Section 2: USMLE Step 1 and Step 2 CK Score Uses and Interpretations (Specialty Studies)

Section Overview

This section includes summaries of articles published between 2004 and 2019. The articles focus on individual specialty areas of practice and predominately seek to examine relationships among USMLE scores (mainly Step 1 and Step 2 CK) and measures of performance in specialty-specific clinical clerkships, residency training programs, and professional practice. The most common specialty area represented is surgery (including various surgical sub-specialties). Other specialty areas covered include anesthesiology, emergency medicine, internal medicine, obstetrics and gynecology, pathology, pediatrics, and psychiatry. Most studies stem from and focus on a single institution, and thus an overall limitation of this body of work is that results may not be generalizable.

Research Summaries and Abstract Links

What was the primary purpose of the study or the research question(s) that the study sought to answer?

The primary purpose of this study was to identify factors that would assist in making decisions about residency selection in radiology. In particular, it focused on the associations between various pieces of information provided in residency applications, including USMLE Step 1 and Step 2 CK scores, and measures of resident clinical performance.

How was the study conducted (i.e., what data and methodologies were used)?

A retrospective cohort study of 27 residents who completed the radiology residency at the University of Pittsburgh Medical Center in 2015 and 2016 was conducted. The dependent measure represented an indicator of clinical performance and was measured by a discordance rate calculated as the percent of discrepant on-call patient case interpretations between the resident and attending with the potential to impact patient care. A total of 67,145 on-call patient cases were analyzed. The following independent variables were examined: medical school grades, Alpha Omega Alpha Honor Society membership, Step 1 and Step 2 CK scores, publication in peer-reviewed journals, and whether the applicant was from a peer institution. Multivariate logistic regression techniques were used to determine significant relationships. Post-hoc analyses investigated the directions of significant findings.

What were the primary results of the study?

The mean major discordance rate was 1.08% (range 0.34%–2.54%). Higher Step 1 scores, publication before residency, and election to the honor society were all statistically significant predictors of lower major discordance rates (P values 0.01, 0.01, and <0.001, respectively). For all significant variables except third-year medicine clerkship grade, higher values were associated with lower discordance rates. Step 2 CK scores were unrelated to discordance rates.

What are the implications of the findings?

Higher Step 1 scores, publications during medical school, and election to an honor society may be indicators of future success in radiology residency training, and as such may provide useful pieces of information for making residency selection decisions. Step 2 CK may be less useful.

What are the limitations of the study?

The study sample was small and thus it may be that limited variation in the independent variables calls into question the reliability of the estimates (e.g., only 2 residents with Step 1 scores in the lowest score category). Relatedly, the range restriction for the independent and dependent variables call into question the practical significance of the findings. For example, 25 of the 27 residents had Step 1 scores that were at least 30 points above the cut-score. Univariate analyses of discordance rates often showed discordance rate differences of a couple tenths of a percent.

Andriole DA, Jeffe DB, Whelan A. What predicts surgical internship performance? American Journal of Surgery. 2004;188(2):161–64.
https://doi.org/10.1016/j.amjsurg.2004.03.003

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This paper examines the performance of surgical residents during their first postgraduate year (PGY-1) and identifies possible predictors of this performance.

How was the study conducted (i.e., what data and methodologies were used)?

At Washington University School of Medicine, a survey was sent to surgical residency program directors who trained 106 surgical graduates from 1997 to 2001. The survey included 21 items and asked program directors to evaluate their trainees based on such constructs as factual knowledge, data collection, interpersonal skills, ethics, and clinical judgment. Survey responses from 14 items were used to compute a single mean composite PGY-1 score for each resident. Multivariate linear regression techniques were utilized to examine the relationships between the PGY-1 composite score and six independent variables: Step 1 score, Step 2 CK score, third-year clerkship GPA, sex, membership in Alpha Omega Alpha, and surgical sub-specialty choice.

What were the primary results of the study?

Responses from program directors who trained 88 out of the 106 surgical graduates (82%) were included in the study sample. The final regression model explained 14% of the total variation in the PGY-1 composite score. Third-year clerkship GPA and Step 2 CK score were the only statistically significant predictors of variation in the composite score. Specifically, Step 2 CK scores were positively related to PGY-1 composite scores (standardized Beta = 0.366, p = 0.020).

What are the implications of the findings?

Trainees with higher Step 2 CK scores performed better during their first year of surgical residency than trainees with lower Step 2 CK scores. When Step 2 CK scores are known, Step 1 scores may not provide additional information with respect to understanding performance during PGY-1 of surgical residency training.

What are the limitations of the study?

This is an older study from 2004 and testing patterns for both Step 1 and Step 2 CK have changed since it was conducted.

Cortez AR, Winer LK, Kim Y, Hanseman DJ, Athota KP, Quillin RC III. Predictors of medical student success on the surgery clerkship. American Journal of Surgery. 2019;217(1):169–74. https://doi.org/10.1016/j.amjsurg.2018.09.021

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study asked the question: What factors predict honors grades in the surgery clerkship at the University of Cincinnati, College of Medicine?

How was the study conducted (i.e., what data and methodologies were used)?

A pre- and post-clerkship survey was sent to 166 third-year medical students. The pre-clerkship survey included 23 questions related to students' expectations about the surgical clerkship and their interest in the field of surgery; it also collected demographic information. The post-clerkship survey included 26 questions focused on students' experiences during the surgery clerkship and their preferred instruction methods. The survey also included items to assess student learning styles and grit (defined as perseverance and passion for long-term goals). The main outcome variable was final clerkship grade, a composite formed from clinical evaluations (50%), NBME subject examination scores (25%), quiz average (10%), and required course activities (15%). Final numeric grades were converted to "Honors" (90%+), "High Pass" (85-89%), or "Pass" (61-84%). Group differences for the primary independent measures (Step 1 score, grit, and learning style) by honors/non-honors grade were compared using t-tests and chi-square analyses. Multivariate logistic regression analyses were performed with a dichotomous variable representing "Honors" (or not "Honors") as the dependent variable and Step 1 score, grit, and learning style as primary independent variables.

What were the primary results of the study?

Sixty-two students (37.4%) completed both the pre- and post-clerkship surveys and were included in the study sample. Compared to students earning a high pass or pass grade, those with honors grades (a) had higher USMLE Step 1 scores (249.6 vs. 239.0, p = .006), (b) had higher grit scores (3.76 vs. 3.42, p=.03), and (c) were more often individual-based learners (75% vs. 45.7%, p=.039). Examining the components of the final clerkship grade, associations were found between Step 1 scores and performance on quizzes and NBME subject examination scores, but not with clinical evaluation scores. Neither grit scores nor learning style were related to performance on the individual components of the final clerkship grade.

