



Trends in Milestones Data

Readers' Guide

The abstracts in this document are organized by year of publication (in descending order). Readers with a particular interest in a subject may use the search function in the PDF to quickly find topics of their choosing. Please note that since the following content comes from a variety of sources, there may be variations in style and spelling. For official ACGME terminology and terminology usage, refer to the ACGME Glossary of Terms.

Background:

The purpose of milestones is to achieve the objectives of outcomes-based education, where the desired outcomes of training, rather than the content or duration of specific rotations, shape the curriculum design. The Milestones program, launched in 2013, provides a comprehensive analysis of performance within the six ACGME competencies. While competencies specify the essential outcomes required for practice across medical specialties, Milestones outline the specialty-specific competencies and knowledge that indicate advancement within each competency.

What's in the literature:

One trend that has emerged is the use of data analytics to track and analyze Milestones data. This allows program directors and educators to identify trends and patterns in resident and fellow performance, which can help inform curriculum and assessment development. For example, they can identify areas where residents and fellows may be struggling and develop targeted interventions to help them improve their skills and knowledge in those areas. Other Milestones trends that readers can find in the bibliography include:

- Assessing the construct and response process validity of milestones in fellowship programs
- Analyzing a residents' self-evaluation and faculty evaluation using the Milestones
- Racial and Ethnic Differences in assessments
- Are CCCs consistent in applying milestones ratings for learners' overtime?
- Predicting performance at graduation from early ACGME Milestones ratings
- Changes in Milestones levels during the transition from residency to fellowship training

Racial and Ethnic Differences in Internal Medicine Residency Assessments

Boatright D, Anderson N, Kim JG, Holmboe ES, McDade WA, Fancher T, Gross CP, Chaudhry S, Nguyen M, Nguemini Tiako MJ, Colson E, Xu Y, Li F, Dziura JD, Saha S. JAMA Netw Open. 2022 Dec 1;5(12):e2247649. doi: 10.1001/jamanetworkopen.2022.47649. PMID: 36580337.

IMPORTANCE:

Previous studies have demonstrated racial and ethnic inequities in medical student assessments, awards, and faculty promotions at academic medical centers. Few data exist about similar racial and ethnic disparities at the level of graduate medical education.

OBJECTIVE:

To examine the association between race and ethnicity and performance assessments among a national cohort of internal medicine residents.

DESIGN:

This retrospective cohort study evaluated assessments of performance for 9026 internal medicine residents from the graduating classes of 2016 and 2017 at Accreditation Council of Graduate Medical Education (ACGME)-accredited internal medicine residency programs in the US. Analyses were conducted between July 1, 2020, and June 31, 2022.

MAIN OUTCOMES AND MEASURES:

The primary outcome was midyear and year-end total ACGME Milestone scores for underrepresented in medicine (URiM [Hispanic only; non-Hispanic American Indian, Alaska Native, or Native Hawaiian/Pacific Islander only; or non-Hispanic Black/African American]) and Asian residents compared with White residents as determined by their Clinical Competency Committees and residency program directors. Differences in scores between Asian and URiM residents compared with White residents were also compared for each of the 6 competency domains as supportive outcomes.

RESULTS:

The study cohort included 9026 residents from 305 internal medicine residency programs. Of these residents, 3994 (44.2%) were female, 3258 (36.1%) were Asian, 1216 (13.5%) were URiM, and 4552 (50.4%) were White. In the fully adjusted model, no difference was found in the initial midyear total Milestone scores between URiM and White residents, but there was a difference between Asian and White residents, which favored White residents (mean [SD] difference in scores for Asian residents: -1.27 [0.38]; $P < .001$). In the second year of training, White residents received increasingly higher scores relative to URiM and Asian residents. These racial disparities peaked in postgraduate year (PGY) 2 (mean [SD] difference in scores for URiM residents, -2.54 [0.38]; $P < .001$; mean [SD] difference in scores for Asian residents, -1.9 [0.27]; $P < .001$). By the final year 3 assessment, the gap between White and Asian and URiM residents' scores narrowed, and no racial or ethnic differences were found. Trends in racial and ethnic differences among the 6 competency domains mirrored total Milestone scores, with differences peaking in PGY2 and then decreasing in PGY3 such that parity in assessment was reached in all competency domains by the end of training.

CONCLUSIONS:

In this cohort study, URiM and Asian internal medicine residents received lower ratings on performance assessments than their White peers during the first and second years of training, which may reflect racial bias in assessment. This disparity in assessment may limit opportunities for physicians from minoritized racial and ethnic groups and hinder physician workforce diversity.

Simulation as Soothsayer: Simulated Surgical Skills MMIs During Residency Interviews are Associated With First Year Residency Performance

Lund S, D'Angelo JD, Baloul M, Yeh VJ, Stulak J, Rivera M. *J Surg Educ.* 2022 Nov-Dec;79(6):e235-e241. doi: 10.1016/j.jsurg.2022.06.002. Epub 2022 Jun 18. PMID: 35725725.

OBJECTIVE:

The main consideration during residency recruitment is identifying applicants who will succeed during residency. However, few studies have identified applicant characteristics that are associated with competency development during residency, such as the Accreditation Council for Graduate Medical Education milestones. As mini multiple interviews (MMIs) can be used to assess various competencies, we aimed to determine if simulated surgical skills MMI scores during a general surgery residency interview were associated with Accreditation Council for Graduate Medical Education milestone ratings at the conclusion of intern year.

DESIGN:

Retrospective cohort study. Interns' Step 1 and 2 clinical knowledge (CK) scores, interview day simulated surgical skills MMI overall score, traditional faculty interview scores, average overall milestone ratings in the spring of residency, and intern American Board of Surgery In-Training Examination (ABSITE) percentile scores were gathered. Two multiple linear regression were performed analyzing the association between Step 1, Step 2 CK, MMI, and traditional faculty interview scores with (1) average overall milestone rating and (2) ABSITE percentile scores, controlling for categorical/preliminary intern classification.

SETTING:

One academic medical center PARTICIPANTS: General surgery interns matriculating in 2020-2021 RESULTS: Nineteen interns were included. Multiple linear regression revealed that higher overall simulated surgical skills MMI score was associated with higher average milestone ratings ($\beta = .45$, $p = 0.03$) and higher ABSITE score ($\beta = .43$, $p = 0.02$) while neither Step 1, Step 2 CK, nor faculty interview scores were significantly associated with average milestone ratings.

CONCLUSIONS:

Surgical residency programs invest a tremendous amount of effort into training residents, thus metrics for predicting applicants that will succeed are needed. Higher scores on a simulated surgical skills MMIs are associated with higher milestone ratings 1 year into residency and higher intern ABSITE percentiles. These results indicate a noteworthy method, simulated surgical skills MMIs, as an additional metric that may select residents that will have early success in residency.

Association Between Parental Leave and Ophthalmology Resident Physician Performance

Huh DD, Wang J, Flitsos MJ, Beal CJ, Boente CS, Wisely CE, De Andrade LM, Lorch AC, Ramanathan S, Reinoso MA, Swamy RN, Waxman EL, Woreta FA, Srikumaran D. JAMA Ophthalmol. 2022 Nov 1;140(11):1066-1075. doi: 10.1001/jamaophthalmol.2022.3778. PMID: 36173610; PMCID: PMC9523550.

IMPORTANCE:

Although parental leave is essential in enhancing resident wellness and fostering inclusive workplace environments, residents may often feel discouraged from using parental leave owing to perceived stigma and concerns about possible negative effects on their training.

OBJECTIVE:

To examine parental leave usage across multiple institutions and compare residency performance metrics between residents who took parental leave vs their peers who did not take leave.

SETTINGS:

This was a retrospective cross-sectional analysis conducted from April 1, 2020, to July 28, 2022, of educational records. Multicenter data were obtained from 10 Accreditation Council for Graduate Medical Education (ACGME)-accredited ophthalmology programs across the US. Included ophthalmology residents graduated between 2015 and 2019. Data were analyzed from August 15, 2021, to July 25, 2022.

EXPOSURES:

Performance metrics of residents who used parental leave during residency were compared with those of residents who did not take parental leave.

OUTCOMES:

Measures of performance included the Ophthalmic Knowledge Assessment Program (OKAP) scores, ACGME milestones scores, board examination pass rates, research activity, and surgical volumes.

RESULTS:

Of the 283 ophthalmology residents (149 male [52.7%]) included in the study, 44 (15.5%) took a median (IQR) parental leave of 4.5 (2-6) weeks. There were no differences in average OKAP percentiles, research activity, average ACGME milestones scores, or surgical volume between residents who took parental leave and those who did not. Residents who pursued fellowship were less likely to have taken parental leave (odds ratio [OR], 0.43; 95% CI, 0.27-0.68; $P < .001$), and residents who practiced in private settings after residency were more likely to have taken parental leave (OR, 3.56; 95% CI, 1.79-7.08; $P < .001$). When stratified by sex, no differences were identified in performance between female residents who took parental leave compared with residents who did not take leave, except a mild surgical number difference in 1 subspecialty category of keratorefractive procedures (difference in median values, -2; 95% CI, -3.7 to -0.3; $P = .03$).

CONCLUSION:

In this multicenter cross-sectional study, no differences in performance metrics were identified between residents taking parental leave compared with their peers. These findings may provide reassurance to trainees and program directors regarding the unlikelihood, on average, that taking adequate parental leave will affect performance metrics adversely.

Predicting Performance at Graduation From Early ACGME Milestone Ratings: Longitudinal Learning Analytics in Professionalism and Communication in Vascular Surgery

Smith BK, Yamazaki K, Luman A, Tekian A, Holmboe E, Mitchell EL, Park YS, Hamstra SJ. *J Surg Educ.* 2022 Sep 28:S1931-7204(22)00234-3. doi: 10.1016/j.jsurg.2022.09.002. Epub ahead of print. PMID: 36182635.

OBJECTIVE:

Program directors in surgical disciplines need more tools from the ACGME to help them use Milestone ratings to improve trainees' performance. This is especially true in competencies that are notoriously difficult to measure, such as professionalism (PROF) and interpersonal and communication skills (ICS). It is now widely understood that skills in these two areas have direct impact on patient care outcomes. This study investigated the potential for generating early predictors of final Milestone ratings within the PROF and ICS competency categories.

DESIGN:

This retrospective cohort study utilized Milestone ratings from all ACGME-accredited vascular surgery training programs, covering residents and fellows who completed training in June 2019. The outcome measure studied was the rate of achieving the recommended graduation target of Milestone Level 4 (possible range: 1-5), while the predictors were the Milestone ratings attained at earlier stages of training. Predictive probability values (PPVs) were calculated for each of the 3 PROF and two ICS sub-competencies to estimate the probability of trainees not reaching the recommended graduation target based on their previous Milestone ratings.

SETTING:

All ACGME-accredited vascular surgery training programs within the United States.

PARTICIPANTS:

All trainees completing a 2 year vascular surgery fellowship (VSF) in June 2019 (n = 119) or a 5 year integrated vascular surgery residency (IVSR) in June 2019 (n = 52) were included in the analyses.

RESULTS:

The overall rate of failing to achieve the recommended graduation target across all PROF and ICS sub-competencies ranged from 7.7% to 21.8% of all trainees. For trainees with a Milestone rating at ≤ 2.5 with 1 year remaining in their training program, the predictive probability of not achieving the recommended graduation target ranged from 37.0% to 71.5% across sub-competencies, with the highest risks observed under PROF for "Administrative Tasks" (71.5%) and under ICS for "Communication with the Healthcare Team" (56.7%).

CONCLUSIONS:

As many as 1 in 4 vascular surgery trainees did not achieve the ACGME vascular surgery Milestones targets for graduation in at least one of the PROF and ICS sub-competencies. Biannual ACGME Milestone assessment ratings of PROF and ICS during early training can be used to predict achievement of competency targets at time of graduation. Early clues to problems in PROF and ICS enable programs to address potential deficits early in training to ensure competency in these essential non-technical skills prior to entering unsupervised practice.

Development of entrustable professional activities for regional anesthesia and pain medicine fellowship training

Porter S, Prendiville E, Allen BFS, Booth G, Boublik J, Burnett GW, Elkassabany N, Hausman J, Klesius L, Le-Wendling L, Machi AT, Maniker R, Parra M, Rosenquist R, Spofford CM, Suresh S, Tedore T, Wilson EH, Zhou JY, Woodworth G. *Reg Anesth Pain Med*. 2022 Jul 25:rapm-2022-103854. doi: 10.1136/rapm-2022-103854. Epub ahead of print. PMID: 35878963.

INTRODUCTION:

The Accreditation Council for Graduate Medical Education (ACGME) offers descriptions of competencies and milestones but does not provide standardized assessments to track trainee competency. Entrustable professional activities (EPAs) and special assessments (SAs) are emerging methods to assess the level of competency obtained by regional anesthesiology and acute pain medicine (RAAPM) fellows.

METHODS:

A panel of RAAPM physicians with experience in education and competency assessment and one medical student were recruited to participate in a modified Delphi method with iterative rounds to reach consensus on: a list of EPAs, SAs, and procedural skills; detailed definitions for each EPA and SA; a mapping of the EPAs and SAs to the ACGME milestones; and a target level of entrustment for graduating US RAAPM fellows for each EPA and procedural skill. A gap analysis was performed and a heat map was created to cross-check the EPAs and SAs to the ACGME milestones.

RESULTS:

Participants in EPA and SA development included 19 physicians and 1 medical student from 18 different programs. The Delphi rounds yielded a final list of 23 EPAs, a defined entrustment scale, mapping of the EPAs to ACGME milestones, and graduation targets. A list of 73 procedural skills and 7 SAs were similarly developed.

DISCUSSION:

A list of 23 RAAPM EPAs, 73 procedural skills, and 7 SAs were created using a rigorous methodology to reach consensus. This framework can be utilized to help assess RAAPM fellows in the USA for competency and allow for meaningful performance feedback.

The Impact of Global Health Experiences on the Emergency Medicine Residency Milestones

Hayward AS, Lee SS, Douglass K, Jacquet GA, Hudspeth J, Walrath J, Dreifuss BA, Baird J, Tupesis JP. *J Med Educ Curric Dev.* 2022 May 11;9:23821205221083755. doi: 10.1177/23821205221083755. PMID: 35572845; PMCID: PMC9102119.

OBJECTIVES:

Identify the impact of experiences in global health (GH) on the Accreditation Council for Graduate Medical Education (ACGME) competencies in emergency medicine (EM) residents and describe the individual characteristics of EM residents with global health experience compared to those without.

METHODS:

From 2015 to 2018, 117 residents from 13 nationally accredited United States EM residency training programs were surveyed. Specifically, the survey gathered demographic data and information regarding timing, type, location and duration of short term experiences in global health (STEGH). The survey collected both qualitative and quantitative data regarding resident experiences, including number of procedures performed and self-assessment of the impact on their residency milestones. ACGME milestone data from survey respondents was collected from each resident's training program coordinators. Chi-squared analysis and t-tests were conducted to assess differences between residents with STEGH and those without. A generalized linear model (GLM) was utilized to assess the effects of time and experience with interaction on achieving milestones in each of the competency domains, to compare milestone achievement over time between residents with STEGH and those without.

RESULTS:

Out of 117 EM residents, 60 were female (44%), the mean age was 30 years (standard deviation = 3.1), and 84 (71.8%) reported STEGH in general, including prior to residency (64.5%). 33 (28.2%) reported having completed STEGH during residency. The results of the GLM analysis showed that residents with STEGH during residency had significantly higher scores compared to those without the experience or STEGH pre-residency across all six competencies.

CONCLUSIONS:

STEGH in EM residents was associated with higher milestone achievement in certain ACGME competency domains including medical knowledge, practice-based learning and improvement, and professionalism. Participation in STEGH during residency appeared to show the strongest effect, with higher scores across all six competencies.

Longitudinal Reliability of Milestones-Based Learning Trajectories in Family Medicine Residents

Park YS, Hamstra SJ, Yamazaki K, Holmboe E. JAMA Netw Open. 2021 Dec 1;4(12):e2137179. doi: 10.1001/jamanetworkopen.2021.37179. PMID: 34874406; PMCID: PMC8652607.

IMPORTANCE:

Longitudinal Milestones data reported to the Accreditation Council for Graduate Medical Education (ACGME) can be used to measure the developmental and educational progression of learners. Learning trajectories illustrate the pattern and rate at which learners acquire competencies toward unsupervised practice.

