JAMA | Original Investigation

Associations of Internal Medicine Residency Milestone Ratings and Certification Examination Scores With Patient Outcomes

Bradley M. Gray, PhD; Jonathan L. Vandergrift, MS; Jennifer P. Stevens, MD; Rebecca S. Lipner, PhD; Furman S. McDonald, MD, MPH; Bruce E. Landon, MD, MBA

IMPORTANCE Despite its importance to medical education and competency assessment for internal medicine trainees, evidence about the relationship between physicians' milestone residency ratings or the American Board of Internal Medicine's initial certification examination and their hospitalized patients' outcomes is sparse.

OBJECTIVE To examine the association between physicians' milestone ratings and certification examination scores and hospital outcomes for their patients.

DESIGN, SETTING, AND PARTICIPANTS Retrospective cohort analyses of 6898 hospitalists completing training in 2016 to 2018 and caring for Medicare fee-for-service beneficiaries during hospitalizations in 2017 to 2019 at US hospitals.

MAIN OUTCOMES AND MEASURES Primary outcome measures included 7-day mortality and readmission rates. Thirty-day mortality and readmission rates, length of stay, and subspecialist consultation frequency were also assessed. Analyses accounted for hospital fixed effects and adjusted for patient characteristics, physician years of experience, and year.

EXPOSURES Certification examination score quartile and milestone ratings, including an overall core competency rating measure equaling the mean of the end of residency milestone subcompetency ratings categorized as low, medium, or high, and a knowledge core competency measure categorized similarly.

RESULTS Among 455 120 hospitalizations, median patient age was 79 years (IQR, 73-86 years), 56.5% of patients were female, 1.9% were Asian, 9.8% were Black, 4.6% were Hispanic, and 81.9% were White. The 7-day mortality and readmission rates were 3.5% (95% CI, 3.4%-3.6%) and 5.6% (95% CI, 5.5%-5.6%), respectively, and were 8.8% (95% CI, 8.7%-8.9%) and 16.6% (95% CI, 16.5%-16.7%) for mortality and readmission at 30 days. Mean length of stay and number of specialty consultations were 3.6 days (95% CI, 3.6-3.6 days) and 1.01 (95% CI, 1.00-1.03), respectively. A high vs low overall or knowledge milestone core competency rating was associated with none of the outcome measures assessed. For example, a high vs low overall core competency rating was associated with a nonsignificant 2.7% increase in 7-day mortality rates (95% CI, −5.2% to 10.6%; $P = .51$). In contrast, top vs bottom examination score quartile was associated with a significant 8.0% reduction in 7-day mortality rates (95% CI, −13.0% to −3.1%; $P = .002$) and a 9.3% reduction in 7-day readmission rates (95% CI, −13.0% to −5.7%; $P < .001$). For 30-day mortality, this association was −3.5% (95% CI, −6.7% to −0.4%; $P = .03$). Top vs bottom examination score quartile was associated with 2.4% more consultations (95% CI, 0.8%-3.9%; $P < .003$) but was not associated with length of stay or 30-day readmission rates.

CONCLUSIONS AND RELEVANCE Among newly trained hospitalists, certification examination score, but not residency milestone ratings, was associated with improved outcomes among hospitalized Medicare beneficiaries.

JAMA. 2024;332(4):300-309. doi:10.1001/jama.2024.5268
Published online May 6, 2024.

Supplemental content

CME Quiz at jamacmelookup.com

Author Affiliations: Assessment and Research, American Board of Internal Medicine, Philadelphia, Pennsylvania (Gray, Vandergrift, Lipner); Division of Pulmonary, Sleep, and Critical Care Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts (Stevens); Harvard Medical School, Boston, Massachusetts (Stevens); J. Edwin Wood Clinic of the Pennsylvania Hospital, Philadelphia (McDonald); Academic and Medical Affairs, American Board of Internal Medicine, Philadelphia, Pennsylvania (McDonald); Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts (Landon); Division of General Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts (Landon).

Corresponding Author: Bradley M. Gray, PhD, American Board of Internal Medicine, Evaluation Research and Development, 510 Walnut St, Ste 1700, Philadelphia, PA 19106 (bgray@abim.org).

Graduates of nearly all US internal medicine residency programs are evaluated with 2 different methods, the Accreditation Council for Graduate Medical Education/American Board of Medical Specialties milestone assessment and the American Board of Internal Medicine (ABIM) internal medicine certification examination.[1]

The internal medicine milestones were implemented in 2013 to create a standardized approach to systematically assessing a resident's progression toward competency and providing feedback to them throughout the course of their training.[2] This program was introduced to address shortcomings in existing residency assessment and feedback methods that lacked clear descriptions of performance expectations, were overly reliant on program directors to evaluate residents, lacked standardized criteria, and did not encourage improvement in competence during residency training.[2] The milestone framework is centered around 6 core competencies evaluated with specific subcompetencies (patient care, medical knowledge, practice-based learning and improvement, systems-based practice, interpersonal and communication skills, and professionalism)[3] that were intended to provide comprehensive feedback throughout residency training.[4]

A second method of assessing the competency of newly trained residents is the ABIM internal medicine board certification examination, which is taken soon after completion of residency and is designed to assess clinical judgment drawn from medical knowledge. In contrast to the milestones, the certification examination does not incorporate direct observations of residents caring for patients. Nonetheless, examination questions are designed by committees composed of diverse groups of practicing physicians to reflect actual clinical scenarios in which physicians are asked to make judgments about patient care.[5]

Despite the importance of assessing the competency for independent practice of newly trained internists, there are few studies assessing the association between internal medicine milestone ratings or the ABIM's initial internal medicine certification examination performance and patient outcomes. We address these gaps after accounting for hospital fixed effects and patient characteristics by examining the relationships between both internal medicine milestone ratings and ABIM's initial certification examination score with hospital outcomes among patients treated by newly trained internists practicing as hospitalists.

Key Points

Question Internal medicine residents' competency is evaluated with milestone ratings and the American Board of Internal Medicine's certification examination. Is physicians' performance on either of these measures related to their hospitalized patients' outcomes?

Findings We analyzed 6898 newly trained hospitalists treating Medicare fee-for-service beneficiaries during 455 120 hospitalizations occurring in 2017 to 2019. We found no associations between overall milestone ratings or medical knowledge ratings and hospitalization outcomes, but certification examination score was associated with reduced 7-day mortality and readmissions.

Meaning Among newly trained hospitalists, certification examination score, but not residency milestone ratings, was associated with improved outcomes among hospitalized Medicare beneficiaries.