What are the implications of the findings?

Step 1 scores were associated with higher final surgery clerkship grades. However, Step 1 scores were not associated with clinical evaluations of student performance. Step 1 scores may not capture certain elements of student performance, particularly those related to demonstrating skills in a clinical setting.

What are the limitations of the study?

This study had a relatively low response rate, thus potentially limiting the generalizability of the results. Also, clerkship grades favored students who excelled on standardized assessments (35% quizzes/NBME subject examinations), potentially confounding the relationship between Step 1 score and clerkship performance. Lastly, no details about the design of the clinical evaluations are provided, but a suggestion is made that they could be improved through the use of EPAs. This raises the question of how reflective the clinical evaluations are to actual student performance.

de Oliveira GS Jr, Akikwala T, Kendall MC, Fitzgerald PC, Sullivan JT, Zell C, McCarthy RJ. Factors affecting admission to anesthesiology residency in the United States: Choosing the future of our specialty. Anesthesiology. 2012;117(2):243–51.
http://anesthesiology.pubs.asahq.org/article.aspx?articleid=1933717#.XFrml-mnH2Q.email

What was primary purpose of the study or the research question(s) that the study sought to answer?
The purpose of this study was to examine factors associated with admission to anesthesiology residency in the United States. In addition to measures of academic success, the study also aims to examine bias in the admission process related to age and gender.

How was the study conducted (i.e., what data and methodologies were used)?
A retrospective cohort study of 1,976 anesthesiology residency applicants from 2010-2011 at Northwestern University was conducted. The primary outcome variable was successful match to an anesthesiology residency. Independent variables included age, gender, ethnicity, USMLE Step 1 and Step 2 CK scores, medical school origin (US vs international), Alpha Omega Alpha honor society membership, class rank, graduate school education (e.g., master's degree or PhD) and number of publications. Data were analyzed using conditional inference tree analysis, propensity score matching, Fisher's exact test, and logistic regression.

What were the primary results of the study?
Of the 1,976 applications, 66% successfully matched to an anesthesiology program. Admission to a residency program was predicted with an overall accuracy of 0.78 (0.87 sensitivity and 0.69 specificity) based on the above-mentioned independent variables. Overall, the strongest predictors of success were graduation from a US medical school and higher Step 2 CK scores. Prior graduate study and number of publications were not associated with successful matching. Analyses also showed differences with respect to age and gender. For both international and US applicants, being female and younger were positively associating with a successful residency match.

What are the implications of the findings?
Anesthesiology residency acceptance was primarily associated with U.S. medical school attendance and USMLE Step 2 CK scores, suggesting that these factors may be useful for making residency entrance decisions. Furthermore, the age and gender differences found in the selection process raise questions about the need for the development of measures that account for gender and age differences by academic programs before submission of a ranking list to the match system.

What are the limitations of the study?
Information often considered by programs in the selection of applicants, such as letters of recommendation and personal statements, was not included in the analysis. This is primarily due to the lack of reliable tools to evaluate recommendation letters and personal statements. However, it is likely that these materials played an important role in the residency selection decision-making process. The paper also ignores the potentially important confounding factors of race and ethnicity.

Dillon GD, 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. Journal of Graduate Medical Education. 2013;6:276-283. https://doi.org/10.4300/JGME-D-12-00205.1

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study examined the relationships between scores on the USMLE (Step 1, Step 2 CK, and Step 3) and first-time pass rates on the American Board of Anesthesiology (ABA) Part 1 Certification Examination. The ABA Part 1 Certification Examination assesses a candidate's basic and clinical knowledge of medical reasoning as it pertains to anesthesiology.

How was the study conducted (i.e., what data and methodologies were used)?

ABA Part 1 Certification Examination first-time pass rates and USMLE scores were gathered for 7,008 examinees who took the ABA Part 1 Certification Examination from 2002 and 2007. Correlations between the four examinations were presented, both in aggregate form and only for USMG students (from medical schools in the US). Basic descriptive statistics were reported for each of the three USMLE examinations by ABA Part 1 Certification Examination first-time pass rate; these basic descriptive statistics were reported in aggregate, as well as comparatively between USMG students and IMG students (from non-US medical schools).

What were the primary results of the study?

Higher scores on each of the three USMLE examinations were associated with higher ABA Part 1 Certification passing rates, for both USMG and IMG students. Correlations between ABA Part 1 Certification Examination scores and USMLE Step 1, Step 2 CK, and Step 3 scores were .59, .56, and .53, respectively.

What are the implications of the findings?

Individuals with higher USMLE scores have higher first-time pass rates for the ABA Part 1 Certification Examination, suggesting that scores from all three USMLE examinations may provide some indication of later performance on specialty-specific certification examinations in anesthesiology.

What are the limitations of the study?

The relationships examined were only descriptive in nature, with each USMLE examination analyzed in isolation. The unique (or at least rank-order) contribution of each part of the USMLE examination was not reported. Results were aggregated across anesthesiology programs; no report of cross-program variation in the results was provided.

Dyrstad BW, Pope D, Milbrandt JC, Beck RT, Weinhoeft AL, Idusuyi OB. Predictive measures of a resident's performance on written orthopaedic board scores. The Iowa Orthopaedic Journal. 2011; 31:238–43.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3215143/

What was the primary purpose of the study or the research question(s) that the study sought to answer?
The study seeks to examine how well USMLE Step 1 scores, USMLE Step 2 CK scores, and Orthopaedic In-training Examination (OITE) scores predict performance on the American Board of Orthopaedic Surgery (ABOS) Part 1 Certification Examination.

How was the study conducted (i.e., what data and methodologies were used)?
Retrospective data from a total of 30 residents from a single residency program were collected across 10 years. Within each year, there were 3 residents. Residents were classified into two groups, depending on whether they scored above or below the 35th percentile on the ABOS exam. Differences in USMLE and OITE scores by ABOS Part 1 Certification Examination group were examined using t-tests. Linear regression analyses were used to examine the relationships between USMLE scores and ABOS Part 1 Certification Examination scores, and OITE scores and ABOS Part 1 Certification Examination scores.

