

# An Examination of the Associations Among USMLE Step 3 Scores and the Likelihood of Disciplinary Action in Practice

Monica M. Cuddy, PhD, Chunyan Liu, PhD, Wenli Ouyang, MS,  
Michael A. Barone, MD, MPH, Aaron Young, PhD, and David A. Johnson, MA

## Abstract

### Purpose

As the last examination in the United States Medical Licensing Examination (USMLE) sequence, Step 3 provides a safeguard before physicians enter into unsupervised practice. There is, however, little validity research focusing on Step 3 scores beyond examining its associations with other educational and professional assessments thought to cover similar content. This study examines the associations between Step 3 scores and subsequent receipt of disciplinary action taken by state medical boards for problematic behavior in practice. It analyzes Step 3 total, Step 3 computer-based case simulation (CCS), and Step 3 multiple-choice question (MCQ) scores.

### Method

The final sample included 275,392 board-certified physicians who graduated from MD-granting medical schools and who passed Step 3 between 2000 and 2017. Cross-classified multilevel logistic regression models were used to examine the effects of Step 3 scores on the likelihood of receiving a disciplinary action, controlling for other USMLE scores and accounting for jurisdiction and specialty.

### Results

Results showed that physicians with higher Step 3 total, CCS, and MCQ scores tended to have lower chances of receiving a disciplinary action, after accounting for other USMLE scores. Specifically, a 1-standard-deviation

increase in Step 3 total, CCS, and MCQ score was associated with a 23%, 11%, and 17% decrease in the odds of receiving a disciplinary action, respectively. The effect of Step 2 CK score on the likelihood of receiving a disciplinary action was statistically significant, while the effect of Step 1 score became statistically nonsignificant when other Step scores were included in the analysis.

### Conclusions

Physicians who perform better on Step 3 are less likely to receive a disciplinary action from a state medical board for problematic behavior in practice. These findings provide some validity evidence for the use of Step 3 scores when making medical licensure decisions in the United States.

**T**he United States Medical Licensing Examination (USMLE) aims to protect the public by providing medical licensing authorities with information about a physician's competence in the knowledge and skills important for the provision of safe and effective patient care. As the final examination in the USMLE sequence, Step 3 assesses the application of foundational and clinical medical knowledge required "for the unsupervised practice of medicine, with an emphasis on patient management in ambulatory settings."<sup>1</sup> It includes both a multiple-choice question (MCQ) and computer-based case simulation (CCS) component. Overall, CCS is designed to evaluate an examinee's approach to

patient management, including elements of provisional diagnosis, treatment, and monitoring.<sup>1</sup> Through a combination of both content and format, CCS is designed to capture unique information about a physician that other aspects of the licensing examination sequence may not.<sup>2</sup> More specifically, CCS requires that examinees manage a series of virtual patients that present with diverse histories and symptoms in dynamic, interactive, simulated patient care settings.<sup>3</sup>

As with any examination program, the ongoing collection and evaluation of validity evidence for USMLE score interpretations is a central component of ensuring that the USMLE fulfills its mission. One source of validity evidence focuses on an analysis of the relationships among examination outcomes and other external criterion measures thought to measure similar constructs<sup>4</sup> and speaks to the extrapolation element of a validity argument.<sup>5</sup> The extrapolation element of a validity argument focuses on the extent to which the inferences made from examination scores can be

extended to an examinee's ability outside of the test setting, an important aspect of any licensure examination aimed at protecting public safety.

For medical licensure examinations, disciplinary actions may be a particularly useful external criterion measure, as disciplinary actions can be received for behavior that threatens patient safety and can include sanctions up to and including license revocations. Responsible for disciplining physicians for professionally improper or clinically incompetent behavior, state medical boards receive complaints about physicians from patients, hospitals, health care professionals, and other concerned individuals.<sup>6</sup> Once a complaint is received, state medical boards conduct thorough investigations and determine whether a disciplinary action is warranted and if it is, the severity of the action to be taken against the physician.<sup>6</sup> Past investigations of medical licensure examination scores and disciplinary actions generally reveal negative associations, providing some validity evidence for the use and

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

Correspondence should be addressed to Monica M. Cuddy, NBME, 3750 Market St., Philadelphia, PA 19104; telephone: (215) 590-9810; email: mcuddy@nbme.org.

*Acad Med.* 2022;97:1504–1510.

First published online June 7, 2022

doi: 10.1097/ACM.00000000000004775

Copyright © 2022 by the Association of American Medical Colleges

interpretation of licensure examination scores.<sup>7-9</sup> In addition to medical licensure examination performance, other studies identify negative associations between board certification in a specialty area and disciplinary actions, providing some validity evidence for the use and interpretation of certification examination scores.<sup>10-17</sup>

With respect to Step 3, previous validity research has highlighted associations between Step 3 scores and scores from other educational and professional assessments thought to cover similar content. For example, several specialty-specific studies note positive associations between Step 3 scores and performance on in-training examinations<sup>18-20</sup> and board certification examinations.<sup>21-23</sup> Less attention, however, has focused on the extent to which Step 3 scores may relate to future behavior in clinical practice. Moving beyond Step 3's associations with educational and professional assessments, such work would help to expand the types of external criterion measures examined in Step 3 validity research and in doing so provide a more robust evaluation of various sources of validity evidence for Step 3 score uses and interpretations. This in turn would provide some assurance that Step 3 captures information relevant to the licensure process.