OBJECTIVE:

To investigate the reliability of learning trajectories and patterns of learning progression that can support meaningful intervention and remediation for residents.

DESIGN:

This national retrospective cohort study included Milestones data from residents in family medicine, representing 6 semi-annual reporting periods from July 2016 to June 2019.

INTERVENTIONS:

Longitudinal formative assessment using the Milestones assessment system reported to the ACGME.

MAIN OUTCOMES TO MEASURE:

To estimate longitudinal consistency, growth rate reliability (GRR) and growth curve reliability (GCR) for 22 subcompetencies in the ACGME family medicine Milestones were used, incorporating clustering effects at the program level. Latent class growth curve models were used to examine longitudinal learning trajectories.

RESULTS:

This study included Milestones ratings from 3872 residents in 514 programs. The Milestones reporting system reliably differentiated individual longitudinal patterns for formative purposes (mean [SD] GRR, 0.63 [0.03]); there was also evidence of precision for model-based rates of change (mean [SD] GCR, 0.91 [0.02]). Milestones ratings increased significantly across training years and reporting periods (mean [SD] of 0.55 [0.04] Milestones units per reporting period; $P < .001$); patterns of developmental progress varied by subcompetency. There were 3 or 4 distinct patterns of learning trajectories for each of the 22 subcompetencies. For example, for the professionalism subcompetency, residents were classified to 4 groups of learning trajectories; during the 3-year family medicine training period, trajectories diverged further after postgraduate year (PGY) 1, indicating a potential remediation point between the end of PGY 1 and the beginning of PGY 2 for struggling learners, who represented 16% of learners (620 residents). Similar inferences for learning trajectories were found for practice-based learning and improvement, systems-based practice, and interpersonal and communication skills. Subcompetencies in medical knowledge and patient care demonstrated more consistent patterns of upward growth.

CONCLUSIONS:

These findings suggest that the Milestones reporting system provides reliable longitudinal data for individualized tracking of progress in all subcompetencies. Learning trajectories with supporting reliability evidence could be used to understand residents' developmental progress and tailored for individualized learning plans and remediation.

Longitudinal Milestone Assessment Extending Through Subspecialty Training: The Relationship Between ACGME Internal Medicine Residency Milestones and Subsequent Pulmonary and Critical Care Fellowship Milestones

Heath JK, Wang T, Santhosh L, Denson JL, Holmboe E, Yamazaki K, Clay AS, Carlos WG. Acad Med. 2021 Nov 1;96(11):1603-1608. doi: 10.1097/ACM.0000000000004165. PMID: 34010863.

PURPOSE:

Accreditation Council for Graduate Medical Education (ACGME) milestones were implemented across medical subspecialties in 2015. Although milestones were proposed as a longitudinal assessment tool potentially providing opportunities for early implementation of individualized fellowship learning plans, the association of subspecialty fellowship ratings with prior residency ratings remains unclear. This study aimed to assess the relationship between internal medicine (IM) residency milestones and pulmonary and critical care medicine (PCCM) fellowship milestones.

METHOD:

A multicenter retrospective cohort analysis was conducted for all PCCM trainees in ACGME- accredited PCCM fellowship programs, 2017-2018, who had complete prior IM milestone ratings from 2014 to 2017. Only professionalism and interpersonal and communication skills (ICS) were included based on shared anchors between IM and PCCM milestones. Using a generalized estimating equations model, the association of PCCM milestones ≤ 2.5 during the first fellowship year with corresponding IM subcompetencies was assessed at each time point, nested by program. Statistical significance was determined using logistic regression.

RESULTS:

The study included 354 unique PCCM fellows. For ICS and professionalism subcompetencies, fellows with higher IM ratings were less likely to obtain PCCM ratings ≤ 2.5 during the first fellowship year. Each ICS subcompetency was significantly associated with future lapses in fellowship (ICS01: $\beta = -0.67$, $P = .003$; ICS02: $\beta = -0.70$, $P = .001$; ICS03: $\beta = -0.60$, $P = .004$) at various residency time points. Similar associations were noted for PROF03 ($\beta = -0.57$, $P = .007$).

CONCLUSIONS:

Findings demonstrated an association between IM milestone ratings and low milestone ratings during PCCM fellowship. IM trainees with low ratings in several professionalism and ICS subcompetencies were more likely to be rated ≤ 2.5 during the first PCCM fellowship year. This highlights a potential use of longitudinal milestones to target educational gaps at the beginning of PCCM fellowship.

Trainee Sex and Accreditation Council for Graduate Medical Education Milestone Assessments During General Surgery Residency

Landau SI, Syvyk S, Wirtalla C, Aarons CB, Butts S, Holmboe E, Kelz RR. *JAMA Surg.* 2021 Oct 1;156(10):925-931. doi: 10.1001/jamasurg.2021.3005. PMID: 34232269; PMCID: PMC8264749.

IMPORTANCE:

In evaluating the effectiveness of general surgery (GS) training, an unbiased assessment of the progression of residents with attention to individual learner factors is imperative.

OBJECTIVE:

To evaluate the role of trainee sex in milestone achievement over the course of GS residency using national data from the Accreditation Council for Graduate Medical Education (ACGME).

DESIGN:

This cross-sectional study evaluated female and male GS residents enrolled in ACGME- accredited programs in the US from 2014 to 2018 with reported variation in milestones performance across years in training and representation. Data were analyzed from November 2019 to June 2021.

MAIN OUTCOMES AND MEASURES:

Mean reported milestone score at initial and final assessment, and predicted time-to-attainment of equivalent performance by sex.

RESULTS:

Among 4476 GS residents from 250 programs who had milestone assessments at any point in their clinical training, 1735 were female (38.8%). Initially, female and male residents received similar mean (SD) milestone scores (1.95 [0.50] vs 1.94 [0.50]; $P = .69$). At the final assessment, female trainees received overall lower mean milestone scores than male trainees (4.25 vs 4.31; $P < .001$). Significantly lower mean milestone scores were reported for female residents at the final assessment for several subcompetencies in both univariate and multivariate analyses, with only medical knowledge 1 (pathophysiology, diagnosis, and initial management) common to both. Multilevel mixed-effects linear modeling demonstrated that female trainees had significantly lower rates of monthly milestone attainment in the subcompetency of medical knowledge 1, which was associated with a significant difference in training time of approximately 1.8 months.

CONCLUSIONS:

Both female and male GS trainees achieved the competency scores necessary to transition to independence after residency as measured by the milestones assessment system. Initially, there were no sex differences in milestone score. By graduation, there were differences in the measured assessment of female and male trainees across several subcompetencies. Careful monitoring for sex bias in the evaluation of trainees and scrutiny of the training process is needed to ensure that surgical residency programs support the educational needs of both female and male trainees.

Using Learning Analytics to Examine Achievement of Graduation Targets for Systems- Based Practice and Practice-Based Learning and Improvement: A National Cohort of Vascular Surgery Fellows

Smith BK, Luman A, Yamazaki K, Tekian A, Hamstra SJ, Holmboe E, Mitchell EL, Park YS. *Ann Vasc Surg.* 2021 Oct;76:463-471. doi: 10.1016/j.avsg.2021.03.046. Epub 2021 Apr 24. PMID: 33905852.

BACKGROUND:

Surgeons provide patient care in complex health care systems and must be able to participate in improving both personal performance and the performance of the system. The Accreditation Council for Graduate Medical Education (ACGME) Vascular Surgery Milestones are utilized to assess vascular surgery fellows' (VSF) achievement of graduation targets in the competencies of Systems Based Practice (SBP) and Practice Based Learning and Improvement (PBLI). We investigate the predictive value of semiannual milestones ratings for final achievement within these competencies at the time of graduation.

METHODS:

National ACGME milestones data were utilized for analysis. All trainees entering the 2-year vascular surgery fellowship programs in July 2016 were included in the analysis (n = 122). Predictive probability values (PPVs) were obtained for each SBP and PBLI sub-competencies by biannual review periods, to estimate the probability of VSFs not reaching the recommended graduation target based on their previous milestones ratings.

RESULTS:

The rate of nonachievement of the graduation target level 4.0 on the SBP and PBLI sub-competencies at the time of graduation for VSFs was 13.1-25.4%. At the first time point of assessment, 6 months into the fellowship program, the PPV of the SBP and PBLI milestones for nonachievement of level 4.0 upon graduation ranged from 16.3-60.2%. Six months prior to graduation, the PPVs across the 6 sub-competencies ranged from 14.6-82.9%.

CONCLUSIONS:

A significant percentage of VSFs do not achieve the ACGME Vascular Surgery Milestone targets for graduation in the competencies of SBP and PBLI, suggesting a need to improve curricula and assessment strategies in these domains across vascular surgery fellowship programs. Reported milestones levels across all time point are predictive of ultimate achievement upon graduation and should be utilized to provide targeted feedback and individualized learning plans to ensure graduates are prepared to engage in personal and health care system improvement once in unsupervised practice.

Exploring the Association Between USMLE Scores and ACGME Milestone Ratings: A Validity Study Using National Data From Emergency Medicine

Hamstra SJ, Cuddy MM, Jurich D, Yamazaki K, Burkhardt J, Holmboe ES, Barone MA, Santen SA. Acad Med. 2021 Sep 1;96(9):1324-1331. doi: 10.1097/ACM.0000000000004207. PMID: 34133345; PMCID: PMC8378430.

PURPOSE:

The United States Medical Licensing Examination (USMLE) sequence and the Accreditation Council for Graduate Medical Education (ACGME) milestones represent 2 major components along the continuum of assessment from undergraduate through graduate medical education. This study examines associations between USMLE Step 1 and Step 2 Clinical Knowledge (CK) scores and ACGME emergency medicine (EM) milestone ratings.

METHOD:

In February 2019, subject matter experts (SMEs) provided judgments of expected associations for each combination of Step examination and EM subcompetency. The resulting sets of subcompetencies with expected strong and weak associations were selected for convergent and discriminant validity analysis, respectively. National-level data for 2013-2018 were provided; the final sample included 6,618 EM residents from 158 training programs. Empirical bivariate correlations between milestone ratings and Step scores were calculated, then those correlations were compared with the SMEs' judgments. Multilevel regression analyses were conducted on the selected subcompetencies, in which milestone ratings were the dependent variable, and Step 1 score, Step 2 CK score, and cohort year were independent variables.

RESULTS:

Regression results showed small but statistically significant positive relationships between Step 2 CK score and the subcompetencies (regression coefficients ranged from 0.02 [95% confidence interval (CI), 0.01-0.03] to 0.12 [95% CI, 0.11-0.13]; all $P < .05$), with the degree of association matching the SMEs' judgments for 7 of the 9 selected subcompetencies. For example, a 1 standard deviation increase in Step 2 CK score predicted a 0.12 increase in MK-01 milestone rating, when controlling for Step 1. Step 1 score showed a small statistically significant effect with only the MK-01 subcompetency (regression coefficient = 0.06 [95% CI, 0.05-0.07], $P < .05$).

CONCLUSIONS:

These results provide incremental validity evidence in support of Step 1 and Step 2 CK score and EM milestone rating uses.

Milestone Learning Trajectories of Residents at Five Anesthesiology Residency Programs

Tanaka P, Park YS, Roby J, Ahn K, Kakazu C, Udani A, Macario A. Teach Learn Med. 2021 Jun-Jul;33(3):304-313. doi: 10.1080/10401334.2020.1842210. Epub 2020 Dec 17. PMID: 33327788.

CONSTRUCT:

Every six months, residency programs report their trainees' Milestones Level achievement to the Accreditation Council for Graduate Medical Education (ACGME). Milestones should enable the learner and training program to know an individual's competency development trajectory.

BACKGROUND:

Milestone Level ratings for residents grouped by specialty (e.g., Internal Medicine and Emergency Medicine) show that, in aggregate, senior residents receive higher ratings than junior residents. Anesthesiology Milestones, as assessed by both residents and faculty, also have a positive linear relationship with postgraduate year. However, these studies have been cross-sectional rather than longitudinal cohort studies, and studies of how individual residents progress during the course of training are needed. Longitudinal data analysis of performance assessment trajectories addresses a relevant validity question for the Next Accreditation System. We explored the application of learning analytics to longitudinal Milestones data to: 1) measure the frequency of "straight-lining"; 2) assess the proportion of residents that reach "Level 4" (ready for unsupervised practice) by graduation for each subcompetency; 3) identify variability among programs and individual residents in their baseline Milestone Level and rates of improvement; and 4) determine how hypothetically constructed growth curve models fit to the Milestones data reported to ACGME.

APPROACH:

De-identified Milestone Level ratings in each of the 25 subcompetencies submitted semiannually to the ACGME from July 1, 2014 to June 30, 2017 were retrospectively analyzed for graduating residents (n = 67) from a convenience sample of five anesthesia residency programs. The data reflected longitudinal resident Milestone progression from the beginning of the first year to the end of the third and final year of clinical anesthesiology training. The frequency of straight-lining, defined as the resident receiving the same exact Milestone Level rating for all 25 subcompetencies on a given 6-month report, was calculated for each program. Every resident was evaluated six times during training with the possibility of six straight-lined ratings.

FINDINGS:

The number of residents in each program ranged from 5-21 (Median 13, range 16). Mean Milestone Level ratings for subcompetencies were significantly different at each six-month assessment ($p < 0.001$). Frequency of straight-lining varied significantly by program from 9% - 57% (Median 22%). Depending on the program, 53%-100% (median 86%) of residents reached the graduation target Level 4 or higher in all 25 anesthesiology subcompetencies. Nine to 18% of residents did not achieve a Level 4 rating for at least one subcompetency at any time during their residency. Across programs, significant variability was found in first-year clinical anesthesia training Milestone Levels, as well in the rate of improvement for five of the six core competencies.

CONCLUSIONS:

Anesthesia residents' Milestone Level growth trajectories as reported to the ACGME vary significantly across individual residents as well as by program. The present study offers a case example that raises concerns regarding the validity of the Next Accreditation System as it is currently used by some residency programs.

Milestone Level Changes from Residency to Fellowship: A Multicenter Cohort Study

Sawyer T, Gray M, Chabra S, Johnston LC, Carbajal MM, Gillam-Krakauer M, Brady JM, French H. J Grad Med Educ. 2021 Jun;13(3):377-384. doi: 10.4300/JGME-D-20-00954.1. Epub 2021 Jun 14. PMID: 34178263; PMCID: PMC8207935.

BACKGROUND:

A vital element of the Next Accreditation System is measuring and reporting educational Milestones. Little is known about changes in Milestones levels during the transition from residency to fellowship training.

OBJECTIVE:

Evaluate the Accreditation Council for Graduate Medical Education (ACGME) Milestones' ability to provide a linear trajectory of professional development from general pediatrics residency to neonatal-perinatal medicine (NPM) fellowship training.

METHODS:

We identified 11 subcompetencies that were the same for general pediatrics residency and NPM fellowship. We then extracted the last residency Milestone level and the first fellowship Milestone level for each subcompetency from the ACGME's Accreditation Data System on 89 subjects who started fellowship training between 2014 and 2018 at 6 NPM fellowship programs. Mixed-effects models were used to examine the intra-individual changes in Milestone scores between residency and fellowship after adjusting for the effects of the individual programs.

RESULTS:

A total of 1905 subcompetency Milestone levels were analyzed. The average first fellowship Milestone levels were significantly lower than the last residency Milestone levels (residency, mean 3.99 [SD = 0.48] vs fellowship 2.51 [SD = 0.56]; $P < .001$). Milestone levels decreased by an average of -1.49 (SD = 0.65) from the last residency to the first fellowship evaluation. Significant differences in Milestone levels were seen in both context-dependent subcompetencies (patient care and medical knowledge) and context-independent subcompetencies (professionalism).

CONCLUSIONS:

Contrary to providing a linear trajectory of professional development, we found that Milestone levels were reset when trainees transitioned from general pediatrics residency to NPM fellowship.

Gender Differences in Milestone Ratings and Medical Knowledge Examination Scores Among Internal Medicine Residents

Hauer KE, Jurich D, Vandergrift J, Lipner RS, McDonald FS, Yamazaki K, Chick D, McAllister K, Holmboe ES. Acad Med. 2021 Jun 1;96(6):876-884. doi: 10.1097/ACM.0000000000004040. PMID: 33711841.