What were the primary results of the study?
The analyses revealed that residents who scored above the 35th ABOS Part 1 Certification Examination percentile had higher Step 1, Step 2 CK, and OITE scores. Step 1 scores, Step 2 CK scores, and OITE scores were positively associated with ABOS Part 1 Certification Examination scores.

What are the implications of the findings?
Results suggest that standardized tests taken before and during residency may be useful for understanding a resident's performance on subsequent certification examinations in orthopaedic surgery. Close monitoring of OITE scores may be helpful for identifying at-risk residents.

What are the limitations of the study?
The analyzed data came from 30 participants in a single residency program, which is a limitation both in terms of sample size and representativeness. Dichotomizing ABOS exam performance at the 35th percentile loses meaningful variation in examinee performance above and below the threshold. The study examines the relationship between tests scores, without considering other qualities that a good orthopaedic surgeon must have that could be better measured by metrics unrelated to examination performance.

Fryer JP, Corcoran N, George B, Wang E, DaRosa D. Does resident ranking during recruitment accurately predict subsequent performance as a surgical resident? Journal of Surgical Education. 2012;69(6):724–30.
https://doi.org/10.1016/j.jsurg.2012.06.010

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study sought to examine the effectiveness of the process by which surgery residents are selected at Northwestern. The aim of the study was to see if the selection processes, particularly the ranking patterns, predict subsequent success in surgery residency.

How was the study conducted (i.e., what data and methodologies were used)?

The sample included 46 candidates who applied to the general surgery residency at Northwestern between 2001 and 2011. Three methods for ranking candidates were compared to 4 measures of resident performance. The ranking methods included: (1) a ranking based on USMLE Step 1 score, (2) an unadjusted ranking based on an academic profile score, program directors' review score, and interview score, and (3) an adjusted ranking based on the elements included in method 2, plus selection committee review. The resident performance measures included: (1) American Board of Surgery In-Training Examination score, PGY-1 resident evaluation grade, overall resident evaluation grade, and faculty ratings. Correlations were used to examine the relationships.

What were the primary results of the study?

The unadjusted ranking (based on academic profile, program directors' review score, and interview score) was a better predictor of resident performance than the adjusted ranking that additionally included selection committee input. Step 1 score was related to the in-training examination scores only.

What are the implications of the findings?

Committee adjustments to rankings based on candidate performance data may not be beneficial for predicting subsequent performance in surgical residency, and as such may not be the best method for making surgery residency selection decisions.

What are the limitations of the study?

The study sample was small, so findings may be unstable. The study's publication date may not reflect changes to USMLE since 2012.

Gelinne A, Zuckerman S, Benzil D, Grady S, Callas P, Durham S. United States Medical Licensing Exam Step I Score as a predictor of neurosurgical career beyond residency. Neurosurgery, 2018;0:1-7.
https://doi.org/10.1093/neuros/nyy313

What was the primary purpose of the study or the research question(s) that the study sought to answer?

The purpose of this study was to examine if USMLE Step 1 scores are predictive of a successful neurosurgical career (after residency). In particular, the study uses Step 1 score to predict certification status, practice type, academic rank, and research productivity.

How was the study conducted (i.e., what data and methodologies were used)?

USMLE Step 1 score data for 1,503 neurosurgery resident applicants from 1997 to 2007 were examined. Of the 1,503 applicants, 460 held an academic position, 871 practiced in community hospitals, and 172 were not practicing neurosurgery. Measures of career success for the neurosurgeons in academia were h-index, National Institutes of Health (NIH) grant funding, and academic rank. Measures of career success for the non-academic group were practice type and American Board of Neurological Surgery (ABNS) certification status (these metrics were also used for the academic group). The study used nonparametric tests and least-squares linear regression to examine the relationships between the independent variable (Step 1 score) and the dependent variables listed above.

What were the primary results of the study?

Neurosurgeons in academic positions had higher Step 1 scores than neurosurgeons in community practice and individuals not practicing neurosurgery. Step 1 scores were unrelated to differences in academic rank, h-index scores, NIH grants, and ABNS certification status.

What are the implications of the findings?

The lack of associations between Step 1 scores and career success in neurosurgery raises questions about the usefulness of Step 1 scores for making selection decisions about entry into neurosurgery residency training programs.

What are the limitations of the study?

One limitation relates to possible ceiling effects, as some of the differences in Step 1 scores, albeit significant, were within the margin of a few questions, in addition to the generally high Step 1 scores observed among examinees entering neurosurgery.

Guffey RC, Rusin K, Chidiac EJ, Marsh HM. The utility of pre-residency standardized tests for anesthesiology resident selection: The place of United States Medical Licensing Examination scores. Anesthesia and Analgesia. 2011;112(1):201–6.
https://doi.org/10.1213/ANE.0b013e3181fcfacd

What was the primary purpose of the study or the research question(s) that the study sought to answer?
This study sought to answer the question: Are USMLE Step 1 and Step 2 CK scores predictive of success on the American Board of Anesthesiology (ABA) Part I Certification Examination?

How was the study conducted (i.e., what data and methodologies were used)?
The sample included 69 anesthesiology residents. USMLE scores (Step 1 and Step 2 CK) and performance on both the ABA Part 1 examination and the ABA/American Society of Anesthesiologists (ASA) in-training examination were examined using regression, adjusted Pearson partial correlation, and analysis of variance.

What were the primary results of the study?
Step 1 and Step 2 CK scores (individually and averaged) were positively associated with ABA Part 1 examination scores and ABA/ASA in-training examination scores.

What are the implications of the findings?
If Step 1 and Step 2 CK scores are used to make selection decisions for anesthesiology residency programs, programs' average ABA Part 1 examination and ABA/ASA in-training examination scores may increase.

What are the limitations of the study?
The study bases its conclusions on a relatively small sample of participants without testing for other confounding factors (e.g. prior anesthesiology training, society memberships). In addition, there is no discussion about the relationships between USMLE scores and other types of knowledge and skills not well measured by multiple-choice standardized tests.

Harmouche E, Goyal N, Pinawin A, Nagarwala J, Bhat R. USMLE scores predict success in ABEM initial certification: A multicenter study. The Western Journal of Emergency Medicine. 2017;18(3):544–49.
https://doi.org/10.5811/westjem.2016.12.32478

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study sought to address whether or not USMLE Step 1, USMLE Step 2 CK, or a Step 1 and Step 2 CK composite score predicted passing scores on the American Board of Emergency Medicine (ABEM) qualifying (written) and oral certification examinations.

How was the study conducted (i.e., what data and methodologies were used)?

