How to Use Milestones Data to Improve your Program

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:

Milestones can be used to spot trends and identify underperformers. The assessment data collated by Clinical Competency Committees (CCCs) lead to assigning appropriate Milestones levels to their trainees. The Milestones data are summarized in the ACGME Accreditation Data System (ADS) which provides graphic representations of Milestones trends for each training level. The data can be analyzed to identify trends within a program and between institutions, and can be used to identify program strengths, gaps, and areas for improvement.

What's in the literature?

The bibliography contains examples of how programs can use Milestones longitudinal data to measure the developmental and educational progression of learners, predict written board passage, using assessment data to detect underperforming trainees, examine achieving graduation targets in certain competencies, and exploring national trends in Milestones data for different specialties. Below are more examples of what's in the literature:

- How to use Milestones data to provide a unique opportunity for individualization of training pathways
- How Milestones data can lead CCC discussions to assign appropriate Milestones levels to trainees and help guide individualized learning plans
- How Milestones data can suggest areas for program improvement
- How to use predictive analytics (predictive probability values [PPVs]) to identify the percentage of learners who are unlikely to achieve a target Milestone level by the time of graduation
- How to spot trends and identify underperformers
- How to use Milestone data to identify curricular needs

Racial and Ethnic Differences in Internal Medicine Residency Assessments

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

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 (β = .45, p = 0.03) and higher ABSITE score (β = .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, Fliotsos 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.

Evaluation of house staff candidates for program fit: a cohort-based controlled study

BACKGROUND:
Medical school academic achievements do not necessarily predict house staff job performance. This study explores a selection mechanism that improves house staff-program fit that enhances the Accreditation Council for Graduate Medical Education Milestones performance ratings.

OBJECTIVE:
Traditionally, house staff were selected primarily on medical school academic performance. To improve residency performance outcomes, the Program designed a theory-driven selection tool to assess house staff candidates on their personal values and goals fit with Program values and goals. It was hypothesized cohort performance ratings will improve because of the intervention.

METHODS:
Prospective quasi-experimental cohort design with data from two house staff cohorts at a university-based categorical Internal Medicine Residency Program. The intervention cohort, comprising 45 house staff from 2016 to 2017, was selected using a Behaviorally Anchored Rating Scales (BARS) tool for program fit. The control cohort, comprising 44 house staff from the prior year, was selected using medical school academic achievement scores. House staff performance was evaluated using ACGME Milestones indicators. The mean scores for each category were compared between the intervention and control cohorts using Student's t-tests with Bonferroni correction and Cohen's d for effect size.

RESULTS: The cohorts were no different in academic performance scores at time of Program entry. The intervention cohort outperformed the control cohort on all 6 dimensions of Milestones by end-PGY1 and 3 of 6 dimensions by mid-PGY3.

CONCLUSION: Selecting house staff based on compatibility with Residency Program values and objectives may yield higher job performance because trainees benefit more from a better fit with the training program.

Beyond opioid prescribing: Evaluation of a substance use disorder curriculum for OBGYN residents

OBJECTIVE:
Amidst the current opioid crisis, there is a need for better integration of substance use disorder screening and treatment across specialties. However, there is no consensus regarding how to best instruct OBGYN trainees in the clinical skills related to opioid and other substance use disorders (SUD). Study objectives were (1) to assess the effectiveness a SUD curriculum to improve self-reported competence among OBGYN residents and (2) to explore its effectiveness to improve attending evaluations of residents' clinical skills as well as its feasibility and acceptability from the resident perspective.

METHODS:
A pilot 3-session curriculum was developed and adapted to SUD screening and treatment which included readings, didactics, and supervised outpatient clinical experiences for OBGYN post-graduate year 1 (PGY-1) residents rotating through an integrated OBGYN-SUD clinic. Eighteen residents completed pre and post clinical skills self-assessments (SUD screening, counseling, referring, Motivational Interviewing) using an adapted Zwisch Rating Scale (range 1-5). Scores were compared between time points using paired t-tests. Sub-samples also (a) were evaluated by the attending on three relevant Accreditation Council for Graduate Medical Education Milestones (ACGME) milestone sets using the web-based feedback program, myTIPreport (n = 10) and (b) completed a qualitative interview (n = 4).

RESULTS:
All PGY-1s (18/18) across three academic years completed the 3-session SUD curriculum. Clinical skill self-assessments improved significantly in all areas [SUD Screening (2.44 (0.98) vs 3.56 (0.62), p = <0.01); Counseling (1.81 (0.71) vs 3.56 (0.51), p = < .01; Referring (2.03 (0.74) vs 3.17 (0.71), p = < .01; Motivational Interviewing (1.94 (1.06) vs 3.33 (0.69), p = < .01)]. Milestone set levels assigned by attending evaluations (n = 10) also improved. Qualitative data (n = 4) revealed high acceptability; all curriculum components were viewed positively, and feedback was provided (e.g., desire for more patient exposures).

CONCLUSION:
A pilot SUD curriculum tailored for OBGYN PGY-1 residents that goes beyond opioid prescribing to encompass SUD management is feasible, acceptable and likely effective at improving SUD core clinical skills.

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

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.

Concurrent Validity Evidence for Entrustable Professional Activities in General Surgery Residents

BACKGROUND:
A subset of Entrustable Professional Activities (EPAs) has been developed for general surgery. We aim to contribute validity evidence for EPAs as an assessment framework for general surgery residents, including concurrent validity compared to ACGME milestones, the current gold standard for evaluating competency.

STUDY DESIGN:
This is a cross-sectional study in a general surgery training program within a tertiary academic medical center. EPA assessments were submitted using a mobile app and scored on a numerical scale, mirroring milestones. EPA score distribution was analyzed with respect to post-graduate year (PGY) level and phase of care. Proportional odds logistic regression identified significant predictors. Spearman rank and Wilcoxon rank tests were used for comparisons with milestone ratings.

RESULTS:
From August 2018 to December 2019, 320 assessments were collected. EPA scores increased by PGY level. Operative phase EPA scores were significantly lower than nonoperative phase scores. PGY level, operative phase, and case difficulty significantly influenced entrustment scoring. EPA scores demonstrated strong correlation with nonoperative milestones patient care-1, medical knowledge-1, interpersonal and communication skills-2, interpersonal and communication skills-3, professionalism-1, professionalism-3, and practice-based learning and improvement-2 ($\rho > 0.5$, $p < 0.05$) and a weaker correlation with operative milestones patient care-3 and medical knowledge-2 ($\rho < 0.5$, $p < 0.05$).

CONCLUSIONS:
The influence of PGY level and operative phase on entrustment scoring supports the validity of EPAs as a formative evaluation framework for general surgery resident performance. In addition, evident correlations between EPA scores and respective milestone ratings provide concurrent validity evidence.

Identifying Strategies for Struggling Surgery Residents

BACKGROUND:
Struggling residents are not uncommon in general surgery. Early identification of these residents and effective remediation remain imperfect.

METHODS:
We performed a survey of program directors (PD) across all general surgery residencies. Survey questions included the following: demographic information about the program and PD, 10 vignettes about hypothetical residents struggling in various ACGME milestones to assess how PDs would address these deficiencies, and self-reported PD preparedness and availability of resources to support struggling residents.