Methods

Physician and Hospitalization Sample

Our study sample consisted of internal medicine residents who completed their third year of training in 2016 to 2018, who had valid milestone ratings, and for whom ABIM was able to ascertain their National Provider Identifiers. We identified Medicare fee-for-service beneficiaries older than 65 years who were hospitalized in 2017 to 2019 to whom these physicians had provided care in the hospital by using their National Provider Identifiers to link to their inpatient claims from Medicare's carrier file according to the dates of service. We excluded elective hospitalizations to reduce the possibility of nonrandom assignment of physicians to hospitalizations. We also excluded patients who were in hospice care at admission. In accordance with prior research, we assigned a physician to a hospitalization if that physician had a plurality of inpatient evaluation and management contacts between admission and discharge of that hospitalized patient among physicians with a generalist specialty[6] or, in the case of ties, to the physician with the most evaluation and management charges.[7-9]

We further restricted our sample to hospitalized patients assigned to physicians in our sample after their residency training and before any fellowship training who were practicing as a hospitalist in a particular year according to a validated criterion: having 90% of evaluation and management contacts as inpatient and having at least 100 evaluation and management claims.[10] We then limited these hospitalizations to those in acute short-stay hospitals with 1 of 25 common diagnosis related groups (see eTable 1 in Supplement 1 for diagnosis related group categories) and those occurring in hospitals with at least 100 beds to allow estimation of within-hospital differences.[6] Last, we required that the attributed physicians had cared for the patient within the first 3 days of the stay because this is the period when care quality has the greatest effect on outcomes.

Outcome Measures

Our primary outcomes of interest were 7-day postadmission mortality and 7-day postdischarge readmission. We chose this period because it is more reflective of care processes that occurred during the hospitalization rather than overall illness burden or social determinants.[11-14] Nonetheless, we also examined these outcomes at 30 days, as well as associations with length of stay and subspecialist consultation frequency. Consultation frequency was defined as the number of other unique subspecialists who cared for the patient during the hospital stay.[6] Patients admitted or discharged during the last week or month of the study period were excluded from the relevant mortality and readmission measures to ensure sufficient follow-up time.

Physician Competency Assessments

Milestone ratings measures were based on the 22 subcompetencies for the original milestone program and assessed at the end of training. For these subcompetencies, raters were asked to check 1 of 9 categories that we converted to a numeric scale (eTable 2 in Supplement 1 lists details). For each of the 6 core competency ratings categories, we calculated the mean of the applicable subcompetencies and constructed an overall core competency rating measure that equaled the mean of these 6 core competency mean ratings. We further categorized each of these core competency rating measures as low (<7), medium (≥7 to <8), or high (≥8) based on the distribution of ratings, as well as the category definitions. Using the same approach, we constructed the medical knowledge core competency rating categories from the 2 medical knowledge subcompetencies (see eTable 2 in Supplement 1 for context in terms category meaning).[3]

Our examination measure was the yearly quartile of the 2018 to 2022 internal medicine certification examination score at first attempt. Examination quartile was based on all physicians first attempting the internal medicine certification examination in a particular year. These scores were standardized to account for differences in examination form difficulty.[15] The first examination attempt was chosen because it was taken closest temporally to the residency ratings, and an indicator was added for physicians (N = 279) who did not attempt an examination by 2022 so their milestone ratings could be included.

Additional Variables

For each hospitalization, we analyzed patient demographic information, including age and age squared (to capture nonlinearity), indicators for sex, Medicaid eligibility, and the Research Triangle Institute race and ethnicity categories (Black, Hispanic, and White) included in the Medicare Master Beneficiary Summary File, as well as the physician years of experience, calculated as the difference between the hospital admission year and the residency completion year. Other measures included comorbid conditions identified with the Elixhauser Comorbidity Index and individual indicators for the conditions that are the basis for the index and the 25 diagnosis related group categories.[16] We also determined zip code–level median annual household income from US Census data.

Statistical Analysis

Our study design used the pseudorandom assignment of patients to hospitalists within hospitals and fully accounted for hospital-level quality differences by estimating associations with separate multivariate linear regressions for each outcome that included hospital fixed effects, which effectively compared outcomes among physicians within the same hospital.[17,18] Additional variables accounted for in the regression analyses included the measures noted earlier; an indicator for years since residency (1, 2, or 3 years); indicators for hospitalization year, region, and measures of emergency department visits and hospitalizations during the prior 30, 180, and 365 days; and measures of physician first patient contact timing (day of first patient contact).

We calculated regression-adjusted outcome means by core competency rating and examination score quartile category, as well as percentage difference comparisons using simulations in which we predicted outcomes based on regression coefficient estimates for each hospitalization in our sample, assuming different rating categories or examination score quartiles while holding all other characteristics constant. Percentage difference comparisons were computed as the adjusted mean difference in the outcome of interest between a rating or examination quartile and the applicably lowest rating or examination quartile category divided by the adjusted mean for this lowest (ie, reference) category. This method measures the hypothetical percentage change in the outcome measure associated with an improvement from the lowest rating or examination score category, holding all other factors constant. The SEs considered nonlinearity with the Delta method and accounted for correlated errors due to the nesting of hospitalizations within physicians.[18,19]

Sensitivity Analyses

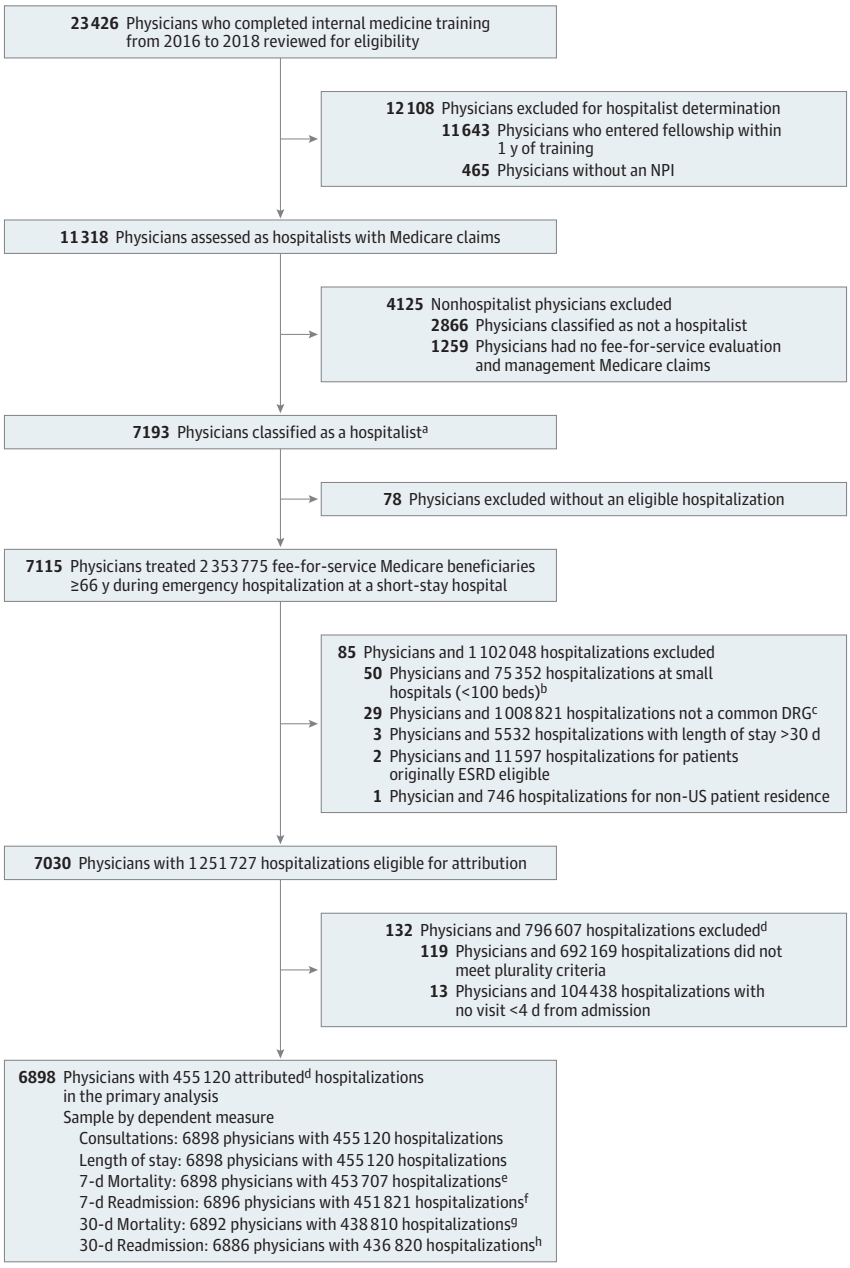
We conducted the following sensitivity analyses: (1) attribution sensitivities (plurality of charges rather than plurality of evaluation and management visits,[7,20,21] majority rather than plurality of evaluation and management visits, and dropping the first physician visit within 3 days of the hospitalization admission criterion); (2) control variable sensitivities (adding control variables observable to stakeholders [physician international medical school attendance and sex]; controlling for residency program 3-year pass rate, which might be related to quality; and including all 6 core competency examination regression controls); and (3) evaluation measure construction sensitivities (using a continuous measure of examination score, a first examination attempt pass/fail, a continuous measure of ratings, and a mid third-year rating to account for social pressure related to terminal-year ratings[22]; and using each subcompetency rating, as well as combining emergency department visits and readmissions or death and readmissions).