This study used a retrospective design to examine the records of 197 residents drawn from nine emergency medicine residency programs. The composite score represented the simple sum of Step 1 and Step 2 CK scores. The Wilcoxon rank sum test (aka the Mann-Whitney U test) was used examine the relationships between USMLE scores and first-time ABEM pass rates. A logistic regression was used to determine the USMLE scores associated with 90%, 95%, and 99% first-time pass rates.

What were the primary results of the study?

The Step 1 scores, Step 2 CK scores, and Step 1/Step 2 CK composite scores were all positively associated with first-time ABEM written board certification passing rates. Higher Step 2 CK and composite scores were associated with first-time ABEM written *and* oral passing rates; Step 1 scores were not. Accordingly, cutoff scores predicting success of first-time pass rates were higher for residents passing both the written and oral boards than just the written boards.

What are the implications of the findings?

The results suggest that emergency medicine residency programs may want to consider Step 1 and Step 2 CK scores as part of the residency selection criteria. Passing the ABEM certification examination on a resident's first attempt is important, as ACGME standards have at times required that 80% of graduates of an emergency medicine residency program pass the ABEM certification examination on the first attempt.

What are the limitations of the study?

A large number of residents (n = 89) were excluded from the analysis due to missing either one of the USMLE scores or a record of their first-time ABEM qualification exam pass status. Additionally, very few residents failed to pass the qualification exam on their first attempt (n = 10) or their qualification and oral exam on the first attempt (n = 14). The range of Step 1 and Step 2 CK scores for applications that were not offered admission were not reported, so it us unknown how the scores of residents compare to their applicant peers.

Kay C, Jackson JL, Frank M. The relationship between internal medicine residency graduate performance on the ABIM certifying examination, yearly in-service training examinations, and the USMLE Step 1 examination. Academic Medicine. 2015;90(1):100–104. https://journals.lww.com/academicmedicine/Fulltext/2015/01000/The_Relationship_Between_Internal_Medicine.28.aspx

What was the primary purpose of the study or the research question(s) that the study sought to answer?
The purpose of this study was to examine the relationships between USMLE Step 1 scores, internal medicine in-training examination scores, and American Board of Internal Medicine (ABIM) Certification Examination scores.

How was the study conducted (i.e., what data and methodologies were used)?
This retrospective study was conducted at the University of Wisconsin and the sample included 241 emergency medicine residents who graduated between 2004 and 2012. Pearson rho, chi-square, and logistic regression were used to examine to extent to which Step 1 scores and in-training examination scores related to ABIM certification examination performance.

What were the primary results of the study?
Step 1 scores were modestly correlated with ABIM Certification Examination scores (.59), as were in-training examination scores (.48-.67). Failing Step 1 increased the likelihood of failing the ABIM certification examination (RR: 2.4; 95% CI: 1.0–5.9. Being in the bottom half of the Step 1 score distribution also increased the likelihood of failing the certification exam (RR: 3.2; 95% CI: 1.7–6.1).

What are the implications of the findings?
Although the relationships are moderate, Step 1 scores and in-training examination scores may be useful for predicting passing scores on subsequent certification examinations in internal medicine. However, other factors not examined by the current study may also help explain variation in passing rates.

What are the limitations of the study?
There are some missing data concerns, as only 81% of the original sample had Step 1 scores available, and only 76% had ABIM certification examination scores available.

Maker VK, Zahedi MM, Villines D, Maker AV. Can we predict which residents are going to pass/fail the oral boards? Journal of Surgical Education. 2012;69(6):705–13. https://doi.org/10.1016/j.jsurg.2012.08.009

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study asked the following question: Can residents at risk of not passing the oral portion of the American Board of Surgery Certifying Examination (ABSCE) on their first try be identified?

How was the study conducted (i.e., what data and methodologies were used)?

Residents that graduated from a large surgical residency program in Chicago over a 10 year period (2001 – 2010) were included in the study, resulting in a sample of 30 participants. The study was a retrospective study that examined the influence of "knowledge" (USMLE Step 1 score, USMLE Step 2 CK score, ABSQE pass rate [ABS written examination prior to oral boards], ABSITE scores [ABS practice examinations], in-house mock oral board scores, city-wide mock oral board scores), "professional communication" (in-house mock oral board scores, city-wide mock oral board scores), and "spoken English" (native vs non-native speaker, US or International medical school graduate) scores on first-time pass rates of the ABSCE oral boards. There were 17 residents that passed the board on their first try and 13 that did not. Two-way ANOVAs (first-time pass status x independent variable category) and chi-square tests were used to examine differences between the groups.

What were the primary results of the study?

Among the "knowledge" factors, Step 2 CK scores and in-house mock oral board scores were positively related to passing the ASBCE on the first try. Among the "professional communication" factors, in-house mock oral board scores and city-wide mock oral board scores were positively related to passing the ASCBE on the first try. "Spoken English" factors were unrelated to passing the ASCBE on the first try.

What are the implications of the findings?

Residency selections committees may consider using Step 2 CK scores as a factor for making entrance decisions. This is important, as the Residency Review Committee requires a 65% first-time pass rate for accreditation. In-house and city-wide mock oral examinations may be used to identify weaknesses in current residents to help them better prepare for the ABSCE, potentially improving first-time pass rates.

What are the limitations of the study?

The study has an extremely small sample size, and some comparisons (e.g., "professional communication") were available for only a subset of the included sample (n = 22). No adjustments were made to account for small samples sizes, imbalanced group size comparisons, or non-equal group variances to provide more robust results.

Miller BJ, Sexson S, Shevitz S, Peeples D, Van Sant S, McCall WV. US Medical Licensing Exam scores and performance on the psychiatry resident in-training examination. Academic Psychiatry. 2014;38(5):627–31.
https://doi.org/10.1007/s40596-014-0130-y

What was the primary purpose of the study or the research question(s) that the study sought to answer?

The purpose of this study was to examine the relationships between USMLE Step 1 and Step 2 CK scores and performance on the Psychiatry resident in-training examination (PRITE).

How was the study conducted (i.e., what data and methodologies were used)?

Retrospective analysis of Step 1 and Step 2 CK scores for 51 psychiatry residents in a single residency program from 2003-2012 was conducted. Multiple linear regression techniques were used to examine the relationships between USMLE sores and PRITE percentile scores. PRITE scores were stratified by global psychiatry and neurology content with covariates including age, sex, and medical school location (US or international).

What were the primary results of the study?