PURPOSE:

To examine whether there are group differences in milestone ratings submitted by program directors working with clinical competency committees (CCCs) based on gender for internal medicine (IM) residents and whether women and men rated similarly on milestones perform comparably on subsequent in-training and certification examinations.

METHOD:

This national retrospective study examined end-of-year medical knowledge (MK) and patient care (PC) milestone ratings and IM In-Training Examination (IM-ITE) and IM Certification Examination (IM-CE) scores for 2 cohorts (2014-2017, 2015-2018) of U.S. IM residents at ACGME-accredited programs. It included 20,098/21,440 (94%) residents, with 9,424 women (47%) and 10,674 men (53%). Descriptive statistics and differential prediction techniques using hierarchical linear models were performed.

RESULTS:

For MK milestone ratings in PGY-1, men and women showed no statistical difference at a significance level of .01 ($P = .02$). In PGY-2 and PGY-3, men received statistically higher average MK ratings than women ($P = .002$ and $P < .001$, respectively). In contrast, men and women received equivalent average PC ratings in each PGY ($P = .47$, $P = .72$, and $P = .80$, for PGY-1, PGY-2, and PGY-3, respectively). Men slightly outperformed women with similar MK or PC ratings in PGY-1 and PGY-2 on the IM-ITE by about 1.7 and 1.5 percentage points, respectively, after adjusting for covariates. For PGY-3 ratings, women and men with similar milestone ratings performed equivalently on the IM-CE.

CONCLUSIONS:

Milestone ratings were largely similar for women and men. Generally, women and men with similar MK or PC milestone ratings performed similarly on future examinations. Although there were small differences favoring men on earlier examinations, these differences disappeared by the final training year. It is questionable whether these small differences are educationally or clinically meaningful. The findings suggest fair, unbiased milestone ratings generated by program directors and CCCs assessing residents.

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

Golden BP, Henschen BL, Liss DT, Kiely SL, Didwania AK. J Grad Med Educ. 2021 Apr;13(2):213-222. doi: 10.4300/JGME-D-20-00603.1. Epub 2021 Apr 16. PMID: 33897955; PMCID: PMC8054584.

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.

Simulation-Based Assessments and Graduating Neurology Residents' Milestones: Status Epilepticus Milestones

Mikhaeil-Demo Y, Holmboe E, Gerard EE, Wayne DB, Cohen ER, Yamazaki K, Templer JW, Bega D, Culler GW, Bhatt AB, Shafi N, Barsuk JH. *J Grad Med Educ.* 2021 Apr;13(2):223-230. doi: 10.4300/JGME-D-20-00832.1. Epub 2021 Apr 16. PMID: 33897956; PMCID: PMC8054597.

BACKGROUND:

The American Board of Psychiatry and Neurology and the Accreditation Council for Graduate Medical Education (ACGME) developed Milestones that provide a framework for residents' assessment. However, Milestones do not provide a description for how programs should perform assessments.

OBJECTIVES:

We evaluated graduating residents' status epilepticus (SE) identification and management skills and how they correlate with ACGME Milestones reported for epilepsy and management/treatment by their program's clinical competency committee (CCC).

METHODS:

We performed a cohort study of graduating neurology residents from 3 academic medical centers in Chicago in 2018. We evaluated residents' skills identifying and managing SE using a simulation-based assessment (26-item checklist). Simulation-based assessment scores were compared to experience (number of SE cases each resident reported identifying and managing during residency), self-confidence in identifying and managing these cases, and their end of residency Milestones assigned by a CCC based on end-of-rotation evaluations.

RESULTS:

Sixteen of 21 (76%) eligible residents participated in the study. Average SE checklist score was 15.6 of 26 checklist items correct (60%, SD 12.2%). There were no significant correlations between resident checklist performance and experience or self-confidence. The average participant's level of Milestone for epilepsy and management/treatment was high at 4.3 of 5 (SD 0.4) and 4.4 of 5 (SD 0.4), respectively. There were no significant associations between checklist skills performance and level of Milestone assigned.

CONCLUSIONS:

Simulated SE skills performance of graduating neurology residents was poor. Our study suggests that end-of-rotation evaluations alone are inadequate for assigning Milestones for high-stakes clinical skills such as identification and management of SE.

Is there a Gender Bias in Milestones Evaluations in General Surgery Residency Training?

Kwasny L, Shebrain S, Munene G, Sawyer R. Am J Surg. 2021 Mar;221(3):505-508. doi: 10.1016/j.amjsurg.2020.12.020. Epub 2020 Dec 13. PMID: 33358140.

BACKGROUND:

Studies of gender disparity in surgical training have yielded conflicting results. We hypothesize that there is no influence of gender on resident self-evaluation Milestone (SEM) scores and those assigned by the Clinical Competency Committee (CCC).

METHODS:

42 residents (25 male & 17 female) and faculty completed 300 Accreditation Council for Graduate Medical Education (ACGME) Milestone evaluations over a 4-year period. Two-way ANOVA, intraclass correlations coefficients, and general linear mixed models were used for analysis.

RESULTS:

CCC Milestone scores from 150 evaluations, 51 (34%) for female residents and 99 (66%) for male residents, were compared to corresponding SEM scores. There is a high interrater reliability (self vs. CCC). There was a significant increase in scores with advancing PGY levels ($p < 0.001$). No effect of gender on Milestones scores ($p > 0.05$) was noted.

CONCLUSIONS:

We found no significant differences in Milestones scores between male and female residents as determined by the CCC. Both scores improved significantly as residents progressed in training.

Do Milestone Ratings Predict Physical Medicine and Rehabilitation Board Certification Examination Scores?

Francisco GE, Yamazaki K, Raddatz M, Sabharwal S, Robinson L, Kinney C, Holmboe E. *Am J Phys Med Rehabil.* 2021 Feb 1;100(2S Suppl 1):S34-S39. doi: 10.1097/PHM.0000000000001613. PMID: 33048889.

ABSTRACT:

The Accreditation Council of Graduate Medical Education developed the Milestones to assist training programs in assessing resident physicians in the context of their participation in Accreditation Council of Graduate Medical Education-accredited training programs. Biannual assessments are done over a resident's entire training period to define the trajectory in achieving specialty-specific competencies. As part of its process of initial certification, the American Board of Physical Medicine and Rehabilitation requires successful completion of two examinations administered approximately 9 mos apart. The Part I Examination measures a single dimensional construct, physical medicine and rehabilitation medical knowledge, whereas Part II assesses the application of medical and psychiatric knowledge to multiple domains, including data acquisition, problem solving, patient management, systems-based practice, and interpersonal and communication skills through specific patient case scenarios. This study aimed to investigate the validity of the Milestones by demonstrating its association with performance in the American Board of Physical Medicine and Rehabilitation certifying examinations. A cohort of 233 physical medicine and rehabilitation trainees in 3-yr residency programs (postgraduate year 2 entry) in the United States from academic years 2014-2016, who also took the American Board of Physical Medicine and Rehabilitation Parts I and II certifying examinations between 2016 and 2018, were included in the study. Milestones ratings in four distinct observation periods were correlated with scores in the American Board of Physical Medicine and Rehabilitation Parts I and II Examinations. Milestones ratings of medical knowledge (but not patient care, professionalism, problem-based learning, interpersonal and communication skills, and systems-based practice) predicted performance in subsequent Part I American Board of Physical Medicine and Rehabilitation Examination, but none of the Milestone ratings correlated with Part II Examination scaled scores.

Female Residents Give Themselves Lower Scores Than Male Colleagues and Faculty Evaluators on ACGME Milestones

Brady JM, Bray A, Kim P, Schneider B, Lippe J, Mercer D, Sutton K. *J Surg Educ.* 2020 Dec 18:S1931- 7204(20)30471-2. doi: 10.1016/j.jsurg.2020.12.003. Epub ahead of print. PMID: 33349566.

OBJECTIVE:

Orthopedic surgery is one of the specialties with the lowest number of women residents and practicing surgeons. The gender discrepancy in orthopedic residency training may drive a competency bias. We asked whether female orthopedic surgery residents score themselves lower on the Accreditation Council for Graduate Medical Education (ACGME) Milestones than their male counterparts, and lower than their faculty evaluators.

DESIGN:

We conducted a retrospective review of ACGME Milestone data from faculty and residents over a 4-year period. The data were analyzed using a snapshot of PGY2 (n = 20 residents) and PGY4 (n = 19 residents) scores, and using a Generalized Estimation Equation (GEE) to account for additional data points from the same residents over the 4-year data collection period.

SETTING:

Assessment scores were compiled from a single orthopedic surgery residency at Oregon Health & Science University from 2014 to 2017.

PARTICIPANTS:

The residency program has 5 residents in each program year (PGY1 through PGY5); a total of 25 residents during each year of the study were included.

RESULTS:

On average, female residents scored themselves lower than both their male counterparts and their faculty mentors. Female PGY2 self-evaluation scores were lower than males in both patient care ($p = 0.005$) and medical knowledge ($p < 0.001$). When the GEE model was applied to 99 responses from 41 residents over a 4-year period, there were no gender-related differences in resident self-evaluation scores and in faculty scores of male and female residents, with the exception of meniscal tear. For this milestone, faculty rated female residents lower than males. Furthermore, the differences between faculty evaluation scores and resident self-evaluation scores were significantly lower for males than for females for 4 of the clinical domains, as well as the systems-based practice domains of cost and communication.

CONCLUSIONS:

Our results indicate female residents are at risk for a competency bias during training, as reflected by evaluations using the ACGME Milestones.

Milestone Learning Trajectories of Residents at Five Anesthesiology Residency Programs

Tanaka P, Park YS, Roby J, Ahn K, Kakazu C, Udani A, Macario A. Teach Learn Med. 2020 Dec 17:1-10. doi: 10.1080/10401334.2020.1842210. Epub ahead of print. PMID: 33327788.

CONSTRUCT:

Every six months, residency programs report their trainees' Milestones Level achievement to the Accreditation Council for Graduate Medical Education (ACGME). Milestones should enable the learner and training program to know an individual's competency development trajectory.

BACKGROUND:

Milestone Level ratings for residents grouped by specialty (e.g., Internal Medicine and Emergency Medicine) show that, in aggregate, senior residents receive higher ratings than junior residents.

Anesthesiology Milestones, as assessed by both residents and faculty, also have a positive linear relationship with postgraduate year. However, these studies have been cross-sectional rather than longitudinal cohort studies, and studies of how individual residents progress during the course of training are needed. Longitudinal data analysis of performance assessment trajectories addresses a relevant validity question for the Next Accreditation System. We explored the application of learning analytics to longitudinal Milestones data to: 1) measure the frequency of "straight-lining"; 2) assess the proportion of residents that reach "Level 4" (ready for unsupervised practice) by graduation for each subcompetency; 3) identify variability among programs and individual residents in their baseline Milestone Level and rates of improvement; and 4) determine how hypothetically constructed growth curve models fit to the Milestones data reported to ACGME.

APPROACH:

De-identified Milestone Level ratings in each of the 25 subcompetencies submitted semiannually to the ACGME from July 1, 2014 to June 30, 2017 were retrospectively analyzed for graduating residents (n = 67) from a convenience sample of five anesthesia residency programs. The data reflected longitudinal resident Milestone progression from the beginning of the first year to the end of the third and final year of clinical anesthesiology training. The frequency of straight-lining, defined as the resident receiving the same exact Milestone Level rating for all 25 subcompetencies on a given 6-month report, was calculated for each program. Every resident was evaluated six times during training with the possibility of six straight-lined ratings.

FINDINGS:

The number of residents in each program ranged from 5-21 (Median 13, range 16). Mean Milestone Level ratings for subcompetencies were significantly different at each six-month assessment ($p < 0.001$). Frequency of straight-lining varied significantly by program from 9% - 57% (Median 22%). Depending on the program, 53%-100% (median 86%) of residents reached the graduation target Level 4 or higher in all 25 anesthesiology subcompetencies. Nine to 18% of residents did not achieve a Level 4 rating for at least one subcompetency at any time during their residency. Across programs, significant variability was found in first-year clinical anesthesia training Milestone Levels, as well in the rate of improvement for five of the six core competencies.

CONCLUSIONS:

Anesthesia residents' Milestone Level growth trajectories as reported to the ACGME vary significantly across individual residents as well as by program. The present study offers a case example that raises concerns regarding the validity of the Next Accreditation System as it is currently used by some residency programs.

Is there a gender bias in milestones evaluations in general surgery residency training?

Kwasny L, Shebrain S, Munene G, Sawyer R. Am J Surg. 2020 Dec 13:S0002-9610(20)30798-4. doi: 10.1016/j.amjsurg.2020.12.020. Epub ahead of print. PMID: 33358140.

BACKGROUND:

Studies of gender disparity in surgical training have yielded conflicting results. We hypothesize that there is no influence of gender on resident self-evaluation Milestone (SEM) scores and those assigned by the Clinical Competency Committee (CCC).

METHODS:

42 residents (25 male & 17 female) and faculty completed 300 Accreditation Council for Graduate Medical Education (ACGME) Milestone evaluations over a 4-year period. Two-way ANOVA, intraclass correlations coefficients, and general linear mixed models were used for analysis.

RESULTS:

CCC Milestone scores from 150 evaluations, 51 (34%) for female residents and 99 (66%) for male residents, were compared to corresponding SEM scores. There is a high interrater reliability (self vs. CCC). There was a significant increase in scores with advancing PGY levels ($p < 0.001$). No effect of gender on Milestones scores ($p > 0.05$) was noted.

CONCLUSIONS:

We found no significant differences in Milestones scores between male and female residents as determined by the CCC. Both scores improved significantly as residents progressed in training.

Association Between Entrustable Professional Activities and Milestones Evaluations: Real-time Assessments Correlate With Semiannual Reviews

Albright JB, Meier AH, Ruangvoravat L, VanderMeer TJ. *J Surg Educ.* 2020 Nov-Dec;77(6):e220- e228. doi: 10.1016/j.jsurg.2020.07.027. Epub 2020 Jul 31. PMID: 32747323.

OBJECTIVE:

Entrustable professional activities (EPAs) have been developed to refine competency-based education. The American Board of Surgery has initiated a 2-year pilot study to evaluate the impact of EPAs on the evaluation and feedback of surgical residents. The ACGME Milestones in Surgery is a semiannual competency-based evaluation program to measure resident progression through 16 professional attributes across 8 practice domains. The correlation between these 2 evaluation tools remains unclear. The purpose of this study is to evaluate this correlation through comparison of an EPA with the corresponding elements of the ACGME Milestones.

DESIGN:

From July, 2018 to October, 2019, all residents submitting EPA evaluations for gall bladder disease were evaluated for preoperative, intraoperative, and/or postoperative entrustability. The ratings were converted to a numerical rank from 0 to 4. Milestones scores from May 2019 and November 2019 were obtained for each resident, with scores ranging from 0 to 4. The gall bladder EPA incorporates the operative PC3 and MK2 and nonoperative PC1, PC2, and ICS3 components. Spearman rank correlation was conducted to evaluate the association between each resident's median EPA ranking and his/her milestones scores.

SETTING:

SUNY Upstate Medical University, Syracuse, NY, a university-based hospital.

PARTICIPANTS:

General surgery residents.

RESULTS:

Among 24 residents, 106 intraoperative EPA evaluations were. For both the May and November milestones, significant positive correlations were noted for PC3 (correlation coefficient $\rho = 0.690$, $p < 0.001$; $\rho = 0.876$, $p < 0.001$). Similarly, for MK2, a significant positive correlation was noted ($\rho = 0.882$, $p < 0.001$; $\rho = 0.759$, $p < 0.001$). Interestingly, significant positive correlations were also identified between the 3 nonoperative milestones and the intraoperative entrustability ranking.

CONCLUSIONS:

We observed significant correlations between EPAs for cholecystectomy and associated milestones evaluation scores. These findings indicate that EPAs may provide more timely and specific feedback than existing tools and, on aggregate, may improve upon existing formative feedback practices provided through the biannual evaluation of surgical residents.

A Not So Perfect Score: Factors Associated with the Rate of Straight Line Scoring in Oncology Training Programs

Hinchcliff E, Gunther J, Ponnio AE, Bednarski B, Onstad M, Shafer A, Frumovitz M, Jazaeri A, Urbauer D, Bodurka DC. *J Cancer Educ.* 2020 Aug 25. doi: 10.1007/s13187-020-01855-6. Epub ahead of print. PMID: 32839894.

