

Multiple United States Medical Licensing Examination Attempts and the Estimated Risk of Disciplinary Actions Among Graduates of U.S. and Canadian Medical Schools

Katie L. Arnhart, PhD, Monica M. Cuddy, MA, David Johnson, MA, Michael A. Barone, MD, MPH, and Aaron Young, PhD

Abstract

Purpose

The United States Medical Licensing Examination (USMLE) recently announced 2 policy changes: shifting from numeric score reporting on the Step 1 examination to pass/fail reporting and limiting examinees to 4 attempts for each Step component. In light of these policies, exam measures other than scores, such as the number of examination attempts, are of interest. Attempt limit policies are intended to ensure minimum standards of physician competency, yet little research has explored how Step attempts relate to physician practice outcomes. This study examined the relationship between USMLE attempts and the likelihood of receiving disciplinary actions from state medical boards.

Method

The sample population was 219,018 graduates from U.S. and Canadian MD-granting medical schools who passed all USMLE Step examinations by 2011 and obtained a medical license in the United States, using data from the NBME and the Federation of State Medical Boards. Logistic regressions estimated how attempts on Steps 1, 2 Clinical Knowledge (CK), and 3 examinations influenced the likelihood of receiving disciplinary actions by 2018, while accounting for physician characteristics.

Results

A total of 3,399 physicians (2%) received at least 1 disciplinary action. Additional attempts needed to pass Steps 1, 2 CK,

and 3 were associated with an increased likelihood of receiving disciplinary actions (odds ratio [OR]: 1.07, 95% confidence interval [CI]: 1.01, 1.13; OR: 1.09, 95% CI: 1.03, 1.16; OR: 1.11, 95% CI: 1.04, 1.17, respectively), after accounting for other factors.

Conclusions

Physicians who took multiple attempts to pass Steps 1, 2 CK, and 3 were associated with higher estimated likelihood of receiving disciplinary actions. This study offers support for licensure and practice standards to account for physicians' USMLE attempts. The relatively small effect sizes, however, caution policy makers from placing sole emphasis on this relationship.

In the United States and its territories, the legal right to practice medicine is conferred at the state level through statutorily established medical licensing authorities collectively referred to as state medical boards. In order for graduates of U.S. and Canadian MD-granting medical schools to be eligible for an unrestricted medical license, they must obtain a medical degree, complete 1 to 3 years of training in an Accreditation Council for Graduate Medical Education (ACGME)-approved residency program, and successfully pass the multipart United States Medical Licensing Examination (USMLE).¹

The USMLE includes 3 steps: Step 1, Step 2, and Step 3. Since 2004, Step 2 has

comprised 2 components: Step 2 Clinical Knowledge (CK) and Step 2 Clinical Skills (CS). As of January 2021, the Step 2 CS examination has been discontinued permanently.² Each of the 3 steps is intended to measure distinct knowledge domains and skill sets relevant to the independent practice of medicine. In aggregate, the USMLE program is designed to evaluate a physician's ability to apply medical knowledge, concepts, and principles and to demonstrate patient-centered skills that foster safe and effective patient care.³⁻⁵

While state medical boards have the ultimate authority to enforce their state's requirements for the licensed practice of medicine, the USMLE develops policy guidelines for the examination sequence to help ensure that physicians who receive a license to practice possess the level of competence required for the provision of safe and effective patient care. In February 2020, the USMLE announced 2 policy changes that may place additional focus on the number of times it takes examinees to pass a Step examination.

First, the USMLE plans to transition Step 1 results from a 3-digit numeric score to a pass/fail only outcome beginning January 26, 2022.⁶ Entities that have traditionally used scores as a way to assess physicians may look to replace scores with other measures, such as the number of attempts it takes to pass the examination sequence.

