

Association Between Internal Medicine Residency Applicant Characteristics and Performance on ACGME Milestones During Intern Year

Blair P. Golden, MD, MS

Bruce L. Henschen, MD, MPH

David T. Liss, PhD

Sara L. Kiely, MS

Aashish K. Didwania, MD

ABSTRACT

Background Residency programs apply varying criteria to the resident selection process. However, it is unclear which applicant characteristics reflect preparedness for residency.

Objective We determined the applicant characteristics associated with first-year performance in internal medicine residency as assessed by performance on Accreditation Council for Graduate Medical Education (ACGME) Milestones.

Methods We examined the association between applicant characteristics and performance on ACGME Milestones during intern year for individuals entering Northwestern University's internal medicine residency between 2013 and 2018. We used bivariate analysis and a multivariable linear regression model to determine the association between individual factors and Milestone performance.

Results Of 203 eligible residents, 198 (98%) were included in the final sample. One hundred fourteen residents (58%) were female, and 116 residents (59%) were White. Mean Step 1 and Step 2 CK scores were 245.5 (SD 12.0) and 258 (SD 10.8) respectively. Step 1 scores, Alpha Omega Alpha membership, medicine clerkship grades, and interview scores were not associated with Milestone performance in the bivariate analysis and were not included in the multivariable model. In the multivariable model, overall clerkship grades, ranking of the medical school, and year entering residency were significantly associated with Milestone performance ($P \leq .04$).

Conclusions Most traditional metrics used in residency selection were not associated with early performance on ACGME Milestones during internal medicine residency.

Introduction

Residency programs devote considerable thought and apply varying criteria to resident selection. However, prior studies have yielded inconsistent results regarding applicant factors (such as clerkship grades or standardized test scores) associated with residency performance in internal medicine¹⁻⁴ and other specialties.⁵⁻⁹ The predictive value of such factors is low. Fine and Hayward reported that residency selection committee ranking was only moderately correlated with subsequent performance assessments,¹ and Neely et al found that applicant characteristics explained a minority of the variance in third-year resident performance rating in internal medicine.² Prior studies have relied largely on internally developed benchmarks, limiting generalizability and reproducibility.^{1,2}

In 2013, the Accreditation Council for Graduate Medical Education (ACGME) introduced Milestones for internal medicine residents, outlining competency-

based expectations across programs and training years.^{10,11} The Milestones represent an attempt to standardize expected educational outcomes, enabling educators to study factors that contribute to performance.¹²⁻¹⁵ While studies have examined the relationship between applicant factors and performance at the conclusion of residency,^{1,2} factors during residency that are difficult to quantify (such as mentorship and peer support) also contribute to resident success.

The aim of this study was to determine the resident applicant characteristics associated with preparedness for internal medicine internship using performance on ACGME Milestones during intern year.

Methods

We performed a retrospective cohort study examining the association between residency application characteristics and subsequent performance on ACGME Milestones among internal medicine residents at McGaw Medical Center, Northwestern University. All data were deidentified prior to analysis.

DOI: <http://dx.doi.org/10.4300/JGME-D-20-00603.1>

The study population consisted of residents entering our categorical internal medicine residency program from 2013 to 2018. These classes were selected because the 2013–2014 intern class was the first assessed using the ACGME Milestones and competency frameworks. Individuals with incomplete data or who transferred in or out of the program outside of the match process were excluded.

We examined resident factors that are used in the Northwestern internal medicine residency selection process or that have been shown in prior studies to be influential.^{1,2,4,7,16–18} We obtained applicant characteristics from residency files derived from Electronic Residency Application Service (ERAS) applications. Self-reported demographic data were obtained from residency records.

Because the incremental difference of a 1-point increase on USMLE examinations is likely small, Step 1 and Step 2 CK scores were defined as categorical variables with 10-point ranges (TABLE 1). Step 2 CK scores are not required as part of the residency application; individuals who did not submit scores were categorized as “unknown.”

As part of the residency selection process, each applicant’s internal medicine clerkship grade was assigned a value from 5 to 90 (with 5 representing the top fifth percentile) using information available in the medical student performance evaluation to account for the variability in grade distribution between schools. This number was based on the individual’s medicine clerkship grade compared to the overall distribution of grades within their medical school class. Applicants were assumed to be in the median percentile within a given grade category (ie, if the top 30% of a medical school class earns honors, an individual earning honors was assumed to be at the top 15th percentile). Numbers were adjusted if more granular data were available within the medicine clerkship letter. A similar process was used to determine each applicant’s average grade across all core clinical clerkships.

We used *US News & World Report (USNWR)* “Best Medical Schools: Research Rankings” as a surrogate for perceived medical school competitiveness.¹⁹ We used the most recent rankings for consistency because ranking volatility is relatively low. Medical school ranking was treated as a categorical variable with 20 schools in each category.

Alpha Omega Alpha (AOA) membership is an optional question in ERAS, and non-responders were assumed to be non-members. The few residents who reported that their school held elections senior year or did not have a chapter were classified as non-members given that some students from these schools explicitly indicated this while others left the question blank.

