.

19 Moore DS, McCabe GP, Craig B. Introduction to the Practice of Statistics. 10th ed. Macmillan Learning: Austin, TX; 2021.

20 Weisman SM, Graham DY. Evaluation of the benefits and risks of low-dose aspirin in the secondary prevention of cardiovascular and cerebrovascular events. Arch Intern Med. 2002;162:2197–2202.

21 Kane MT. Validation. In: Brennan RL, ed. Educational Measurement. 4th ed. Westport, CT: Praeger; 2006:17–64.

22 Kane MT. Validating the interpretations and uses of test scores. J Educ Meas. 2013;50:1–73.