To begin to address this gap, this study examines the associations between Step 3 scores and subsequent receipt of disciplinary actions taken by state medical boards for problematic behavior in practice. It analyzes Step 3 total, CCS, and MCQ scores separately. This approach allows for an examination of the total test scores used to make licensure decisions of which CCS is a component, as well as an examination of the scores for the unique competencies that CCS is intended to measure—above and beyond what the MCQs measure—many of which are central to the successful independent practice of medicine.

## Method

### Data and sample

The dataset used in this study was obtained by merging physicians' USMLE performance information from NBME and associated licensure and disciplinary action information from the Federation of State Medical Boards (FSMB). The

licensure and disciplinary data for this study came from FSMB's Physician Data Center, a national repository for state medical boards that provides information about all licensed physicians in the United States and its territorial jurisdictions. All state medical boards provide updated licensure data to the Physician Data Center, with almost all boards providing information monthly. Disciplinary data is provided continuously throughout the year by medical boards typically when action information is finalized and made public.

The final sample included 275,392 board-certified physicians who graduated from MD-granting medical schools and who passed Step 3 between 2000 and 2017. Some physicians passed Step 3 the first time that they took the examination, while others passed it on a repeat attempt. The physicians in the sample represented 50 practice jurisdictions, including Washington, DC, reflecting a national sample. Pertinent data were unavailable for South Dakota which has fewer than 5,600 physicians, representing 0.34% of the total licensed physician population in the United States; this state was removed from the sample. In addition, the physicians in the sample practiced in 17 major medical specialty areas representing a range of training and expertise. As part of a data sharing agreement with the American Board of Medical Specialties, the FSMB receives specialty certification data directly from the American Board of Medical Specialties. The specialty information used in this study stems from this arrangement and as such all physicians in the study sample were board certified.

This study was reviewed and approved by the American Institutes for Research Institutional Review Board.

### Variables

In all analyses, the dependent variable was a binary measure indicating whether a physician had ever received a punitive disciplinary action from a state medical board (0 = no action, 1 = at least one action). A binary measure was used to understand potential differences in Step 3 scores for physicians with disciplinary actions compared with those without them. Here, behaviors that rise to the level of formal disciplinary actions are considered to be meaningfully different,

in general, from behaviors that do not warrant actions.

The primary independent variables included Step 3 total score, Step 3 CCS score, and Step 3 MCQ score. Although separate scores are computed for the CCS and MCQ components of Step 3, it is the total score that is used to make pass/fail decisions and that is reported to examinees. Step 3 total scores are standardized scores that range from 1 to 300. CCS performance is evaluated based on case-specific scoring algorithms that represent codified expert physician-defined criteria.<sup>1,24</sup> CCS cases are scored between 1 and 9, with 9 being the optimum performance. For this study, Step 3 CCS scores were calculated as the average of the case-level 1-to-9 scores. A measure of examinee performance on just the MCQ portion of Step 3 was estimated using the Rasch model<sup>25</sup> to make scores comparable across years; MCQ scores mainly range from -3 to 3, where higher scores indicate better performance. Step 3 total, CCS, and MCQ scores were converted to standardized *z* scores for analysis purposes with a mean of 0 and a standard deviation (SD) of 1 within year. This was done to account for possible differences in scores across a 17-year time period and for ease of interpretation.

To account for possible differences in prior ability and achievement, Step 1 and Step 2 Clinical Knowledge (CK) scores were included in the analyses as covariates. From a validity perspective, this allowed for comparisons of combined and relative effects of each of the steps in the USMLE sequence. Other independent variables treated as covariates included physician gender (0 = female, 1 = male), medical school location (United States = 0, outside of the United States = 1), Step 3 attempt (0 = first-time taker, 1 = repeater), and the number of years a physician had been practicing medicine. Time in practice was used as a proxy for exposure to receive a disciplinary action; in this way, time was controlled for in the analyses. Several studies have shown that male physicians<sup>8-11,13-17,26,27</sup> and physicians who have practiced in medicine longer<sup>8,10,26,27</sup> tend to have a higher risk of receiving a disciplinary action from a state medical board. With respect to medical school location, inconsistent findings have been reported in the literature. For example, several studies

have found that graduates of medical schools outside of the United States are *more* likely to be disciplined by state medical boards.<sup>11,13,27</sup> Yet, other research has found that graduates of medical schools outside of the United States are *less* likely to be disciplined<sup>15</sup> or has demonstrated no significant difference between graduates of U.S. medical schools and graduates of medical schools outside of the United States in terms of the likelihood of receiving a disciplinary action.<sup>10</sup>

### Statistical analyses

Descriptive statistics were computed based on the full sample. In addition, descriptive statistics were computed for USMLE scores and years in practice by physician subgroups (e.g., gender, medical school location). These results provide basic information for understanding the observed patterns as well as for interpreting the results of the inferential statistics.