Objectives

The aim of this study is to identify the characteristics of internal medicine residency applicants that predict performance on ACGME Milestones during intern year.

Findings

Some factors (such as overall clerkship grades and ranking of an applicant’s medical school) were modestly correlated with ACGME Milestones scores, but most traditional metrics used to rank residency applicants were not associated with Milestone performance.

Limitations

This was a single institution study in which a summative assessment tool averaging rotation-based ratings served as an anchor for the final Milestone assessments.

Bottom Line

The disconnect between applicant characteristics and subsequent performance on ACGME Milestones may reflect limitations in the residency selection process, in the use of the Milestones, or both.

Gold Humanism Honor Society (GHHS) membership was determined using a publicly available database of GHHS members and chapters.²⁰ A resident was considered to be eligible for GHHS if a chapter existed at their medical school at least 1 year prior to their medical school graduation. Because many residents attended a school without a chapter, membership was categorized into 3 groups: members, non-members who were eligible, and non-members who were not eligible.

Each applicant was interviewed by 2 faculty. Interviewers were not provided with applicant grades or test scores. Each interviewer gave an overall interview score from 1 to 5 using a standardized rubric, with 1 being the strongest score. We averaged interview scores for each applicant.

Age and gender are not used in our residency selection process, but have been correlated with resident performance assessments elsewhere.^{7,21–25} Race was not used as a predictor but was included as a covariate in the multivariable model to account for possible implicit bias in the assessment process.^{21,26}

The primary outcome was mean performance across all 22 ACGME subcompetencies on the midyear assessment in December of intern year. For the 2013–2014 intern class, we used the year-end assessment because a midyear assessment was not completed.

As part of the ongoing resident development process, the clinical competency committee assessed each resident across the 22 ACGME subcompetencies. Performance for each subcompetency was determined using attending evaluations, resident evaluations, and a summative assessment tool generated from the electronic assessment system. Further input was derived

TABLE 1

Demographic and Academic Characteristics of the Study Population (N = 198)

Characteristic	Frequency
Age	Mean (SD)
Average age	27.3 (1.8)
Gender	N (%)
Female	114 (58)
Race/ethnicity	N (%)
Non-Hispanic White	116 (59)
Asian	55 (28)
Non-Hispanic Black	10 (5)
Hispanic	17 (9)
Intern year	N (%)
2013–2014	31 (16)
2014–2015	34 (17)
2015–2016	31 (16)
2016–2017	35 (18)
2017–2018	33 (17)
2018–2019	34 (17)
Step 1 score	N (%)
< 230	23 (12)
230–240	36 (18)
240–250	68 (34)
250–260	57 (29)
> 260	14 (7)
Step 2 CK score	N (%)
< 245	23 (12)
245–254	37 (19)
255–264	71 (36)
≥ 265	49 (25)
Not submitted in ERAS	23 (12)
Medical school USNWR research ranking	N (%)
1–20	92 (47)
21–40	44 (22)
41–60	32 (16)
> 60 or unranked	30 (15)
Medical school grades	Mean (SD)
Medicine clerkship (5–90)	30.9 (19)
All third-year clerkships (5–90)	33.2 (12)
Alpha Omega Alpha membership	N (%)
Yes	82 (41)
Gold Humanism Honor Society membership	N (%)
Yes	24 (12)
No, chapter at medical school	101 (51)
No, no chapter at medical school	73 (37)
Interview score	Mean (SD)
Mean score (1–5)	2.05 (0.5)

Abbreviations: CK, clinical knowledge; ERAS, Electronic Residency Application Service; USNWR, *US News & World Report*.

from nurse evaluations, conference attendance, evaluation completion rate, scholarly productivity, and extracurricular activities. Residents were rated on each subcompetency using the Milestone-based scale (1–9), with 9 representing the aspirational Milestone. The program director, an associate program director, and a program coordinator approved the final Milestone ratings for each resident.

Each of the 22 ACGME subcompetencies is grouped under 1 of 6 core competencies (patient care, medical knowledge, systems-based practice, practice-based learning and improvement, professionalism, and interpersonal and communication skills). Mean performance for each core competency was assessed as a secondary outcome by averaging the performance on the subcompetencies that comprise the broader core competency.

We first performed an exploratory bivariate analysis to understand the impact of individual predictor variables in isolation. We used Pearson's correlation coefficients to evaluate the correlation between continuous predictors and outcome measures and 2-sample *t* tests to assess the association between binary predictors (gender and AOA membership) and outcome measures. One-way analysis of variance (ANOVA) tested for differences in mean outcome scores between non-binary categorical predictor groups.

We then performed a multivariable linear regression analysis. Given the moderate sample size, we initially included all predictors in our model to account for potentially unmeasured confounders, irrespective of whether predictors were statistically significant in bivariate analysis. We then used a backward stepwise approach to refine our model, removing individual predictors sequentially. We determined the Akaike information criterion (an estimate of model fit) for each sequential model to select the set of covariates that demonstrated the best regression model fit for the primary outcome.²⁷ Gender, age, intern academic year, Step 2 CK score, overall clerkship grades, and USNWR rankings were retained in the final model. This set of covariates consistently demonstrated relatively high goodness of fit across regression models for secondary outcomes.