RESULTS:
In total, we received 82 responses to our survey. All PDs who participated in our study reported having struggling residents in their program. The three most common ways struggling residents are identified were faculty word-of-mouth, formal evaluations such as milestones and ABSITE performance, and resident word-of-mouth. Over 18% of PDs reported having little to no relevant training in addressing the needs of a struggling resident, and 65.9% of PDs did not feel that their program had 'completely adequate' resources to address these needs. In the majority of cases, PDs offer mentorship with themselves or other faculty as a remediation strategy with infrequent use of other resources.

CONCLUSIONS:
Strategies to identify struggling residents and remediation strategies varied widely across programs. Diversifying remediation approaches should be considered for more effective remediation.

Impact of autonomic regulation on burnout and performance in thoracic surgery residents

OBJECTIVE:
This study sought to determine the feasibility of collecting physiologic data in thoracic surgery residents and whether it would correlate with burnout and burnout with performance.

METHODS:
This was a prospective study of thoracic surgery residents over a 5-month period. Participants were evaluated with a wearable biometric device (heart rate variability and sleep) and the Maslach Burnout Inventory. Resident performance was quantified using Accreditation Council for Graduate Medical Education Milestones (scale, 1-5) normalized to program-designated targets (3 for postgraduate year 6 or lower residents and 4 for postgraduate year 7 residents).

RESULTS:
The cohort consisted of 71% female participants (5/7) with 86% of residents having 1 or more children. High levels of emotional exhaustion (median, 30 [interquartile range, 20-36], where >26 is high) and high levels of depersonalization (median, 16 [interquartile range, 14-22], where >12 is high) were common, but personal accomplishment was also uniformly high (median, 43 [interquartile range, 41-46], where >38 is high). There was a significant correlation between heart rate variability and emotional exhaustion ($r(12) = 0.65$, $P = .01$) but not depersonalization ($P = .28$) or personal accomplishment ($P = .24$). Depersonalization and personal accomplishment did not correlate with resident performance ($P = .12$ and $P = .75$, respectively); however, increased emotional exhaustion showed a significant correlation with higher resident performance during periods when burnout was reported ($r(6) = 0.76$, $P = .047$).

CONCLUSIONS:
Dynamic measurement of resting heart rate variability may offer an objective measure of burnout in thoracic surgery residents. Thoracic surgery residents who report high levels of burnout in this cohort maintained the ability to meet program-designated milestones at or above the level expected of their postgraduate year.

Diversity, Equity, and Inclusion Milestones: Creation of a Tool to Evaluate Graduate Medical Education Programs

BACKGROUND:
As the Accreditation Council for Graduate Medical Education (ACGME) began to ask programs to report their efforts surrounding diversity, equity, and inclusion (DEI), program directors felt ill prepared to evaluate their programs and measure change.

OBJECTIVE:
To develop a tool that would allow graduate medical education (GME) programs to evaluate the current state of DEI within their residencies, identify areas of need, and track progress; to evaluate feasibility of using this assessment method within family medicine training programs; and to analyze and report pilot data from implementation of these milestones within family medicine residency programs.

METHODS:
The Association of Family Medicine Residency Directors (AFMRD) Diversity and Health Equity (DHE) Task Force developed a tool for program DEI evaluation modeled after the ACGME Milestones. These milestones focus on DEI assessment in 5 key domains: Institution, Curriculum, Evaluation, Resident Personnel, and Faculty Personnel. After finalizing a draft, a pilot implementation of the milestones was conducted by a convenience sample of 10 AFMRD DHE Task Force members for their own programs.

RESULTS:
Scores varied widely across surveyed programs for all milestones. Highest average scores were seen for the Curriculum milestone (2.65) and the lowest for the Faculty Personnel milestone (2.0). Milestone assessments were completed within 10 to 40 minutes using various methods.

CONCLUSIONS:
The AFMRD DEI Milestones were developed for program assessment, goal setting, and tracking of progress related to DEI within residency programs. The pilot implementation showed these milestones were easily used by family medicine faculty members in diverse settings.

Milestones as a Faculty Development Tool for Career Academic Physicians

BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) has implemented milestones for progression of residents. Career academic physicians would benefit from similar concrete guidance for scholarly activity and faculty development. After developing milestones across six recognized competencies among our family medicine academicians, we acknowledged the potential benefit of expanding the development of milestones throughout the academic medical center.

METHODS:
Milestones that we previously developed were modified by departmental leaders within our institution reflecting levels of career development based on benchmarks in each field. These objective measures for guiding maturation of clinical and academic skill sets were then circulated to clinicians in five residency programs throughout our academic medical center for self-evaluation. We analyzed the completed surveys to determine if an association exists between years in academics and rank across each area of competency.

RESULTS:
We received fifty-three responses from the 91 faculty invited. We noted a significant association in the competency of medical knowledge with progression from assistant to full professor, and we noted a trend toward significance in professionalism and progression from assistant to full professor. These objective measures of clinician development and competency suggest association with levels of academic career development by rank within the institution.

CONCLUSIONS:
This rubric can be helpful for directing faculty development and faculty mentorship. These milestones are general enough that other physician specialties may be able to adopt them for their own needs.

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

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

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: β = -0.67, P = .003; ICS02: β = -0.70, P = .001; ICS03: β = -0.60, P = .004) at various residency time points. Similar associations were noted for PROF03 (β = -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.

Can ACGME Milestones Predict Surgical Specialty Board Passage: An Example in Obstetrics and Gynecology

BACKGROUND:
Multiple tools including Accreditation Council for Graduate Medical Education (ACGME) standardized milestones can be utilized to assess trainee and residency program performance. However, little is known regarding the objective validation of these tools in predicting written board passage.

METHODS:
In this retrospective study, data was gathered on n = 45 Wayne State University Obstetrics and Gynecology program graduates over the five-year period ending July 2018. United States Medical Licensing Examination (USMLE) scores, Council on Resident Education in Obstetrics and Gynecology (CREOG) in-training scores and ACGME milestones were used to predict American Board of Obstetrics and Gynecology (ABOG) board passage success on first attempt. Significance was set at $p < 0.05$.

RESULTS:
Written board passage was associated with average CREOGs ($p = 0.01$) and milestones ($p = 0.008$) while USMLE1 was not significantly associated ($p = 0.055$). USMLE1 <217 (Positive predictive value (PPV) = 96%). CREOGs <197 (PPV = 100%) and milestones <3.25 (PPV = 100%), particularly practice-based learning and systems-based practice milestones were most strongly correlated with board failure. Using a combination of these two milestones, it is possible to correctly predict board passage using our model (PPV = 86%).

DISCUSSION:
This study is the first validating the utility of milestones in a surgical specialty by demonstrating their ability to predict board passage. Residents with CREOGs or milestones below thresholds are at risk for board failure and may warrant early intervention.

The Use of Learning Analytics to Enable Detection of Underperforming Trainees: An Analysis of National Vascular Surgery Trainee ACGME Milestones Assessment Data

OBJECTIVE:
This study aims to investigate at-risk scores of semiannual Accreditation Council for Graduate Medical Education (ACGME) Milestone ratings for vascular surgical trainees' final achievement of competency targets.

BACKGROUND:
ACGME Milestones assessments have been collected since 2015 for Vascular Surgery. It is unclear whether milestone ratings throughout training predict achievement of recommended performance targets upon graduation.