ABSTRACT:

Straight line scoring (SLS), defined as trainee assessments with the same score for all evaluation items, is statistically improbable and potentially indicates inaccurate assessment. Factors contributing to higher SLS rates are unknown, and knowledge of SLS prevalence within oncologic training is lacking. SLS frequency was measured for evaluations from all Accreditation Council for Graduate Medical Education (ACGME)-accredited programs at a single cancer care institution between 2014 and 2018. SLS prevalence was estimated using hierarchical linear models (HLM) that considered characteristics of evaluator, trainee, and evaluation potentially related to SLS. Results were compared with national SLS rates. Six thousand one hundred sixty evaluations were included from 476 evaluators. Overall prevalence of SLS was 12.1% (95% CI 4.5–28.8). Residents (vs fellows) were less likely to have SLS evaluations (OR 0.5, 95% CI 0.4–0.8), though for all trainees increasing training year corresponded with increasing SLS frequency (OR 1.5, 95% CI 1.3–1.7). SLS was more common in procedural specialties compared with medical specialties (OR 2.1, 95% CI 1.1–3.8). Formative evaluations had lower SLS rates (OR 0.6, 95% CI 0.5–0.9) than summative evaluations, while milestone-based evaluations had higher rates than those that were not milestone-based (OR 1.5, 95% CI 1.03–2.2). Features of evaluators, such as subspecialty within oncology, and of trainees, such as seniority or trainee type, were related to SLS. Summative intent and milestone-based evaluations were more likely to be straight line scored. Specific evaluation scenarios at higher risk of SLS should be further examined.

Do Pediatric Emergency Medicine Fellows Meet the Milestone Targets for Graduation?

Roskind, C. G., Leonard, K., Baghdassarian, A., Kou, M., Levasseur, K., Rose, J., Shabanova, V., Vu, T., Zuckerbraun, N. S., & Langan, M. L. *Academic Pediatrics*. 2020;20(7):e35-e36.

ABSTRACT:

The ACGME Milestone Project is a competency-based assessment tool. Subcompetencies (SC) are scored on a 5-point scale, and level 4 is recommended for graduation. The 2018 Milestones Report found that across subspecialties, not all graduates are attaining a level 4 for every SC. To describe the number of pediatric emergency medicine (PEM) fellows who achieve level 4 in all 23 SC at graduation and to identify SC and predictive factors where a level 4 is not achieved. This is a multicenter, retrospective cohort study of PEM fellows. Program directors provided de-identified milestone reports from 2015- 2018. Descriptive analysis of milestone scores at graduation was performed. Demographics were compared between fellows who did and did not meet level 4 at graduation for each SC. Sub-analyses assessed differences in residency and first year milestone scores and the rate of milestone attainment between fellows who did and did not attain level 4 at graduation. Data from 48 PEM fellowship programs yielded graduation scores for 392 fellows (62% of total). 87% completed pediatric residency and 60% were female. Residency scores were available for 45 fellows. There were no SC in which all fellows attained at least level 4 at graduation; the range of fellows scoring < level 4 per SC was 7-39%. (Table 1) 67% of fellows did not attain level 4 on one or more of the 23 SC at graduation. While some fellows failed to attain a level 4 on up to all 23 SC, 26% failed to meet level 4 on only 1 or 2 SC. In 19/23 SC, residency and/or first year milestones scores were significantly lower for those who did not attain level 4 at graduation compared to those who did (mean difference 0.74 points). Those who did not attain level 4 at graduation had a significantly faster rate of improvement in milestone scores for 10/23 SC compared to those who did attain level 4. In our sample, 67% of PEM fellows did not attain level 4 for at least 1 of the 23 SC at graduation. Low scores during residency or early in fellowship may predict difficulty in meeting level 4 by fellowship completion.

Milestones Achievement from Residency to Fellowship: A Continuum?

Rose JA. 74. *Academic Pediatrics*. 2020;(7). doi:10.1016/j.acap.2020.06.095

ABSTRACT:

Background In 2013, the ACGME implemented the Milestones as a competency-based evaluation framework, spanning the continuum from novice to expert. Some subcompetencies from residency programs were adopted into subspecialty subcompetencies with the same 5-point scores. ACGME Milestones reports indicate lower achievement in identical subcompetencies for first year Pediatric Emergency Medicine (PEM) fellows compared with graduating pediatric residents. While differences between residency and fellowship programs exist, it is unclear why demonstrated skills would be 'lost' by residents transitioning to PEM fellowship as suggested by lower milestones scores.

AIM:

To investigate the trajectory of milestone scores from residency to PEM fellowship. Methods We completed a multicenter retrospective cohort study of a national sample of PEM fellows. Participating programs submitted de-identified Milestones data for PEM fellows (2015-2018). Of 23 PEM subcompetencies, 10 were adopted from pediatric and 7 from EM residency subcompetencies. We compared first year PEM fellow performance for these 17 subcompetencies to end-of-residency performance, using Wilcoxon signed rank tests to evaluate the difference in fellows paired scores. A 1- point decline in milestone score was deemed a priori to be clinically significant.

RESULTS:

We collected data for 639 PEM fellows from 48 fellowships. End-of-residency scores were available for 218 fellows from 42 programs. Most (210/218, 96%) completed pediatric residencies; 8 (4%) completed EM training. Declines in median milestones scores between end-of-residency and first year of PEM fellowship were observed for all fellows. These declines were statistically significant for pediatric- trained fellows; clinically significant declines were seen in two subcompetencies (Table 1).

CONCLUSIONS:

Our study found significant declines across adopted pediatric subcompetencies for pediatric residency graduates assessed early in their PEM fellowships. It is unclear whether this observed decline in achievement represents a true loss of skills, or a reset of faculty expectations and variance of Milestones interpretation by fellowships. Future studies are warranted to examine whether Milestones accurately assess trainee development across the continuum, or if they are applied differently as physicians transition from residency to subspecialty training.

Developing a Novel Scoring System to Objectively Track Orthopaedic Resident Educational Performance and Progression

Krueger CA, Rivera JC, Bhullar PS, Osborn PM. J Surg Educ. 2020 Mar - Apr;77(2):454-460. doi: 10.1016/j.jsurg.2019.09.009. Epub 2019 Dec 27.

OBJECTIVE:

Objectively determining orthopedic resident competence remains difficult and lacks standardization across residency programs. We sought to develop a scoring system to measure resident educational activity to stratify participation and performance in particular aspects of training and the effect of these measures on board certification.

DESIGN:

A weighted scoring system (Average Resident Score, ARS) was developed using the number of logged cases, clinic notes dictated, OITE PGY percentile, case minimums met, and scholarly activity completed each academic year (AY), with clinical activity being more heavily weighted. The Resident Effectiveness Score (RES), a z-score showing the number of standard deviations from the mean, was determined using the ARS. The RES effect on the Accreditation Council for Graduate Medical Education (ACGME) Milestones and American Board of Orthopedic Surgery (ABOS) Part 1 percentile score was determined using a Spearman correlation.

SETTING:

Large academic orthopedic residency.

PARTICIPANTS:

Thirty one orthopedic residents graduating between 2011 and 2016 were included.

RESULTS:

The RES did not differ between classes in the same AY, nor change significantly for individual residents during their training. Milestone z-scores increased as residents progressed in their education. The RES correlated with each Milestone competency subscore. The PGY5 OITE score and achieving ACGME minimums correlated with passing ABOS Part 1 (28/31 1st time pass), but the RES did not predict passing the board examination.

CONCLUSIONS:

This study demonstrates a scoring system encompassing multiple facets of resident education to track resident activity and progress. The RES can be tailored to an individual program's goals and aims and help program directors identify residents not maximizing educational opportunities compared to their peers. Monitoring this score may allow tailoring of educational efforts to individual resident needs. This RES may also allow residents to measure their performance and educational accomplishments and adjust their focus to obtain competence and board certification.

Mid-Year Medical Knowledge Milestones and ABSITE Scores in First-Year Surgery Residents

Cassaro S, Jarman BT, Joshi ART, Goldman-Mellor S, Hope WW, Johna S, Kaufman T, Grannan KJ; Association of Program Directors in Surgery Independent Surgery Programs Collaborative. *J Surg Educ.* 2020 Mar - Apr; 77(2):273-280. doi: 10.1016/j.jsurg.2019.09.012. Epub 2019 Sep 28.

OBJECTIVE:

Accreditation Council for Graduate Medical Education (ACGME) Surgery milestone ratings in the "Knowledge of Diseases and Conditions" (MK1) sub competency have been shown to correlate with American Board of Surgery In Training Examination (ABSITE) scores, and hypothesized to predict them. To better assess the predictive value of the MK1 milestone and avoid the potential bias caused by previous years' ABSITE scores, we designed a study including only first-year (PGY-1) residents and analyzed the correlation between their mid-year MK1 ratings and their scores in the ABSITE they took approximately a month later.

METHODS:

De-identified United States Medical Licensing Examination (USMLE) Step 1 and Step 2 scores, mid-year MK1 milestone ratings and the subsequent ABSITE standard scores for the five academic years from 2014-2015 to 2018-2019 were collected and tabulated for 247 PGY-1 preliminary- and categorical-track residents from ten ACGME-accredited surgery residency programs.

RESULTS:

The mid-year rating of PGY-1 residents' MK1 was predictive of their subsequent first ABSITE score for the entire cohort and for the categorical residents' subset. Notably, controlling for all other independent predictors, each half-point increase in MK1 rating was associated with a 25-point increase in ABSITE score. Preliminary residents performed significantly worse on the ABSITE, and their scores did not correlate significantly with their MK1 ratings.

CONCLUSIONS:

The mid-year rating of PGY-1 residents' MK1 was predictive of their subsequent first ABSITE score for the entire cohort and for the categorical but not the preliminary residents. This finding suggests that evaluators correctly rated MK1 higher in the categorical residents who did perform better on the subsequent ABSITE.

Cytopathology Milestones: Can You Get to Level 5?

Dyhdalo KS, Oshilaja O, Chute DJ, Booth CN, Suchy P, Smith K, Procop GW, Reynolds JP. J Am Soc Cytopathol. 2020 Mar 30. pii: S2213-2945(20)30046-6. doi: 10.1016/j.jasc.2020.03.002. [Epub ahead of print]

INTRODUCTION:

ACGME Milestones describe 6 areas of proficiency, indicating readiness for practice. Each is divided into 5 levels of mastery; Level 1 (new trainees) through Levels 4 (graduation) and 5 (aspirational). Milestones reporting began Spring 2016. We used Milestones to assess graduated fellows.

MATERIALS AND METHODS:

We conducted phone interviews with previous fellows and collected demographic information including practice setting. We asked graduates if they fulfilled each example of mastery and recorded their answers.

RESULTS:

A total of 22 fellows graduated from 2010 to 2017; 15 responded (10 academic, 5 private). Milestones in which nearly all respondents performed well (Level 4+) were: PC1, MK1, SBP2, SBP4, PROF1-4, ICS1-3. Some were more challenging (PC2, MK2, SBP1/3/5, PBL1). For PC2, 2 respondents achieved Level 1 (did not perform fine-needle aspirations). For MK2, 2 respondents achieved Level 1 (did not evaluate Papanicolaou). For SBP1, 80% in private practice achieved Level 5; 50% in academics achieved Level 3. For SBP3, 80% in private practice achieved Level 4+; 100% in academics achieved maximum Level 2. For SBP5, 60% of all respondents achieved maximum Level 3; only 1 achieved Level 5.

CONCLUSIONS:

Many Milestones are attainable. Eleven of 18 yielded Level 4+ from most respondents. Three (PC2, MK1, MK2) yielded rare Level 1 due to scope of practice. Others (SBP1, SBP3) reflect more of an all-or-nothing phenomenon. For SBP5, most respondents achieved Level 3; only 1 achieved Level 5. Some Milestones are highly dependent on practice setting, and others remain aspirational.

Association of intraoperative entrustment with clinical competency amongst general surgery residents

Ji S1, Hwang C, Karmakar M, Matusko N, Thompson-Burdine J, Williams AM, Leininger L, Minter RM3, Sandhu G4. *Am J Surg.* 2020 Feb;219(2):245-252. doi: 10.1016/j.amjsurg.2019.12.012. Epub 2019 Dec 13.

BACKGROUND:

Lack of transparency and meaningful assessment in surgical residency has led to inconsistent intraoperative entrustment and highly variable trainee competence at graduation. The relationship between faculty entrustment and resident entrustability on clinical competency remains unclear. We sought to evaluate the dynamic between entrustment/entrustability and clinical competency in general surgery residency.

METHODS:

Intraoperative observations were conducted across a 22-month period at an academic tertiary center. Entrustment/entrustability were measured using OpTrust. Clinical competencies were appraised via ACGME Milestones and Objective Structured Assessment of Technical Skill (OSATS) scores. Mixed effects linear regression was used to investigate the relationship among overall ACGME Milestone scores, OSATS domain scores, and overall OpTrust scores.

RESULTS:

Overall OpTrust scores significantly correlated with overall Milestone scores and multiple OSATS score domains.

CONCLUSIONS:

OpTrust demonstrated a positive association between ACGME general surgery Milestones and OSATS scores. Overall, OpTrust may help optimize intraoperative faculty entrustment and resident entrustability, facilitating surgical trainee success during residency.

Comparison of Male and Female Resident Milestone Assessments During Emergency Medicine Residency Training: A National Study

Santen SA, Yamazaki K, Holmboe ES, Yarris LM, Hamstra SJ. Acad Med. 2020 Feb;95(2):263-268. doi: 10.1097/ACM.0000000000002988.

PURPOSE:

A previous study found that milestone ratings at the end of training were higher for male than female residents in emergency medicine (EM). However, that study was restricted to a sample of 8 EM residency programs, and used individual faculty ratings from milestone reporting forms that were designed for use by the program's Clinical Competency Committee (CCC). The objective of this study was to investigate whether similar results would be found when examining the entire national cohort of EM milestone ratings reported by programs after CCC consensus review.

METHOD:

This study examined longitudinal milestone ratings for all EM residents (n = 1,363; 125 programs) reported to the Accreditation Council for Graduate Medical Education every 6 months from 2014- 2017. A multilevel linear regression model was used to estimate differences in slope for all subcompetencies, and predicted marginal means between genders were compared at time of graduation.

RESULTS:

There were small but statistically significant differences between males' and females' increase in ratings from initial rating to graduation on 6 of the 22 subcompetencies. Marginal mean comparisons at time of graduation demonstrated gender effects for 4 patient care subcompetencies. For these subcompetencies, males were rated as performing better than females; differences ranged from 0.048 to 0.074 milestone ratings.

CONCLUSIONS:

In this national dataset of EM resident milestone assessments by CCCs, males and females were rated similarly at the end of their training for the majority of subcompetencies. Statistically significant but small absolute differences were noted in 4 patient care subcompetencies.

Using Longitudinal Milestones Data and Learning Analytics to Facilitate the Professional Development of Residents: Early Lessons from Three Specialties

Holmboe ES, Yamazaki K, Nasca TJ, Hamstra SJ. Acad Med. 2020 Jan;95(1):97-103. doi: 10.1097/ACM.0000000000002899.

PURPOSE:

To investigate the effectiveness of using national, longitudinal milestones data to provide formative assessments to identify residents at risk of not achieving recommended competency milestone goals by residency completion. The investigators hypothesized that specific, lower milestone ratings at earlier time points in residency would be predictive of not achieving recommended Level (L) 4 milestones by graduation.

METHOD:

In 2018, the investigators conducted a longitudinal cohort study of emergency medicine (EM), family medicine (FM), and internal medicine (IM) residents who completed their residency programs from 2015 to 2018. They calculated predictive values (PVs) and odds ratios (ORs), adjusting for nesting within programs, for specific milestone rating thresholds at 6-month intervals for all subcompetencies within each specialty. They used final milestone ratings (May/June 2018) as the outcome variables, setting L4 as the ideal educational outcome.

RESULTS:

The investigators included 1,386 (98.9%) EM residents, 3,276 (98.0%) FM residents, and 7,399 (98.0%) IM residents in their analysis. The percentage of residents not reaching Level 4 by graduation ranged from 11-31% in EM, 16-53% in FM, and 5-15% in IM. Using a milestone rating of Level 2.5 or lower at the end of PGY2, the predictive probability of not attaining the L4 milestone graduation goal ranged from 32-56% in EM, 32-67% in FM, and 15-36% in IM.