The Advarra institutional review board approved our study protocol, and informed consent was waived because the study was viewed as exempt. Analyses were conducted with Stata version 17 (StataCorp). We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines, and 2-sided $P < .05$, calculated with a t test, was considered statistically significant.[23]

Results

Our final study sample included 6898 physicians practicing as hospitalists who finished residency training in 2016-2018 (4125 internists from these residency classes were excluded because they did not meet our hospitalist criteria), who cared for 455 120 hospitalized fee-for-service beneficiaries (1 898 655 hospitalizations were excluded because of factors described in **Figure 1**) in 1911 hospitals (75 352 hospitalizations were excluded because they occurred in short-stay hospitals with fewer than 100 beds). As shown in Figure 1, major hospitalization exclusions were due to limiting the sample to

Figure 1. Cohort Development of Physicians and Hospitalizations in a Study of the Relationship Between Milestones, Board Certification Examinations, and Patient Outcomes



[a]Defined as greater than 100 total evaluation and management contacts, with greater than 90% being inpatient contacts.

[b]Hospitals with fewer than 100 beds were excluded to accommodate hospital fixed effects (ie, small hospitals would be unlikely to have 2 physicians and therefore would not contribute to estimations of associations with evaluation measures).

[c]Included only common DRG codes (193-195, 291-293, 64-66, 177-179, 190-192, 871-872, 689-690, 391-392, 308-310, 377-379, 640-641, 682-684, 602-603, 280-282, 811-812, 388-390, 299-301, 551-552, 393-395, 286-287, 637-639, 947-948, 371-373, 202-203, and 56-57).

[d]Hospitalizations attributed to physicians with a plurality of evaluation and management contacts during that hospitalization among generalist physicians with at least 1 contact within 3 days of the admission.

[e]Excludes 1413 hospitalizations because the admission date was fewer than 7 days before the end of 2021.

[f]Excludes 2 physicians and 3299 hospitalizations because the discharge date was fewer than 7 days before the end of 2019.

[g]Excludes 6 physicians and 16 310 hospitalizations because the admission date was fewer than 30 days before the end of 2019.

[h]Excludes 12 physicians and 18 300 hospitalizations because the discharge date was fewer than 30 days before the end of 2019.

DRG indicates diagnosis related group; ESRD, end-stage renal disease; and NPI, National Provider Identifier.

common diagnosis related groups (1 008 821 exclusions), having hospitalizations in which a hospitalist in our study sample was the attributed physician (692 169 exclusions), and applying the criterion of having a contact within the first 3 days of admission (104 438 exclusions).

For the final sample, 87.3% of hospitalizations occurred in hospitals with 4 or more physicians; the median patient age was 79 years (IQR, 73-86 years); and 43.5% of patients were male, 56.5% were female, 1.9% were Asian, 9.8% were Black, 4.6% were Hispanic, and 81.9% were White (**Table**). Just over one-third of hospitalizations (35.9%) were attributed to female physicians. The unadjusted mortality rate was 3.5% (95% CI, 3.4%-3.6%) at 7 days and 8.8% (95% CI, 8.7%-8.9%) at 30 days, whereas the readmission rate was 5.6% (95% CI, 5.5%-5.6%) at 7 days and 16.6% (95% CI, 16.5%-16.7%) at 30 days. The mean hospitalization length of stay was 3.60 days (95% CI, 3.57-3.62 days) and the number of subspecialist consultations was 1.01 (95% CI, 1.00-1.03).

Patient Characteristics by Overall Milestone Core Competency Rating Category and Examination Score Quartile

One-quarter of hospitalizations (25.7%) were attributed to physicians in the low overall core competency rating category vs 8.2% in the high category (these figures were 25.4% and 8.4%,

Table. Physician Mean Hospitalization Characteristics[a]

	Total hospitalizations (n = 455 120)
Beneficiary characteristics	
Sex, No. (%)	
Male	198 201 (43.5)
Female	256 919 (56.5)
Age, median (IQR), y	79 (73-86)
Race, No. (%)	
Asian	8545 (1.9)
Black	44 629 (9.8)
Hispanic	21 061 (4.6)
White	372 720 (81.9)
Medicare-Medicaid dual eligible, No. (%)	94 120 (20.7)
Family income by zip code, median (IQR) [No.], $	61 250 (48 506-81 075) [447 359]
Weekend admission, No. (%)[b]	59 258 (13.0)
Elixhauser Comorbidity Index score, median (IQR)[c]	6 (4-9)
Comorbid condition indicators, No. (%)	
Cardiovascular	195 950 (43.1)
Diabetes, uncomplicated	178 910 (39.3)
Kidney failure	157 162 (34.5)
Diabetes, complicated	147 710 (32.5)
Depression	124 962 (27.5)
Neurologic disease	90 341 (19.8)
Pulmonary	58 322 (12.8)
Metastatic cancer	18 780 (4.1)
Physician demographic and training characteristics	
Overall milestone rating, median (IQR)[d]	7.00 (6.97-7.35)
Knowledge rating, median (IQR)[e]	7 (7-7)
Certification examinations score, median (IQR) [No.][f]	473 (419-527) [436 371]
Age, median (IQR) [No.], y	33 (31-36) [454 805]
Sex, No. (%)	
Male	291 376 (64.0)
Female	163 744 (36.0)
Medical school, No. (%)	
US based	176 294 (38.7)
International	264 078 (58.0)
Unknown	14 748 (3.2)
Medical degree, No. (%)	
Allopathic	400 007 (87.9)
Osteopathic	55 113 (12.1)
Residency program characteristics	
3-y Pass rate, median (IQR) [No.]	89.4 (82.7-94.2) [449 685]
Residency class size, median (IQR)	18 (12-26)
Rural location, No. (%)[g]	7034 (1.5)
Hospital characteristics	
Bed size, median (IQR)	356 (230-566)
Ownership, No. (%)	
Nonprofit	350 930 (77.1)
For profit	58 327 (12.8)
Governmental	45 863 (10.1)

[a] See eTables 4.1 and 4.2 in Supplement 1 for patient and physician characteristics by rating category or examination quartile category.