METHODS:
National ACGME Milestones data were utilized for analyses. All trainees completing 2-year vascular surgery fellowships in June 2018 and 5-year integrated vascular surgery residencies in June 2019 were included. A generalized estimating equations model was used to obtain at-risk scores for each of the 31 sub-competencies by semiannual review periods, to estimate the probability of trainees achieving the recommended graduation target based on their previous ratings.

RESULTS:
122 VSFs (95.3%) and 52 IVSRs (100%) were included. VSFs and IVSRs did not achieve level 4.0 competency targets at a rate of 1.6-25.4% across sub-competencies, which was not significantly different between the two groups for any of the sub-competencies (p=0.161-0.999). Trainees were found to be at greater risk of not achieving competency targets when lower milestone ratings were assigned, and at later time-points in training. At a milestone rating of ≤ 2.5, with one year remaining prior to graduation, the at-risk score for not achieving the target level 4.0 milestone ranged from 2.9% - 77.9% for VSFs and 33.3% - 75.0% for IVSRs.

CONCLUSIONS:
The ACGME Milestones provide early diagnostic and predictive information for vascular surgery trainees' achievement of competence at completion of training.

National Resident Matching Program Rank Order and Performance in an Internal Medicine Residency

OBJECTIVES:
Prospective first-year house staff and residency program leaders spend substantial time, effort, and expense preparing a rank order list for the National Resident Matching Program (NRMP). Previous studies have mostly shown minimal or no relation between rank order and subsequent resident performance, raising questions about the value of this process. Furthermore, no previous studies have been done with Internal Medicine residencies. As such, the purpose of this study was to compare NRMP rank order to multiple objective outcomes of an Internal Medicine residency.

METHODS:
A retrospective cohort of Internal Medicine residents from five consecutive graduating classes, trained between July 1, 2013 and July 31, 2020, were evaluated for five objective outcomes: Accreditation Council for Graduate Medical Education (ACGME) milestones, faculty rankings of quality, National In-Training Examination scores, chief resident attainment, and fellowship attainment. Outcomes were analyzed in relation to eight potential predictors: NRMP rank, medical school type and grades, immigration status, added qualifications, sex, age and US Medical Licensing Examination (USMLE) scores, using univariate and multivariate analyses.

RESULTS:
From a cohort of 61 residents, 56 were eligible. All eligible residents' data were included, for a participation rate of 100% (56 of 56). There were no statistically significant univariate or multivariate predictors for the endpoint of fellowship attainment. Higher USMLE scores were predictive of chief resident status in univariate analysis only. NRMP rank was significantly correlated with ACGME milestones in the univariate analysis. The multivariate analysis revealed that higher USMLE score was statistically significantly predictive of more favorable milestones, faculty ranking, and National In-Training Examination score.

CONCLUSIONS:
Higher USMLE score was statistically significantly associated with multiple favorable objective residency outcomes in an Internal Medicine residency. A better NRMP rank was correlated with favorable ACGME milestones in univariate analysis, but USMLE score emerged as the strongest predictor in multivariate analysis.

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

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

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.

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

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.

Development and Pilot Testing of Entrustable Professional Activities for US Anesthesiology Residency Training

Woodworth GE, Marty AP, Tanaka PP, Ambardekar AP, Chen F, Duncan MJ, Fromer IR, Hallman MR, Klesius LL, Ladlie BL, Mitchell SA, Miller Juve AK, McGrath BJ, Shepler JA, Sims C 3rd, Spofford CM, Van Cleve W, Maniker RB.

BACKGROUND:
Modern medical education requires frequent competency assessment. The Accreditation Council for Graduate Medical Education (ACGME) provides a descriptive framework of competencies and milestones but does not provide standardized instruments to assess and track trainee competency over time. Entrustable professional activities (EPAs) represent a workplace-based method to assess the achievement of competency milestones at the point-of- care that can be applied to anesthesiology training in the United States.

METHODS:
Experts in education and competency assessment were recruited to participate in a 6-step process using a modified Delphi method with iterative rounds to reach consensus on an entrustment scale, a list of EPAs and procedural skills, detailed definitions for each EPA, a mapping of the EPAs to the ACGME milestones, and a target level of entrustment for graduating US anesthesiology residents for each EPA and procedural skill. The defined EPAs and procedural skills were implemented using a website and mobile app. The assessment system was piloted at 7 anesthesiology residency programs. After 2 months, faculty were surveyed on their attitudes on usability and utility of the assessment system. The number of evaluations submitted per month was collected for 1 year.

RESULTS:
Participants in EPA development included 18 education experts from 11 different programs. The Delphi rounds produced a final list of 20 EPAs, each differentiated as simple or complex, a defined entrustment scale, mapping of the EPAs to milestones, and graduation entrustment targets. A list of 159 procedural skills was similarly developed. Results of the faculty survey demonstrated favorable ratings on all questions regarding app usability as well as the utility of the app and EPA assessments. Over the 2-month pilot period, 1636 EPA and 1427 procedure assessments were submitted. All programs continued to use the app for the remainder of the academic year resulting in 12,641 submitted assessments.

CONCLUSIONS:
A list of 20 anesthesiology EPAs and 159 procedural skills assessments were developed using a rigorous methodology to reach consensus among education experts. The assessments were pilot tested at 7 US anesthesiology residency programs demonstrating the feasibility of implementation using a mobile app and the ability to collect assessment data. Adoption at the pilot sites was variable; however, the use of the system was not mandatory for faculty or trainees at any site.

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

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.

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

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?

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 physiatric 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.

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

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., & Langhan, 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.

The Implementation of ACGME 'Reporting Milestones' in Internal Medicine Postgraduate Year Training Program - Experience Sharing and Preliminary Outcome

PURPOSE:
Competency-based medical education has been the current trend of education, and milestones assessment is one of the means to practice it. We hereby shared our experience and preliminary results in applying reporting milestones in our postgraduate year (PGY) training program.

METHODS:
We translated the English version of 'reporting milestones' to Chinese one and built the electronic report worksheets on our web-system. Milestones assessments were conducted at the end of every month since August 2018. We analyzed the completion rate and the results of the assessments in the 2018 academic year by using IBM SPSS Statistics version 23.

RESULTS:
The completion rate of milestones assessments was 96.7% (175/181), of which the most frequent unassessed subcompetencies were patient care (PC)-4 and PC-5 (both n = 9). The mean levels of the subcompetencies ranged from 3.5 to 4 on a scale of 1-5. Our PGY physicians demonstrated better in subcompetencies systems-based practice (SBP)-1, practice-based learning and improvement (PBLI)-1, and PBLI-3 (mean levels 4.0 ± 0.4, 4.0 ± 0.4, 4.0 ± 0.5, respectively), and worse in subcompetence SBP-3 (mean levels 3.5 ± 0.6) (p ＜0.05 by ANOVA). The progress of all subcompetencies between the first and third months of rotation course was insignificant.

CONCLUSIONS:
We accomplished the reporting milestones assessments in our hospital. Our experience in implementing milestones can be as a reference for other hospitals. The learning outcomes from milestones assessments may be helpful to improve our PGY program.

Application Factors Associated with Clinical Performance during Pediatric Internship

OBJECTIVE:
Our goal was to identify aspects of residency applications predictive of subsequent performance during pediatric internship.