CONCLUSIONS:

Longitudinal milestones ratings may provide educationally useful, predictive information to help individual residents address potential competency gaps, but the predictive power of the milestones ratings varies by specialty and subcompetency within these three adult care specialties.

Completion of an Individualized Learning Plan for Otology-Related Milestone Subcompetencies Leads to Improved Otology Section Otology Training Exam Scores

Penneck M, Svrakic M, Bent JP 3rd. Otol Neurotol. 2019 Dec;40(10):1392-1398. doi: 10.1097/MAO.0000000000002392.

OBJECTIVE:

To examine the relationships among self-assessment of knowledge in otology via an individualized learning plan (ILP), otology milestone achievement rate, and otology training exam (OTE) otology scores.

STUDY DESIGN:

Prospective study.

SETTING:

One otology residency covering a tertiary care facility, trauma and hospital center, outpatient ambulatory surgery center, and outpatient clinics.

PARTICIPANTS:

Twenty otology residents, four from each class.

METHODS:

Residents identified four milestones from otology-related sub-competencies to achieve in a 3-month rotation via an ILP. During the same rotation, the residents sat for the OTE, and their overall and otology scores were analyzed.

MAIN OUTCOME MEASURES:

Completion of an ILP before and at the end of the rotation, self-reported achievement of otology milestones, and OTE score components including total percent correct, scaled score, group stanine, national stanine, and residency group weighted scores.

RESULTS:

Group stanine OTE otology scores were higher for those residents who completed pre- and post- rotation ILPs compared with those who did not, 4.0 (± 0.348) versus 2.75 (± 0.453), respectively ($p=0.04$). Residents who self-reported achieving all four otology milestones had significantly higher otology group stanine scores than the residents who achieved less, 4.1 (± 0.348) versus 2.9 ± 0.433 , respectively ($p=0.045$). Residents who performed well in their PGY program cohort on the otology OTE 1 year were less inclined to complete an ILP for otology in the subsequent year (Pearson correlation -0.528, $p=0.035$).

CONCLUSION:

In the otology subspecialty, residents who completed ILPs scored better on OTE examinations independent of resident class. Consequently, programs may find ILPs useful in other otology subspecialties and across residencies.

Exploring Gender Bias in Nursing Evaluations of Emergency Medicine Residents

Brucker K, Whitaker N, Morgan ZS, Pettit K, Thinnis E, Banta AM, Palmer MM. Acad Emerg Med. 2019 Nov;26(11):1266-1272. doi: 10.1111/acem.13843. Epub 2019 Sep 23.

OBJECTIVES:

Nursing evaluations are an important component of residents' professional development as nurses are present for interactions with patients and nonphysician providers. Despite this, there has been few prior studies on the benefits, harms, or effectiveness of using nursing evaluations to help guide emergency medicine residents' development. We hypothesized that gender bias exists in nursing evaluations and that female residents, compared to their male counterparts, would receive more negative feedback on the perception of their interpersonal communication skills.

METHODS:

Data were drawn from nursing evaluations of residents between March 2013 and April 2016. All comments were coded if they contained words falling into four main categories: standout, ability, grindstone, and interpersonal. This methodology and the list of words that guided coding were based on the work of prior scholars. Names and gendered pronouns were obscured and each comment was manually reviewed and coded for valence (positive, neutral, negative) and strength (certain or tentative) by at least two members of the research team. Following the qualitative coding, quantitative analysis was performed to test for differences. To evaluate whether any measurable differences in ability between male and female residents existed, we compiled and compared American Board of Emergency Medicine in-training examination scores and relevant milestone evaluations between female and male residents from the same period in which the residents were evaluated by nursing staff.

RESULTS:

Of 1,112 nursing evaluations, 30% contained comments. Chi-square tests on the distribution of valence (positive, neutral, or negative) indicated statistically significant differences in ability and milestone categories based on the gender of the resident. A total of 51% of ability comments about female residents were negative compared to 20% of those about male residents ($\chi^2 = 11.83$, $p < 0.01$). A total of 57% of milestone comments about female residents were negative as opposed 24% of those about male residents ($\chi^2 = 6.03$, $p < 0.01$).

CONCLUSIONS:

Our findings demonstrate that, despite the lack of difference in ability or competence as measured by in-service examination scores and milestone evaluations, nurses evaluate female residents lower in their abilities and work ethic compared to male residents.

Evaluation of Gender Differences in Ultrasound Milestone Evaluations During Emergency Medicine Residency Training: A Multicenter Study

Acuña J, Stolz U, Stolz LA, Situ-LaCasse EH, Bell G, Berkeley RP, Boyd JS, Castle D, Carmody K, Fong T, Grewal E, Jones R, Hilberts S, Kanter C, Kelley K, Leetch SJ, Pazderka P, Shaver E, Stowell JR, Josephson EB, Theodoro D, Adhikari S. *AEM Educ Train*. 2019 Oct 24;4(2):94-102. doi: 10.1002/aet2.10397. eCollection 2020 Apr.

OBJECTIVES:

Prior literature has demonstrated incongruities among faculty evaluation of male and female residents' procedural competency during residency training. There are no known studies investigating gender differences in the assessment of procedural skills among emergency medicine (EM) residents, such as those required by ultrasound. The objective of this study was to determine if there are significant gender differences in ultrasound milestone evaluations during EM residency training.

METHODS:

We used a stratified, random cluster sample of Accreditation Council for Graduate Medical Education (ACGME) EM residency programs to conduct a longitudinal, retrospective cohort analysis of resident ultrasound milestone evaluation data. Milestone evaluation data were collected from a total of 16 ACGME-accredited EM residency programs representing a 4-year period. We stratified milestone data by resident gender, date of evaluation, resident postgraduate year, and cohort (residents with the same starting date).

RESULTS:

A total of 2,554 ultrasound milestone evaluations were collected from 1,187 EM residents (750 men [62.8%] and 444 women [37.1%]) by 104 faculty members during the study period. There was no significant overall difference in mean milestone score between female and male residents [mean difference = 0.01 (95% confidence interval {CI} = -0.04 to 0.05)]. There were no significant differences between female and male residents' mean milestone scores at the first (baseline) PGY1 evaluation (mean difference = -0.04 [95% CI = -0.09 to 0.003]) or at the final evaluation during PGY3 (mean difference = 0.02 [95% CI = -0.03 to 0.06]).

CONCLUSIONS:

Despite prior studies suggesting gender bias in the evaluation of procedural competency during residency training, our study indicates that there were no significant gender-related differences in the ultrasound milestone evaluations among EM residents within training programs throughout the United States.

Family Medicine Residency Graduates' Preparation for Quality Improvement Leadership

Lichkus J, Fang B, Peterson LE. J Grad Med Educ. 2019 Oct;11(5):558-564. doi: 10.4300/JGME-D-18-01060.1.

BACKGROUND:

Training in quality improvement (QI) is a standard component of family medicine residency education. Graduating family medicine residents' ability to lead QI initiatives is unknown.

OBJECTIVE:

We assessed the preparedness of graduating family medicine residents to lead QI projects and to identify factors that may increase such readiness.

METHODS:

Milestone data for all graduating family medicine residents were linked to a practice demographic questionnaire completed by the same residents who registered for the American Board of Family Medicine certification examination between 2014 and 2017. The change in self-assessed QI preparedness over time and its association with faculty-assigned milestone ratings were examined using descriptive and regression analyses.

RESULTS:

The questionnaire had a 100% response rate (12 208 responded). Between 2014 and 2017, the percentage of residents who self-reported being "extremely" or "moderately" prepared to lead QI projects increased from 72.7% (2208 of 3038) to 75.8% (2434 of 3210, $P = .009$). Self-reported QI team leadership was associated with 93% higher odds of feeling extremely prepared compared to moderately prepared (odds ratio 1.93, 95% CI 1.58-2.35). The average midyear faculty-assigned milestone rating for QI among residents who felt "extremely" prepared was 3.28 compared to 3.14 among those who felt "not at all" prepared.

CONCLUSIONS:

Over the past 4 years, family medicine residents' self-assessed preparedness to lead QI projects has barely increased. There was no correlation between self-assessed preparation and faculty-assigned milestone rating. However, we found a small association between self-reported QI leadership and self-assessed QI preparedness.

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

Hamstra SJ, Yamazaki K, Barton MA, Santen SA, Beeson MS, Holmboe ES. Acad Med. 2019 Oct;94(10):1522-1531. doi: 10.1097/ACM.0000000000002820.

PURPOSE:

To investigate whether clinical competency committees (CCCs) were consistent in applying milestone ratings for first-year residents over time or whether ratings increased or decreased.

METHOD:

Beginning in December 2013, the ACGME initiated a phased-in requirement for reporting milestones; emergency medicine (EM), diagnostic radiology (DR), and urology (UR) were among the earliest reporting specialties. The authors analyzed CCC milestone ratings of first-year residents from 2013-2016 from all ACGME-accredited EM, DR, and UR programs for which they had data. The number of first-year residents in these programs ranged from 2,838 to 2,928 over this time period. The program-level average milestone rating for each subcompetency was regressed onto the time of observation using a random coefficient multilevel regression model.

RESULTS:

National average program-level milestone ratings of first-year residents decreased significantly over the observed time period for 32 of the 56 subcompetencies examined. None of the other subcompetencies showed a significant change. National average in-training examination scores for each of the specialties remained essentially unchanged over the time period, suggesting differences between the cohorts was not likely an explanatory factor.

CONCLUSIONS:

The findings indicate that CCCs tend to become more stringent or maintain consistency in their ratings of beginning residents over time. One explanation for these results is that CCCs may become increasingly comfortable in assigning lower ratings when appropriate. This finding is consistent with an increase in confidence with the milestones rating process and the quality of feedback it provides. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

The Value of Numerical USMLE Step 1 Scores in Radiology Resident Selection

Maxfield CM, Grimm LJ. Acad Radiol. 2019 Aug 21. pii: S1076-6332(19)30386-1. doi: 10.1016/j.acra.2019.08.007.

RATIONALE AND OBJECTIVES:

In response to a recent proposal to change scoring on the United States Medical Licensing (USMLE) Step 1 exam to pass/fail, we sought to determine the value of numerical Step 1 scores in predicting success in our radiology residency program.

MATERIALS AND METHODS:

Residency applications for 157 residents entering the program between 2005 and 2017 were retrospectively reviewed. Biographical (gender, sports participation, advanced degree), undergraduate (school, major), and medical school (grades, USMLE Step 1 score, Alpha Omega Alpha membership, letters of recommendation, publications) data were recorded. Multivariate regression analysis was used to examine the relationship between these application factors and subsequent performance as a radiology resident, as determined by completion of the program without requiring corrective action, select Accreditation Council for Graduate Medical Education milestones, and selection as chief resident.

RESULTS:

Corrective action was required for 7% (n = 12) of residents. Of the predictor variables, only Step 1 score was associated with the need for corrective action ($p < 0.001$). The interpretation of exams milestone was associated with higher Step 1 scores ($p = 0.001$) and number of medical school clerkship honors ($p = 0.008$). Selection as chief resident was associated with sports participation ($p = 0.04$), and clerkship honors ($p = 0.02$).

CONCLUSION:

Numerical USMLE Step 1 scores are predictive of successful completion of radiology residency training without the need for corrective action, and of accelerated competence in the interpretation of exams milestone. Continued reporting of numerical Step 1 scores would be valuable in selection of radiology residents.

Are Otolaryngology Milestones Predictive of Otolaryngology Training Examination Scores?

Lozada KN, Ferrandino RM, Teng MS, Colley PM, Bernstein JM. Ear Nose Throat J. 2019 Mar;98 (3):139-142. doi: 10.1177/0145561319832253. Epub 2019 Feb 21.

IMPORTANCE:

Otolaryngology residents take the otolaryngology training examination (OTE) yearly to assess their fund of knowledge. The Accreditation Council for Graduate Medical Education (ACGME) milestone evaluations are also conducted semiannually. Accurate prediction of training examination performance allows identification of residents who are performing well and those who need targeted remediation.

Prior studies in other specialties have attempted to use milestone evaluations to help predict in-training examination scores.

OBJECTIVE:

In this study, we aim to identify whether ACGME milestone evaluation scores predict OTE performance.

DESIGN:

Milestone ratings and OTE scores for residents at 2 US otolaryngology residency programs were collected. Multivariate analysis was achieved using linear mixed modeling. We considered a 2-tailed P value of $\leq .05$ as statistically significant.

SETTING:

Two US otolaryngology residency programs.

PARTICIPANTS:

Forty-eight otolaryngology residents postgraduate years 2 to 5.

RESULTS:

Otolaryngology training examination scores and ACGME milestone evaluations were collected from 48 residents from postgraduate year 2 to 5 between the years 2014 and 2017. One hundred eight OTE scores were available. Linear mixed-effect models were constructed, and after adjusting for level of training and OTE year, the total milestone rating made a negligible impact in estimating OTE percentage correct ($\beta = -.01$, $P = .9$). Similarly, total milestone rating demonstrated a minimal contribution in approximating OTE national stanine score after adjusting for the level of training ($\beta = -.003$, $P = .9$).

CONCLUSIONS AND RELEVANCE:

In our study, ACGME milestone evaluations are not predictive of residents' OTE performance. What these milestone evaluation data mean and how they should be used continues to be an unanswered question. We should aim to identify the most effective applications of the milestone data collected yearly by otolaryngology programs.

Identification of Gender Differences in Ultrasound Milestone Assessments during Emergency Medicine Residency Training: A Pilot Study

Acuña J, Situ-LaCasse EH, Patanwala AE, Stolz LA, Amini R, Friedman L, Adhikari S. *Adv Med Educ Pract*. 2019 Mar 27;10:141-145. doi: 10.2147/AMEP.S196140. eCollection 2019.

OBJECTIVES:

Prior literature suggests that incongruities between male and female resident's procedural competency may be explained by gender bias during the evaluation process. There are no known studies investigating gender differences in the assessment of ultrasound-based procedural skills among emergency medicine (EM) residents. The purpose of this study was to evaluate for gender differences in ultrasound milestone assessments among EM residents.

METHODS:

This is a retrospective study including EM residents. Milestone assessment data were collected from a total of 3 Accreditation Council for Graduate Medical Education (ACGME) EM residency programs representing a 3-year period. The outcome measures included mean milestone levels, milestone levels at baseline and graduation and differences in milestone achievement between female and male EM residents. An unpaired Student's t-test was used to compare milestone scores between female and male residents.

RESULTS:

A total of 456 ultrasound milestone evaluations were collected from 91 EM residents (34 females [37%] and 57 males [63%]). No significant differences were noted in the overall mean milestone level between females (2.3 ± 0.6) and males (2.2 ± 0.6) ($P=0.387$). There were no significant differences noted in the ultrasound milestone level between females (0.8 ± 0.6) and males (0.7 ± 0.7) at baseline ($P=0.754$). Although it did not reach statistical significance ($P=0.197$), the increase in the mean ultrasound milestone level from baseline to graduation was greater in males (3.4 ± 0.7) compared to females (3.1 ± 0.7).

CONCLUSION:

Overall, there were no statistically significant differences in the mean ultrasound milestone levels between females and males. The rate of ultrasound milestone level achievement during EM residency training at our institution had a slight tendency to be higher for males than females in the observed residency programs; however, this also did not reach statistical significance. Possible gender bias while evaluating ultrasound milestone levels needs to be further studied on a larger scale.

Predicting Performance of First-Year Residents: Correlations between Structured Interview, Licensure Exam, and Competency Scores in a Multi-Institutional Study

Marcus-Blank B, Dahlke JA, Braman JP, Borman-Shoap E, Tiryaki E, Chipman J, Andrews JS, Sackett PR, Cullen MJ. *Acad Med.* 2019 Mar;94(3):378-387. doi: 10.1097/ACM.0000000000002429.

PURPOSE:

To determine whether scores on structured interview (SI) questions designed to measure non-cognitive competencies in physicians (1) predict subsequent first-year resident performance on Accreditation Council for Graduate Medical Education (ACGME) milestones and (2) add incremental validity over United States Medical Licensing Examination (USMLE) Step 1 and Step 2 Clinical Knowledge (CK) scores in predicting performance.