A significance level of .05 was used for all analyses. Analyses were conducted in Stata 15.1 (StataCorp LLC, College Station, TX).

This study was determined to be exempt by Northwestern University's Institutional Review Board with a waiver of informed consent.

Results

Of 203 eligible residents, 198 were included in the final analysis. Of the 5 excluded residents, 4

TABLE 2
Unadjusted ACGME Core Competency Milestone Ratings by Applicant Group Entering Northwestern McGaw Medical Center Internal Medicine Residency Over a 6-Year Period
(N = 198)

Characteristic		Primary Outcome		Secondary Outcomes											
		All Milestones		Patient Care		Medical Knowledge		Systems-Based Practice		Practice-Based Learning and Improvement		Professionalism		Interpersonal and Communication Skills	
		Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value
All participants		5.22 (0.51)		4.98 (0.64)		5.04 (0.53)		5.12 (0.54)		5.16 (0.52)		5.54 (0.60)		5.43 (0.63)	
Gender ^a	Female	5.19 (0.51)	.31	4.98 (0.54)	.09	4.90 (0.64)	.033	5.09 (0.53)	.35	5.13 (0.51)	.41	5.54 (0.59)	.94	5.42 (0.66)	.62
	Male	5.27 (0.52)		5.12 (0.51)		5.10 (0.64)		5.16 (0.55)		5.19 (0.54)		5.54 (0.60)		5.46 (0.60)	
Year entering residency ^b	2013–2014	5.81 (0.43)	< .001	5.67 (0.48)	< .001	5.66 (0.45)	< .001	5.68 (0.46)	< .001	5.75 (0.46)	< .001	6.13 (0.54)	< .001	5.96 (0.51)	< .001
	2014–2015	5.34 (0.38)		5.04 (0.48)		5.06 (0.44)		5.24 (0.47)		5.31 (0.45)		5.76 (0.48)		5.54 (0.55)	
	2015–2016	5.38 (0.29)		5.13 (0.23)		5.37 (0.32)		5.15 (0.36)		5.15 (0.25)		5.70 (0.47)		5.89 (0.61)	
	2016–2017	4.98 (0.55)		4.79 (0.59)		4.80 (0.56)		4.88 (0.67)		4.91 (0.61)		5.29 (0.57)		5.08 (0.58)	
	2017–2018	5.10 (0.40)		4.88 (0.44)		4.94 (0.46)		5.00 (0.41)		5.01 (0.44)		5.45 (0.50)		5.29 (0.51)	
	2018–2019	4.81 (0.23)		4.80 (0.36)		4.16 (0.38)		4.81 (0.31)		4.88 (0.30)		4.98 (0.22)		4.94 (0.21)	
Step 1 score ^b	< 230	5.42 (0.46)	.25	5.20 (0.49)	.59	5.15 (0.61)	.33	5.35 (0.46)	.17	5.39 (0.43)	.11	5.72 (0.60)	.57	5.62 (0.58)	.44
	230–240	5.20 (0.38)		5.01 (0.43)		5.01 (0.50)		5.08 (0.33)		5.12 (0.34)		5.51 (0.59)		5.41 (0.54)	
	240–250	5.24 (0.56)		5.05 (0.59)		4.98 (0.70)		5.14 (0.60)		5.19 (0.55)		5.55 (0.60)		5.47 (0.68)	
	250–260	5.13 (0.50)		4.98 (0.53)		4.86 (0.66)		5.02 (0.55)		5.05 (0.57)		5.47 (0.53)		5.33 (0.60)	
	> 260	5.24 (0.63)		5.05 (0.52)		5.14 (0.63)		5.07 (0.66)		5.18 (0.62)		5.55 (0.80)		5.43 (0.79)	
	< 245	5.20 (0.37)	.98	4.96 (0.45)	.95	5.03 (0.58)	.43	5.07 (0.35)	.97	5.18 (0.37)	.99	5.50 (0.50)	.94	5.41 (0.48)	.99
Step 2 CK score ^b	245–254	5.25 (0.51)		5.05 (0.48)		5.01 (0.58)		5.17 (0.56)		5.16 (0.51)		5.59 (0.63)		5.42 (0.69)	
	255–264	5.21 (0.51)		5.03 (0.58)		4.87 (0.68)		5.11 (0.53)		5.16 (0.52)		5.52 (0.58)		5.46 (0.65)	
	≥ 265	5.25 (0.50)		5.08 (0.48)		5.08 (0.62)		5.11 (0.54)		5.17 (0.54)		5.57 (0.62)		5.45 (0.63)	
	Not submitted	5.19 (0.62)		5.03 (0.65)		5.04 (0.71)		5.10 (0.68)		5.10 (0.65)		5.48 (0.66)		5.38 (0.66)	
	1–20	5.31 (0.50)	.12	5.16 (0.51)	.036	5.11 (0.63)	.021	5.20 (0.52)	.28	5.25 (0.51)	.14	5.61 (0.59)	.39	5.51 (0.59)	.46
	21–40	5.14 (0.52)		4.97 (0.57)		4.92 (0.68)		5.03 (0.57)		5.06 (0.51)		5.43 (0.56)		5.39 (0.64)	
Medical school USNWR research ranking ^b	41–60	5.16 (0.61)		4.91 (0.57)		4.94 (0.67)		5.06 (0.57)		5.09 (0.60)		5.55 (0.73)		5.35 (0.78)	
	> 60 or unranked	5.13 (0.38)		4.93 (0.44)		4.72 (0.50)		5.05 (0.37)		5.10 (0.46)		5.48 (0.49)		5.36 (0.59)	