METHODS:
We conducted a retrospective cohort study of graduates of U.S. medical schools who began pediatric internship in a large pediatric residency program in the summers of 2013 through 2017. The primary outcome was the weighted average of subjects' ACGME pediatric milestones scores at the end of pediatric internship. To determine factors independently associated with performance, we conducted multivariate linear mixed-effects models controlling for match year and Milestone grading committee as random effects and the following application factors as fixed effects: letter of recommendation strength, clerkship grades, medical school reputation, master's or PhD degrees, gender, USMLE Step 1 score, Alpha Omega Alpha membership, private medical school, and interview score.

RESULTS:
Our study population included 195 interns. In multivariate analyses, the aspects of applications significantly associated with composite Milestone scores at the end of internship were LOR strength (estimate 0.09, 95% confidence intervals 0.04, 0.15), numbers of clerkship honors (est. 0.05, 95% CI:
0.01-0.09), medical school ranking (est. 0.04, 95% CI: 0.08-0.01), having a master's degree (est. 0.19, 95% CI: 0.03-0.36), and not having a PhD (est. 0.14, 95% CI: 0.02-0.26). Overall the final model explained 18% of the variance in milestone scoring.

CONCLUSION:
Letter of recommendation strength, clerkship grades, medical school ranking, and having obtained a Master's degree were significantly associated with higher clinical performance during pediatric internship.

Utility of Residency Milestones Reported to Fellowship Directors: A National Survey of Pediatric Fellowship Program Directors

Reed S, Mink R, Li ST. Acad Pediatr. 2020 Jan 21. pii: S1876-2859(20)30005-X. doi: 10.1016/j.acap.2020.01.004. [Epub ahead of print]

BACKGROUND:
The Accreditation Council for Graduate Medical Education (ACGME) requires milestone-based assessments of residents and fellows. The ACGME recently allowed fellowship programs access to the final residency milestones for incoming fellows through the ACGME Accreditation Data System. It is unknown if fellowship programs are downloading residency milestones and if fellowship program directors (FPDs) believe they have value.

OBJECTIVE:
Determine how many pediatric FPDs downloaded residency milestones and FPD perspectives on usefulness of residency milestones for first-year fellows.

METHODS:
Cross-sectional survey of pediatric FPDs in the US, with assistance from the Subspecialty Pediatrics Investigator Network (SPIN) Steering Committee. Respondents were asked whether they downloaded residency milestones and their programs' specific use of these milestones for their first-year fellows. FPDs were asked open-ended questions about why residency milestones were or were not useful, how they could be more useful, and if they would be useful in recruitment. Descriptive statistics were used to explore quantitative data and content analysis was used to analyze qualitative data.

RESULTS:
66.5% (532/800) of FPDs responded, representing all 14 pediatric subspecialties. Most programs (60.7%; 323/532) did not download residency milestones for their first-year fellows. Of these, 67.5% (218/323) did not know they could. Of FPDs that downloaded and reviewed residency milestones, only 27% (50/185) used them for individualized education. Only 24% (129/532) of all FPDs thought residency milestones were useful or very useful. 41% (218/532) thought residency milestones would be useful or very useful during recruitment, but some believed this may harm applicants. FPDs felt residency milestones allowed for identification of trainee needs and baseline assessments, but felt that these milestones had limited usefulness during fellowship due to concerns about lack of validity evidence, relevance, and how milestones are assessed and reported (Table).

CONCLUSIONS:
Most pediatric subspecialty programs do not use residency milestones to tailor education for their first- year fellows and most think they have limited usefulness. While more FPDs felt that residency milestones might be useful during recruitment, there was not universal agreement. Further studies to improve validity of residency milestones may make them more useful to fellowship programs.

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

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.

Palliative Care and Communication Training in Neurosurgery Residency: Results of a Trainee Survey

OBJECTIVE:
Neurosurgeons care for critically ill patients near the end of life, yet little is known about how well their training prepares them for this role. We surveyed a random sample of neurosurgery residents to describe the quantity and quality of teaching activities related to serious illness communication and palliative care, and resident attitudes and perceived preparedness to care for seriously ill patients.

METHODS:
A previously validated survey instrument was adapted to reflect required communication and palliative care competencies in the 2015 the Accreditation Council for Graduate Medical Education (ACGME) Milestones for Neurological Surgery. The survey was reviewed for content validity by independent faculty neurosurgeons, piloted with graduating neurosurgical residents, and distributed online in August 2016 to neurosurgery residents in the United States using the American Association of Neurological Surgeons (AANS)/Congress of Neurological Surgeons (CNS) Joint Section on Neurotrauma and Critical Care email listserv. Multiple choice and Likert scale responses were analyzed using descriptive statistics.

RESULTS:
Sixty-two responses were recorded between August 2016 and October 2016. Most respondents reported no explicit teaching on: explaining risks and benefits of intubation and ventilation (69%), formulating prognoses in neurocritical care (60%), or leading family meetings (69%). Compared to performing craniotomies, respondents had less frequent practice leading discussions about withdrawing life-sustaining treatment (61% vs. 90%, p < 0.01, "weekly or more frequently"), and were less often observed (18% vs. 87%, p < 0.01) and given feedback on their performance (11% vs. 58%, p < 0.01). Nearly all respondents (95%) felt
"prepared to discuss withdrawing life-sustaining treatments," however half (48%) reported they "would benefit from more communication training during residency." Most (87%) reported moral distress, agreeing that they "participated in operations and worried whether surgery aligned with patient goals."

CONCLUSIONS:
Residents in our sample reported limited formal training, and relatively less observation and feedback, on required ACGME competencies in palliative care and communication. Most reported preparedness in this domain, but many were receptive to more training. Better quality and more consistent palliative care education in neurosurgery residency could improve competency and help ensure that neurosurgical care aligns with patient goals.

Family Medicine Residency Graduates' Preparation for Quality Improvement Leadership

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.

Resident Education in Complex Obstetric Procedures: Are We Adequately Preparing Tomorrow's Obstetricians?

OBJECTIVES:
The Accreditation Council for Graduate Medical Education (ACGME) milestones for obstetrics and gynecology (OB/GYN) residents include obstetrical technical skills. We sought to describe resident experience with surgical obstetrics and comfort performing procedures independently postgraduation.

STUDY DESIGN:
An anonymous 27-question e-survey was sent to OB/GYN residents in United States in March 2018, using the Council of Resident Education in Obstetrics and Gynecology coordinator listserv. Complex obstetric procedures included: forceps-assisted vaginal delivery (FAVD) and vacuum-assisted vaginal delivery (VAVD), cerclage, breech second twin, breech delivery, perineal repairs, and cesarean hysterectomy. Technical skill questions included experience as primary surgeon, comfort performing procedures independently, and for 4th year residents- comfort performing procedures postresidency. Demographic information was queried. Descriptive statistics was used to analyze responses.

RESULTS:
A total of 417 residents completed the survey. Respondents were 88% female, 75% from academic programs, and 48% postgraduate year 3 and 4. Among all residents, many had been primary surgeon in operative vaginal deliveries (51% FAVD, 72% VAVD), fewer for breech vaginal delivery (21%), breech second twin (34%), cesarean hysterectomy (21%), and 4th degree repairs (37%). All 4th-year respondents stated that they would feel comfortable performing either VAVD or FAVD postresidency. Note that 17, 33, 28, and 74% would not feel comfortable performing a 4th degree repair, cesarean hysterectomy, breech second twin, and breech vaginal delivery, respectively, postresidency.