Physicians are jointly nested in both jurisdiction and specialty in that physicians in the same jurisdiction do not all practice in the same clinical area and physicians in the same specialty do not all practice in the same place. Moreover, previous research shows significant variations in disciplinary actions by medical specialty area<sup>10,11,26,27</sup> and practice jurisdiction.<sup>28–30</sup> Given the data structure and prior research findings, cross-classified multilevel models<sup>31,32</sup> were used to estimate the primary relationships of interest—i.e., between Step 3 scores and receipt of a disciplinary action—while also accounting for other physician-level factors and controlling for both jurisdiction and specialty. To facilitate interpretation of results, all Step 1 and Step 2 CK scores were converted to *z* scores before entry into the models.

Cross-classified multilevel logistic regression models were used to examine the effects of Step 3 total, CCS, and MCQ scores on the likelihood of receiving a disciplinary action. One set of models included Step 3 total score as an independent variable and did not include Step 3 CCS or MCQ scores. This was done because CCS and MCQ performance is already considered in calculations of the Step 3 total score and thus a model including these 3 performance indicators as independent

variables would be conceptually redundant as well as pose possible issues of multicollinearity. A second set of models included Step 3 CCS and MCQ scores as independent variables and did not include Step 3 total score. Including Step 3 CCS and MCQ scores in the same model allows for the estimation of the unique and relative effects of the CCS and MCQ portions of Step 3. All analyses were conducted using R 3.5.1.<sup>33</sup>

### Results

Table 1 summarizes the descriptive statistics for the full sample ( $n = 275,392$ ). Fifty-two percent ( $n = 143,672$ ) of the sample was male, 32% ( $n = 86,984$ ) attended a medical school outside of the United States, and 9% ( $n = 23,485$ ) passed Step 3 on a repeat attempt during the study period. On average, the physicians in the sample had been practicing medicine for 8 years ( $SD = 4$ ). They had an average Step 1 score of 216 ( $SD = 24$ ) and an average Step 2 CK score of 220 ( $SD = 27$ ). Their average Step 3 total score was 212 ( $SD = 20$ ), their average Step 3 CCS score was 5.76 ( $SD = 0.76$ ), and their average Step 3 MCQ score was 1.21 ( $SD = 0.34$ ). One percent ( $n = 2,391$ ) of the sample had received at least one disciplinary action from a state medical board for improper or incompetent patient care.

Table 2 provides the results from the final cross-classified multilevel logistic regression model investigating the association between Step 3 total score and the likelihood of receiving a disciplinary action. Results showed that physicians with higher Step 3 total scores tended to have lower chances of receiving a disciplinary action. Specifically, the odds ratio (OR) was 0.77 ( $P < .001$ ), which indicates that, on average, there was a 23% decrease in the odds that a physician would receive a disciplinary action for each 1-SD increase (the equivalent of about 20 points) in Step 3 total score. The OR for Step 2 CK score was 0.91 ( $P = .004$ ), suggesting a 9% decrease in the odds of receiving a disciplinary action for every 1-SD increase (the equivalent of about 27 points) in Step 2 CK score. With respect to the other covariates included in the model, Step 3 repeaters (OR = 1.23,  $P = .003$ ), physicians who have been practicing for more years (OR = 1.25,  $P < .001$ ), and male physicians (OR = 2.07,  $P < .001$ ) were more likely to receive

a disciplinary action. Step 1 score and medical school location were statistically unrelated to the likelihood of receiving a disciplinary action.

Table 3 presents the results from the final cross-classified multilevel logistic regression model examining the associations between Step 3 CCS and MCQ scores and the likelihood of receiving a disciplinary action. In general, the results mirrored those found for Step 3 total score in that physicians with higher Step 3 CCS and MCQ scores were less likely to receive a disciplinary action. More specifically, the odds that a physician would receive an action, on average, decreased by 11% for each 1-SD increase (the equivalent of about .76 points) in Step 3 CCS score (OR = 0.89,  $P < .001$ ) and by 17% for each 1-SD increase (the equivalent of about .34 logits) in Step 3 MCQ score (OR = 0.83,  $P < .001$ ). The OR for Step 2 CK score was 0.91 ( $P = .003$ ), suggesting a 9% decrease in the odds of receiving a disciplinary action for every 1-SD increase (the equivalent of about 27 points) in Step 2 CK score. Again, Step 3 repeaters (OR = 1.42,  $P < .001$ ), physicians who have been practicing for more years (OR = 1.24,  $P < .001$ ), and male physicians (OR = 2.07,  $P < .001$ ) were more likely to receive a disciplinary action. Consistent with the Step 3 total score model, Step 1 score and medical school location were statistically unrelated to the likelihood of receiving a disciplinary action.