[b] Based on Medicare's MedPAR hospital admission date's being on a Saturday or Sunday.

[c] Elixhauser Comorbidity Index score is the count of 31 conditions recorded in Medicare claim *International Statistical Classification of Diseases and Related Health Problems, 10th Revision* codes in the year before the hospital admission (score range, 0-31).

[d] Computed as the mean of the end of the third-year subcompetency ratings for each of 6 core competencies.

[e] Computed as the mean of the end of the third-year subcompetency ratings for the medical knowledge core competency.

[f] First-attempt internal medicine certification examination standard score, ranging from 200 to 800 points, that is equated to account for differences in difficulty between examination forms.

[g] Based on the National Center for Health Statistics urban-rural continuum code for the program's state and county.

respectively, at the physician level), and 87.9% of hospitalizations were attributed to physicians who passed the certification examination on their first attempt. We observed important mismatches between examination score and core competency ratings measures. For example, approximately two-thirds of physicians in the lowest overall core competency rating category were not in the lowest examination score quartile (64.3%) and 15.3% were in the top quartile (eTable 3 in Supplement 1).

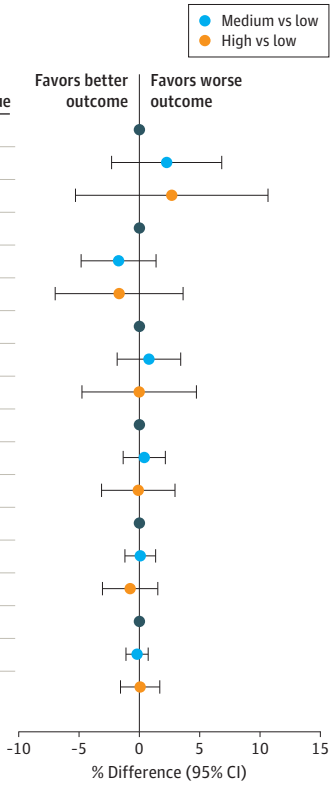
Patient characteristics were evenly distributed across physicians according to their overall core competency rating category and examination score quartile (eTables 4.1 and 4.2 in Supplement 1).

Associations Between Competency Assessments and Outcomes

As shown in **Figure 2** and **Figure 3**, we observed no statistically significant associations between the top vs bottom ratings category for either the overall or medical knowledge core competency rating categories and any hospital outcome measure (see eTables 5.1 and 5.2 in Supplement 1 for unadjusted mean comparison; see eTables 6.1-6.4 in Supplement 1 for regression statistics and coefficient estimates). For example, for the overall core competency ratings measure, the high rating category was associated with a nonsignificant 2.7% increased 7-day mortality rate (95% CI, -5.2% to 10.6%; $P = .51$) and a 1.7% lower 7-day readmission rate (95% CI, -6.9% to 3.6%; $P = .54$) compared with the low rating category. The one exception to this overall pattern was a statistically significant negative association in the rates of 7-day readmissions between the middle vs bottom knowledge rating category (4.3% reduction percentage comparing the middle with the low rating category; 95% CI, −7.7% to −0.8%; $P = .02$).

Figure 2. Overall Core Competency Associations: Adjusted Percentage Difference Compared With the Adjusted Low Ratings Category Outcome

Outcome	Ratings category[a]	Adjusted mean (95% CI)	Adjusted[b] difference vs quartile 1 (95% CI)	*P* value	% Difference[c] (95% CI)	*P* value
7-d Mortality 6898 Physicians 453 707 Hospitalizations	Low	3.4 (3.3 to 3.6)	0 [Reference]	NA	0 [Reference]	NA
	Medium	3.5 (3.4 to 3.6)	0.1 (-0.1 to 0.2)	.32	2.3 (-2.3 to 6.8)	.33
	High	3.5 (3.3 to 3.8)	0.1 (-0.2 to 0.4)	.51	2.7 (-5.2 to 10.6)	.51
7-d Readmissions 6896 Physicians 451 821 Hospitalizations	Low	5.6 (5.5 to 5.8)	0 [Reference]	NA	0 [Reference]	NA
	Medium	5.6 (5.5 to 5.6)	-0.1 (-0.3 to 0.1)	.29	-1.7 (-4.8 to 1.4)	.28
	High	5.6 (5.3 to 5.8)	0.0 (0.0 to 0.0)	.54	-1.7 (-6.9 to 3.6)	.54
30-d Mortality 6892 Physicians 438 810 Hospitalizations	Low	8.7 (8.5 to 8.9)	0 [Reference]	NA	0 [Reference]	NA
	Medium	8.8 (8.7 to 8.9)	0.1 (-0.2 to 0.3)	.55	0.8 (-1.8 to 3.4)	.55
	High	8.7 (8.4 to 9.1)	0.0 (-0.4 to 0.4)	>.99	0 (-4.7 to 4.7)	>.99
30-d Readmissions 6886 Physicians 436 820 Hospitalizations	Low	16.5 (16.3 to 16.8)	0 [Reference]	NA	0 [Reference]	NA
	Medium	16.6 (16.5 to 16.7)	0.1 (-0.2 to 0.4)	.64	0.4 (-1.3 to 2.1)	.64
	High	16.5 (16.1 to 17.0)	0.0 (-0.5 to 0.5)	.96	-0.1 (-3.1 to 2.9)	.96
Consultations 6898 Physicians 455 120 Hospitalizations	Low	1.01 (1.00 to 1.02)	0 [Reference]	NA	0 [Reference]	NA
	Medium	1.01 (1.01 to 1.02)	0.00 (-0.01 to 0.01)	.90	0.1 (-1.2 to 1.3)	.90
	High	1.00 (0.99 to 1.02)	-0.01 (-0.03 to 0.02)	.52	-0.7 (-3.0 to 1.5)	.52
Length of stay 6898 Physicians 455 120 Hospitalizations	Low	3.60 (3.57 to 3.63)	0 [Reference]	NA	0 [Reference]	NA
	Medium	3.59 (3.58 to 3.61)	-0.01 (-0.04 to 0.03)	.70	-0.2 (-1.1 to 0.7)	.70
	High	3.60 (3.55 to 3.65)	0 (-0.06 to 0.06)	.93	0.1 (-1.5 to 1.7)	.93



[a]The overall core competency ratings were categorized as low (<7), medium (≥7 to <8), and high (≥8) based on the mean of 22 subcompetencies assessed at the end of third-year residency; ratings were weighted to give equal weight to each for 6 core competencies.