CONCLUSION:
Despite ACGME recommendations, data suggest that many graduating residents may not be comfortable with these complex procedures.

Milestone Implementation's Impact on Narrative Comments and Perception of Feedback for Internal Medicine Residents: A Mixed Methods Study

BACKGROUND:
Feedback is a critical element of graduate medical education. Narrative comments on evaluation forms are a source of feedback for residents. As a shared mental model for performance, milestone-based evaluations may impact narrative comments and resident perception of feedback.

OBJECTIVE:
To determine if milestone-based evaluations impacted the quality of faculty members' narrative comments on evaluations and, as an extension, residents' perception of feedback.

DESIGN:
Concurrent mixed methods study, including qualitative analysis of narrative comments and survey of resident perception of feedback.

PARTICIPANTS:
Seventy internal medicine residents and their faculty evaluators at the University of Utah.

APPROACH:
Faculty narrative comments from 248 evaluations pre- and post-milestone implementation were analyzed for quality and Accreditation Council for Graduate Medical Education competency by area of strength and area for improvement. Seventy residents were surveyed regarding quality of feedback pre- and post-milestone implementation.

KEY RESULTS:
Qualitative analysis of narrative comments revealed nearly all evaluations pre- and post-milestone implementation included comments about areas of strength but were frequently vague and not related to competencies. Few evaluations included narrative comments on areas for improvement, but these were of higher quality compared to areas of strength (p < 0.001). Overall resident perception of quality of narrative comments was low and did not change following milestone implementation (p = 0.562) for the 86% of residents (N = 60/70) who completed the pre- and post-surveys.

CONCLUSIONS:
The quality of narrative comments was poor, and there was no evidence of improved quality following introduction of milestone-based evaluations. Comments on areas for improvement were of higher quality than areas of strength, suggesting an area for targeted intervention. Residents' perception of feedback quality did not change following implementation of milestone-based evaluations, suggesting that in the post-milestone era, internal medicine educators need to utilize additional interventions to improve quality of feedback.

Feedback with Performance Metric Scorecards Improves Resident Satisfaction but Does Not Impact Clinical Performance

OBJECTIVES:
The Emergency Medicine Milestone Project, a framework for assessing competencies, has been used as a method of providing focused resident feedback. However, the emergency medicine milestones do not include specific objective data about resident clinical efficiency and productivity, and studies have shown that milestone-based feedback does not improve resident satisfaction with the feedback process. We examined whether providing performance metric reports to resident physicians improves their satisfaction with the feedback process and their clinical performance.

METHODS:
We conducted a three-phase stepped-wedge randomized pilot study of emergency medicine residents at a single, urban academic site. In phase 1, all residents received traditional feedback; in phase 2, residents were randomized to receive traditional feedback (control group) or traditional feedback with performance metric reports (intervention group); and in phase 3, all residents received monthly performance metric reports and traditional feedback. To assess resident satisfaction with the feedback process, surveys using 6-point Likert scales were administered at each study phase and analyzed using two-sample t-tests. Analysis of variance in repeated measures was performed to compare impact of feedback on resident clinical performance, specifically patient treatment time (PTT) and patient visits per hour.

RESULTS:
Forty-one residents participated in the trial of which 21 were randomized to the intervention group and 20 in the control group. Ninety percent of residents liked receiving the report and 74% believed that it better prepared them for expectations of becoming an attending physician. Additionally, residents randomized to the intervention group reported higher satisfaction ($p = 0.01$) with the quality of the feedback compared to residents in the control group. However, receiving performance metric reports, regardless of study phase or postgraduate year status, did not affect clinical performance, specifically PTT (183 minutes vs. 177 minutes, $p = 0.34$) or patients visits per hour (0.99 vs. 1.04, $p = 0.46$).

CONCLUSIONS:
While feedback with performance metric reports did not improve resident clinical performance, resident physicians were more satisfied with the feedback process, and a majority of residents expressed liking the reports and felt that it better prepared them to become attending physicians. Residency training programs could consider augmenting feedback with performance metric reports to aide in the transition from resident to attending physician.

Effectiveness of the Adolescent Medicine Rotation in Improving Pediatric Residents Self- Assessed Skill and Confidence Caring for Youth

PURPOSE:
Practicing and resident pediatricians report inadequate skill in caring for adolescents, despite adolescents comprising roughly one-quarter of most general and subspecialty practices. This study examined the effectiveness of participation in an adolescent medicine rotation at improving pediatric residents' self-perceived skills and confidence across nine key adolescent health domains. We also evaluated the impact of didactic instruction during the rotation.

METHODS:
Resident and recent-graduate participants (n = 34) completed milestone-based self-assessment of their skill and confidence caring for adolescent patients in nine key adolescent health-related domains. This study employed a post-test then retrospective pretest, an educational study design used to minimize response-shift bias whereby participants rate their skill and confidence at the end of the intervention (post-test), and then reflect back to retrospectively rate their preintervention skill (retrospective pretest). Additionally, differences in gains between those who did and did not participate in didactic instruction were evaluated. Didactic instruction was delivered during the adolescent medicine rotation utilizing a flipped-classroom model; participants received standardized preparatory materials and participated in active-learning workshops.

RESULTS:
Participants demonstrated a significant (p ≤ .0001) increase in self-perceived skill levels for all assessed domains after the rotation as compared to before the rotation, whether or not they received didactic instruction. Participation in didactic instruction did not yield significant (p ≤ .05) additional benefit for any of the assessed domains.

CONCLUSIONS:
Participation in an adolescent medicine rotation is of value to pediatric resident trainees and leads to increased self-assessed skill and confidence in caring for youth.

The Use of Patient-Specific Three-Dimensional Printed Surgical Models Enhances Plastic Surgery Resident Education in Craniofacial Surgery

PURPOSE:
A significant challenge in surgical education is to provide a meaningful hands-on experience with the pathology the trainee will see in independent practice. Craniofacial anatomy is challenging and unfamiliar to the learner.

METHODS:
Using preoperative computed tomography data, the authors produced an accurately sized, three- dimensional (3D) printed model of the congenital craniofacial anatomy of patients treated by the same attending surgeon-PGY4 resident surgeon pair over the course of a 6-month rotation. A preoperative stepwise surgical plan was written by the attending and resident, and the plan was marked on the 3D model by the attending and resident separately. The written and marked plans were measured for accuracy and time to completion. The resident surgeon's applicable milestone levels were assessed.

RESULTS:
Seven congenital craniofacial anomalies met criteria for inclusion: 4 craniosynostosis cases, 2 mandibular distractions, and 1 LeFort I distraction. The number of inaccuracies of the written plan improved from 5 to 0 for sagittal synostosis and 4 to 0 for mandibular distraction. The time to complete the written plan decreased by 22% for sagittal synostosis and 45% for mandibular distraction. The number of inaccuracies of the marked plan decreased from 5 to 0 for sagittal synostosis and 2 to 0 for mandibular distraction. Time to completion of the marked plan decreased by 76% for sagittal synostosis and 50% for mandibular distraction. Milestone scores increased an average of 1.875 levels.

CONCLUSION:
Three-dimensional printed craniofacial models are a positive addition to resident training and have been objectively quantified to improve the accuracy and time to completion of the surgical plan as well as progression in the plastic surgery milestones.

Factors that Contribute to Resident Teaching Effectiveness

BACKGROUND:
One of the key components of residency training is to become an educator. Resident physicians teach students, advanced practice providers, nurses, and even faculty on a daily basis.