Second, the USMLE plans to limit examinees to no more than 4 attempts on each step or its component beginning July 1, 2021.⁷ This amends the current USMLE policy, which is set at a 6-attempt limit. Beyond the attempt limits outlined by the USMLE, most state medical boards further restrict how many times an examinee can take a Step examination to be eligible for initial licensure. These additional requirements can vary from state to state. For example, some state medical boards enforce a limit of 2 or 3 attempts per Step examination, while others have a limit of 5 attempts for all Step examinations combined.⁸

Attempt limit policies aim to balance providing individuals sufficient opportunities for demonstrating

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

Correspondence should be addressed to Katie L. Arnhart, 400 Fuller Wisser Rd., Suite 300, Euless, TX 76039; telephone: (817) 868-4000; email: karnhart@fsmb.org.

Acad Med. 2021;96:1319-1323.

First published online June 15, 2021

doi: 10.1097/ACM.0000000000004210

Copyright © 2021 by the Association of American Medical Colleges

competence with protecting the integrity of the examination. Each examination attempt provides individuals with greater exposure to specific USMLE content, which potentially jeopardizes the security of the examination and ultimately the validity of scores. Equally important, given the measurement error associated with any test, additional attempts increase the likelihood that an unqualified examinee who failed on an earlier attempt receives a subsequent passing score.^{9,10} Although instances of these scenarios are relatively low, the use of attempt limits aids in upholding the standards of the USMLE and helps to ensure that individuals who pass have reached minimum competency—both of which are critical to the public protection role of state medical boards.

Focusing primarily on examination scores, previous research has explored how performance on medical licensure examinations relates to practice performance, including patient outcomes and problematic behaviors resulting in disciplinary actions from state medical boards. For example, higher USMLE Step 2 CK scores have been associated with lower patient mortality among international medical school graduates¹¹ and a lower likelihood of receiving disciplinary actions among U.S. medical school graduates.¹² Similarly, physicians with higher Level 3 scores on the Comprehensive Osteopathic Medical Licensing Examination of the United States are less likely to receive disciplinary actions by state medical boards.¹³ In Canada, higher medical licensing examination scores have also been associated with more positive peer assessments of physicians' quality of care¹⁴ and fewer patient complaints.¹⁵ Though not specifically focused on licensure, better performance on American Board of Medical Specialties (ABMS) certification examinations is also associated with lower risk of receiving a disciplinary action.¹⁶⁻¹⁸ Specifically, one recent study found that physicians who obtain American Board of Surgery specialty certification on their first attempt of the examination tend to also have a lower rate of severe disciplinary actions.¹⁹

Despite the body of work related to examination scores and practice performance, there is a dearth of peer-review literature examining how USMLE

attempts might relate to future practice performance, including the sanctioning of physicians through disciplinary actions by state medical boards for improper or incompetent behavior. This study begins to bridge this gap by extending past research and examining how the number of times that it takes an individual to pass Step 1, Step 2 CK, and Step 3 of the USMLE relates to the likelihood that a physician will receive a disciplinary action from a state medical board. In general, better understanding the associations between USMLE attempts and receipt of disciplinary actions in practice may help stakeholders such as state medical boards evaluate and inform their policies and practices related to the USMLE program.

Method

Data and sample

We obtained data used for this study from the cosponsors of the USMLE program, the NBME and the Federation of State Medical Boards (FSMB). Data were linked by examinees' USMLE identification number and other internal identifiers. Specifically, we included USMLE and disciplinary action data for a sample of physicians who graduated from an MD-granting medical school in the United States or Canada; took 1 or more attempts to pass Step 1, Step 2 CK, and Step 3 between 1994 and 2011; and were granted a license from one of the state medical boards in the United States or District of Columbia by 2018. USMLE performance information included the number of attempts and the passing scores for Steps 1, 2 CK, and 3. Step 2 CS was not included in this analysis, primarily because the scope of the study covers USMLE performance patterns from 1994 to 2011, while Step 2 CS was not implemented until 2004.²⁰

Disciplinary action data from 1994 to 2018 came from the FSMB's Physician Data Center, which collects and stores data on disciplinary actions taken by state medical boards in the United States and its territories. These actions are the formal means by which state medical boards regulate physicians after a finding of professional misconduct or competency deficiency. We excluded board actions that were categorized as administrative or other (e.g., terms of a prior order of the board satisfied, probation terminated, prior agreement modified) from the

analysis in an effort to analyze only actions related to problematic behavior.