Step 1 and Step 2 CK scores were positively related to PRITE percentile scores in both global psychiatry and neurology across all levels of training (PGY-1 through PGY-4).

What are the implications of the findings?

The study suggests that Step 1 and 2 CK scores may be useful measures for predicting subsequent performance on in-training exams in a psychiatry residency program.

What are the limitations of the study?

The use of percentile scores, rather than percent correct, on the PRITE exam is a limitation. Another limitation is the amount of missing data, as complete data were available for only 42 PGY1 residents, 36 PGY2 residents, 30 PGY3 residents, and 19 PGY4 residents.

Nagasawa DT, Beckett JS, Lagman C, Chung LK, Schmidt B, Safaee M, Bergsneider M, Martin N, Gaonkar B, Macyszyn L, Yang I. United States Medical Licensing Examination Step 1 scores directly correlate with American Board of Neurological Surgery scores: A single-institution experience. World Neurosurgery. 2017;98: 427–31. https://doi.org/10.1016/j.wneu.2016.11.091

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study asked the question: What is the relationship between USMLE Step 1 scores and American Board of Neurological Surgery (ABNS) Certification Examination scores?

How was the study conducted (i.e., what data and methodologies were used)?

This study was a retrospective review of performance data (Step 1 and ABNS scores) for 21 graduating residents from the UCLA neurosurgery residency program. Linear regression was used to examine the relationship between Step 1 scores and ABNS Certification Examination scores.

What were the primary results of the study?

The mean Step 1 score for the sample was 242.1 (range, 184-271). All residents passed the ABNS Certification Examine, with a mean score of 493.2 (range, 341-651). Step 1 was a positive predictor of ABNS Certification Examination score (R^2 = 0.429, B = 2.587, $p < 0.01$). Step 1 score explained 43% of the variance in ABNS Certification Examination scores.

What are the implications of the findings?

Step 1 scores may provide useful information for predicting scores on the ABNS Certification Examination. Although, since the majority of variation in ABNS Certification Examination scores was not explained by Step 1 score, other factors not examined in the current study may be more associated with ABNS Certification Examination scores. Notably, while a range of Step 1 scores were represented in the sample (184 to 271), all residents still passed the ABNS Certification Examination.

What are the limitations of the study?

The study sample was small, thus limiting the generalizability of the results. No other predictive variables other than Step 1 score were examined with regard to understanding variation in ABNS Certification Examination scores. Since all residents included in the sample passed the ABNS Certification Examination, understanding failure rates was not possible. The relationship between Step 1 scores and ABNS Certification Examination scores may be different than the relationship between Step 1 scores and ABNS Certification Examination pass rates, a potentially important measure of performance to consider.

Perez JA, Greer S. Correlation of United States Medical Licensing Examination and internal medicine in-training examination performance. Advances in Health Sciences Education. 2009;14:753–758
https://doi.org/10.1007/s10459-009-9158-2

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study investigated the relationships between performance on the Internal Medicine In-Training Examination (IM-ITE) and USMLE Step 1, Step 2 CK, and Step 3.

How was the study conducted (i.e., what data and methodologies were used)?

The study sample included 31 residents from an internal medicine residency program who graduated between 1999 and 2006 and had complete data for the ITE and each USMLE examination investigated. The residents also took the ITE in each of their post-graduate training years (PGY). ITE percent correct scores were correlated with Step 1, Step 2 CK, and Step 3 scores within and across PGYs.

What were the primary results of the study?

All correlations between IM-ITE and USMLE scores were statistically significant within each PGY. IM-ITE showed the strongest relationships with Step 2 CK, with correlation coefficients of .79, .70, and .72 for PGY 1, 2, and 3, respectively. IM-ITE score showed lower, but still strong correlations, with Step 1 (.46, .55, .51) and Step 3 (.51, .27, .51) for PGY 1, 2, and 3, respectively. Not surprisingly, IM-ITE scores from each PGY were highly correlated at approximately .80.

What are the implications of the findings?

USMLE scores may provide useful information for internal medicine residency program directors attempting to identify candidates with the potential for achieving high ITE scores. Step 2 CK may be more useful than Step 1 scores.

What are the limitations of the study?

The study's generalizability is limited because the sample consisted of only 31 students. In addition, correlations can be heavily influenced by outliers or other idiosyncratic cases with small samples. Percent correct scores were analyzed across different years of the examination. If the IM-ITE used different forms at different times, it is possible that the percent correct scores may not be comparable if forms varied in difficulty and scores were not equated.

Puscas L, Chang CWD, Lee HJ, Diaz R, Miller R. USMLE and otolaryngology: Predicting board performance. Otolaryngology--Head and Neck Surgery. 2017;156(6):1130–35. https://doi.org/10.1177/0194599817702874

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study investigated the relationships between USMLE (Step 1 and Step 2 CK) performance and performance on the American Board of Otolaryngology Written Qualifying Examination (ABO-WQE).

How was the study conducted (i.e., what data and methodologies were used)?

The study sample included residents who entered otolaryngology-head and neck surgery (OHNS) residency between 2007 and 2009 and took the ABO-WQE for the first time between 2012 and 2014. Of these residents, 611 had Step 1 scores and 402 had both Step 1 and Step 2 CK scores. The study employed a longitudinal retrospective study design. Three separate logistic regression models were used to examine 1) the relationship between Step 1 score and passing the ABO-WQE on the first attempt, 2) the relationship between Step 2 CK score and passing the ABO-WQE on the first attempt, and 3) the combined effect of Step 1 score and Step 2 CK score on passing the ABO-WQE on the first attempt. Additional logistic regression models were used to examine the relationship between Step 1 score and passing the ABO-WQE for residents without Step 2 CK scores and for residents with higher Step 2 CK scores than Step 1 scores.

What were the primary results of the study?

Higher Step 1 and Step 2 CK scores were associated with a greater likelihood of passing the ABO-WQE on the first attempt, but Step 1 demonstrated a stronger relationship. Specially, a 1-point increase in Step 1 score corresponded to a 7% increase in the likelihood of passing the ABO-WQE and a 1-point increase in Step 2 CK score corresponded to a 4% increase. Having a higher Step 2 CK score than Step 1 score was unrelated to the first-time ABO-WQE pass rate.

What are the implications of the findings?

Step 1 score was a better predictor than Step 2 CK for successful passage of the ABO-WQE on the first attempt. Despite these findings, the model showed low sensitivity in predicting failure, indicating that factors beyond USMLE scores may play a role in passing the WQE.