METHOD:

The authors developed 18 behavioral description questions to measure key non-cognitive competencies (e.g., teamwork). In 2013-2015, 14 programs (13 residency, 1 fellowship) from 6 institutions used subsets of these questions in their selection processes. The authors conducted analyses to determine the validity of SIs and USMLE scores in predicting first-year resident milestone performance in the ACGME's core competency domains and overall.

RESULTS:

SI scores predicted mid-year and year-end overall performance ($r = .18$ and $.19$, respectively, $P < .05$) and year-end performance on patient care, interpersonal and communication skills, and professionalism competencies ($r = .23$, $r = .22$, and $r = .20$, respectively, $p < .05$). SI scores contributed incremental validity over USMLE scores in predicting year-end performance on patient care ($\Delta R = .05$), interpersonal and communication skills ($\Delta R = .09$), and professionalism ($\Delta R = .09$; all $P < .05$). USMLE scores contributed incremental validity over SI scores in predicting year-end performance overall and on patient care and medical knowledge.

CONCLUSIONS:

SI scores predict first-year resident year-end performance in the interpersonal and communication skills, patient care, and professionalism competency domains. Future research should investigate whether SIs predict a range of clinically relevant outcomes.

ACGME Milestones within Subspecialty Training Programs: One Institution's Experience

Heath JK, Dine CJ. J Grad Med Educ. 2019 Feb;11(1):53-59. doi: 10.4300/JGME-D-18-00308.1.

BACKGROUND:

The Accreditation Council for Graduate Medical Education Milestones were created as a criterion- based framework to promote competency-based education during graduate medical education. Despite widespread implementation across subspecialty programs, extensive validity evidence supporting the use of milestones within fellowship training is lacking.

OBJECTIVE:

We assessed the construct and response process validity of milestones in subspecialty fellowship programs in an academic medical center.

METHODS:

From 2014-2016, we performed a single center retrospective cohort analysis of milestone data from fellows across 5 programs. We analyzed summary statistics and performed multivariable linear regression to assess change in milestone ratings by training year and variability in ratings across fellowship programs. Finally, we examined a subset of Professionalism and Interpersonal and Communication Skills subcompetencies from the first 6 months of training to identify the proportion of fellows deemed "ready for independent practice" in these domains.

RESULTS:

Milestone data were available for 68 fellows, with 75933 unique subcompetency ratings. Multivariable linear regression, adjusted for subcompetency and subspecialty, revealed an increase of 0.17 (0.16-0.19) in ratings with each postgraduate year level increase ($P < .005$), as well as significant variation in milestone ratings across subspecialties. For the Professionalism and Interpersonal and Communication Skills domains, mean ratings within the first 6 months of training were 3.78 and 3.95, respectively.

CONCLUSIONS:

We noted a minimal upward trend of milestone ratings in subspecialty training programs, and significant variability in implementing milestones across differing subspecialties. This may suggest possible difficulties with the construct validity and response process of the milestone system in certain medical subspecialties.

Surgical Simulation: Markers of Proficiency

Binkley J, Bukoski AD, Doty J, Crane M, Barnes SL, Quick JA. J Surg Educ. 2019 Jan - Feb;76 (1): 234-241. doi: 10.1016/j.jsurg.2018.05.018. Epub 2018 Jul 6.

OBJECTIVE:

Surgical simulation has become an integral component of surgical training. Simulation proficiency determination has been traditionally based upon time to completion of various simulated tasks. We aimed to determine objective markers of proficiency in surgical simulation by comparing novel assessments with conventional evaluations of technical skill.

DESIGN:

Categorical general surgery residents completed 10 laparoscopic cholecystectomy modules using a high-fidelity simulator. We recorded and analyzed simulation task times, as well as number of hand movements, instrument path length, instrument acceleration, and participant affective engagement during each simulation. Comparisons were made to Objective Structured Assessment of Technical Skill (OSATS) and Accreditation Council for Graduate Medical Education Milestones, as well as previous laparoscopic experience, duration of laparoscopic cholecystectomies performed by participants, and postgraduate year. Comparisons were also made to Fundamentals of Laparoscopic Surgery task times. Spearman's rho was utilized for comparisons, significance set at >0.50 .

SETTING:

University of Missouri, Columbia, Missouri, an academic tertiary care facility.

PARTICIPANTS:

Fourteen categorical general surgery residents (postgraduate year 1-5) were prospectively enrolled.

RESULTS:

One hundred forty simulations were included. The number of hand movements and instrument path lengths strongly correlated with simulation task times (ρ 0.62-0.87, $p < 0.0001$), FLS task completion times (ρ 0.50-0.53, $p < 0.0001$), and prior real-world laparoscopic cholecystectomy experience (ρ -0.51 to -0.53, $p < 0.0001$). No significant correlations were identified between any of the studied markers with Accreditation Council for Graduate Medical Education Milestones, Objective Structured Assessment of Technical Skill evaluations, total previous laparoscopic experience, or postgraduate year level. Neither instrument acceleration nor participant engagement showed significant correlation with any of the conventional markers of real-world or simulation skill proficiency.

CONCLUSIONS:

Simulation proficiency, measured by instrument and hand motion, is more representative of simulation skill than simulation task time, instrument acceleration, or participant engagement.

Application Factors Associated with Clinical Performance during Pediatric Internship: A 5-Year Single Center Retrospective Cohort Study

Gross C J, O'Halloran C, Deshpande S, Lux S, Sectish T, Michelson C, Winn A, Sox C. 64. *Academic Pediatrics*. 2019;(6). doi:10.1016/j.acap.2019.05.078.

BACKGROUND:

The specific components of an application to residency that predict clinical performance during training in pediatrics remain unknown.

METHODS:

Retrospective cohort study of all pediatric interns who matched into the Boston Combined Residency Program from 2013-2017. Demographics, subspecialty track, medical school ranking, USMLE scores, advanced degrees, clerkship grades, Alpha-Omega-Alpha (AOA) and Gold Humanism Honor Society membership, interview day performance, letters of recommendation (LOR) strength, and number of publications were extracted from application materials. The primary outcome was clinical performance at the end of internship, measured as a weighted average of existing ACGME pediatric milestones scores. Linear mixed effects modeling with random effects for grading committee and match year was used to identify factors independently associated with clinical performance. Variables with p-values <0.2 in bivariate analysis were included in the final model.

RESULTS:

223 interns were included in the study. In the model (Table 1), higher average LOR score ($B=.07$, $p=.01$), having a master's degree ($B=.19$, $p=.03$), and not having a PhD ($B=.13$, $p=.03$) were associated with more advanced clinical performance at the end of pediatric internship. AOA membership, medical school ranking, public medical school attendance, time off prior to medical school, number of clerkship honors, and interview score were included in the model, but not significant predictors of clinical performance. The fixed effects explained 15% of the variance in milestones score, while the random effects (match year and grading committee) explained 8% of the variance (marginal $R^2=.15$, conditional $R^2=.23$).

CONCLUSIONS:

Strong letters of recommendation, having a master's degree and not having a PhD are associated with more advanced clinical performance during pediatric internship. However, much of the variance in clinical performance remains unexplained by quantifiable application variables.

Timing of Milestone Competency Acquisition in Neurology Residency: What by When?

Jones LK Jr, Eggers SDZ, Capobianco DJ, Boes CJ. Neurology. 2018 Oct 16;91(16):748-754. doi: 10.1212/WNL.0000000000006361. Epub 2018 Sep 14.

OBJECTIVE:

To determine the stage of training at which neurology residents should achieve individual elements of the Accreditation Council for Graduate Medical Education neurology Milestones and to examine the relationship between perceived importance of Milestones and the stage by which they should be achieved.

METHODS:

A modified Delphi technique was used to establish consensus postgraduate year (PGY) expectations for neurology Milestone competencies across 3 geographically and administratively distinct Mayo Clinic neurology residency programs. Timing expectations were examined for relationships to perceived importance of the individual Milestones and effects of participant characteristics.

RESULTS:

PGY expectations for neurology Milestone elements ranged from PGY 1.3 to PGY 4.1. Extent of rater educational seniority had no effect on PGY competency expectations. There was a moderate inverse relationship between perceived importance of the Milestone element and the PGY by which it should be achieved ($r_s = -0.74$, $p < 0.0001$).

CONCLUSIONS AND RELEVANCE:

Expectations for neurology Milestone competency acquisition can be measured and may help inform individual program design, educational expectations, and future Milestone design.

Visuospatial Aptitude Testing Differentially Predicts Simulated Surgical Skill

Hinchcliff E, Green I, Destephano C, Cox M, Smink D, Kumar A, Hokenstad E, Bengtson J, Cohen SL. J Minim Invasive Gynecol. 2018 Sep - Oct;25(6):1044-1050. doi: 10.1016/j.jmig.2018.01.031. Epub 2018 Feb 5.

OBJECTIVE:

To determine whether visuospatial perception (VSP) testing is correlated to simulated or intraoperative surgical performance as rated by the American College of Graduate Medical Education (ACGME) milestones.

DESIGN: (Canadian Task Force classification II-2).

SETTING: Two academic training institutions.

PARTICIPANTS:

Forty-one residents, including 19 from Brigham and Women's Hospital and 22 from the Mayo Clinic, from 3 different specialties: obstetrics and gynecology, general surgery, and urology.

INTERVENTION:

Participants underwent 3 different tests: visuospatial perception testing (VSP), Fundamentals of Laparoscopic Surgery (FLS) peg transfer, and da Vinci robotic simulation peg transfer. Surgical grading from the ACGME milestones tool was obtained for each participant. Demographic and background information was also collected, including specialty, year of training, previous experience with simulated skills, and surgical interest. Standard statistical analyses were performed using Student's t test, and correlations were determined using adjusted linear regression models.

MEASUREMENTS AND MAIN RESULTS:

In univariate analysis, Brigham and Women's Hospital and Mayo Clinic training programs differed in times and overall scores for both the FLS peg transfer and da Vinci robotic simulation peg transfer tests ($p < .05$ for all). In addition, type of residency training affected time and overall score on the robotic peg transfer test. Familiarity with tasks correlated with higher score and faster task completion ($p = .05$ for all except VSP score). There were no differences in VSP scores by program, specialty, or year of training. In adjusted linear regression modeling, VSP testing was correlated only to robotic peg transfer skills (average time, $p = .006$; overall score, $p = .001$). Milestones did not correlate to either VSP or surgical simulation testing.

CONCLUSION:

VSP score was correlated with robotic simulation skills, but not with FLS skills or ACGME milestones. This suggests that the ability of VSP score to predict competence differs between tasks. Therefore, further investigation of aptitude testing is needed, especially before its integration as an entry examination into a surgical subspecialty.

Pediatric Program Director Minimum Milestone Expectations before Allowing Supervision of Others and Unsupervised Practice

Li ST, Tancredi DJ, Schwartz A, Guillot A, Burke AE, Trimm RF, Guralnick S, Mahan JD, Gifford K. *Acad Pediatr*. 2018 Sep - Oct;18(7):828-836. doi: 10.1016/j.acap.2018.04.010. Epub 2018 Apr 25.

BACKGROUND:

The Accreditation Council for Graduate Medical Education requires semiannual Milestone reporting on all residents. Milestone expectations of performance are unknown.

OBJECTIVE:

Determine pediatric program director (PD) minimum Milestone expectations for residents prior to being ready to supervise and prior to being ready to graduate.

METHODS:

Mixed methods survey of pediatric PDs on their programs' Milestone expectations before residents are ready to supervise and before they are ready to graduate, and in what ways PDs use Milestones to make supervision and graduation decisions. If programs had no established Milestone expectations, PDs indicated expectations they considered for use in their program. Mean minimum Milestone level expectations adjusted for program size, region, and clustering of Milestone expectations by program were calculated for prior to supervise and prior to graduate. Free-text questions were analyzed using thematic analysis.

RESULTS:

The response rate was 56.8% (113/199). Most programs had no required minimum Milestone level before residents are ready to supervise (80%; 76/95) or ready to graduate (84%; 80/95). For readiness to supervise, minimum Milestone expectations PDs considered establishing for their program were highest for humanism (2.46, 95% CI: 2.21-2.71) and professionalization (2.37, 2.15-2.60). Minimum Milestone expectations for graduates were highest for help-seeking (3.14, 2.83-3.46). Main themes included the use of Milestones in combination with other information to assess learner performance and Milestones are not equally weighted when making advancement decisions.

CONCLUSIONS:

Most PDs have not established program minimum Milestones, but would vary such expectations by competency.

Challenges in Measuring ACGME Competencies: Considerations for Milestones

Natesan P, Batley NJ, Bakhti R, El-Doueih PZ. *Int J Emerg Med*. 2018 Sep 28;11(1):39. doi: 10.1186/s12245-018-0198-3.

BACKGROUND:

Measuring milestones, competencies, and sub-competencies as residents progress through a training program is an essential strategy in Accreditation Council for Graduate Medical Education (ACGME)'s attempts to ensure graduates meet expected professional standards. Previous studies have found, however, that physicians make global ratings often by using a single criterion.

METHODS:

We use advanced statistical analysis to extend these studies by examining the validity of ACGME International competency measures for an international setting, across emergency medicine (EM) and neurology, and across evaluators. Confirmatory factor analysis (CFA) models were fitted to both EM and neurology data. A single-factor CFA was hypothesized to fit each dataset. This model was modified based on model fit indices. Differences in how different EM physicians perceived the core competencies were tested using a series of measurement invariance tests.

RESULTS:

Extremely high alpha reliability coefficients, factor coefficients ($> .93$), and item correlations indicated multicollinearity, that is, most items being evaluated could essentially replace the underlying construct itself. This was true for both EM and neurology data, as well as all six EM faculty.

CONCLUSIONS:

Evaluation forms measuring the six core ACGME competencies did not possess adequate validity. Severe multicollinearity exists for the six competencies in this study. ACGME is introducing milestones with 24 sub-competencies. Attempting to measure these as discrete elements, without recognizing the inherent weaknesses in the tools used will likely serve to exacerbate an already flawed strategy. Physicians likely use their "gut feelings" to judge a resident's overall performance. A better process could be conceived in which this subjectivity is acknowledged, contributing to more meaningful evaluation and feedback.

National Internal Medicine Milestone Ratings: Validity Evidence from Longitudinal Three-Year Follow-Up

Hauer KE, Vandergrift J, Lipner RS, Holmboe ES, Hood S, McDonald FS. Acad Med. 2018 Aug;93(8):1189-1204. doi: 10.1097/ACM.0000000000002234.

PURPOSE:

To evaluate validity evidence for internal medicine milestone ratings across programs for three resident cohorts by quantifying "not assessable" ratings; reporting mean longitudinal milestone ratings for individual residents; and correlating medical knowledge ratings across training years with certification examination scores to determine predictive validity of milestone ratings for certification outcomes.

METHOD:

This retrospective study examined milestone ratings for postgraduate year (PGY)-1-3 residents in United States internal medicine residency programs. Data sources included milestone ratings, program characteristics, and certification examination scores.

RESULTS:

Among 35,217 participants, there was a decreased percentage with "not assessable" ratings across years: 1,566 (22.5%) PGY-1s in 2013-2014 versus 1,219 (16.6%) in 2015-2016 ($P = .01$), and 342 (5.1%) of PGY-3s in 2013-2014 versus 177 (2.6%) in 2015-2016 ($P = .04$). For individual residents with three years of ratings, mean milestone ratings increased from around 3 (behaviors of an early learner or advancing resident) in PGY-1 (ranging from a mean of 2.73 to 3.19 across subcompetencies) to around 4 (ready for unsupervised practice) in PGY-3 (mean of 4.00 to 4.22 across subcompetencies, $P < .001$ for all subcompetencies). For each increase of 0.5 units in two medical knowledge (MK1, MK2) subcompetency ratings, the difference in examination scores for PGY-3s was 19.5 points for MK1 ($P < .001$) and 19.0 for MK2 ($P < .001$).

CONCLUSIONS:

These findings provide evidence of validity of the milestones by showing how training programs have applied them over time and how milestones predict other training outcomes.

A Survey on Recent Medical School Graduate Comfort with the Level 1 Milestones

Petravick ME, Marsh JL, Karam MD, Dirschl DR. J Surg Educ. 2018 Jul - Aug;75(4):911-917. doi: 10.1016/j.jsurg.2017.10.004. Epub 2017 Nov 7.