OBJECTIVE:
The goal of this study was to identify the objective characteristics of residents, which correlate with perceived overall teaching effectiveness.

METHODS:
We conducted a one-year, retrospective study to identify factors that were associated with higher resident teaching evaluations. Senior emergency medicine (EM) teaching residents are evaluated by medical students following clinical teaching shifts. Eighteen factors pertaining to resident teaching effectiveness were chosen. Two items from the medical students' evaluations were analyzed against each factor: teaching effectiveness was measured on a five-point Likert scale and an overall teaching score (1-75).

RESULTS:
A total of 46 EM residents and 843 medical student evaluations were analyzed. The ACGME milestones for systems-based practice (p = 0.02) and accountability (p = 0.05) showed a statistically significant association with a rating of "five" on the Likert scale for teaching effectiveness. Three other ACGME milestones, systems-based practice (p = 0.01), task switching (p = 0.04), and team management (p = 0.03) also showed a statically significant association of receiving a score of 70 or greater on the overall teaching score.

CONCLUSION:
Residents with higher performance associated with system management and accountability were perceived as highly effective teachers. USMLE and in-service exams were not predictive of higher teaching evaluations. Our data also suggest that effective teachers are working in both academic and community settings, providing a potential resource to academic departments and institutions.

Resident Case Volume Correlates with Clinical Performance: Finding the Sweet Spot

RATIONALE AND OBJECTIVES:
To determine whether the total number of studies interpreted during radiology residency correlates with clinical performance as measured by objective criteria.

MATERIALS AND METHODS:
We performed a retrospective cohort study of three graduating classes of radiology residents from a single residency program between the years 2015-2017. The total number of studies interpreted by each resident during residency was tracked. Clinical performance was determined by tracking an individual resident's major discordance rate. A major discordance was recorded when there was a difference between the preliminary resident interpretation and final attending interpretation that could immediately impact patient care. Accreditation council for graduate medical education milestones at the completion of residency, Diagnostic radiology in- training scores in the third year, and score from the American board of radiology core exam were also tabulated. Pearson correlation coefficients and polynomial regression analysis were used to identify correlations between the total number of interpreted films and clinical, test, and milestone performance.

RESULTS:
Thirty-seven residents interpreted a mean of 12,709 studies (range 8898-19,818; standard deviation [SD] 2351.9) in residency with a mean major discordance rate of 1.1% (range 0.34%- 2.54%; stand dev 0.49%). There was a nonlinear correlation between total number of interpreted films and performance. As the number of interpreted films increased to approximately 16,000, clinical performance (p = 0.004) and test performance (p = 0.01) improved, but volumes over 16,000 correlated with worse performance.

CONCLUSION:
The total number of studies interpreted during radiology training correlates with performance. Residencies should endeavor to find the "sweet spot": the amount of work that maximizes clinical exposure and knowledge without overburdening trainees.

Intern Self-Reported Preparedness for Residency: An ACGME Milestone-Based Study

McLean ME, Huls TA, Park JC, Anana MC, Chen AS, Chien GK, Cygan L, Gupta SJ, Husain A, Mishra DN, Ng KM, Russell JT, Surles RT, Kulkami ML. Western Journal of Emergency Medicine: Integrating Emergency Care with Population Health. 2019;20:S10.

OBJECTIVE:
Transitioning from medical school to internship is challenging. While several curricula for medical students and interns have been proposed during this transition period, there has not been a large- scale self assessment of incoming emergency medicine (EM) interns' preparedness for EM milestones. While many medical schools and EM residencies host "boot camps" or other intensive orientation programs for EM-bound students, having knowledge of incoming EM residents' self- perceived strengths and weaknesses will help clerkship directors and EM residency leadership better serve this group of learners. Milestones in EM are used in the United States to measure residents' progress and determine competence at residency completion.4 Incoming interns are expected to have achieved level 1 milestones by the time they enter residency, to have achieved level 2 milestones between the first and second year, and to have achieved level 4 milestones before completion of residency. We reached out to 151 newly- matched interns at 11 different sites to ask about their self- perceived "preparedness" for levels 1, 2, and 3 of our eight selected EM milestones (numbers 1, 3, 4,
5, 7, 9, 10, and 12).

DESIGN AND METHOD:
This was a prospective, cross-sectional study of 151 newly-matched pre-interns at 11 EM residency programs. We included all newly-matched interns at each program. Interns were invited via email from their programs to complete a voluntary, anonymous survey prior to the start of residency. The survey used a Likert scale (1 = very unprepared to 5 = extremely prepared) to assess self-reported preparedness to perform levels 1 and 2 of milestones 1, 3, 4, 5, 7, 9, 10, and 12. Milestones were chosen based on ease of teaching in an EM case curriculum that was later implemented.

RESULTS:
A total of 126 pre-interns completed the survey (response rate 83.4%).

CONCLUSION:
Subjects reported highest level of preparedness for emergency stabilization (PC1), and lowest levels of preparedness for airway management (PC10) and pharmacological management (PC5). The data suggest that teachers of fourth-year medical students and new EM interns may want to emphasize milestones 5 and 10 early in internship or late in medical school.

A Milestone-Based Pediatric Intern Boot Camp: An Educational Intervention to Minimize the July Effect

INTRODUCTION:
The transition from student to intern is difficult and highlighted by performance missteps often referred to as the July Effect. Some pediatric institutions have implemented intern boot camps (IBC) to better prepare interns at the start of residency. Such pediatric boot camps described in the literature have not specifically targeted the ACGME/ABP Pediatric Milestones. We implemented an IBC that utilized these milestones to improve the interns' confidence, knowledge, and skills. Methods 19 new interns participated in the IBC at the HDVCH/MSU Pediatric Residency Program. We used Kerns Six-Step Approach as a conceptual framework and targeted 3 levels of Kirkpatrick's level of evaluation (reaction, learning, and behavior). A needs assessment from residents and faculty was used to identify specific milestones. We designed our IBC to include lectures, workshops and clinical experiences to target these milestones. A questionnaire containing 15-confidence (Likert Scale 1-5) and 10 knowledge-based questions was given before and after the IBC. The paired t-test was used to assess total confidence scores and pre/post knowledge measures. The sign test was used to compare individual confidence questions. Block 1 milestone evaluations were analyzed for pre-IBC (2016, 2017) and post-IBC interns (2018). Significance was assessed at $p<0.05$.

RESULTS:
Interns demonstrated a significant improvement in their overall confidence score (Pre: 47.7+/-4.1, Post: 58.6+/-5.3; $p<0.001$). All individual confidence questions showed increases. Interns demonstrated a significant improvement in perceived pediatric knowledge on the post-IBC test (Pre: 5.2+/-1.5, Post: 6.8+/-1.3; $p=0.004$). Block 1 evaluations from 7/2018 did not show improved evaluations when compared to pre-IBC cohorts.

CONCLUSIONS:
Incoming interns demonstrated a significant improvement in confidence and perceived knowledge of the targeted pediatric milestones after participating in the IBC. Our innovative approach of targeting pediatric milestones in an IBC suggests that such a targeted curriculum helps the difficult transition for interns.