[b]Adjusted for hospital fixed effects; training and practice year; physician years of experience; hospital day; and patient race and ethnicity, age, sex, Medicaid eligibility, zip code (rural), median income, US Department of Health and Human Services region, Elixhauser Comorbidity Index score and indicators, and diagnosis related group.

[c]Percentage difference equals the difference between the rating category adjusted mean outcome minus the low rating category adjusted mean outcome divided by the low rating category adjusted outcome mean, holding other characteristics constant. NA indicates not applicable.

In contrast, the top examination score quartile was associated with a significant 8.0% reduction in 7-day mortality rate (95% CI, −13.0% to −3.1%; *P* = .002) and 9.3% reduction in 7-day readmission rate (95% CI, −13.0% to −5.7%; *P* < .001) compared with the bottom examination quartile (**Figure 4**). This reduction was 3.5% for the 30-day mortality rate (95% CI, −6.7% to −0.4%; *P* = .03). The absolute percentage point difference in the 30-day mortality measure was similar to that of the 7-day mortality (−0.3 [95% CI, −0.6 to 0.0] vs −0.3 [95% CI, −0.5 to −0.1]), indicating that there was little additional benefit in terms of mortality risk reduction beyond 7 days. Top vs bottom quartile examination score was also associated with a 2.4% increase in consultations (95% CI, 0.8%-3.9%; *P* = .003). We did not observe associations between examination score quartile and 30-day readmission or length of stay.

Sensitivity Analyses

Ratings Sensitivity

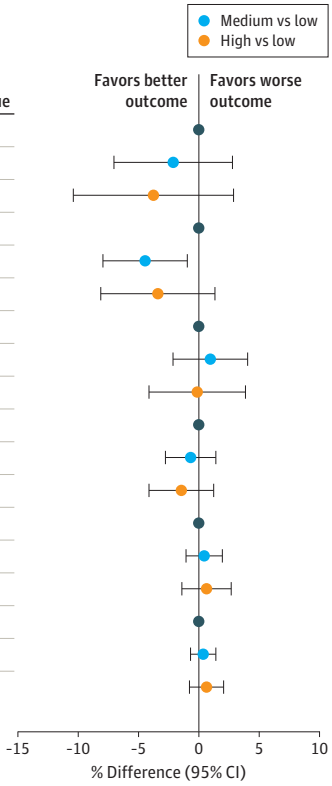
Our results were robust to control variable, most notably controlling for all core competencies, and to attribution choices described in the Methods section (eTables 7.1-7.3 in Supplement 1). Results also were similar when we used continuous rather than discrete measures of ratings and examination score (eTables 7.4 and 7.5 in Supplement 1). Additionally, readmission results were similar when we combined readmission and death or readmission and emergency department visits. We did not observe associations with discharge to hospice care but did observe a positive association between discharge long-term care and examination score quartile but not milestone ratings (eTables 7.6-7.8 in Supplement 1). In our sensitivity analysis, we did not observe consistent associations for each of the 6 core competency milestone ratings measures (eTable 7.9 in Supplement 1) or each of 22 subcompetency measures (eTable 7.10 in Supplement 1) and our base case outcome measures. Sensitivity analysis also indicated that passing the initial certification examination, although correctly signed, was not a statistically significant predictor of any outcomes measure (eTable 7.11 in Supplement 1).

Discussion

Currently, there are 2 principal methods for assessing physician education and competency in internal medicine: the ABIM

Figure 3. Knowledge Core Competency Associations: Adjusted Percentage Difference Compared With the Adjusted Low Ratings Outcome Category

Outcome	Ratings category[a]	Adjusted mean (95% CI)	Adjusted[b] difference vs quartile 1 (95% CI)	*P* value	% Difference[c] (95% CI)	*P* value
7-d Mortality 6898 Physicians 453 707 Hospitalizations	Low	3.6 (3.4 to 3.7)	0 [Reference]	NA	0 [Reference]	NA
	Medium	3.5 (3.4 to 3.6)	-0.1 (-0.2 to 0.1)	.42	-2.0 (-6.8 to 2.8)	.41
	High	3.4 (3.3 to 3.6)	-0.1 (-0.4 to 0.1)	.28	-3.6 (-10.1 to 2.8)	.27
7-d Readmissions 6896 Physicians 451 821 Hospitalizations	Low	5.8 (5.6 to 6.0)	0 [Reference]	NA	0 [Reference]	NA
	Medium	5.5 (5.4 to 5.6)	-0.2 (-0.5 to 0)	.02	-4.3 (-7.7 to -0.8)	.02
	High	5.6 (5.4 to 5.8)	-0.2 (-0.5 to 0.1)	.18	-3.2 (-7.8 to 1.4)	.17
30-d Mortality 6892 Physicians 438 810 Hospitalizations	Low	8.7 (8.5 to 8.9)	0 [Reference]	NA	0 [Reference]	NA
	Medium	8.8 (8.7 to 8.9)	0.1 (-0.2 to 0.3)	.50	1.0 (-2.0 to 4.0)	.51
	High	8.7 (8.5 to 9.0)	0.0 (-0.3 to 0.3)	.97	-0.1 (-4.0 to 3.9)	.97
30-d Readmissions 6886 Physicians 436 820 Hospitalizations	Low	16.7 (16.4 to 17.0)	0 [Reference]	NA	0 [Reference]	NA
	Medium	16.6 (16.5 to 16.7)	-0.1 (-0.4 to 0.2)	.55	-0.6 (-2.6 to 1.4)	.55
	High	16.5 (16.2 to 16.8)	-0.2 (-0.7 to 0.2)	.32	-1.3 (-4.0 to 1.3)	.32
Consultations 6898 Physicians 455 120 Hospitalizations	Low	1.01 (0.99 to 1.02)	0 [Reference]	NA	0 [Reference]	NA
	Medium	1.01 (1.01 to 1.02)	0.01 (-0.01 to 0.02)	.51	0.5 (-1.0 to 2.0)	.51
	High	1.01 (1.00 to 1.03)	0.01 (-0.01 to 0.03)	.49	0.7 (-1.3 to 2.7)	.49
Length of stay 6898 Physicians 455 120 Hospitalizations	Low	3.58 (3.55 to 3.61)	0 [Reference]	NA	0 [Reference]	NA
	Medium	3.60 (3.58 to 3.61)	0.01 (-0.02 to 0.05)	.44	0.4 (-0.6 to 1.4)	.44
	High	3.61 (3.57 to 3.64)	0.02 (-0.02 to 0.07)	.33	0.7 (-0.7 to 2.1)	.33



[a]The knowledge core competency ratings were categorized as low (<7), medium (≥7 to <8), and high (≥8) based on the mean of 2 applicable third-year subcompetency ratings.