### Discussion

This study examines the associations between physicians' performance on Step 3—the last step in the USMLE licensing examination sequence that physicians take before entering independent practice—and the likelihood of being disciplined by a state medical board for improper or incompetent behavior in practice. To our knowledge, this is the first study to focus on validity evidence for Step 3 scores as they relate to future practice behaviors. Using a large national sample of physicians representing a range of practice jurisdictions and specialty areas, our results show an average expected 23%, 11%, and 17% decrease in the chances of receiving a disciplinary action from a state medical board for each 1-SD increase in Step 3 total, Step 3 CCS, and Step 3 MCQ score, respectively.

**Table 1**  
**Descriptive Statistics for the Study Sample (n = 275,392 Board-Certified Physicians Who Graduated From MD-Granting Medical Schools and Passed USMLE Step 3 Between 2000 and 2017)**

Variable	No. (%)	Years in practice, <sup>a</sup> average (SD)	USMLE Step 1 score, <sup>b</sup> average (SD)	USMLE Step 2 CK score, <sup>b</sup> average (SD)	USMLE Step 3 total score, <sup>b</sup> average (SD)	USMLE Step 3 CCS score, <sup>c</sup> average (SD)	USMLE Step 3 MCQ score, <sup>d</sup> average (SD)
<b>Disciplinary action</b>							
At least one action	2,391 (1)	12 (3)	204 (26)	201 (30)	200 (22)	5.61 (0.80)	1.05 (0.29)
No action	273,001 (99)	8 (4)	216 (24)	220 (27)	212 (20)	5.76 (0.76)	1.21 (0.34)
<b>Gender</b>							
Female	131,720 (48)	8 (4)	214 (24)	221 (27)	213 (19)	5.77 (0.76)	1.23 (0.34)
Male	143,672 (52)	9 (4)	219 (24)	219 (28)	211 (20)	5.75 (0.76)	1.19 (0.34)
<b>Medical school location</b>							
Outside of US	86,984 (32)	8 (4)	208 (28)	210 (32)	200 (20)	5.70 (0.77)	1.01 (0.28)
US	188,408 (68)	9 (4)	220 (22)	225 (23)	217 (17)	5.79 (0.75)	1.30 (0.33)
<b>USMLE Step 3 attempt</b>							
Repeater	23,485 (9)	10 (4)	189 (24)	182 (28)	173 (11)	5.13 (0.84)	0.85 (0.17)
First-time taker	251,907 (91)	8 (4)	219 (23)	224 (25)	215 (16)	5.82 (0.72)	1.24 (0.34)
<i>Total</i>	275,392 (100)	8 (4)	216 (24)	220 (27)	212 (20)	5.76 (0.76)	1.21 (0.34)

Abbreviations: USMLE, United States Medical Licensing Examination; SD, standard deviation; CK, Clinical Knowledge; CCS, computer-based case simulation; MCQ, multiple-choice question; US, United States.

<sup>a</sup>Possible range = 0 to 17.

<sup>b</sup>Possible range = 1 to 300.

<sup>c</sup>Possible range = 1 to 9.

<sup>d</sup>Possible range = -3 to 3.

Because Step 3 is intended to assess application of foundational and clinical medical knowledge essential for the unsupervised practice of medicine, disciplinary actions related to both incompetent and unprofessional behavior may be related to Step 3 performance. Moreover, a portion of Step 3 content covers the following competencies—communication and interpersonal skills,

professionalism, legal and ethical issues, systems-based practice, and patient safety—suggesting that the content measured by Step 3 may overlap with the reasons for which physicians may receive disciplinary actions.

In general, the results of this study provide some validity evidence in support of Step 3 scores for use in determining

medical licensure in the United States, given that licensure is intended to signal readiness for the safe and effective unsupervised practice of medicine. Our findings suggest that how well a physician does on Step 3 may, on average, provide useful information about the safety and effectiveness of their subsequent clinical practice as measured by disciplinary actions.