A total of 219,086 physicians who graduated from U.S. and Canadian MD-granting medical colleges were identified as having taken and passed USMLE Step 1, Step 2 CK, and Step 3 between 1994 and 2011 and obtained a medical license in 49 states or District of Columbia by 2018 (physicians receiving their first medical license in South Dakota were excluded as that state's data did not include an issue date for the medical licenses). We excluded physicians who had missing information regarding their gender and birth year, leaving a total of 219,018 physicians in the final sample. The American Institutes for Research Institutional Review Board exempted this study from further human subjects review.

Variables

In all analyses, the dependent variable was a binary measure indicating if a physician received a disciplinary action from a state medical board between 1994 and 2018. The primary independent variables included the number of attempts that it took a physician to pass Step 1, Step 2 CK, and Step 3. In addition, to help account for possible differences in competency levels, we included the final passing scores for Step 1, Step 2 CK, and Step 3 in the analyses. We transformed step scores into z-scores for the analyses for ease of interpretation. Other physician-level factors were also accounted for including gender, age when having passed Step 3, year of first state medical license, and current or prior ABMS specialty certification as of 2018. Year of first state medical license was included as a proxy to measure how long a physician had been practicing, as time in practice can affect the probability of receiving an action.

Analyses

We used logistic regression techniques²¹ to estimate the associations between the number of attempts it took to pass Steps 1, 2 CK, and 3 and receipt of a disciplinary action in practice. Specifically, for each Step examination, separate models estimated the likelihood that a physician would receive a disciplinary action in practice as a function of the number of times it took to pass the particular Step examination, after accounting for gender, age when having passed Step 3, license

year, ABMS certification, and score for that respective Step examination. We used SAS Enterprise software, version 6.1, for the analyses (SAS Institute; Cary, North Carolina).

Results

Table 1 provides summary statistics for the study sample. The sample of physicians ($n = 219,018$) was 54% male ($n = 118,921$) and on average 30 years old (standard deviation [SD] = 3 years) when passing Step 3. Physicians in the sample received their first medical license between 1994 and 2018, with 2005 as the average issue year (SD = 5 years). The vast majority ($n = 209,486$, 96%) of the physicians had a current or prior ABMS certification. The mean USMLE scores for Steps 1, 2 CK, and 3 were 218 (SD = 20), 219 (SD = 22), and 214 (SD = 16), respectively, out of a possible score of 1 to 300. While the mean number of attempts hovered around 1 for Steps 1, 2 CK, and 3 (1.08, SD = 0.40; 1.06, SD = 0.32; 1.06, SD = 0.33, respectively), the range of times it took individuals to pass each examination varied greatly from 1 to 14 attempts for Step 1 and Step 3 and from 1 to 11 for Step 2 CK. For Step 1, 6% of physicians passed the examination after multiple attempts ($n = 12,202$) and 4% of physicians passed Step 2 CK and Step 3 after multiple attempts ($n = 9,178$; $n = 9,053$, respectively). A total of 2% ($n = 3,399$) of the physicians received at least 1 disciplinary action from a state medical board between 1994 and 2018.

Table 2 shows the results of the 3 step-specific logistic regression models. Findings indicate positive relationships among Step examination attempts and the estimated likelihood of receiving a disciplinary action. With each additional Step 1 attempt, the likelihood of receiving a disciplinary action increased by an estimated 7% (odds ratio [OR]: 1.07, 95% confidence interval [CI]: 1.01, 1.13). Similarly, the likelihood of receiving a disciplinary action increased by an estimated 9% for each additional attempt needed to pass Step 2 CK (OR: 1.09, 95% CI: 1.03, 1.16). For Step 3, each additional attempt was associated with an estimated 11% increase in the likelihood of receiving a disciplinary action (OR: 1.11, 95% CI: 1.04, 1.17).