TABLE 2
Unadjusted ACGME Core Competency Milestone Ratings by Applicant Group Entering Northwestern McGaw Medical Center Internal Medicine Residency Over a 6-Year Period
(continued)

Characteristic		Primary Outcome		Secondary Outcomes											
		All Milestones		Patient Care		Medical Knowledge		Systems-Based Practice		Practice-Based Learning and Improvement		Professionalism		Interpersonal and Communication Skills	
		Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value	Mean (SD)	P Value
AOA membership ^a	Yes	5.23 (0.54)	.84	5.03 (0.57)	.88	4.93 (0.66)	.31	5.13 (0.55)	.69	5.18 (0.54)	.62	5.58 (0.63)	.46	5.44 (0.60)	.93
	No	5.22 (0.49)		5.05 (0.51)		5.02 (0.63)		5.10 (0.53)		5.14 (0.51)		5.13 (0.57)		5.43 (0.68)	
Gold Humanism Honor Society membership ^b	Yes	5.25 (0.54)	.19	5.06 (0.60)	.25	4.98 (0.63)	.009	5.19 (0.61)	.48	5.19 (0.54)	.77	5.56 (0.62)	.16	5.43 (0.54)	.16
	No—chapter at medical school	5.16 (0.50)		4.98 (0.50)		4.86 (0.65)		5.07 (0.53)		5.13 (0.50)		5.46 (0.58)		5.36 (0.65)	
No—no chapter at medical school	5.30 (0.51)		5.12 (0.55)		5.16 (0.59)		5.15 (0.52)		5.18 (0.55)		5.64 (0.60)		5.54 (0.63)		

Abbreviations: USNWR, US News & World Report; AOA, Alpha Omega Alpha.

^a 2-sample *t* test performed.

^b ANOVA test performed.

transferred in or out of the program and 1 had incomplete data. Residents' demographic and academic characteristics are summarized in TABLE 1. The study population was majority female (114 residents, 58%) and White (116 residents, 59%). Mean Step 1 and Step 2 CK scores were 245.5 (SD 12) and 258.0 (SD 10.8), respectively. Approximately 41% (82 of 198) of residents were in AOA. Only 12% (24 of 198) of residents were in GHHS, but approximately 37% (73 of 198) of students attended a school without a chapter.

The mean score across all subcompetencies was 5.22 (SD 0.51), and there was an approximately normal distribution of scores. For individual core competencies, the mean score ranged from 4.98 (SD 0.64) for patient care to 5.54 (SD 0.60) for professionalism.

The results of the bivariate analysis are presented in TABLE 2. The year entering residency had a statistically significant association with the primary outcome and all secondary outcomes ($P < .001$ for all), although no trend was observed across years. The 2013–2014 intern class had the highest mean Milestone score (5.81), whereas the 2018–2019 class had the lowest (4.81). Women were assessed as having lower performance on Medical Knowledge Milestones compared to men (4.90 vs 5.10, $P = .033$). USNWR medical school ranking was associated with performance on patient care and medical knowledge competencies, with students attending a school ranked 1 to 20 having the highest mean score (5.16, $P = .036$ and 5.11, $P = .021$, respectively). Milestone performance was not significantly associated with Step 1, Step 2 CK, or AOA membership.

For continuous variables, interview score, age, and performance in the medicine clerkship were not associated with Milestone performance. Performance across all core clerkships was correlated with performance on professionalism subcompetencies ($r = -0.14$, $P = .045$), but was not significantly associated with other outcomes.

In multivariable regression analysis, only a few predictors were significantly associated with Milestone performance (TABLE 3). There were statistically significant differences between each year entering residency compared to the referent group (2018–2019 intern year) for the primary outcome ($P < .001$ to .03) and many of the secondary outcomes. Male gender was associated with 0.14 points higher performance on medical knowledge Milestones (95% CI 0.01–0.26, $P = .031$). Core clerkship grades were significantly associated with the primary outcome as well as performance on professionalism, interpersonal communication skills, and practice-based learning ($P = .01$ to .02). Each 1 percentile point worsening in clerkship

grades was associated with a -0.01 change in overall Milestone score (95% CI -0.01 to -0.001).