As noted, the CCS portion of Step 3 represents a unique element within the USMLE sequence that captures examinee behaviors related to patient management in a simulated environment. Overall, this study suggests that CCS scores may contribute important discrete information, above and beyond the information that can be gleaned from Step 3 MCQ scores, for understanding practice patterns as characterized by disciplinary actions received from a state medical board for unprofessional or incompetent behavior. It is possible that both the content and format of CCS provide examinees an opportunity to demonstrate application of knowledge and skills relevant to safe and effective practice in ways that the MCQ portion

**Table 2**  
**Results of Final Cross-Classified Multilevel Logistic Regression Model Predicting Disciplinary Action Using USMLE Step 3 Total Score (n = 275,392 Board-Certified Physicians Who Graduated From MD-Granting Medical Schools and Passed USMLE Step 3 Between 2000 and 2017)**

Variable	B	SE <sub>B</sub>	Odds ratio	95% CI	P value
USMLE Step 3 total score	-0.26	0.03	0.77	0.73–0.83	< .001
USMLE Step 2 CK score	-0.10	0.03	0.91	0.85–0.97	.004
USMLE Step 1 score	-0.04	0.03	0.96	0.90–1.02	.211
USMLE Step 3 repeater	0.20	0.07	1.23	1.07–1.41	.003
More years in practice	0.22	0.01	1.25	1.23–1.26	< .001
Male	0.73	0.05	2.07	1.88–2.29	< .001
Medical school outside of US	-0.00	0.05	1.00	0.90–1.11	.992
Intercept	-7.66				< .001

Abbreviations: USMLE, United States Medical Licensing Examination; SE, standard error; CI, confidence interval; CK, Clinical Knowledge; US, United States.

Downloaded from http://journals.lww.com/academicmedicine by bX+ShHh3+eXkYlbnhYk1trfG9YmoRN60NB0J7WK+68Vh88N9VACV1Yr/DGGeAZwXy8dbhzNLKd+FPvGEdm+qixpeQZ6t5f99+IqemQ6mImvFEI9mFHZpPISNO7XIA2XNRC/+8ZWIzAJXCR/E3BBWbPWONId9sqCX4= on 09/27/2024



Table 3

**Results of Final Cross-Classified Multilevel Logistic Regression Model Predicting Disciplinary Action Using USMLE Step 3 CCS and MCQ Scores (n = 275,392 Board-Certified Physicians Who Graduated From MD-Granting Medical Schools and Passed USMLE Step 3 Between 2000 and 2017)**

Variable	B	SE <sub>B</sub>	Odds ratio	95% CI	P value
USMLE Step 3 CCS score	-0.12	0.02	0.89	0.85–0.92	< .001
USMLE Step 3 MCQ score	-0.18	0.03	0.83	0.78–0.89	< .001
USMLE Step 2 CK score	-0.10	0.03	0.91	0.85–0.97	.003
USMLE Step 1 score	-0.04	0.03	0.96	0.90–1.03	.235
USMLE Step 3 repeater	0.35	0.06	1.42	1.25–1.61	< .001
More years in practice	0.22	0.01	1.24	1.22–1.26	< .001
Male	0.73	0.05	2.07	1.88–2.28	< .001
Medical school outside of US	0.01	0.05	1.01	0.91–1.13	.803
Intercept	-7.66				< .001

Abbreviations: USMLE, United States Medical Licensing Examination; CCS, computer-based case simulation; MCQ, multiple-choice question; SE, standard error; CI, confidence interval; CK, Clinical Knowledge; US, United States.

of Step 3 does not. While CCS may contribute unique information, our results indicate that the negative associations between Step 3 total and MCQ scores and receipt of disciplinary actions may be stronger than the comparable negative associations for CCS scores. Past research documents potential threats to the reliability of CCS scores,<sup>24,34–36</sup> which provides one possible explanation for the smaller association for CCS scores. Arguably, the associations might be greater if CCS scores were more reliable.

It is important to note that the negative association between Step 3 scores and the likelihood of disciplinary action were found after accounting for physicians' performance on Step 1 and Step 2 CK. Consistent with previous research,<sup>8</sup> the negative effect of Step 1 score became statistically indistinguishable from zero after Step 2 CK and Step 3 scores were included in the analysis. This may reflect Step 1's focus on the scientific foundations of medicine, which may be less influential once competence in certain knowledge and skill domains is gleaned from Step 2 CK and Step 3 scores. In addition, the temporal order in which the Step examinations tend to be taken may play a role, as Step 1 is typically taken first, thus, its completion tends to be the most distant from entry into independent practice.

Unlike Step 1, the effect for Step 2 CK scores continued to be statistically distinguishable from zero when Step 3 scores were included in the

analysis. This suggests that both Step 2 CK and Step 3 provide unique information about the likelihood that a physician will receive a disciplinary action for problematic behavior in practice. Furthermore, our results indicate that Step 3 total scores (which, as noted, encompass CCS performance) are more strongly related to the chance of disciplinary action than Step 2 CK scores. This may be due to examination content and timing. Step 2 CK focuses on the knowledge and skills needed for supervised practice, whereas Step 3 focuses on the knowledge and skills needed for unsupervised practice. Therefore, the content included on the Step 3 examination may be more representative of what is required of physicians once they enter independent practice. Additionally, Step 3 is the last licensing examination physicians take before entering independent practice; thus, its stronger effect may be due to its proximity to when disciplinary actions are most likely to be taken.