[b]Adjusted for hospital fixed effects; training or practice year; physician years of experience; hospital day; and patient race and ethnicity, age, sex, Medicaid eligibility, zip code (rural), median income, HHS region, Elixhauser Comorbidity Index score and indicators, and DRG.

[c]Percentage difference equals the difference between the rating category adjusted mean outcome minus the low rating category adjusted mean outcome divided by the low rating category adjusted outcome mean, holding other characteristics constant. NA indicates not applicable.

certification examination and the milestone residency ratings.[24] In this cross-sectional study of newly trained internists practicing as hospitalists, we found no difference in mortality or readmission rates among hospitalized Medicare beneficiaries cared for by hospitalists who received high vs low residency milestone ratings either overall or specifically for medical knowledge. In contrast, we observed a strong negative association between internal medicine certification examination score quartiles and these outcomes. Overall, results suggest that resident evaluations may be improved by better incorporation of results of standardized examinations such as the in-training examinations.[25]

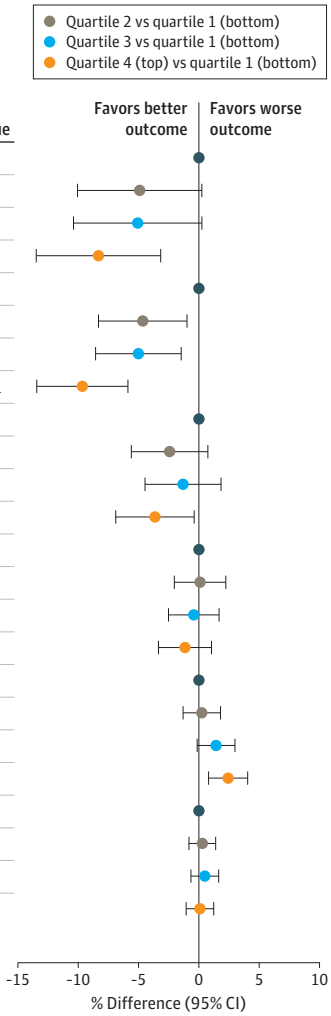
Although to our knowledge our study is unique in examining associations between internal medicine milestone ratings and patient outcomes, 1 prior study reported negative associations between the professionalism and interpersonal communication milestone ratings and patient complaints for newly trained physicians.[26] Another study found no association between surgical milestone ratings and adverse surgical outcomes.[22]

We examined ratings from the end of training, when most residents were deemed to have "crossed the threshold" of competence wherein quality gaps may have been reduced through training, to the point of undetectability in practice. There may be social pressure to give higher terminal ratings to marginal candidates, as was argued in the surgical study, but we found similar associations when we used the midyear rating.[22,26]

Several studies have found midcareer internal medicine maintenance of certification examination score to be associated with outpatient care quality and that both initial certification and maintenance of certification performance were related to disciplinary actions or that certification status and maintenance of certification status, more generally, are related to quality of care, including hospital mortality. Nevertheless, to our knowledge, our study is unique for its assessment of the association for both milestone ratings and certification examination score with hospital outcomes for newly certified internists.[27-39] Although not examining hospital outcomes, one study found that performance on diagnostic related questions in ABIM's maintenance of certification examination was associated with reductions in the risk of hospitalizations and death among patients treated in the outpatient setting for a complaint related to a condition prone to misdiagnosis. In terms of standardized examination score more generally, another study reported associations between performance on the College of Family Physicians Canada

Figure 4. Certifying Examination Quartile Associations: Adjusted Percentage Difference Compared With the Adjusted Bottom Quartile Outcome

Outcome	Quartile[a]	Adjusted[b] mean (95% CI)	Adjusted[b] difference vs quartile 1 (95% CI)	*P* value	% Difference[c] (95% CI)	*P* value
7-d Mortality 6898 Physicians 453 707 Hospitalizations	1	3.7 (3.5 to 3.8)	0 [Reference]	NA	0 [Reference]	NA
	2	3.5 (3.4 to 3.6)	-0.2 (-0.4 to 0.0)	.07	-4.7 (-9.7 to 0.2)	.06
	3	3.5 (3.4 to 3.6)	-0.2 (-0.4 to 0.0)	.07	-4.9 (-10.0 to 0.3)	.06
	4	3.4 (3.2 to 3.5)	-0.3 (-0.5 to -0.1)	.002	-8.0 (-13.0 to -3.1)	.002
7-d Readmissions 6896 Physicians 451 821 Hospitalizations	1	5.8 (5.7 to 6.0)	0 [Reference]	NA	0 [Reference]	NA
	2	5.6 (5.4 to 5.7)	-0.3 (-0.5 to -0.1)	.02	-4.5 (-8.0 to -0.9)	.01
	3	5.6 (5.4 to 5.7)	-0.3 (-0.5 to -0.1)	.007	-4.8 (-8.3 to -1.4)	.006
	4	5.3 (5.1 to 5.4)	-0.5 (-0.8 to -0.3)	<.001	-9.3 (-13.0 to -5.7)	<.001
30-d Mortality 6892 Physicians 438 810 Hospitalizations	1	8.9 (8.8 to 9.1)	0 [Reference]	NA	0 [Reference]	NA
	2	8.7 (8.6 to 8.9)	-0.2 (-0.5 to 0.1)	.14	-2.3 (-5.4 to 0.7)	.014
	3	8.8 (8.7 to 9.0)	-0.1 (-0.4 to 0.2)	.42	-1.3 (-4.3 to 1.8)	.42
	4	8.6 (8.4 to 8.8)	-0.3 (-0.6 to 0.0)	.03	-3.5 (-6.7 to -0.4)	.03
30-d Readmissions 6886 Physicians 436 820 Hospitalizations	1	16.6 (16.4 to 16.9)	0 [Reference]	NA	0 [Reference]	NA
	2	16.6 (16.4 to 16.9)	0 (-0.3 to 0.4)	.92	0.1 (-2.0 to 2.2)	.92
	3	16.6 (16.3 to 16.8)	-0.1 (-0.4 to 0.3)	.70	-0.4 (-2.4 to 1.6)	.70
	4	16.4 (16.2 to 16.7)	-0.2 (-0.5 to 0.2)	.31	-1.1 (-3.2 to 1.0)	.31
Consultations 6898 Physicians 455 120 Hospitalizations	1	1.00 (0.99 to 1.01)	0 [Reference]	NA	0 [Reference]	NA
	2	1.01 (0.99 to 1.02)	0.2 (-1.3 to 1.8)	.75	0.2 (-1.3 to 1.7)	.75
	3	1.02 (1.01 to 1.03)	1.4 (-0.1 to 2.9)	.07	1.4 (-0.1 to 2.9)	.07
	4	1.03 (1.02 to 1.04)	2.4 (0.8 to 3.9)	.003	2.4 (0.8 to 3.9)	.003
Length of stay 6898 Physicians 455 120 Hospitalizations	1	3.59 (3.56 to 3.61)	0 [Reference]	NA	0 [Reference]	NA
	2	3.60 (3.57 to 3.62)	1.0 (-2.8 to 4.9)	.60	0.3 (-0.8 to 1.4)	.60
	3	3.61 (3.58 to 3.63)	1.7 (-2.2 to 5.7)	.39	0.5 (-0.6 to 1.6)	.39
	4	3.59 (3.56 to 3.62)	0.3 (-3.6 to 4.3)	.87	0.1 (-1.0 to 1.2)	.87



[a]Examination score quartiles were based on the initial internal medicine certification examination first-attempt score ranking for the year of this attempt.