Compared to attending a medical school ranked in the top 20 by *USNWR*, attending a school ranked 20 to 40 was associated with lower performance for the primary outcome and all core competencies except patient care and systems-based practice. Attending the lowest ranked category of school (> 60 or unranked) was also associated with lower overall performance on the Milestones (-0.23; 95% CI -0.42 to -0.04; $P = .019$) as well as lower performance on medical knowledge (-0.36; 95% CI -0.57 to -0.15; $P = .001$), patient care (-0.23; 95% CI -0.44 to -0.02; $P = .034$), and practice-based learning and improvement (-0.22; 95% CI -0.43 to -0.005; $P = .045$).

Discussion

Most internal medicine residency applicant factors (including Step 1 scores, medicine clerkship grades, interview performance, and AOA membership) were not associated with Milestone performance during intern year. Attending a medical school ranked in the top 20 by *USNWR* was associated with a statistically significant improvement in overall performance on ACGME Milestones, but the absolute difference was minimal (0.19 points higher compared to those who attended a school ranked 20–40) and was not statistically significant when compared to individuals who attended a school ranked 41 to 60. Core clerkship grades were significantly associated with mean Milestone performance. While the effect size (a .01 change in Milestone score per 1 percentile improvement in grade) appears small, it may suggest meaningful differences in residency performance: a student ranked in the middle of the first quartile (12.5 percentile) of their medical school may perform 0.5 points higher on Milestones as intern compared to a student ranked in the middle of the third quartile (62.5 percentile).

This study builds on prior studies of internal medicine residency programs, suggesting that most traditional residency selection criteria do not predict resident performance.^{1,2,4} A University of Michigan study found that only internal medicine clerkship honors and medical school were significantly associated with third-year resident performance in a multivariable model.¹ A prior study of Northwestern internal medicine residents graduating between 2000 and 2005 found that medical school quality and overall clerkship grades were most strongly associated with residency performance.² Research from other clinical specialties also has suggested that commonly used metrics (such as USMLE scores and interview

performance) are not strongly predictive of residency success.^{28–30}

These findings may also reflect the challenges of using the ACGME Milestones as an assessment tool. These Milestones, while reflecting the theoretical educational outcomes of a program, may also be an imperfect measure of resident performance in the real world.^{31,32} Further work is needed to understand how Milestone performance correlates to patient outcomes and other measures of clinical competency.¹⁵

Female gender was negatively associated with performance on medical knowledge subcompetencies. This may reflect gender bias within the assessment process. Studies from internal medicine and other specialties on gender bias within the resident assessment process have had mixed findings.^{21,22,33,34} Dayal et al found potential gender bias in faculty assessments of emergency medicine residents,²¹ but Santen et al subsequently found in a larger national study that male and female emergency medicine residents had similar Milestone ratings for the majority of competencies.³⁵

This study has several limitations. First, this is a single institution study where the residency selection process favors competitive applicants whose characteristics (eg, USMLE scores, percent of students in AOA) do not mirror the general population. Second, midyear Milestone scores were used to estimate an intern's initial performance in residency; faculty may not have sufficient interactions at that point to make accurate assessments.¹⁸ However, we found the same conclusions in the year that used year-end assessments in the absence of midyear assessments (2013–2014). Third, we excluded residents who joined or left the program outside of the Match; although this was a small number of trainees, these individuals may be important outliers. Fourth, our summative assessment tool averages rotation-based global ratings and served as an anchor for the final Milestone assessment; the tool may not be as accurate as a deconstructed rating system. It is also possible that Milestone assessments are proxies for preexisting global assessments of competence that are influenced by criteria other than the Milestones themselves.³⁶ Fifth, we defined AOA as a binary variable given limitations in ERAS data. However, we performed a sensitivity analysis in which students who indicated that they attended a school with no AOA chapter or with elections during senior year were excluded and found that the final multivariable regression model had similar findings. Finally, many residents were not eligible for GHHS, limiting our ability to assess this factor.

This study supports the need for reform within the medical student assessment and residency admissions

TABLE 3
Multivariable Linear Regression Model Results Assessing Potential Predictors of ACGME Milestone Ratings Over a 6-Year Period (N = 198)