In terms of the other characteristics examined as covariates in this study, findings support previous research showing that male physicians<sup>8–11,13–17,26,27</sup> and physicians who have been in practice longer<sup>8,10,26,27</sup> are more likely to receive disciplinary actions. In addition, the effect of medical school location was statistically indistinguishable from zero, which one other study has also found.<sup>10</sup> However, the broader findings in the literature regarding differences in

disciplinary actions for graduates of U.S. medical schools compared with graduates of medical schools outside of the United States remain mixed.<sup>11,13,15,27</sup>

This study does have limitations. First, determinations about whether individuals who pass Step 3 provide better patient care than individuals who fail the examination cannot be drawn, since failure prevents the possibility of legal practice. Focusing on individuals who have obtained a medical license, however, allows for an examination of the underlying score scale used to determine pass/fail standards and as such provides an approach for evaluating the relevance of competence in the knowledge and skills measured by Step 3 for performance in unsupervised medical practice.

Second, given that board-certified physicians tend to have higher USMLE scores,<sup>19–23</sup> it may be that our inclusion of only board-certified physicians underestimated the extent of the associations found among Step 3 performance measures and subsequent receipt of disciplinary actions. To examine this possibility, future research should study similar relationships as those examined in the current study for both board-certified and non-board-certified physicians when specialty area data are available for the noncertified group.

Finally, this study treated all disciplinary actions the same and considered neither the types of offenses for which physicians were sanctioned nor the severity of the punishments that they received. It is possible that for offenses linked more closely to the content assessed by Step 3 (e.g., mismanagement of a patient resulting in undue patient harm), the effect of Step 3 scores on the likelihood that a physician will receive a disciplinary action may be greater. As an initial investigation, this study demonstrates that overall, physicians with higher Step 3 scores are less likely to be disciplined for problematic behavior in practice. Future studies should explore the associations between Step 3 scores and the types of behaviors for which physician receive disciplinary actions and the severity of the sanctions that they receive.

In summary, this study analyzed the associations between 2 medical

Downloaded from http://journals.lww.com/academicmedicine by bX+ShHeh3+eXkyIbnhYkTrt69YmoRN60N8BDJ7WK+68Vh88N9VACV1YrDGGeAZwXy8d8hzNkKd+FPvGEEdnqkxpeQZ6t5f9+IqemQ6mJmVFEI9mFHZpplSN07XlIA2XNRC/+82WZAJXCR/E3B8WbPW0NI0d9qCX4= on 09/27/2024

regulatory practices in the United States—licensure and discipline—both of which have implications for patient safety. By examining the connections between licensure and discipline, it provides a lens through which to view the overarching role of medical regulation in ensuring that physicians are adequately prepared and able to provide safe and effective patient care across the continuum of their learning and career paths. The results of this study suggest that physicians who perform better on Step 3 (i.e., who have higher total, CCS, and MCQ scores) are less likely to receive a disciplinary action from a state medical board for problematic behavior in practice. The results also imply that CCS scores provide unique information above and beyond the information gleaned from the MCQ component of Step 3. As part of ongoing research efforts, this study provides some validity evidence for the use of USMLE Step 3 scores for making medical licensure decisions in the United States.

**Funding/Support:** None reported.

**Other disclosures:** None reported.

**Ethical approval:** This study was reviewed and approved by the American Institutes for Research Institutional Review Board.

**Disclaimers:** The authors are employed by NBME and the Federation of State Medical Boards, co-sponsors of the United States Medical Licensing Examination.

**M.M. Cuddy** is measurement scientist, NBME, Philadelphia, Pennsylvania; ORCID: <https://orcid.org/0000-0002-5756-9113>.