OBJECTIVE:

The Next Accreditation System implemented 5 levels of milestones for orthopedic surgery residents in 2013. The Level 1 milestones were noted as those "expected of an incoming resident." While the milestones were intended for assessing resident progression and readiness for independent practice, this designation can also be used to assess how well prepared graduating medical students are for beginning an orthopedic surgery residency. The primary objective of this paper is to measure recent medical school graduate comfort with the Level 1 milestones.

DESIGN, SETTING, AND PARTICIPANTS:

In June 2015, the program directors for the Midwest Orthopaedic Surgical Skills (MOSS) Consortium affiliated residency programs were sent an online survey for distribution to the recent medical school graduates who matched at their respective programs. The survey was about recent graduate comfort with the Level 1 milestone handles associated with 16 orthopedic milestones spanning multiple subspecialties. Responses were grouped based on comfort with individual milestone handles with orthopedic conditions (e.g., carpal tunnel) or with broader categories spanning orthopedic milestones (e.g., imaging).

RESULTS:

In all, 66 of 112 graduates (58.9%) responded. Of 60 milestone handles surveyed, respondents were "Comfortable" with an average of 31.6 ± 14.2 handles with some conditions performing much better than others. The median "Comfortable" response rate was 31 handles. The 8 broader categories had "Comfortable" response rates between 35% and 70%. All 8 orthopedic conditions had significantly higher "Comfortable" response rates for "Evaluation & Knowledge" handles than for "Decision Making & Treatment" handles.

CONCLUSIONS:

Most recent medical student graduates who matched into an orthopedic surgery residencies are only comfortable with about half of the Level 1 milestone handles even though they are expected to meet the Level 1 milestones upon beginning residency. This finding suggests the development of an assessment based on the Level 1 milestones would be appropriate to better inform both graduate and undergraduate medical education in orthopedic surgery.

Surgical Simulation: Markers of Proficiency

Binkley J, Bukoski AD, Doty J, Crane M, Barnes SL, Quick JA. J Surg Educ. 2018 Jul 5. pii: S1931- 7204(18)30055-2. doi: 10.1016/j.jsurg.2018.05.018.

OBJECTIVE:

Surgical simulation has become an integral component of surgical training. Simulation proficiency determination has been traditionally based upon time to completion of various simulated tasks. We aimed to determine objective markers of proficiency in surgical simulation by comparing novel assessments with conventional evaluations of technical skill.

DESIGN:

Categorical general surgery residents completed 10 laparoscopic cholecystectomy modules using a high-fidelity simulator. We recorded and analyzed simulation task times, as well as number of hand movements, instrument path length, instrument acceleration, and participant affective engagement during each simulation. Comparisons were made to Objective Structured Assessment of Technical Skill (OSATS) and Accreditation Council for Graduate Medical Education Milestones, as well as previous laparoscopic experience, duration of laparoscopic cholecystectomies performed by participants, and postgraduate year. Comparisons were also made to Fundamentals of Laparoscopic Surgery task times. Spearman's rho was utilized for comparisons, significance set at >0.50 .

SETTING:

University of Missouri, Columbia, Missouri, an academic tertiary care facility.

PARTICIPANTS:

Fourteen categorical general surgery residents (postgraduate year 1-5) were prospectively enrolled.

RESULTS:

One hundred forty simulations were included. The number of hand movements and instrument path lengths strongly correlated with simulation task times (ρ 0.62-0.87, $p < 0.0001$), FLS task completion times (ρ 0.50-0.53, $p < 0.0001$), and prior real-world laparoscopic cholecystectomy experience (ρ -0.51 to -0.53, $p < 0.0001$). No significant correlations were identified between any of the studied markers with Accreditation Council for Graduate Medical Education Milestones, Objective Structured Assessment of Technical Skill evaluations, total previous laparoscopic experience, or postgraduate year level. Neither instrument acceleration nor participant engagement showed significant correlation with any of the conventional markers of real-world or simulation skill proficiency.

CONCLUSIONS:

Simulation proficiency, measured by instrument and hand motion, is more representative of simulation skill than simulation task time, instrument acceleration, or participant engagement.

An Analysis of Residents' Self Evaluation and Faculty Evaluation in Internal Medicine Standardized Residency Training Program using Milestones Evaluation System

Zhang Y, Chu XT, Zeng XJ, Li H, Zhang FC, Zhang SY, Shen T. Zhonghua Nei Ke Za Zhi. 2018 Jun 1;57(6):440-445. doi: 10.3760/cma.j.issn.0578-1426.2018.06.009.

OBJECTIVE:

To assess the value of internal medicine residency training program at Peking Union Medical College Hospital (PUMCH), and the feasibility of applying revised Milestones evaluation system.

METHODS:

Postgraduate-year-one to four (PGY-1 to PGY-4) residents in PUMCH finished the revised Milestones evaluation scales in September 2017. Residents' self-evaluation and faculty- evaluation scores were calculated. Statistical analysis was conducted on the data.

RESULTS:

A total of 207 residents were enrolled in this cross-sectional study. Both self and faculty scores showed an increasing trend in senior residents. PGY-1 residents were assessed during their first month of residency with scores of 4 points or higher, suggesting that residents have a high starting level. More strikingly, the mean score in PGY-4 was 7 points or higher, proving the career development of residency training program. There was no statistically significant difference between total self- and faculty-evaluation scores. Evaluation scores of learning ability and communication ability were lower in faculty group ($t=-2.627, -4.279$, all $P<0.05$). The scores in graduate students were lower than those in standardized training residents.

CONCLUSIONS:

The goal of national standardized residency training is to improve the quality of healthcare and residents' career development. The evaluation results would guide curriculum design and emphasize the importance and necessity of multi-level teaching. Self-evaluation contributes to the understanding of training objectives and personal cognition.

Case Reporting, Competence, and Confidence: A Discrepancy in the Numbers

Shah D, Haisch CE, Noland SL. *J Surg Educ.* 2018 Mar - Apr;75(2):304-312. doi: 10.1016/j.jsurg.2018.01.007. Epub 2018 Feb 1.

PURPOSE:

The Accreditation Council for Graduate Medical Education (ACGME) continues to play an integral role in accreditation of surgical programs. The institution of case logs to demonstrate competency of graduating residents is a key component of evaluation. This study compared the number of vascular cases a surgical resident has completed according to the ACGME operative log to their operative proficiency, quality of anastomosis, operative experience, and confidence in both a simulation and operative setting.

MATERIALS AND METHODS:

General surgery residents ranging from PGY 1 to 5 participated in a simulation laboratory in which they completed an end-to-side vascular anastomosis. Each participant was given a weighted score based on technical proficiency and anastomosis quality using a previously validated Global Rating Scale (Duran et al, 2014). These scores were correlated to the General Surgery Milestones. Participants completed preoperative and postoperative surveys assessing resident operative experience using the 4-level Zwisch scale (DaRosa et al., 2013), confidence with vascular procedures and confidence performing simulated anastomoses. Confidence was assessed on a scale from 1 to 9 (not confident to extremely confident). Case logs were recorded for each participant. An IRB approved questionnaire was distributed to assess preoperative and postoperative roles of both the resident physician and faculty, with a defined goal. Univariate and multivariate analysis was performed.

RESULTS:

Twenty-one general surgery residents were evaluated in the simulation laboratory and 8 residents were assessed intraoperatively. The residents were evenly distributed throughout clinical years. Groups of residents were divided into quartiles based upon the number of vascular cases recorded in the ACGME database. No correlation was found between number of cases, Milestones score and the weighted score ($p = 0.94$). No statistical significance was found between confidence and quality of anastomosis ($p = 0.1$). Resident operative experience per the Zwisch scale was categorized most commonly as "Smart Help" by both the trainee and attending surgeon, despite mean resident confidence ratings of $6.67 (\pm 1.61)$ with vascular procedures.

CONCLUSIONS:

ACGME case logs, which are utilized to assess readiness for completion of general surgery residency, may not be indicative of a resident's operative competency and technical proficiency. Confidence is not correlated with technical ability. Faculty and resident insight as to their role in a procedure differ, as faculty feel that they are providing less help than the resident perceives. Careful examination of resident operative technique is the best measure of competency.

Simulation for Assessment of Milestones in Emergency Medicine Residents

Hart D, Bond W, Siegelman JN, Miller D, Cassara M, Barker L, Anders S, Ahn J, Huang H, Strother C, Hui J. *Acad Emerg Med.* 2018 Feb;25(2):205-220. doi: 10.1111/acem.13296. Epub 2017 Nov 9.

OBJECTIVES:

All residency programs in the United States are required to report their residents' progress on the milestones to the Accreditation Council for Graduate Medical Education (ACGME) biannually. Since the development and institution of this competency-based assessment framework, residency programs have been attempting to ascertain the best ways to assess resident performance on these metrics. Simulation was recommended by the ACGME as one method of assessment for many of the milestone subcompetencies. We developed three simulation scenarios with scenario-specific milestone-based assessment tools. We aimed to gather validity evidence for this tool.

METHODS:

We conducted a prospective observational study to investigate the validity evidence for three mannequin-based simulation scenarios for assessing individual residents on emergency medicine (EM) milestones. The subcompetencies (i.e., patient care [PC]1, PC2, PC3) included were identified via a modified Delphi technique using a group of experienced EM simulationists. The scenario-specific checklist (CL) items were designed based on the individual milestone items within each EM subcompetency chosen for assessment and reviewed by experienced EM simulationists. Two independent live raters who were EM faculty at the respective study sites scored each scenario following brief rater training. The inter-rater reliability (IRR) of the assessment tool was determined by measuring intraclass correlation coefficient (ICC) for the sum of the CL items as well as the global rating scales (GRSs) for each scenario. Comparing GRS and CL scores between various postgraduate year (PGY) levels was performed with analysis of variance.

RESULTS:

Eight subcompetencies were chosen to assess with three simulation cases, using 118 subjects. Evidence of test content, internal structure, response process, and relations with other variables were found. The ICCs for the sum of the CL items and the GRSs were >0.8 for all cases, with one exception (clinical management GRS = 0.74 in sepsis case). The sum of CL items and GRSs ($p < 0.05$) discriminated between PGY levels on all cases. However, when the specific CL items were mapped back to milestones in various proficiency levels, the milestones in the higher proficiency levels (level 3 [L3] and 4 [L4]) did not often discriminate between various PGY levels. L3 milestone items discriminated between PGY levels on five of 12 occasions they were assessed, and L4 items discriminated only two of 12 times they were assessed.

CONCLUSION:

Three simulation cases with scenario-specific assessment tools allowed evaluation of EM residents on proficiency L1 to L4 within eight of the EM milestone subcompetencies. Evidence of test content, internal structure, response process, and relations with other variables were found. Good to excellent IRR and the ability to discriminate between various PGY levels was found for both the sum of CL items and the GRSs. However, there was a lack of a positive relationship between advancing PGY level and the completion of higher-level milestone items (L3 and L4).

Competency Assessment in Family Medicine Residency: Observations, Knowledge-Based Examinations, and Advancement

Mainous AG 3rd, Fang B, Peterson LE. J Grad Med Educ. 2017 Dec;9(6):730-734. doi: 10.4300/JGME-D-17-00212.1.

BACKGROUND:

The Family Medicine (FM) Milestones are competency-based assessments of residents in key dimensions relevant to practice in the specialty. Residency programs use the milestones in semiannual reviews of resident performance from the time of entry into the program to graduation.

OBJECTIVE:

Using a national sample, we investigated the relationship of FM competency-based assessments to resident progress and the complementarity of milestones with knowledge-based assessments in FM residencies.

METHODS:

We used midyear and end-of-year milestone ratings for all FM residents in Accreditation Council for Graduate Medical Education-accredited programs during academic years 2014-2015 and 2015-2016. The milestones contain 22 items across 6 competencies. We created a summative index across the milestones. The American Board of Family Medicine database provided resident demographics and in-training examination (ITE) scores. We linked information to the milestone data.

RESULTS:

The sample encompassed 6630 FM residents. The summative milestone index increased, on average, for each cohort (postgraduate year 1 [PGY-1] to PGY-2 and PGY-2 to PGY-3) at each assessment. The correlation between the milestone index that excluded the medical knowledge milestone and ITE scores was $r = .195$ ($P < .001$) for PGY-1 to PGY-2 cohort and $r = .254$ ($P < .001$) for PGY-2 to PGY-3 cohort. For both cohorts, ITE scores and composite milestone assessments were higher for residents who advanced than for those who did not.

CONCLUSIONS:

Competency-based assessment using the milestones for FM residents seems to be a viable multidimensional tool to assess the successful progression of residents.

A Multicenter Prospective Comparison of the Accreditation Council for Graduate Medical Education Milestones: Clinical Competency Committee vs. Resident Self-Assessment

Watson RS, Borgert AJ, O Heron CT, Kallies KJ, Sidwell RA, Mellinger JD, Joshi AR, Galante JM, Chambers LW, Morris JB, Josloff RK, Melcher ML, Fuhrman GM, Terhune KP, Chang L, Ferguson EM, Auyang ED, Patel KR, Jarman BT. *J Surg Educ.* 2017 Nov - Dec;74(6):e8-e14. doi: 10.1016/j.jsurg.2017.06.009. Epub 2017 Jun 27.

OBJECTIVE:

The Accreditation Council for Graduate Medical Education requires accredited residency programs to implement competency-based assessments of medical trainees based upon nationally established Milestones. Clinical competency committees (CCC) are required to prepare biannual reports using the Milestones and ensure reporting to the Accreditation Council for Graduate Medical Education. Previous research demonstrated a strong correlation between CCC and resident scores on the Milestones at 1 institution. We sought to evaluate a national sampling of general surgery residency programs and hypothesized that CCC and resident assessments are similar.

DESIGN:

Details regarding the makeup and process of each CCC were obtained. Major disparities were defined as an absolute mean difference of ≥ 0.5 on the 4-point scale. A negative assessment disparity indicated that the residents evaluated themselves at a lower level than did the CCC. Statistical analysis included Wilcoxon rank sum and Sign tests.

SETTING:

CCCs and categorical general surgery residents from 15 residency programs completed the Milestones document independently during the spring of 2016.

RESULTS:

Overall, 334 residents were included; 44 (13%) and 43 (13%) residents scored themselves ≥ 0.5 points higher and lower than the CCC, respectively. Female residents scored themselves a mean of 0.08 points lower, and male residents scored themselves a mean of 0.03 points higher than the CCC. Median assessment differences for postgraduate year (PGY)-1-5 were 0.03 (range: -0.94 to 1.28), -0.11 (range: -1.22 to 1.22), -0.08 (range: -1.28 to 0.81), 0.02 (range: -to 1.00), and -0.19 (range: -1.16 to 0.50), respectively. Residents in university vs. independent programs had higher rates of negative assessment differences in medical knowledge (15% vs. 6%; $P = 0.015$), patient care (17% vs. 5%; $P = 0.002$), professionalism (23% vs. 14%; $P = 0.013$), and system-based practice (18% vs. 9%; $P = 0.031$) competencies. Major assessment disparities by sex or PGY were similar among individual competencies.

CONCLUSIONS:

Surgery residents in this national cohort demonstrated self-awareness when compared to assessments by their respective CCCs. This was independent of program type, sex, or level of training. PGY 5 residents, female residents, and those from university programs consistently rated themselves lower than the CCC, but these were not major disparities and the significance of this is unclear.

Validity Evidence from Ratings of Pediatric Interns and Subinterns on a Subset of Pediatric Milestones

Turner TL, Bhavaraju VL, Luciw-Dubas UA, Hicks PJ, Multerer S, Osta A, McDonnell J, Poynter S, Schumacher DJ, Tenney-Soeiro R, Waggoner-Fountain L, Schwartz A; and the Association of Pediatric Program Directors Longitudinal Educational Assessment Research Network–National Board of Medical Examiners Pediatrics Milestones Assessment Group. *Acad Med.* 2017 Jun;92(6):809-819. doi: 10.1097/ACM.0000000000001622.

PURPOSE:

To investigate evidence for validity of faculty members' pediatric milestone (PM) ratings of interns (first-year residents) and subinterns (fourth-year medical students) on nine subcompetencies related to readiness to serve as a pediatric intern in the inpatient setting.

METHOD:

The Association of Pediatric Program Directors Longitudinal Educational Assessment Research Network (APPD LEARN) and the National Board of Medical Examiners collaborated to investigate the utility of assessments of the PMs for trainees' performance. Data from 32 subinterns and 179 interns at 17 programs were collected from July 2012 through April 2013. Observers used several tools to assess learners. At each site, a faculty member used these data to make judgments about the learner's current developmental milestone in each subcompetency. Linear mixed models were fitted to milestone judgments to examine their relationship with learner's rank and subcompetency.