Characteristics		Primary Outcome		Secondary Outcomes					
		All Milestones, β (95% CI)	Patient Care, β (95% CI)	Medical Knowledge, β (95% CI)	Systems-Based Practice, β (95% CI)	Practice-Based Learning and Improvement, β (95% CI)	Professionalism, β (95% CI)	Interpersonal and Communication Skills, β (95% CI)	
Gender	Male	0.07 (-0.04, 0.18)	0.11 (-0.02, 0.23)	0.14 (0.01-0.26) ^a	0.07 (-0.06, 0.21)	0.07 (-0.05, 0.20)	-0.01 (-0.15, 0.12)	0.06 (-0.09, 0.21)	
	Female	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	
Year entering residency	2013–2014	1.00 (0.81, 1.20) ^c	0.86 (0.64, 1.08) ^c	1.47 (1.26, 1.69) ^c	0.87 (0.64, 1.11) ^c	0.87 (0.66, 1.09) ^c	1.19 (0.95, 1.42) ^c	1.01 (0.76, 1.27) ^c	
	2014–2015	0.57 (0.38, 0.76) ^c	0.28 (0.07, 0.49) ^a	0.90 (0.70, 1.11) ^c	0.48 (0.26, 0.71) ^c	0.50 (0.29, 0.71) ^c	0.85 (0.62, 1.08) ^c	0.66 (0.41, 0.90) ^c	
	2015–2016	0.59 (0.40, 0.78) ^c	0.31 (0.10, 0.53) ^b	1.21 (1.00, 1.43) ^c	0.38 (0.15, 0.62) ^b	0.31 (0.10, 0.53) ^b	0.78 (0.55, 1.01) ^c	1.01 (0.77, 1.26) ^c	
	2016–2017	0.20 (0.01, 0.38) ^a	0.01 (-0.19, 0.22)	0.65 (0.44, 0.86) ^c	0.11 (-0.12, 0.34)	0.10 (-0.11, 0.31)	0.38 (0.15, 0.60) ^b	0.20 (-0.04, 0.44)	
	2017–2018	0.32 (0.14, 0.51) ^b	0.12 (-0.08, 0.33)	0.81 (0.61, 1.02) ^c	0.23 (0.01, 0.45) ^a	0.16 (-0.04, 0.37)	0.51 (0.29, 0.73) ^c	0.40 (0.16, 0.64) ^b	
	2018–2019	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	
Age		-0.03 (-0.06, 0.005)	-0.03 (-0.07, 0.00)	-0.03 (-0.06, 0.01)	-0.02 (-0.06, 0.01)	-0.01 (-0.05, 0.02)	-0.03 (-0.07, 0.01)	-0.02 (-0.06, 0.02)	
Step 2 CK score	< 245	-0.09 (-0.31, 0.13)	-0.14 (-0.38, 0.10)	-0.12 (-0.36, 0.12)	-0.08 (-0.34, 0.18)	0.03 (-0.22, 0.27)	-0.10 (-0.36, 0.16)	-0.16 (-0.44, 0.12)	
	245–254	0.02 (-0.13, 0.18)	0.00 (-0.17, 0.18)	0.06 (-0.11, 0.24)	0.06 (-0.13, 0.25)	0.001 (-0.17, 0.17)	0.07 (-0.11, 0.26)	-0.06 (-0.26, 0.14)	
	255–264	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	
	≥ 265	-0.06 (-0.20, 0.09)	-0.03 (-0.19-0.13)	0.07 (-0.09, 0.23)	-0.06 (-0.24, 0.11)	-0.07 (-0.23, 0.09)	-0.07 (-0.24, 0.10)	-0.13 (-0.32, 0.05)	
Medical School USNWR Research Ranking	Not submitted	-0.26 (-0.45, -0.07) ^b	-0.21 (-0.42, 0.01)	-0.18 (-0.39, 0.04)	-0.23 (-0.46, 0.05)	-0.31 (-0.52, -0.09) ^b	-0.32 (-0.55, -0.09) ^b	-0.32 (-0.57, -0.07) ^a	
	1–20	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	0 (Ref)	
	21–40	-0.19 (-0.33, -0.04) ^a	-0.15 (-0.32, 0.01)	-0.18 (-0.34, -0.01) ^a	-0.16 (-0.34, 0.02)	-0.22 (-0.39, -0.06) ^a	-0.21 (-0.38, -0.03) ^a	-0.20 (-0.39, -0.01) ^a	
	41–60	-0.09 (-0.26, 0.09)	-0.14 (-0.33, 0.06)	-0.06 (-0.25, 0.13)	-0.05 (-0.26, 0.16)	-0.12 (-0.31, 0.07)	-0.02 (-0.23, 0.19)	-0.11 (-0.34, 0.11)	
	> 60 or unranked	-0.23 (-0.42, -0.04) ^a	-0.23 (-0.44, -0.02) ^a	-0.36 (-0.57, -0.15) ^b	-0.18 (-0.41, 0.05)	-0.22 (-0.43,-0.00) ^{a,f}	-0.21 (-0.44, 0.01)	-0.24 (-0.49, 0.00)	
Overall clerkship grades		-0.01(-0.01, -0.00) ^{a,d}	-0.01 (-0.01, 0.00)	-0.01 (-0.01, 0.00)	-0.01 (-0.01, 0.00)	-0.01 (-0.01, -0.00) ^{a,d}	-0.01 (-0.02, -0.00) ^{a,e}	-0.01 (-0.02, -0.00) ^{a,d}	

Abbreviations: ACGME, Accreditation Council for Graduate Medical Education; USNWR, US News & World Report.

^a $P \leq .05$.

^b $P \leq .01$.

^c $P \leq .001$.

^d Lower confidence bound -0.001.

^e Lower confidence bound -0.002.

^f Lower confidence bound -0.005.

processes.³⁷ The USMLE recently announced that Step 1 will transition to a pass/fail format, underscoring the limitations of this assessment. “Traditional” metrics, such as standardized test scores and AOA membership, produce anxiety for medical students without delivering reliable assessment information. Novel and holistic assessment methods of medical students (eg, those assessing entrustable professional activities) have the potential to benefit both students and residency programs alike.³⁸

Conclusions

Most selection criteria for internal medicine residency applicants are poorly predictive of intern year performance as measured by performance on the ACGME Milestones. This may be due to imperfect selection criteria, the limitations of the Milestones as measurements of intern year performance (in our residency program, or perhaps globally), or both.