What are the limitations of the study?

One potential limitation relates to missing data: 30% of residents were missing Step 2 CK scores. Another limitation is the lack of data beyond USMLE scores. In other words, the low sensitivity of the model based on USMLE scores alone shows that other factors beyond USMLE may influence performance on the ABO-WQE.

Raman T, Alrabaa RG, Sood A, Maloof P, Benevenia J, Berberian W. Does residency selection criteria predict performance in orthopaedic surgery residency? Clinical Orthopaedics and Related Research. 2016;474(4):908–14.
https://doi.org/10.1007/s11999-015-4317-7

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This paper sought to identify the criteria used to select orthopaedic surgery residency candidates that best predict future success in orthopaedic residency.

How was the study conducted (i.e., what data and methodologies were used)?

This study focused on a single institution and utilized a retrospective study design to analyze the associations between residency selection criteria and residency performance in orthopaedic surgery. The selection criteria included USMLE Step 1 scores, USMLE Step 2 CK scores, MCAT scores, clerkship honors, away rotations, letters of recommendation, Alpha Omega Alpha (AOA) honor medical society membership, completion of sub-internship at the study institution, and number of research publications. The measures of residency performance included American Board of Surgery (ABOS) Part 1 Certification Examination scores, orthopaedics in-training examination (OITE) scores during the last year of residency, global evaluation scores during the last 6 months of residency, and faculty rank of residents. Charts from 60 residents were included in the sample and correlation analyses were used.

What were the primary results of the study?

USMLE Step 2 CK scores and number of clerkship honors received were moderately positively related to ABOS Part I Certification Examination scores. Step 2 CK scores and clerkship honors were weakly positively related to OITE scores. AOA membership was strongly positively related to scores from the global evaluation.

What are the implications of the findings?

The findings suggest that USMLE Step 2 CK scores, clerkship honors, and AOA membership may be useful for predicting success in orthopedic residency as defined in the study. Step 1 scores, and the other residency selection criteria examined, may be less useful.

What are the limitations of the study?

One limitation of the study is the small sample size, which may mean that the correlations found here are unstable. Other limitations include a lack of information provided about the training that faculty received for the global evaluation and ranking, and lack of clarity around the didactic experience of the residency related to OITS and ABOS preparation. Finally, use of evaluations and OITE from the final 6 months of a residency may not be an accurate picture of overall resident performance.

Schrock JB, Kraeutler MJ, Dayton MR, McCarty EC. A cross-sectional analysis of minimum USMLE Step 1 and 2 criteria used by orthopaedic surgery residency programs in screening residency applications. The Journal of the American Academy of Orthopaedic Surgeons. 2017;25(6):464–68.
https://doi.org/10.5435/JAAOS-D-16-00725

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study examined how orthopaedic surgery program directors use USMLE Step 1 and Step 2 CK scores in screening applicants for entry into orthopaedic surgery residencies.

How was the study conducted (i.e., what data and methodologies were used)?

In 2015, a survey was administered to 151 program directors that asked if they used a Step 1 and/or Step 2 CK cut score when reviewing residency applications. If a program director indicated that they did use a cut score, they were then asked to specify the cut score, as well as how they used it and if it had changed over time.

What were the primary results of the study?

A total of 113 program directors responded to the survey, of which 107 provided usable data. Eighty-nine (83%) program directors reported that they used a Step 1 cut score when making residency selection decisions, with mean/median values of 230/231. Seventy-one (80%) program directors reported that the Step 1 score was used as a "hard cutoff." Forty-nine (55%) program directors indicated an increase in the Step 1 and/or Step 2 CK cut score in recent years or indicated that such an increase would occur in the future.

What are the implications of the findings?

The practice of using Step 1 scores to screen orthopaedic surgical residency applicants appears to be widespread, with many programs using a particular Step 1 cut score to make initial decisions. Knowing these cut scores may make the matching process more efficient and set realistic expectations for potential applicants.

What are the limitations of the study?

While the authors received usable data from program directors at 107 of 151 institutions, the practices at 44 institutions remain unidentified and thus results may be impacted by response bias. Also, despite specification by program directors of Step 1 cut scores, the practical implementation of these cut scores is unknown.

Spellacy WN, Downes KL. United States Medical Licensing Examination scores as a predictor of performance on the annual council of resident education in obstetrics and gynecology examinations. The Journal of Reproductive Medicine. 2014:59(1-2):17–19. https://www.ncbi.nlm.nih.gov/pubmed/24597281

What was the primary purpose of the study or the research question(s) that the study sought to answer?
This study examined the relationships between USMLE Step 1 and Step 2 CK scores and scores from the Council of Resident Education in Obstetrics and Gynecology (CREOG) examination.

How was the study conducted (i.e., what data and methodologies were used)?
The sample included 61 obstetrics and gynecology residents who took the CREOG examination in their second and third years of training between 2000 and 2011. Correlations and linear regression models were used to examine the relationships between USMLE and CREOG exam scores. Separate analyses were conducted for CREOG exam scores from the second and third years of training. ROC curves were also used to compare the relative value of Step 1 and Step 2 CK scores for predicting CREOG exam performance during residency training.

What were the primary results of the study?
Overall, results indicate positive relationships between USMLE scores and CREOG exam scores, suggesting that higher USMLE scores are associated with higher CREOG exam scores. Specifically, the correlations between Step 1 scores and CREOG scores at the second and third years of training were .588 and .630, respectively, and the correlations between Step 2 CK scores and CREOG exam scores at the second and third years of training were .667 and .640, respectively. The Step 1 regression coefficients at the second and third years of training were .34 and .41, respectively, and the Step 2 CK regression coefficients at the second and third years of training were .47 and .36, respectively. For CREOG exam scores from the second year, Step 2 CK captured slightly more of the area under the curve (89%), but for CREOG exam scores from the third year, Step 1 scores captured slightly more of the area under the curve (84%).

What are the implications of the findings?
Given the correlation between standardized formative assessments in residency and ultimate board certification, program directors may view screening residency applicants on USMLE scores as useful.

What are the limitations of the study?
The sample is small and the study was conducted at a single institution, so results may not generalize. Data span just over a decade and potential longitudinal effects were not examined. This likely is due to the limitations of very small within-year samples. The regression models have limited covariates, which if examined might have influenced the findings.