Moving toward Milestone-Based Assessment in Osteopathic Manipulative Medicine

ABSTRACT:
Osteopathic medicine is continuing to move toward competency-based education at undergraduate and graduate levels. Competencies and Entrustable Professional Activities (EPAs) have been implemented to guide educators on the skills and abilities that osteopathic medical students and residents should be able to perform as physicians. Unfortunately, many of these skills have not been well described, and the threshold of "competence" or "entrustability" for each of these tasks remains elusive. The author presents an approach to measuring competence in the domain of osteopathic manipulative medicine using a milestone rubric to assess skills related to osteopathic screening, diagnosis, technique, and explanation. This rubric can be applied to all levels of osteopathic training and across many diagnostic and treatment modalities. Clearly defining and assessing the individual skills composing competence in osteopathic manipulative medicine will be increasingly important as medical education continues to evolve and modernize.

Electronic Health Records as an Educational Tool: Viewpoint

BACKGROUND:
Electronic health records (EHRs) have been adopted by most hospitals and medical offices in the United States. Because of the rapidity of implementation, health care providers have not been able to leverage the full potential of the EHR for enhancing clinical care, learning, and teaching. Physicians are spending an average of 49% of their working hours on EHR documentation, chart review, and other indirect tasks related to patient care, which translates into less face time with patients.

OBJECTIVE:
The purpose of this article is to provide a preliminary framework to guide the use of EHRs in teaching and evaluation of residents.

METHODS:
First we discuss EHR educational capabilities that have not been reviewed in sufficient detail in the literature and expand our discussion for each educational activity with examples. We emphasize quality improvement of clinical notes as a basic foundational skill using a spreadsheet-based application as an assessment tool. Next, we integrate the six Accreditation Council for Graduate Medical Education (ACGME) Core Competencies and Milestones (CCMs) framework with the Reporter-Interpreter- Manager-Educator (RIME) model to expand our assessments of other areas of resident performance related to EHR use. Finally, we discuss how clinical utility, clinical outcome, and clinical reasoning skills can be assessed in the EHR.

RESULTS:
We describe a pilot conceptual framework-CCM framework-to guide and demonstrate the use of the EHR for education in a clinical setting.

CONCLUSIONS:
As EHRs and other supporting technologies evolve, medical educators should continue to look for new opportunities within the EHR for education. Our framework is flexible to allow adaptation and use in most training programs. Future research should assess the validity of such methods on trainees' education.

Competencies, Milestones, and a Level of Supervision Scale for Entrustable Professional Activities for Scholarship

Mink RB, Myers AL, Turner DA, Carraccio CL. Acad Med. 2018 Nov;93(11):1668- 1672. doi: 10.1097/ACM.0000000000002353.

PROBLEM:
Scholarship is an important element of both undergraduate and graduate medical education, and scholarly activity is required for all pediatric fellows. However, despite the creation of entrustable professional activities (EPA) for scholarship, the specific progressive levels of performance and the appropriate level of supervision for a given performance level have not been defined. The authors developed competencies and milestones for the scholarship EPA to provide a framework for assessment across the continuum; a level of supervision scale was also developed.

APPROACH:
The Vitae Researcher Development Framework served as a template to create the competencies and milestones for the scholarship EPA. Beginning in September 2015 and using a modified Delphi approach, three distinct drafts were circulated to individuals with expertise in various types of scholarship until broad agreement was achieved. Then, in October 2016, the Steering Committee of the Subspecialty Pediatrics Investigator Network created a level of supervision scale, modeled after one it had previously developed.

OUTCOMES:
Eight competencies were identified as important in making entrustment decisions related to scholarship. For each competency, five milestone levels that span the continuum from novice to expert were created. A supervision scale with five progressive levels of entrustment was also created.

NEXT STEPS:
Next steps include a study to obtain validity evidence for the supervision scale and determine the correlation between milestone and supervision levels. These competencies, milestones, and supervision levels can potentially serve as a roadmap for trainees and junior faculty and also play a role in the assessment of physician-scientists.

Radiology Resident Assessment and Feedback Dashboard

ABSTRACT:
Assessment of residents is optimally performed through processes and platforms that provide daily feedback, which can be immediately acted on. Given the documentation required by the Accreditation Council for Graduate Medical Education (ACGME), effective data management, integration, and presentation are crucial to ease the burden of manual documentation and increase the timeliness of actionable information. To this end, the authors modeled the learning activities of residents using the Experience Application Programming Interface (xAPI) framework, which is a standard framework for the learning community. On the basis of the xAPI framework and using open-source software to extend their existing infrastructure, the authors developed a Web-based dashboard that provides residents with a more holistic view of their educational experience. The dashboard was designed around the ACGME radiology milestones and provides real-time feedback to residents using various assessment metrics derived from multiple data sources. The purpose of this article is to describe the dashboard's architecture and components, the design and technical considerations, and the lessons learned in implementing the dashboard.

Analysis of Milestone-Based End-of-Rotation Evaluations for Ten Residents Completing a Three-Year Anesthesiology Residency

INTRODUCTION:
Faculty are required to assess the development of residents using educational milestones. This descriptive study examined the end-of-rotation milestone-based evaluations of anesthesiology residents by rotation faculty directors. The goals were to measure: (1) how many of the 25 Accreditation Council for Graduate Medical Education (ACGME) anesthesiology subcompetency milestones were included in each of the residency's rotations evaluations, (2) the percentage of evaluations sent to the rotation director that were actually completed by the director, (3) the length of time between the end of the residents' rotations and completion of the evaluations, (4) the frequency of straight line scoring, defined as the resident receiving the same milestone level score for all subcompetencies on the evaluation, and (5) how often a resident received a score below a Level 4 in at least one subcompetency in the three months prior to graduating.

METHODS:
In 2013, the directors for each the 24 anesthesia rotations in the Stanford University School of Medicine Anesthesiology Residency Program created new milestone-based evaluations to be used at the end of rotations to evaluate residents. The directors selected the subcompetencies from the list released by the ACGME that were most appropriate for their rotation. End-of-rotation evaluations for the post graduate year (PGY)-2 to PGY-4 from July 1, 2014 to June 30, 2017 were retrospectively analyzed for a sample of 10 residents randomly selected from 22 residents in the graduating class.

RESULTS:
The mean number of subcompetencies evaluated by each of the 24 rotations in the residency equaled 17.88 (standard deviation (SD): 3.39, range 10-24, median 18.5) from the available possible total of 25 subcompetencies. Three subcompetencies (medical knowledge, communication with patients and families, and coordination of patient care within the healthcare system) were included in the evaluation instruments of all 24 rotations. The three least frequently listed subcompetencies were: "acute, chronic, and cancer-related pain consultation/management" (25% of rotations had this on the end-of-rotation evaluation), "triage and management of critically ill patient in non- operative setting" (33%), and "education of patient, families, students, residents, and others" (38%). Overall, 418 end of rotation evaluations were issued and 341 (82%) completed, with 63% completed within one month, 22% between month one and two, and 15% after two months. The frequency of straight line scoring varied, from never occurring (0%) in three rotations to always occurring (100%) in two rotations, with an overall average of 51% (SD: 33%). Sixty-one percent of straight line scoring corresponded to the residents' postgraduate year whereby, for example, a post-graduate year two resident received an ACGME Level 2 proficiency for all subcompetencies. Thirty-one percent of the straight line scoring was higher than the resident's year of training (e.g., a PGY-2 received Level 3 or higher for all the subcompetencies). The remaining 7% of straight line scoring was below the expected level for the year of training. Three of seven residents had at least one subcompetency rated as below a Level 4 on one of the evaluations during the three months prior to finishing residency.