References

1. Fine PL, Hayward RA. Do the criteria of resident selection committees predict residents' performances? *Acad Med*. 1995;70(9):834–838.
2. Neely D, Feinglass J, Wallace WH. Developing a predictive model to assess applicants to an internal medicine residency. *J Grad Med Educ*. 2010;2(1):129–132. doi:10.4300/JGME-D-09-00044.1.
3. Cullen MW, Reed DA, Halvorsen AJ, et al. Selection criteria for internal medicine residency applicants and professionalism ratings during internship. *Mayo Clin Proc*. 2011;86(3):197–202. doi:10.4065/mcp.2010.0655.
4. Sharma A, Schauer DP, Kelleher M, Kinnear B, Sall D, Warm E. USMLE Step 2 CK: best predictor of multimodal performance in an internal medicine residency. *J Grad Med Educ*. 2019;11(4):412–419. doi:10.4300/JGME-D-19-00099.1.
5. Dawkins K, Ekstrom RD, Maltbie A, Golden RN. The relationship between psychiatry residency applicant evaluations and subsequent residency performance. *Acad Psychiatry*. 2005;29(1):69–75. doi:10.1176/appi.ap.29.1.69.
6. Schaverien MV. Selection for surgical training: an evidence-based review. *J Surg Educ*. 2016;73(4):721–729. doi:10.1016/j.jsurg.2016.02.007.
7. Tolan AM, Kaji AH, Quach C, Hines OJ, de Virgilio C. The electronic residency application service application can predict Accreditation Council for Graduate Medical Education competency-based surgical resident performance. *J Surg Educ*. 2010;67(6):444–448. doi:10.1016/j.jsurg.2010.05.002.
8. Stohl HE, Hueppchen NA, Bienstock JL. Can medical school performance predict residency performance? Resident selection and predictors of successful performance in obstetrics and gynecology. *J Grad Med Educ*. 2010;2(3):322–326. doi:10.4300/JGME-D-09-00101.1.
9. Hartman ND, Lefebvre CW, Manthey DE. A narrative review of the evidence supporting factors used by residency program directors to select applicants for interviews. *J Grad Med Educ*. 2019;11(3):268–273. doi:10.4300/JGME-D-18-00979.3.
10. 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.
11. Accreditation Council for Graduate Medical Education and American Board of Internal Medicine. The Internal Medicine Milestone Project. <https://www.acgme.org/Portals/0/PDFs/Milestones/InternalMedicineMilestones.pdf>. Accessed January 13, 2021.
12. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012;366(11):1051–1056. doi:10.1056/NEJMs1200117.
13. 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.
14. Conforti LN, Yaghmour NA, Hamstra SJ, et al. The effect and use of Milestones in the assessment of neurological surgery residents and residency programs. *J Surg Educ*. 2018;75(1):147–155. doi:10.1016/j.jsurg.2017.06.001.
15. Hauer KE, Vandergrift J, Lipner RS, Holmboe ES, Hood S, McDonald FS. National internal medicine Milestone ratings: validity evidence from longitudinal three-year follow-up. *Acad Med*. 2018;93(8):1189–1204. doi:10.1097/ACM.0000000000002234.
16. Agarwal V, Bump GM, Heller MT, et al. Do residency selection factors predict radiology resident performance? *Acad Radiol*. 2018;25(3):397–402. doi:10.1016/j.acra.2017.09.020.
17. Rosenthal S, Howard B, Schluskel YR, et al. Does medical student membership in the gold humanism honor society influence selection for residency? *J Surg Educ*. 2009;66(6):308–313. doi:10.1016/j.jsurg.2009.08.002.
18. Raman T, Alrabaa RG, Sood A, Maloof P, Benevenia J, Berberian W. Does residency selection criteria predict performance in orthopaedic surgery residency? *Clin Orthop Relat Res*. 2016;474(4):908–914. doi:10.1007/s11999-015-4317-7.