**C. Liu** is senior psychometrician, NBME, Philadelphia, Pennsylvania.

**W. Ouyang** is data analyst III, NBME, Philadelphia, Pennsylvania.

**M.A. Barone** is vice president, Competency-Based Assessment, NBME, Philadelphia, Pennsylvania.

**A. Young** is vice president, Research and Data Integration, Federation of State Medical Boards, Eules, Texas.

**D.A. Johnson** is chief assessment officer, Federation of State Medical Boards, Eules, Texas.

## References

- United States Medical Licensing Examination. Step 3 Overview. Accessed April 12, 2022. <https://www.usmle.org/step-3>.
- Clauser BE, Margolis MJ, Swanson DB. An examination of the contribution of computer-based case simulations to the USMLE Step 3 examination. *Acad Med.* 2002;77:S80–S82.
- Margolis MJ, Clauser BE. A regression-based procedure for automated scoring of a complex medical performance assessment. In: Williamson DM, Mislevy RJ, Bejar II, eds. *Automated Scoring of Complex Tasks in Computer-Based Testing*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.; 2006:123–167.
- American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.
- Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas.* 2013;50:1–73. doi:10.1111/jedm.12000.
- Federation of State Medical Boards. U.S. Medical Regulatory Trends and Actions 2018. Published 2018. Accessed April 12, 2022. <https://www.fsmb.org/siteassets/advocacy/publications/us-medical-regulatory-trends-actions.pdf>.
- Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *N Engl J Med.* 2005;353:2673–2682. doi:10.1056/NEJMsa052596.
- 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. doi:10.1097/ACM.0000000000001747.
- Roberts WL, Gross GA, Gimpel JR, et al. An investigation of the relationship between COMLEX-USA licensure examination performance and state licensing board disciplinary actions. *Acad Med.* 2020;95:925–930. doi:10.1097/ACM.0000000000003046.
- Morrison J, Wickersham P. Physicians disciplined by a state medical board. *JAMA.* 1998;279:1889–1893. doi:10.1001/jama.279.23.1889.
- Kohatsu ND, Gould D, Ross LK, Fox PJ. Characteristics associated with physician discipline: A case-control study. *Arch Intern Med.* 2004;164:653–658. doi:10.1001/archinte.164.6.653.
- Kocher MS, Dichtel L, Kasser JR, Gebhardt MC, Katz JN. Orthopedic board certification and physician performance: An analysis of medical malpractice, hospital disciplinary action, and state medical board disciplinary action rates. *Am J Orthop.* 2008;37:73–75.
- Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Ann Intern Med.* 2008;148:869–876. doi:10.7326/0003-4819-148-11-200806030-00009.
- Lipner RS, Young A, Chaudhry HJ, Duhigg LM, Papadakis MA. Specialty certification status, performance ratings, and disciplinary actions of internal medicine residents. *Acad Med.* 2016;91:376–381. doi:10.1097/ACM.0000000000001055.
- Peabody MR, Young A, Peterson LE, et al. The relationship between board certification and disciplinary actions against board-eligible family physicians. *Acad Med.* 2019;94:847–852. doi:10.1097/ACM.0000000000002650.
- Kopp JP, Ibanez B, Jones AT, et al. Association between American Board of Surgery initial certification and risk of receiving severe disciplinary actions against medical licenses. *JAMA Surg.* 2020;155:e200093. doi:10.1001/jamasurg.2020.0093.
- Nelson LS, Duhigg LM, Arnold GK, Lipner RS, Harvey AL, Reisdorff EJ. The association between maintaining American Board of Emergency Medicine certification and state medical board disciplinary actions. *J Emerg Med.* 2019;57:772–779. doi:10.1016/j.jemermed.2019.08.028.
- Perez JA, Jr, Greer S. Correlation of United States Medical Licensing Examination and Internal Medicine In-Training Examination performance. *Adv Health Sci Educ.* 2009;14:753–758.
- McDonald FS, Zeger SL, Kolars JC. Associations between United States Medical Licensing Examination (USMLE) and Internal Medicine In-Training Examination (IM-ITE) scores. *J Gen Intern Med.* 2008;23:1016–1019.
- Peterson LE, Boulet JR, Clauser BE. Associations between medical education assessments and American Board of Family Medicine certification examination score and failure to obtain certification. *Acad Med.* 2020;95:1396–1403. doi:10.1097/ACM.0000000000003344.
- Fish DE, Radfar-Baublitz L, Choi H, Felsenthal G. Correlation of standardized testing results with success on the 2001 American Board of Physical Medicine and Rehabilitation Part 1 Board Certification Examination. *Am J Phys Med Rehab.* 2003;82:686–691. doi:10.1097/01.PHM.0000083688.23340.EA.
- Dillon GF, Swanson DB, McClintock JC, Gravlee GP. The relationship between the American Board of Anesthesiology Part 1 Certification Examination and the United States Medical Licensing Examination. *J Grad Med Educ.* 2013;5:276–283. doi:10.4300/JGME-D-12-00205.1.
- Klein GR, Austin MS, Randolph S, Sharkey PF, Hilibrand AS. Passing the boards: Can USMLE and Orthopaedic In-Training Examination scores predict passage of the ABOS Part-I examination?. *J Bone and Joint Surg.* 2004;86:1092–1095.
- Harik P, Baldwin P, Clauser BE. Comparison of automated scoring methods for a computerized performance assessment of clinical judgement. *Appl Psychol Meas.* 2013;37:587–597. doi:10.1177/0146621613493829.
- Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.
- Clay SW, Conatser RR. Characteristics of physicians disciplined by the state medical board of Ohio. *J Am Osteopath Assoc.* 2003;103:81–88. doi:10.7556/jaoa.2003.103.2.81.
- Khaliq AA, Dimassi H, Huang C, Narine L, Smego RA. Disciplinary action against