RESULTS:

On a 5-point developmental scale, mean milestone levels for interns ranged from 3.20 (for the subcompetency Work effectively as a member of a team) to 3.72 (Humanism) and for subinterns from 2.89 (Organize and prioritize care) to 3.61 (Professionalization). Mean milestone ratings were significantly higher for the Professionalism competency (3.59-3.72) for all trainees compared with Patient Care (2.89- 3.24) and Personal and Professional Development (3.33-3.51). Mean intern ratings were significantly higher than mean subintern ratings for all nine subcompetencies except Professionalization, Humanism, and Trustworthiness.

CONCLUSIONS:

The PMs had a coherent internal structure and could distinguish between differing levels of trainees, which supports their validation for documenting developmental progression of pediatric trainees.

Examining the Functioning and Reliability of the Family Medicine Milestones

Peabody MR, O'Neill TR, Peterson LE. J Grad Med Educ. 2017 Feb;9(1):46-53. doi: 10.4300/JGME-D-16-00172.1.

BACKGROUND:

The Family Medicine (FM) Milestones are a framework designed to assess development of residents in key dimensions of physician competency. Residency programs use the milestones in semiannual reviews of resident performance from entry toward graduation.

OBJECTIVE:

To examine the functioning and reliability of the FM Milestones and to determine whether they measure the amount of a latent trait (eg, knowledge or ability) possessed by a resident or simply indicate where a resident falls along the training sequence.

METHODS:

This study utilized the Rasch Partial Credit model to examine academic year 2014-2015 ratings for 10563 residents from 476 residency programs (postgraduate year [PGY] 1 = 3639; PGY-2 = 3562; PGY-3 = 3351; PGY-4 = 11).

RESULTS:

Reliability was exceptionally high at 0.99. Mean scores were 3.2 (SD = 1.3) for PGY-1; 5.0 (SD = 1.3) for PGY-2; 6.7 (SD = 1.2) for PGY-3; and 7.4 (SD = 1.0) for PGY-4. Keyform analysis showed a rating on 1 item was likely to be similar for all other items.

CONCLUSIONS:

Our findings suggest that FM Milestones seem to largely function as intended. Lack of spread in item difficulty and lack of variation in category probabilities show that FM Milestones do not measure the amount of a latent trait possessed by a resident, but rather describe where a resident falls along the training sequence. High reliability indicates residents are being rated in a stable manner as they progress through residency, and individual residents deviating from this rating structure warrant consideration by program leaders.

Correlations between Ratings on the Resident Annual Evaluation Summary and the Internal Medicine Milestones and Association with ABIM Certification Examination Scores among US Internal Medicine Residents, 2013-2014

Hauer KE, Vandergrift J, Hess B, Lipner RS, Holmboe ES, Hood S, Iobst W, Hamstra SJ, McDonald FS. JAMA. 2016 Dec 6;316(21):2253-2262. doi: 10.1001/jama.2016.17357.

IMPORTANCE:

US internal medicine residency programs are now required to rate residents using milestones. Evidence of validity of milestone ratings is needed.

OBJECTIVE:

To compare ratings of internal medicine residents using the pre-2015 resident annual evaluation summary (RAES), a non-developmental rating scale, with developmental milestone ratings.

DESIGN, SETTING, AND PARTICIPANTS:

Cross-sectional study of US internal medicine residency programs in the 2013-2014 academic year, including 21 284 internal medicine residents (7048 postgraduate-year 1 [PGY-1], 7233 PGY-2, and 7003 PGY-3).

EXPOSURES:

Program director ratings on the RAES and milestone ratings.

MAIN OUTCOMES AND MEASURES:

Correlations of RAES and milestone ratings by training year; correlations of medical knowledge ratings with American Board of Internal Medicine (ABIM) certification examination scores; rating of unprofessional behavior using the 2 systems.

RESULTS:

Corresponding RAES ratings and milestone ratings showed progressively higher correlations across training years, ranging among competencies from 0.31 (95% CI, 0.29 to 0.33) to 0.35 (95% CI, 0.33 to 0.37) for PGY-1 residents to 0.43 (95% CI, 0.41 to 0.45) to 0.52 (95% CI, 0.50 to 0.54) for PGY-3 residents (all P values <.05). Linear regression showed ratings differed more between PGY-1 and PGY-3 years using milestone ratings than the RAES (all P values <.001). Of the 6260 residents who attempted the certification examination, the 618 who failed had lower ratings using both systems for medical knowledge than did those who passed (RAES difference, -0.9; 95% CI, -1.0 to -0.8; P < .001; milestone medical knowledge 1 difference, -0.3; 95% CI, -0.3 to -0.3; P < .001; and medical knowledge 2 difference, -0.2; 95% CI, -0.3 to -0.2; P < .001). Of the 26 PGY-3 residents with milestone ratings indicating deficiencies on either of the 2 medical knowledge subcompetencies, 12 failed the certification examination. Correlation of RAES ratings for professionalism with residents' lowest professionalism milestone ratings was 0.44 (95% CI, 0.43 to 0.45; P <.001).

CONCLUSIONS AND RELEVANCE:

Among US internal medicine residents in the 2013-2014 academic year, milestone-based ratings correlated with RAES ratings but with a greater difference across training years. Both rating systems for medical knowledge correlated with ABIM certification examination scores. Milestone ratings may better detect problems with professionalism. These preliminary findings may inform establishment of the validity of milestone-based assessment.

Using the ACGME Milestones for Resident Self-Evaluation and Faculty Engagement

Meier AH, Gruessner A, Cooney RN. J Surg Educ. 2016 Nov-Dec;73(6):e150-e157. doi: 10.1016/j.jsurg.2016.09.001.

BACKGROUND:

Since July 2014 General Surgery residency programs have been required to use the Accreditation Council for Graduate Medical Education milestones twice annually to assess the progress of their trainees. We felt this change was a great opportunity to use this new evaluation tool for resident self- assessment and to furthermore engage the faculty in the educational efforts of the program.

METHODS:

We piloted the milestones with postgraduate year (PGY) II and IV residents during the 2013/2014 academic year to get faculty and residents acquainted with the instrument. In July 2014, we implemented the same protocol for all residents. Residents meet with their advisers quarterly. Two of these meetings are used for milestones assessment. The residents perform an independent self- evaluation and the adviser grades them independently. They discuss the evaluations focusing mainly on areas of greatest disagreement. The faculty member then presents the resident to the clinical competency committee (CCC) and the committee decides on the final scores and submits them to the Accreditation Council for Graduate Medical Education website. We stored all records anonymously in a MySQL database. We used Anova with Tukey post hoc analysis to evaluate differences between groups. We used intraclass correlation coefficients and Krippendorff's α to assess interrater reliability.

RESULTS:

We analyzed evaluations for 44 residents. We created scale scores across all Likert items for each evaluation. We compared score differences by PGY level and raters (self, adviser, and CCC). We found highly significant increases of scores between most PGY levels ($p < 0.05$). There were no significant score differences per PGY level between the raters. The interrater reliability for the total score and 6 competency domains was very high (ICC: 0.87-0.98 and α : 0.84-0.97). Even though this milestone evaluation process added additional work for residents and faculty we had very good participation (93.9% by residents and 92.9% by faculty) and feedback was generally positive.

CONCLUSION:

Even though implementation of the milestones has added additional work for general surgery residency programs, it has also opened opportunities to furthermore engage the residents in reflection and self- evaluation and to create additional venues for faculty to get involved with the educational process within the residency program. Using the adviser as the initial rater seems to correlate closely with the final CCC assessment. Self-evaluation by the resident is a requirement by the RRC and the milestones seem to be a good instrument to use for this purpose. Our early assessment suggests the milestones provide a useful instrument to track trainee progression through their residency.

The Internal Medicine Reporting Milestones: Cross-Sectional Description of Initial Implementation in US Residency Programs

Hauer KE, Clauser J, Lipner RS, Holmboe ES, Caverzagie K, Hamstra SJ, Hood S, Iobst W, Warm E, McDonald FS. *Ann Intern Med.* 2016 Sep 6;165(5):356-62. doi: 10.7326/M15-2411. Epub 2016 May 10.

BACKGROUND:

High-quality assessment of resident performance is needed to guide individual residents' development and ensure their preparedness to provide patient care. To facilitate this aim, reporting milestones are now required across all internal medicine (IM) residency programs.

OBJECTIVE:

To describe initial milestone ratings for the population of IM residents by IM residency programs.

DESIGN:

Cross-sectional study.

SETTING:

IM residency programs.

PARTICIPANTS:

All IM residents whose residency program directors submitted milestone data at the end of the 2013-2014 academic year.

MEASUREMENTS:

Ratings addressed 6 competencies and 22 subcompetencies. A rating of "not assessable" indicated insufficient information to evaluate the given subcompetency. Descriptive statistics were calculated to describe ratings across competencies and training years.

RESULTS:

Data were available for all 21 774 U.S. IM residents from all 383 programs. Overall, 2889 residents (1621 in postgraduate year 1 [PGY-1], 902 in PGY-2, and 366 in PGY-3) had at least 1 subcompetency rated as not assessable. Summaries of average ratings by competency and training year showed higher ratings for PGY-3 residents in all competencies. Overall ratings for each of the 6 individual competencies showed that fewer than 1% of third-year residents were rated as "unsatisfactory" or "conditional on improvement." However, when subcompetency milestone ratings were used, 861 residents (12.8%) who successfully completed training had at least 1 competency with all corresponding subcompetencies graded below the threshold of "readiness for unsupervised practice."

LIMITATION:

Data were derived from a point in time in the first reporting period in which milestones were used.

CONCLUSION:

The initial milestone-based evaluations of IM residents nationally suggest that documenting developmental progression of competency is possible over training years. Subcompetencies may identify areas in which residents might benefit from additional feedback and experience. Future work is needed to explore how milestones are used to support residents' development and enhance residency curricula.

A First Look at the Accreditation Council for Graduate Medical Education Anesthesiology Milestones: Implementation of Self-Evaluation in a Large Residency Program

Ross FJ, Metro DG, Beaman ST, Cain JG, Dowdy MM, Apfel A, Jeong JH, Ibinson JW. J Clin Anesth. 2016 Aug;32:17-24. doi: 10.1016/j.jclinane.2015.12.026. Epub 2016 Mar 22.

STUDY OBJECTIVE:

The objective was to determine if there is a correlation between resident postgraduate year (PGY) of training and self-evaluation of performance using the Accreditation Council for Graduate Medical Education milestones.

DESIGN:

Survey.

SETTING:

Residency program at a large academic center.

PATIENTS:

Residents and Faculty Clinical Competency Committee (CCC).

INTERVENTIONS:

None.

MEASUREMENTS:

Resident and CCC milestone scores.

MAIN RESULTS:

Correlation coefficients for average score for each milestone vs PGY level ranged from 0.80 for receiving and giving feedback to 0.95 for anesthetic choice and conduct. All milestones showed a relatively linear relationship with PGY of training, and none were found to be consistently reached very late or very early in training. When examining variation across the scores for the individual residents, the distributions for PGY-2 and -3 appeared to be wider than those for PGY-1 and -4. The intraclass correlation coefficients ranged from 0.718 to 0.928.

CONCLUSIONS:

There was a remarkable degree of consistency in the relationship between level of training and resident self-assessment score for every milestone, as well as strong agreement between the resident and CCC faculty scores. Examination of the variance in the scores, when interpreted in light of our particular training program's characteristics, suggests that the milestones accurately reflect the progression in skill across the residency. In addition, given the concordance between the self-evaluation scores and the CCC faculty scores, self-evaluation may be a reasonable starting point as programs begin the daunting task of determining scores for each of the 25 milestones as part of the biannual evaluation process.

Pathology Milestones: Assessing Clinical Competency by Committee

Klutts JS, Guerin LA, Bruch LA, Firchau DJ, Knudson CM, Rosenthal NS, Samuelson MI, Jensen CS, Delwiche JL, Krasowski MD. Acad Pathol. 2015 Oct 29;2(4):2374289515614003. doi: 10.1177/2374289515614003. eCollection 2015 Oct-Dec.

ABSTRACT:

All Accreditation Council for Graduate Medical Education accredited pathology residency training programs are now required to evaluate residents using the new Pathology Milestones assessment tool. Similar to implementation of the 6 Accreditation Council for Graduate Medical Education competencies a decade ago, there have been challenges in implementation of the new milestones for many residency programs. The pathology department at the University of Iowa has implemented a process that divides the labor of the task in rating residents while also maintaining consistency in the process. The process is described in detail, and some initial trends in milestone evaluation are described and discussed. Our experience indicates that thoughtful implementation of the Pathology Milestones can provide programs with valuable information that can inform curricular changes.

Practicing Emergency Physicians Report Performing Well on Most Emergency Medicine Milestones

Peck TC, Dubosh N, Rosen C, Tibbles C, Pope J, Fisher J. J Emerg Med. 2014 Oct;47(4):432-40. doi: 10.1016/j.jemermed.2014.04.032. Epub 2014 Jul 8.

BACKGROUND:

The Accreditation Council for Graduate Medical Education's Next Accreditation System endorsed specialty-specific milestones as the foundation of an outcomes-based resident evaluation process. These milestones represent five competency levels (entry level to expert), and graduating residents will be expected to meet Level 4 on all 23 milestones. Limited validation data on these milestones exist. It is unclear if higher levels represent true competencies of practicing emergency medicine (EM) attendings.

OBJECTIVE:

Our aim was to examine how practicing EM attendings in academic and community settings self-evaluate on the new EM milestones.

METHODS:

An electronic self-evaluation survey outlining 9 of the 23 EM milestones was sent to a sample of practicing EM attendings in academic and community settings. Attendings were asked to identify which level was appropriate for them.

RESULTS:

Seventy-nine attendings were surveyed, with an 89% response rate. Sixty-one percent were academic. Twenty-three percent (95% confidence interval [CI] 20%-27%) of all responses were Levels 1, 2, or 3; 38% (95% CI 34%-42%) were Level 4; and 39% (95% CI 35%-43%) were Level 5. Seventy-seven percent of attendings found themselves to be Level 4 or 5 in eight of nine milestones. Only 47% found themselves to be Level 4 or 5 in ultrasound skills ($p = 0.0001$).

CONCLUSIONS:

Although a majority of EM attendings reported meeting Level 4 milestones, many felt they did not meet Level 4 criteria. Attendings report less perceived competence in ultrasound skills than other milestones. It is unclear if self-assessments reflect the true competency of practicing attendings. The study design can be useful to define the accuracy, precision, and validity of milestones for any medical field.

The Pathology Milestones and the Next Accreditation System

Naritoku WY, Alexander CB, Bennett BD, Black-Schaffer WS, Brissette MD, Grimes MM, Hoffman RD, Hunt JL, Iezzoni JC, Johnson R, Kozel J, Mendoza RM, Post MD, Powell SZ, Procop GW, Steinberg JJ, Thorsen LM, Nestler SP. Arch Pathol Lab Med. 2014 Mar;138(3):307-15. doi: 10.5858/arpa.2013-0260-SA.

BACKGROUND:

In the late 1990s, the Accreditation Council for Graduate Medical Education developed the Outcomes Project and the 6 general competencies with the intent to improve the outcome of graduate medical education in the United States. The competencies were used as the basis for developing learning goals and objectives and tools to evaluate residents' performance. By the mid-2000s the stakeholders in resident education and the general public felt that the Outcomes Project had fallen short of expectations.

OBJECTIVE:

To develop a new evaluation method to track trainee progress throughout residency using benchmarks called milestones. A change in leadership at the Accreditation Council for Graduate Medical Education brought a new vision for the accreditation of training programs and a radically different approach to the evaluation of residents.

DATA SOURCES:

The Pathology Milestones Working Group reviewed examples of developing milestones in other specialties, the literature, and the Accreditation Council for Graduate Medical Education program requirements for pathology to develop pathology milestones. The pathology milestones are a set of objective descriptors for measuring progress in the development of competency in patient care, procedural skill sets, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice.

CONCLUSIONS:

The milestones provide a national standard for evaluation that will be used for the assessment of all residents in Accreditation Council for Graduate Medical Education-accredited pathology training programs.