[b]Adjusted for hospital fixed effects; training or practice year; physician years of experience; hospital day; and patient race and ethnicity, age, sex, Medicaid eligibility, zip code (rural), median income, US Department of Health and Human Services region, Elixhauser Comorbidity Index score and indicators, and DRG.

[c]Percentage difference equals the difference between the quartile adjusted mean outcome minus the first-quartile adjusted mean outcome divided by the adjusted first-quartile outcome mean, holding other characteristics constant.

certification examination and process measures of care.[40] The researchers reported a positive association between examination score and consultations, findings similar to ours. Other studies reported associations between US licensure examination score and both quality and outcomes of care, including a negative association with mortality.[31,40-43] None of these studies simultaneously considered standardized examination score and residency ratings, the 2 primary forms of evaluation for newly trained internists, or compared outcomes among physicians practicing in the same hospital.

Our findings that higher examination score was associated with increased consultation frequency suggest that the greater use of specialty care among high-performing hospitalists may be due to their ability to distinguish what they know and what they do not know, and thus they may have more insight into when they require additional specialty expertise. However, prior work reported that consulting hospitalists with high vs low performance used more resources without clinical benefit to patients.[6] Moreover, sensitivity analysis indicated that, although correctly signed, associations with first-attempt passing of a certification examination was not statically significant. This result may indicate that either the pass/fail criteria were not stringent enough or there is no one clear cut point in examination score that differentiates physicians across the outcomes we studied (eg, physicians who barely passed but still had low performance were included in the pass group). That our findings in terms of associations with examination quartile were stronger for 7-day vs 30-day mortality and readmissions is consistent with past research indicating that beyond 7 days, these outcomes were more likely to be related to

chronic health conditions and household-level factors beyond hospitalists' control.[11]

Limitations

Our study has limitations. Associations with ratings and examination scores are subject to omitted variable bias because we relied on cross-sectional analysis. The quasirandom assignment of hospitalists to patients within hospitals, combined with our application of hospital fixed effects in our models, should fully account for hospital-level differences in care quality unrelated to the physicians under study, as well as any bias related to patient assignment, 2 important sources of potential bias. Yet our ability to infer causality between our evaluation and outcome measures is limited because we cannot say whether the associations we observed or other correlated physician characteristics affected these results. Nonetheless, our study suggests that certification examination score, but not ratings, signals physician care quality. The examination scores finding is consistent with the findings of other studies.[27,28,35-37,40,42,43] We measured average associations within hospitals in our sample, so it is possible that associations differ within specific types of hospitals because of infrastructure support and quality of care teams. Last, the results of our research might not be generalizable to the revised internal medicine milestone rating system that was implemented in 2021.[44]

Conclusions

Among newly trained hospitalists, certification examination score, but not residency milestone ratings, was associated with improved outcomes among hospitalized Medicare beneficiaries, suggesting that formally incorporating standardized examinations into residency ratings might improve their validity.

ARTICLE INFORMATION

Accepted for Publication: March 14, 2024.

Published Online: May 6, 2024. doi:10.1001/jama.2024.5268

Author Contributions: Dr Gray had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Gray, Vandergrift, Stevens, Lipner, Landon.
Acquisition, analysis, or interpretation of data: Gray, Vandergrift, Stevens, McDonald, Landon.
Drafting of the manuscript: Gray, Vandergrift.
Critical review of the manuscript for important intellectual content: All authors.
Statistical analysis: Gray, Vandergrift, Landon.
Administrative, technical, or material support: Stevens, McDonald, Landon.
Supervision: Lipner, McDonald.

Conflict of Interest Disclosures: Dr Gray reported that he is an employee of the American Board of Internal Medicine (ABIM). Dr Vandergrift reported that he is an employee of ABIM. Dr Lipner reported that she is an employee of ABIM. Dr McDonald reported that he is an employee of ABIM. Dr Landon reported receiving consulting fees from ABIM for ongoing work during the conduct of the study. No other disclosures were reported.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Beckman JJ, Speicher MR. *Characteristics of ACGME Residency Programs That Select Osteopathic Medical Graduates*. Vol 12. Accreditation Council for Graduate Medical Education; 2020:435-440.

2. Holmboe ES, Yamazaki K, Edgar L, et al. Reflections on the first 2 years of milestone implementation. *J Grad Med Educ*. 2015;7(3):506-511. doi:10.4300/JGME-07-03-43

3. Iobst W, Aagaard E, Bazari H, et al. Internal medicine milestones. *J Grad Med Educ*. 2013;5(1)(suppl 1):14-23. doi:10.4300/JGME-05-01s1-03

4. Holmboe ES, Yamazaki K, Nasca TJ, Hamstra SJ. Using longitudinal milestones data and learning analytics to facilitate the professional development of residents: early lessons from three specialties. *Acad Med*. 2020;95(1):97-103. doi:10.1097/ACM.0000000000002899

5. American Board of Internal Medicine. How exams are developed. Accessed January 27, 2023. https://www.abim.org/about/exam-information/exam-development

6. Stevens JP, Hatfield LA, Nyweide DJ, Landon B. Association of variation in consultant use among hospitalist physicians with outcomes among Medicare beneficiaries. *JAMA Netw Open*. 2020;3(2):e1921750. doi:10.1001/jamanetworkopen.2019.21750

7. Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK. Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. *JAMA Intern Med*. 2017;177(2):206-213. doi:10.1001/jamainternmed.2016.7875

8. Tsugawa Y, Jena AB, Orav EJ, Jha AK. Quality of care delivered by general internists in US hospitals who graduated from foreign versus US medical schools: observational study. *BMJ*. 2017;356:j273. doi:10.1136/bmj.j273

9. Tsugawa Y, Jha AK, Newhouse JP, Zaslavsky AM, Jena AB. Variation in physician spending and association with patient outcomes. *JAMA Intern Med*. 2017;177(5):675-682. doi:10.1001/jamainternmed.2017.0059

10. Kuo YF, Sharma G, Freeman JL, Goodwin JS. Growth in the care of older patients by hospitalists in the United States. *N Engl J Med*. 2009;360(11):1102-1112. doi:10.1056/NEJMsa0802381

11. Chin DL, Bang H, Manickam RN, Romano PS. Rethinking thirty-day hospital readmissions: shorter intervals might be better indicators of quality of care. *Health Aff (Millwood)*. 2016;35(10):1867-1875. doi:10.1377/hlthaff.2016.0205

12. Graham KL, Auerbach AD, Schnipper JL, et al. Preventability of early versus late hospital readmissions in a national cohort of general medicine patients. *Ann Intern Med*. 2018;168(11):766-774. doi:10.7326/M17-1724

13. Joynt KE, Jha AK. Thirty-day readmissions—truth and consequences. *N Engl J Med*. 2012;366(15):1366-1369. doi:10.1056/NEJMp1201598

14. Cram P, Wachter RM, Landon BE. Readmission reduction as a hospital quality measure: time to move on to more pressing concerns? *JAMA*. 2022;328(16):1589-1590. doi:10.1001/jama.2022.18305

15. Kolen MJ, Brennan RL. *Test Equating, Scaling, and Linking: Methods and Practices*. 2nd ed. Springer-Verlag; 2004. doi:10.1007/978-1-4757-4310-4

16. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27. doi:10.1097/00005650-199801000-00004

17. Adams JL, Mehrotra A, Thomas JW, et al. Physician cost profiling-reliability and risk of misclassification: detailed methodology and sensitivity analyses. *Rand Health Q*. 2012;2(1):3.

18. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. MIT Press; 2010.

19. Cameron AC, Miller DL. A practitioner's guide to cluster-robust inference. *J Hum Resour*. 2015;50(2):317-372. doi:10.3368/jhr.50.2.317

20. Miyawaki A, Jena AB, Gross N, Tsugawa Y. Comparison of hospital outcomes for patients treated by allopathic versus osteopathic hospitalists: an observational study. *Ann Intern Med*. 2023;176(6):798-806. doi:10.7326/M22-3723

21. Tsugawa Y, Newhouse JP, Zaslavsky AM, Blumenthal DM, Jena AB. Physician age and outcomes in elderly patients in hospital in the US: observational study. *BMJ*. 2017;357:j1797. doi:10.1136/bmj.j1797

22. Kendrick DE, Thelen AE, Chen X, et al. Association of surgical resident competency ratings with patient outcomes. *Acad Med*. 2023;98(7):813-820. doi:10.1097/ACM.0000000000005157

23. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457. doi:10.1016/S0140-6736(07)61602-X

24. Gray BM, Vandergrift JL, Stevens JP, Landon BE. Evolving practice choices by newly certified and more senior general internists: a cross-sectional and panel

comparison. *Ann Intern Med*. 2022;175(7):1022-1027. doi:10.7326/M21-4636

25. McDonald FS, Jurich D, Duhigg LM, et al. Correlations between the USMLE Step examinations, American College of Physicians in-training examination, and ABIM internal medicine certification examination. *Acad Med*. 2020;95(9):1388-1395. doi:10.1097/ACM.0000000000003382

26. Han M, Hamstra SJ, Hogan SO, et al. Trainee physician milestone ratings and patient complaints in early posttraining practice. *JAMA Netw Open*. 2023;6(4):e237588. doi:10.1001/jamanetworkopen.2023.7588

27. Vandergrift JL, Weng W, Gray BM. The association between physician knowledge and inappropriate medications for older populations. *J Am Geriatr Soc*. 2021;69(12):3584-3594. doi:10.1111/jgs.17413

28. Vandergrift JL, Gray BM. Physician clinical knowledge, practice infrastructure, and quality of care. *Am J Manag Care*. 2019;25(10):497-503.

29. Reid RO, Friedberg MW, Adams JL, McGlynn EA, Mehrotra A. Associations between physician characteristics and quality of care. *Arch Intern Med*. 2010;170(16):1442-1449. doi:10.1001/archinternmed.2010.307

30. Papadakis MA, Arnold GK, Blank LL, Holmboe ES, Lipner RS. Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Ann Intern Med*. 2008;148(11):869-876. doi:10.7326/0003-4819-148-11-200806030-00009

31. Norcini JJ, Weng W, Boulet J, McDonald F, Lipner RS. Associations between initial American Board of Internal Medicine certification and maintenance of certification status of attending physicians and in-hospital mortality of patients with acute myocardial infarction or congestive heart failure: a retrospective cohort study of hospitalisations in Pennsylvania, USA. *BMJ Open*. 2022;12(4):e055558. doi:10.1136/bmjopen-2021-055558

32. Norcini JJ, Kimball HR, Lipner RS. Certification and specialization: do they matter in the outcome of acute myocardial infarction? *Acad Med*. 2000;75(12):1193-1198. doi:10.1097/00001888-200012000-00016

33. 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(3):376-381. doi:10.1097/ACM.0000000000001055

34. Lipner RS, Hess BJ, Phillips RL Jr. Specialty board certification in the United States: issues and evidence. *J Contin Educ Health Prof*. 2013;33(suppl 1):S20-S35. doi:10.1002/chp.21203

35. Holmboe ES, Wang Y, Meehan TP, et al. Association between maintenance of certification examination scores and quality of care for Medicare beneficiaries. *Arch Intern Med*. 2008;168(13):1396-1403. doi:10.1001/archinte.168.13.1396

36. Gray BM, Vandergrift JL, Weng W, Lipner RS, Barnett ML. Clinical knowledge and trends in physicians' prescribing of opioids for new onset back pain, 2009-2017. *JAMA Netw Open*. 2021;4(7):e2115328. doi:10.1001/jamanetworkopen.2021.15328

37. Gray BM, Vandergrift JL, McCoy RG, Lipner RS, Landon BE. Association between primary care physician diagnostic knowledge and death, hospitalisation and emergency department visits following an outpatient visit at risk for diagnostic error: a retrospective cohort study using Medicare claims. *BMJ Open*. 2021;11(4):e041817. doi:10.1136/bmjopen-2020-041817

38. Fiorilli PN, Minges KE, Herrin J, et al. Association of physician certification in interventional cardiology with in-hospital outcomes of percutaneous coronary intervention. *Circulation*. 2015;132(19):1816-1824. doi:10.1161/CIRCULATIONAHA.115.017523

39. Gray B, Vandergrift J, Landon B, Reschovsky J, Lipner R. Associations between American Board of Internal Medicine maintenance of certification status and performance on a set of Healthcare Effectiveness Data and Information Set (HEDIS) process measures. *Ann Intern Med*. 2018;169(2):97-105. doi:10.7326/M16-2643

40. Tamblyn R, Abrahamowicz M, Brailovsky C, et al. Association between licensing examination scores and resource use and quality of care in primary care practice. *JAMA*. 1998;280(11):989-996. doi:10.1001/jama.280.11.989

41. De Champlain AF, Ashworth N, Kain N, Qin S, Wiebe D, Tian F. Does pass/fail on medical licensing exams predict future physician performance in practice? a longitudinal cohort study of Alberta physicians. *J Med Regul*. 2020;106(4):17-26. doi:10.30770/2572-1852-106.4.17

42. Norcini J, Grabovsky I, Barone MA, Anderson MB, Pandian RS, Mechaber AJ. The associations between United States medical licensing examination performance and outcomes of patient care. *Acad Med*. 2024;99(3):325-330.

43. 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(8):1157-1162. doi:10.1097/ACM.0000000000000310

44. Accreditation Council for Graduate Medical Education (ACGME). ACGME program requirements for graduate medical education in internal medicine. 2021. Accessed December 1, 2021. https://www.acgme.org/globalassets/pfassets/programrequirements/140_internalmedicine_2023.pdf