19. US News and World Report. Best Medical Schools: Research. <https://www.usnews.com/best-graduate-schools/top-medical-schools/research-rankings>. Accessed January 13, 2021.
20. The Gold Foundation. Member directory. <https://www.gold-foundation.org/programs/ghhs/member-directory/>. Accessed January 13, 2021.
21. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident Milestone evaluations by faculty during emergency medicine residency training. *JAMA Intern Med*. 2017;177(5):651–657. doi:10.1001/jamainternmed.2016.9616.
22. Gerull KM, Loe M, Seiler K, McAllister J, Salles A. Assessing gender bias in qualitative evaluations of surgical residents. *Am J Surg*. 2019;217(2):306–313. doi:10.1016/j.amjsurg.2018.09.029.
23. Sulistio MS, Khera A, Squiers K, et al. Effects of gender in resident evaluations and certifying examination pass rates. *BMC Med Educ*. 2019;19(1):10. doi:10.1186/s12909-018-1440-7.
24. Kanna B, Gu Y, Akhuetie J, Dimitrov V. Predicting performance using background characteristics of international medical graduates in an inner-city university-affiliated internal medicine residency training program. *BMC Med Educ*. 2009;9:42. doi:10.1186/1472-6920-9-42.
25. Rand VE, Hudes ES, Browner WS, Wachter RM, Avins AL. Effect of evaluator and resident gender on the American Board of Internal Medicine evaluation scores. *J Gen Intern Med*. 1998;13(10):670–674. doi:10.1046/j.1525-1497.1998.00202.x.
26. Ross DA, Boatright D, Nunez-Smith M, Jordan A, Chekroud A, Moore EZ. Differences in words used to describe racial and gender groups in medical student performance evaluations. *PLoS One*. 2017;12(8):e0181659. doi:10.1371/journal.pone.0181659.
27. Akaike H. A new look at the statistical model identification. *Automatica*. 1978;19(6):465–471.
28. Cohen ER, Goldstein JL, Schroedl CJ, Parlapiano N, McGaghie WC, Wayne DB. Are USMLE scores valid measures for chief resident selection? *J Grad Med Educ*. 2020;12(4):441–446. doi:10.4300/JGME-D-19-00782.1.
29. McGaghie WC, Cohen ER, Wayne DB. Are United States Medical Licensing Exam Step 1 and 2 scores valid measures for postgraduate medical residency selection decisions? *Acad Med*. 2011;86(1):48–52. doi:10.1097/ACM.0b013e3181ffacdb.
30. Stephenson-Famy A, Houmar BS, Oberoi S, Manyak A, Chiang S, Kim S. Use of the interview in resident candidate selection: a review of the literature. *J Grad Med Educ*. 2015;7(4):539–548. doi:10.4300/JGME-D-14-00236.1.
31. Hamstra SJ, Yamazaki K, Barton MA, Santen SA, Beeson MS, Holmboe ES. A national study of longitudinal consistency in ACGME Milestone ratings by clinical competency committees: exploring an aspect of validity in the assessment of residents' competence. *Acad Med*. 2019;94(10):1522–1531. doi:10.1097/ACM.0000000000002820.
32. 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.
33. Galvin SL, Parlier AB, Martino E, Scott KR, Buys E. Gender bias in nurse evaluations of residents in obstetrics and gynecology. *Obstet Gynecol*. 2015;126(suppl 4):7–12. doi:10.1097/AOG.0000000000001044.
34. Holmboe ES, Huot SJ, Brienza RS, Hawkins RE. The association of faculty and residents' gender on faculty evaluations of internal medicine residents in 16 residencies. *Acad Med*. 2009;84(3):381–384. doi:10.1097/ACM.0b013e3181971c6d.
35. Santen SA, Yamazaki K, Holmboe ES, Yarris LM, Hamstra SJ. Comparison of male and female resident Milestone assessments during emergency medicine residency training: a national study. *Acad Med*. 2020;95(2):263–268. doi:10.1097/ACM.0000000000002988.
36. Kennedy TJ, Regehr G, Baker GR, Lingard L. Point-of-care assessment of medical trainee competence for independent clinical work. *Acad Med*. 2008;83(suppl 10):89–92. doi:10.1097/ACM.0b013e318183c8b7.
37. Radabaugh CL, Hawkins RE, Welcher CM, et al. Beyond the United States Medical Licensing Examination score: assessing competence for entering residency. *Acad Med*. 2019;94(7):983–989. doi:10.1097/ACM.0000000000002728.
38. Chen HC, van den Broek WE, ten Cate O. The case for use of entrustable professional activities in undergraduate medical education. *Acad Med*. 2015;90(4):431–436. doi:10.1097/ACM.0000000000000586.



At the time of writing, **Blair P. Golden, MD, MS**, was Chief Resident, Internal Medicine Residency, and Clinical Instructor, Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine, and currently is Assistant Professor, Division of Hospital Medicine, Department of Medicine, University of Wisconsin School of Medicine and Public Health; **Bruce L. Henschen, MD, MPH**, is Assistant Professor, Division of General Internal Medicine and

Geriatrics, Northwestern University Feinberg School of Medicine; **David T. Liss, PhD**, is Research Associate Professor, Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine; **Sara L. Kiely, MS**, is Accreditation Council for Graduate Medical Education Liaison, McGaw Medical Center of Northwestern University Feinberg School of Medicine; and **Aashish K. Didwania, MD**, is Associate Professor, Division of General Internal Medicine and Geriatrics, Program Director, Internal Medicine Residency, and Vice Chair for Education, Department of Medicine, Northwestern University Feinberg School of Medicine.

Funding: The authors report no external funding source for this study.

Conflict of interest: The authors declare they have no competing interests.

The authors would like to thank Drs Diane Wayne, Meghan McHugh, and Joseph Feinglass for reviewing a draft of the manuscript.

Corresponding author: Blair P. Golden, MD, MS, University of Wisconsin School of Medicine and Public Health, bpgolden@medicine.wisc.edu, Twitter @blairgolden

Received June 9, 2020; revisions received October 7, 2020, and December 3, 2020; accepted December 9, 2020.