Spurlock DR Jr, Holden C, Hartranft T. Using United States Medical Licensing Examination (USMLE) examination results to predict later in-training examination performance among general surgery residents. Journal of Surgical Education. 2010;67(6):452–56. https://doi.org/10.1016/j.jsurg.2010.06.010

What was the primary purpose of the study or the research question(s) that the study sought to answer?
This study examined the relationships between USMLE Step 1 and Step 2 CK scores and American Board of Surgery in-training examination (ABS-ITE) scores for PGY 1-5 residents.

How was the study conducted (i.e., what data and methodologies were used)?
The sample included 34 PGY-1, 2, 3, 4, and 5 residents from two community teaching hospital general surgery residency programs in the Midwest. Stepwise regression analyses were used to estimate the relationships between Step 1 and Step 2 CK scores and ABS-ITE percentile scores.

What were the primary results of the study?
Step 2 CK scores were more highly correlated with ABS-ITE scores than Step 1 scores for all years of residency training. Regression analyses demonstrated that when Step 1 and Step 2 CK scores are both included as independent variables, the effect of Step 1 score is not statistically significant, yet the effect of Step 2 CK score is statistically significant in a positive direction. The strongest effects were seen among Step 2 CK scores and ABS-ITE PGY 3 and PGY 5 scores.

What are the implications of the findings?
When Step 2 CK scores are known, Step 1 scores may be less useful for understanding ABS-ITE scores. Surgery program directors who elect to use USMLE scores for screening and selection could weigh Step 2 CK scores more when ranking potential residents for selection into a surgery residency program.

What are the limitations of the study?
The sample size was small and thus results may be unstable. Also, other factors that may influence ABS-ITE scores were not included in the analyses. Lastly, there are discrepancies in Tables 2 and 3 that make it hard to interpret the results. Some of the statistics from the simple linear regression models (Step 2 CK predicting ABS-ITE scores) presented in Table 3 correspond to the statistics from the first-step in the stepwise regression models presented in Table 2, but strangely others do not (e.g., in Table 2, Step 2 CK B = .437 and p = .037, whereas in Table 3, Step 2 CK B =.416, but p still = .037).

Stain SC, Hiatt JR, Ata A, Ashley SW, Roggin KK, Potts JR, Moore RA, Galante JM, Brit LD, Deveney KE, Ellison C. Characteristics of highly ranked applicants to general surgery residency programs. JAMA Surgery. 2013;148(5):413–17.
https://doi.org/10.1001/jamasurg.2013.180

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study sought to identify the characteristics of top-ranked general surgery residents.

How was the study conducted (i.e., what data and methodologies were used)

Academic measures, such as USMLE Step 1 scores, USMLE Step 2 CK scores, Alpha Omega Alpha (AOA) honor society membership, research experience, and number of publications were compared for the 20 top-ranked applicants from 22 general surgery residency programs. Programs were rated as competitive, very competitive, or highly competitive and applicants' personal statements were rated as best, good, or below average. Differences in the characteristics of the ranked applicants by program competitiveness were examined using chi-square tests, t-tests, and multivariate regression. The characteristics of the 5 top-ranked applicants were compared to the characteristics of the other 15 applicants from each program using multivariate regression.

What were the primary results of the study?

On average, programs ranked as more competitive had applicants with higher mean Step 1 and Step 2 CK scores. More competitive programs also had more applicants with Alpha Omega Alpha (AOA) honor society memberships, research experience, and publications. Within a program, the top five applicants had higher Step 1 scores and more AOA memberships than the other 15 applicants.

What are the implications of the findings?

The highest ranked applicants to surgery residency programs and applicants from more competitive surgery residency programs had higher Step 1 scores, research success, and AOA memberships. These factors may provide useful information for understanding what is considered a desirable candidate for a surgical residency position.

What are the limitations of the study?

This study examined a limited and self-selected sample of residency programs. The program ratings were done by the researchers and thus may be subjective. Additional factors which may impact applicant rankings were not considered (e.g., interview quality).

Sutton E, Richardson JD, Ziegler C, Bond J, Burke-Poole M, McMasters KM. Is USMLE Step 1 score a valid predictor of success in surgical residency? American Journal of Surgery. 2014;208(6):1029–34.
https://doi.org/10.1016/j.amjsurg.2014.06.032

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study examined the relationship between USMLE Step 1 performance and subsequent performance in surgical residency training.

How was the study conducted (i.e., what data and methodologies were used)?

Using a sample of 123 residents from a single institution who trained over a 20-year time period, regression models were estimated to examine the associations between Step 1 scores and rotation evaluations, drop-out status, overall America Board of Surgery (ABS) Certification Examination pass rate, first-time taker ABS Certification Examination pass rate, and retrospective comprehensive faculty evaluations.

What were the primary results of the study?

Step 1 scores were unrelated to rotation evaluations and drop-out status. Residents with higher Step 1 scores were more likely to pass the ABS Certification Examination on their first attempt, and Step 1 scores were moderately associated with retrospective comprehensive faculty ratings.

What are the implications of the findings?

Step 1 scores may be useful for predicting a passing score on the ABS Certification Examination on the first try. Notably, all residents in the study ended up passing the ABS Certification Examination regardless of Step 1 score.

What are the limitations of the study?

The sample size is small and thus results may not be stable. Also, retrospective faculty comprehensive evaluations were based on the judgments of16 faculty. While the scale the faculty used is provided, little information is provided about the criteria used to make judgments using the scale or the consistency of those judgments.

Swanson DB, Sawhill A, Holtzman KZ, Bucak SD, Morrison C, Hurwitz S, DeRosa GP. Relationship between performance on Part I of the American Board of Orthopaedic Surgery Certifying Examination and scores on USMLE Steps 1 and 2. Academic Medicine. 2009;84(10):S21-S24. https://doi.org/10.1097/ACM.0b013e3181b37fd2

What was the primary purpose of the study or the research question(s) that the study sought to answer?
This study investigated the relationships between USMLE Step 1 and USMLE Step 2 CK scores and performance on the American Board of Orthopaedic Surgery (ABOS) Certifying Examination Part I. Associations between USMLE scores and ABOS Certifying Examination Part I scores and associations between USMLE scores and ABOS Certifying Examination Part I first-attempt pass/fail outcomes were both investigated.