The models also showed that, on average, male physicians and older physicians

Table 1

Descriptive Summary of Physician Characteristics and USMLE Performance 1994–2011, With Medical License by 2018, From a Study of Examination Attempts and Disciplinary Actions for 219,018 U.S. and Canadian Medical School Graduates

Variable	Measure
Male, no. (%)	118,921 (54)
Age in years, mean (SD)	30 (3)
License year, mean (SD)	2005 (5)
ABMS certified, no. (%)	209,486 (96)
Step 1 score, mean (SD)	218 (20)
Step 2 CK score, mean (SD)	219 (22)
Step 3 score, mean (SD)	214 (16)
Step 1 attempts, mean (SD)	1.08 (0.40)
Step 2 CK attempts, mean (SD)	1.06 (0.32)
Step 3 attempts, mean (SD)	1.06 (0.33)
Disciplinary action, no. (%) ^a	3,399 (2)

Abbreviations: USMLE, United States Medical Licensing Examination; SD, standard deviation; ABMS, American Board of Medical Specialties; CK, Clinical Knowledge.

^aDisciplinary action received by state medical board between 1994 and 2018.

were at an increased risk of receiving disciplinary actions, while physicians who received their licenses more recently, were ABMS certified, and received higher Step scores were associated with a decreased risk of receiving disciplinary actions. Refer to Table 2 for ORs and CIs for these variables.

Discussion

Findings from this study suggest that the more attempts it took physicians to pass Step 1, Step 2 CK, and Step 3 examinations, the more likely it was that they would receive a disciplinary action from a state medical board, after accounting for other physician-level factors such as Step examination scores. Stated another way, among individuals with the same average passing Step score, those who attempted the examination multiple times to achieve that passing score, on average, were estimated to be more likely to be disciplined by a state medical board. It should be noted, however, that the effect sizes are small, and the analyses demonstrate associations and not causation.

The greatest increase in the likelihood of disciplinary action was estimated for Step 3 attempts (11%), whereas the lowest was estimated for Step 1 attempts (7%). This may speak to the order in which the Step examinations are taken and/or the content included on Step 3. The Step 3 examination is taken only after Step 1 and

Step 2 have been passed and is designed to assess if examinees can apply the medical knowledge and biomedical and clinical science needed for unsupervised practice. This may help to explain why Step 3 attempts, compared with Step 1 and 2 CK attempts, were more related to receiving a disciplinary action in practice. Furthermore, while the effect of examination attempts appears smaller than the effect of examination score for each Step examination, the presence of the attempt effect suggests that knowing how many times an examinee took a Step examination, in combination with their score, may be useful for understating future problematic behaviors that could lead to disciplinary actions.

With the announcement that Step 1 will be transitioning to a pass/fail only outcome,⁶ entities that have traditionally used scores may start to look at the number of attempts to pass the USMLE as 1 alternative way to evaluate physicians. Although our findings show that additional attempts on Steps 1, 2 CK, and 3 were related to an increase in the estimated likelihood of receiving a disciplinary action, the results also can serve as a caution against placing sole emphasis on the number of attempts to evaluate physicians due to the small effect sizes found. The USMLE program also plans to implement a policy that reduces examinees to attempt each Step or its component from no more than 6 to 4 attempts.⁷ This study can

Table 2

Odds Ratio Estimates for Receiving a Disciplinary Action 1994–2018, From a Study of Examination Attempts and Disciplinary Actions for 219,018 U.S. and Canadian Medical School Graduates