- physicians: Who is likely to get disciplined? *Am J Med.* 2005;118:773–777. doi:10.1016/j.amjmed.2005.01.051.
- 28 Harris JA, Byhoff E. Variations by state in physician disciplinary actions by US medical licensure boards. *BMJ Qual Saf.* 2017;26:200–208.
- 29 Law MT, Hansen Z. Medical licensing board characteristics and physician discipline: An empirical analysis. *J Health Polit Policy Law.* 2010;35:63–93. doi:10.1215/03616878-2009-041.
- 30 Lillis DF, McGrath RJ. Directing discipline: State medical board responsiveness to state legislatures. *J Health Polit Policy Law.* 2017;42:123–165.
- 31 Gelman AB, Hill J. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* New York, NY: Cambridge University Press; 2007.
- 32 Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, CA: Sage Publications; 2002.
- 33 R 3.5.1. R Foundation for Statistical Computing, Vienna, Austria. Accessed April 12, 2022. <https://www.R-project.org>.
- 34 Clauser BE, Swanson DB, Clyman SG. The generalizability of scores from a performance assessment of physicians' patient management skills. *Acad Med.* 1996;71(10 suppl):S109–S111.
- 35 Clauser BE, Margolis MJ, Clyman SG, Ross LP. The generalizability of scores for a performance assessment scored with a computer-automated scoring system. *J Educ Meas.* 1997;37:245–261. doi:10.1111/j.1745-3984.2000.tb01085.x.
- 36 Margolis MJ, Clauser BE, Harik P. Scoring the computer-based case simulation component of USMLE Step 3: A comparison of preoperational and operational data. *Acad Med.* 2004;79(10 suppl):S62–S64.

## Teaching and Learning Moments

### Reading Beyond the Medical Chart



How can one understand a human life by reading a medical chart? Only so much can be conveyed through written words and the rest in conversation. While a patient's social history is often overlooked, these details are the most revealing.

It is late afternoon, and I am a physical medicine and rehabilitation resident working at a rehabilitation hospital. My pager buzzes, telling me that a new admission will arrive in one hour. The admissions liaison—the nurse in charge of identifying patients suitable for our hospital—brings me a large stack of documents made up of daily progress notes, routine lab reports, repeat imaging, and detailed medication changes. Thus, I begin a 90-day journey through my patient's hospital course, one that I have only an hour to traverse. I sift through weeks of data, highlighting what I must know: how to dose his daily medications, how long to continue his antibiotics, and which physicians he needs to follow in the outpatient setting. The details of his complicated hospital course—the emergency surgery, complicated coagulopathy, and unexpected bacteremia—infuse my brain.

When he arrives, I run upstairs to meet him. He is lying in bed talking on the phone, and when I enter, he pauses to tell the person on the line that he has to go. For the most part, patients look to be in better health than their doctors' notes suggest, and I am impressed by the jolly voice that greets me after reading that he had been intubated for weeks.

As our conversation begins, it becomes apparent that he has no interest in quizzing me on his medical complexities. He brushes over the details of his hospitalization and rushes to tell me that he wants to go home soon. He lives in a mobile home, and it scares him to think that this hospital is in the city because he has lived his entire life in the country. He cares for his 95-year-old father whose memory fades by the day—much like his once bright peonies, which have since withered in the summer heat. He is very busy managing his shop, but every week, he takes his father to Sunday brunch. Brunch helps his father remember that the day is Sunday. But for the past 90 days, there have been no more Sundays. His father calls daily, wondering what day it is and wondering when he will come home.

His voice trails off, and I notice him blinking. His snowy white hair is overgrown, but it cannot hide the sadness in his heart. He hastily pushes his spectacles up his nose and reaches into the stash of peppermints hiding in his lap. He offers me a piece of candy as he nibbles on his own. “Remember, that life can be sweet,” he says. I accept the peppermint, and I tell him that we will work to get him home.

Back in the workroom, I pause for a moment. It is an immense task to transcribe a patient's story into a standard note. So often, we get caught up in studying and updating the medical chart. When we get to know our patients

as human beings, however, we realize that providing care goes beyond triple-checking medication dosages. Providing care involves understanding the social context in which our patients live and appreciating their routines outside of the hospital setting. Human beings are gregarious creatures by nature, and a human being is the product of every person, every success, and every misfortune he or she has encountered. Thus, to truly care for a human being, we need to appreciate the social context in which they live. This is especially important in situations where patients are not afforded visitations. This patient taught me that having a plan to care for the loved ones who depend on the patient is equally important as caring for the patient.

The lightness that he felt after our conversation eased my fingers as they clattered across the keyboard. I tap my fingers and shuffle again through his discharge documents, making sure his note is up to date. As I finish my note, I add one directive to his disposition plans: Please call to update his father every Sunday.

#### Theodora Lananh Swenson, MD

**T.L. Swenson** is a fourth-year resident physician, Department of Physical Medicine and Rehabilitation, Vanderbilt University Medical Center, Nashville, Tennessee; email: [Theodora.tran@vumc.org](mailto:Theodora.tran@vumc.org); Twitter: [@doctorTheaLS](https://twitter.com/doctorTheaLS); ORCID: <https://orcid.org/0000-0001-9067-3919>.

An Academic Medicine Podcast episode featuring this article is available wherever you get your podcasts.