How was the study conducted (i.e., what data and methodologies were used)?
The sample included 2,654 examinees that took ABOS Certifying Examination Part 1 for the first time from 2002 to 2006 and took USMLE Step 1 and Step 2 CK from 1992 to 2004. Graduates of international medical schools were excluded from the study. Correlational analyses and multiple linear regression models were used to examine the associations among Step 1 scores, Step 2 CK sores, and ABOS Certifying Examination Part I scores, with mode of Step 1 or Step 2 CK administration (paper-and-pencil vs CBT) as an additional main effect and all interaction effects estimated by the model. Prediction equations were produced for ABOS scores from Step 1 alone, Step 2 CK alone, and Step 1 and Step 2 CK combined. Logistic regression analyses were used to predict ABOS Certifying Examination Part I pass/fail outcomes from Step 1 scores alone and Step 2 CK scores alone.

What were the primary results of the study?
Step 1 scores alone predicted 29% of the variation in ABOS Certifying Examination Part I scores; Step 2 CK scores alone similarly predicted 29% of the variation in ABOS Certifying Examination Part I scores. The combined model predicted 34% of ABOS Certifying Examination Part I score variance. Higher Step 1 and higher Step 2 CK scores were each associated with a greater likelihood of passing ABOS Certifying Examination Part I on the first attempt. Examinees with Step 1 scores of 180 were 10 times more likely to fail ABOS Certifying Exam Part I than those with Step 1 scores of 220, and findings were similar for Step 2 CK scores.

What are the implications of the findings?
Both Step 1 and Step 2 CK scores may be useful to orthopaedic surgery residency program directors in predicting which graduates will pass certifying examinations in orthopaedic surgery.

What are the limitations of the study?
The use of a specialty board examination score as a criterion measure is a limitation, as this may not be the gold standard for assessing clinical competence.

Welch TR, Olson BG, Nelsen E, Dallaghan GLB, Kennedy GA, Botash A. United States Medical Licensing Examination and American Board of Pediatrics Certification Examination results: Does the residency program contribute to trainee achievement. The Journal of Pediatrics. 2017;188:270-274.e3.
https://doi.org/10.1016/j.jpeds.2017.05.057

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study aimed to understand the predictive value of residency training site, USMLE Step 1 score, and USMLE Step 2 CK score on how pediatric residents perform on their first attempt on the American Board of Pediatrics (ABP) Certification Examination.

How was the study conducted (i.e., what data and methodologies were used)?

Data from 1,110 graduates studying at 15 pediatric residency programs were analyzed using t-tests, hierarchical linear models, and linear regression. In the hierarchical models ABP Certification Examination scores were nested within residency programs.

What were the primary results of the study?

Training program accounted for about 10% of the variance in ABP Certification Examination scores. Both Step 1 and Step 2 CK scores contributed to the prediction of ABP Certification Examination performance, although Step 2 CK explained more of the variance in scores than Step 1 alone.

What are the implications of the findings?

Step 1 and Step 2 CK scores may provide program directors with useful information for predicting subsequent specialty examination performance in pediatrics.

What are the limitations of the study?

Data were from a convenience sample of participating residencies, and therefore may not be reflective of the full range of pediatric residencies.

Yousem IJ, Liu L, Aygun N, Yousem DM. United States Medical Licensing Examination Step 1 and 2 scores predict neuroradiology fellowship success. Journal of the American College of Radiology. 2016;13(4):438-444.e2.
https://doi.org/10.1016/j.jacr.2015.10.024

What was the primary purpose of the study or the research question(s) that the study sought to answer?
This study examined whether USMLE scores (Step 1, Step 2 CK, and Step 3) predict neuroradiology fellowship success.

How was the study conducted (i.e., what data and methodologies were used)?
Data was collected for 73 neuroradiology fellows training from 2004-2014 at a single site, including Step 1 score, Step 2 CK score, Step 3 score, and average global assessment rating over the course of fellowship training. The global assessment instrument evaluates a fellow's performance across a host of characteristics (e.g., effectively councils patients, creates clear reports, presentation skills, knowledge of relevant literature, etc.). Fellows were also grouped into three tiers (high, middle, low) based on their quarterly global assessment ratings, with tier membership used as an outcome variable. Correlations were computed and regression analyses were used to examine the relationships between USMLE scores and fellowship success.

What were the primary results of the study?
Step 1, Step 2 CK, and Step 3 scores were strongly associated with higher average resident global assessment ratings. Step 1 and Step 2 CK also are positive predictors of global rating tier membership, even after controlling for gender and medical school attended (US or international).

What are the implications of the findings?
USMLE scores, although quite removed from the time of the fellowship application (~6 years prior), were strongly associated with neuroradiology fellowship success, and therefore may be of use to program directors in neuroradiology.

What are the limitations of the study?
How raters were trained and information about the reliability of ratings from the global assessment instrument were not described.

Zuckerman SL, Kelly PD, Dewan MC, Monrone PJ, Yengo-Kahn AM, Magarik JA, Baticulon RE, Zusman EE, Solomon GS, Wellons JC III. Predicting resident performance from preresidency factors: A systematic review and applicability to neurosurgical training. World Neurosurgery. 2018;110:475-484.
https://www.ncbi.nlm.nih.gov/pubmed/29174240

What was the primary purpose of the study or the research question(s) that the study sought to answer?

This study aimed to identify and categorize existing studies on how pre-residency variables relate to residency success in surgical subspecialty areas.

How was the study conducted (i.e., what data and methodologies were used)?

A systematic review procedure was utilized to search for studies including objective or subjective pre-residency variables and their relationship to residency success, broadly defined, in surgical subspecialty areas. Twenty-one studies met inclusion criteria. Objective measures included examination scores, grades, and dichotomous variables representing whether a degree beyond the MD was obtained and whether applicants were members of an honor society. Subjective measures included letters of recommendation, interview quality, and additional talents. Outcome measures included faculty evaluations, in-training examination scores, board examination scores, and research productivity. Authors summarized the strength of relationships across studies and generated a heat map of results.

What were the primary results of the study?

In general, examination performance in medical school (including USMLE scores), medical school grades, and honors society status were predictive of in-training and board exam performance, and not predictive of faculty evaluations. Rank scores, letters of recommendation, and additional talents were associated with faculty evaluations, yet not highly related to standardized test performance.

What are the implications of the findings?

Prior standardized test performance correlates to in-training and board certification exam performance. More subjective pre-residency factors show moderate levels of correlation with other measures of residency success.

What are the limitations of the study?

This study included articles from across surgical subspecialties; relationships between variables may differ by subspecialty. Reliability estimates are not reported for the study variables, which could impact the strength of the correlations found. Empirical data on other measures of residency success, including patient outcomes, were not analyzed.