Variable	Step 1		Step 2 CK		Step 3	
	Odds ratio	95% CI	Odds ratio	95% CI	Odds ratio	95% CI
Gender (male)	2.09 ^a	1.94, 2.26	1.91 ^a	1.77, 2.06	1.89 ^a	1.75, 2.04
Age	1.05 ^a	1.05, 1.06	1.05 ^a	1.04, 1.06	1.05 ^a	1.05, 1.06
License year	0.90 ^a	0.90, 0.91	0.91 ^a	0.91, 0.92	0.90 ^a	0.89, 0.91
ABMS certified	0.24 ^a	0.22, 0.26	0.25 ^a	0.22, 0.27	0.26 ^a	0.23, 0.28
Step 1 score	0.76 ^a	0.73, 0.79	—	—	—	—
Step 1 attempts	1.07 ^b	1.01, 1.13	—	—	—	—
Step 2 CK score	—	—	0.70 ^a	0.67, 0.73	—	—
Step 2 CK attempts	—	—	1.09 ^c	1.03, 1.16	—	—
Step 3 score	—	—	—	—	0.73 ^a	0.70, 0.75
Step 3 attempts	—	—	—	—	1.11 ^a	1.04, 1.17

Abbreviations: CK, Clinical Knowledge; CI, confidence interval; ABMS, American Board of Medical Specialties.

^a*P* < .001.

^b*P* < .05.

^c*P* < .01.

offer state medical boards and other stakeholders some support for the use of USMLE attempt limits, especially those that already have or are considering implementing attempt limits as a part of their medical licensure requirements.

There are, however, some caveats and limitations that should be noted when considering the relationships between examination attempts and disciplinary actions. While disciplinary actions have potentially large repercussions for a physician's career and patient safety, the number of physicians with disciplinary actions represented a very small percentage of our study population. It is also important to consider that most physicians who require more than 1 attempt on any USMLE Step never receive a disciplinary action. Another limitation of this study is that the physicians included in the sample have been licensed and practicing medicine for a limited period; therefore, their potential exposure to receiving a disciplinary action was restricted. Though we included the year physicians received their first license as a proxy for total time physicians could be disciplined, the findings need to be interpreted with this data restriction in mind.

Overall, while our findings indicate an association between USMLE attempts and disciplinary actions, these findings call for additional research to better

understand the relationships between USMLE Step attempts and education- and practice-related outcomes. Future research could examine the relationships between USMLE attempts and USMLE pass rates. This may be especially useful considering the announcement of Step 1 performance reporting from a number score to only a pass/fail outcome. Other future research could explore how the number of attempts needed to pass each Step examination relates to the type and severity of disciplinary actions to help to better conceptualize the underlying knowledge, skills, and abilities reflected in measures of multiple attempts and disciplinary actions, which in turn could aid in the development of educational programs and policies. Furthermore, additional research could also investigate how certain factors related to undergraduate or graduate medical education (e.g., remediation) might mitigate the risk of receiving a disciplinary action among individuals who took multiple attempts to pass a Step examination. Taken together, these types of studies could give a more complete understanding of the associations between USMLE attempts and other performance indicators and would help provide state medical boards, the USMLE program, and other stakeholders added confidence that examinees who require multiple attempts to pass the USMLE are at a higher risk to receive disciplinary actions.

Though the effect sizes are small, to our knowledge, this is the first study to offer empirical evidence that examinees taking multiple attempts to pass USMLE Steps 1, 2 CK, and 3 are associated with a higher risk of receiving a disciplinary action from a state medical board, after accounting for other factors. In turn, these findings provide some support for the use of attempt limits as suggested by the USMLE program and implemented by state medical boards. When minimum standards for medical licensure and practice are established, reviewed, or modified, stakeholders such as state medical boards and the USMLE program may want to take into consideration that number of USMLE attempts is associated with the likelihood that physicians may receive disciplinary actions.

Funding/Support: None reported.

Other disclosures: The authors are employed by the Federation of State Medical Boards (FSMB) and the NBME, which are cosponsors of the United States Medical Licensing Examination (USMLE).

Ethical approval: The American Institutes for Research (AIR) Institutional Review Board exempted this study from further human subjects review on July 05, 2017 (reference number: EX00429).

Data: Data used in this study were from the FSMB and USMLE and have received the necessary permission and review of the manuscript before submission for publication.

K.L. Arnhart is senior research analyst, Research and Data Integration, Federation of State Medical Boards, Eules, Texas; ORCID: <http://orcid.org/0000-0001-9975-6358>.

M.M. Cuddy is measurement scientist, Center for Advanced Assessment, NBME, Philadelphia, Pennsylvania; ORCID: <http://orcid.org/0000-0002-4724-9113>.

D. Johnson is chief assessment officer, Federation of State Medical Boards, Eules, Texas; ORCID: <http://orcid.org/0000-0003-3669-1838>.

M.A. Barone is vice president, Competency Based Assessment, NBME, Philadelphia, Pennsylvania; ORCID: <http://orcid.org/0000-0002-4724-784X>.

A. Young is vice president, Research and Data Integration, Federation of State Medical Boards, Eules, Texas; ORCID: <http://orcid.org/0000-0002-5517-5874>.

References

- 1 Federation of State Medical Boards (FSMB). U.S. medical regulatory trends and actions, 2018. <https://www.fsmb.org/siteassets/advocacy/publications/us-medical-regulatory-trends-actions.pdf>. Published December 3, 2018. Accessed June 2, 2021.
- 2 United States Medical Licensing Examination (USMLE). Announcements: Work to relaunch USMLE Step 2 CS discontinued. <https://www.usmle.org/announcements/?ContentId=309>. Published January 26, 2021. Accessed June 2, 2021.
- 3 United States Medical Licensing Examination (USMLE). Step 1. <https://www.usmle.org/step-1>. Accessed June 2, 2021.
- 4 United States Medical Licensing Examination (USMLE). Step 2 CK. <https://www.usmle.org/step-2-ck>. Accessed June 2, 2021.
- 5 United States Medical Licensing Examination (USMLE). Step 3. <https://www.usmle.org/step-3>. Accessed June 2, 2021.
- 6 United States Medical Licensing Examination (USMLE). InCUS: Invitational Conference on USMLE Scoring. <https://www.usmle.org/inCUS>. Accessed June 2, 2021.
- 7 United States Medical Licensing Examination (USMLE). Change to the USMLE attempt limit policy. <https://www.usmle.org/attemptlimit.html>. Accessed June 2, 2021.
- 8 Federation of State Medical Boards. State specific requirements for initial medical licensure. <https://www.fsmb.org/step-3/state-licensure>. Accessed June 2, 2021.
- 9 Clauser BE, Nungester RJ. Classification accuracy for tests that allow retakes. *Acad Med.* 2001;76(suppl 10):S108–S110.
- 10 Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. *N Engl J Med.* 1978;299:690–694.
- 11 Norcini JJ, Boulet JR, Opalek A, Dauphinee WD. The relationship between licensing examination performance and the outcomes of care by international medical school graduates. *Acad Med.* 2014;89:1157–1162.
- 12 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.
- 13 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.
- 14 Wenghofer E, Klass D, Abrahamowicz M, et al. Doctor scores on national qualifying examinations predict quality of care in future practice. *Med Educ.* 2009;43:1166–1173.
- 15 Tamblin 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.
- 16 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.
- 17 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.
- 18 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.
- 19 Kopp JP, Ibáñez 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:E7–E7.
- 20 Hallock JA, Melnick DE, Thompson JN. The Step 2 Clinical Skills examination. *JAMA.* 2006;295:1123–1124.
- 21 Agresti A, Finlay B. *Statistical Methods for the Social Sciences.* 4th ed. Upper Saddle River, NJ: Prentice Hall; 2009.