CONCLUSION:
Formal analysis of a residency program's end-of-rotation milestone evaluations may uncover opportunities to improve competency-based evaluations.

Resident-Specific Morbidity Reduced Following ACS NSQIP Data-Driven Quality Program

BACKGROUND:
The Accreditation Council for Graduate Medical Education Milestone Project for general surgery provided a more robust method for developing and tracking residents' competence. This framework enhanced systematic and progressive development of residents' competencies in surgical quality improvement.

STUDY DESIGN:
A 22-month interactive, educational program based on resident-specific surgical outcomes data culminated in a quality improvement project for postgraduate year 4 surgery residents. Self- assessment, quality knowledge test, and resident-specific American College of Surgeons National Surgical Quality Improvement Program Quality In-Training Initiative morbidity were compared before and after the intervention.

RESULTS:
Quality in-training initiative morbidity decreased from 25% (82/325) to 18% (93/517), p = 0.015 despite residents performing more complex cases. All participants achieved level 4 competency (4/4) within the general surgery milestones improvement of care, practice-based learning and improvement competency. Institutional American College of Surgeons National Surgical Quality Improvement Program general surgery morbidity improved from the ninth to the sixth decile. Quality assessment and improvement self-assessment postintervention scores (M = 23.80, SD = 4.97) were not significantly higher than preintervention scores (M = 19.20, SD = 5.26), p = 0.061. Quality Improvement Knowledge Application Tool postintervention test scores (M = 17.4, SD = 4.88), were not significantly higher than pretest scores (M = 13.2, SD = 1.92), p = 0.12.

CONCLUSION:
Sharing validated resident-specific clinical data with participants was associated with improved surgical outcomes. Participating fourth year surgical residents achieved the highest score, a level 4, in the practice based learning and improvement competency of the improvement of care practice domain and observed significantly reduced surgical morbidity for cases in which they participated.

Education Research: The Current State of Neurophysiology Education in Selected Neurology Residency Programs

OBJECTIVE:
Prior research has illustrated there is a knowledge gap in neurology residents' neurophysiology education (EEG and EMG), and we sought to understand whether this is still an issue and to recognize the barriers in order to create solutions and improve education.

METHODS:
Surveys were developed for adult neurology residents and one for program directors asking about confidence in neurophysiology knowledge, percent of graduates reaching level 4 ACGME (American Council of Graduate Medical Education) milestones in EEG and EMG, methods of learning used, interest in the subjects, and suggestions for improvements.

RESULTS:
Twenty-six program directors (19% responder rate) and 55 residents (from at least 16 different programs) completed the survey. Program directors thought that 85% of graduating residents met level 4 milestones in EEG and only 75% in EMG. Structured rotations and more time allocated to education of these topics were frequent barriers mentioned. Postgraduate year 4 residents were 60% and 67% confident in EEG and 64%, 59%, and 62.3% in EMG level 4 milestones. Learning to read EEGs was considered important throughout residents' training; however, this interest and value decreased over time with EMG.

CONCLUSION:
In our study, program directors suspect up to a quarter of residents may graduate not meeting level 4 ACGME milestones, and residents expressed lack of confidence in these areas. The educational methods used to instruct residents in EEG and EMG were similar as were the barriers they face across programs. This information hopefully will help fuel curriculum design and interest in these important neurology techniques.

Using an Alumni Survey to Target Improvements in an Emergency Medicine Training Programme

INTRODUCTION:
The Accreditation Council for Graduate Medical Education (ACGME) is the governing body responsible for accrediting graduate medical training programme in the USA. The Emergency Medicine Milestones (EM-Milestones) were developed by the ACGME and American Board of Emergency Medicine as a guide and monitoring tool for the knowledge, skills, abilities and experiences to be acquired during training. Alumni surveys have been reported as a valuable resource for training programme to identify areas for improvement; however, there are few studies regarding programme improvement in emergency medicine. We aimed to use the EM- Milestones, adapted as an alumni self-assessment survey, to identify areas for training programme improvement.

METHODS:
This study was conducted at an urban, academic affiliated, community hospital in New York city with an emergency medicine training programme consisting of 30 residents over 3 years. Alumni of our emergency medicine training programme were sent an EM-Milestones-based self- assessment survey. Participants evaluated their ability in each EM-Milestones subcompetency on a Likert scale. Data were analysed using descriptive statistics.

RESULTS:
Response rate was 74% (69/93). Alumni reported achieving the target performance in 5/6 general competencies, with Systems-Based Practice falling below the target performance. The survey further identified 6/23 subcompetencies (Pharmacotherapy, Ultrasound, Wound Management, Patient Safety, Systems-Based Management and Technology) falling below the target performance level.

DISCUSSION:
Alumni self-evaluation of competence using the EM-Milestones provides valuable information concerning confidence to practice independently; these data, coupled with regular milestone evaluation of existing trainees, can identify problem areas and provide a blueprint for targeted programme improvement.

Burnout is Associated with Emotional Intelligence but not Traditional Job Performance Measurements in Surgical Residents

OBJECTIVE:
To evaluate whether burnout was associated with emotional intelligence and job performance in surgical residents.

DESIGN:
General surgery residents at a single institution were surveyed using the Maslach Burnout Inventory (MBI) and trait EI questionnaire (TEIQ-SF). Burnout was defined as scoring in 2 of the 3 following domains; Emotional Exhaustion (high), Depersonalization (high), and Personal Accomplishment (low). Job performance was evaluated using faculty evaluations of clinical competency-based surgical milestones and standardized test scores including the American Board of Surgery In-Training Exam (ABSITE) and the United States Medical Licensing Examination (USMLE) Step 3. USMLE Step 1 and USMLE Step 2, which were taken prior to residency training, were included to examine possible associations of burnout with USMLE examinations. Statistical comparison was made using Pearson correlation and simple linear regression adjusting for PGY level.

SETTING:
This study was conducted at the University of Alabama at Birmingham (UAB) general surgery residency program.

PARTICIPANTS:
All current and incoming general surgery residents at UAB were invited to participate in this study.

RESULTS:
Forty residents participated in the survey (response rate 77%). Ten residents, evenly distributed from incoming residents to PGY-4, had burnout (25%). Mean global EI was lower in residents with burnout versus those without burnout (3.71 vs 3.9, p = 0.02). Of the 4 facets of EI, mean self-control values were lower in residents with burnout versus those without burnout (3.3 vs 4.06, p < 0.01). Each component of burnout was associated with global EI, with the strongest correlation being with personal accomplishment (r = 0.64; p < 0.01). Residents with burnout did not have significantly different mean scores for USMLE Step 1 (229 vs 237, p = 0.12), Step 2 (248 vs 251, p = 0.56), Step 3 (223 vs 222, p = 0.97), or ABSITE percentile (44.6 vs 58, p = 0.33) compared to residents without burnout. Personal accomplishment was associated with ABSITE percentile scores (r = 0.35; p = 0.049). None of the 16 surgical milestone scores were significantly associated with burnout.

CONCLUSIONS:
Burnout is present in surgery residents and associated with emotional intelligence. There was no association of burnout with USMLE scores, ABSITE percentile, or surgical milestones. Traditional methods of assessing resident performance may not be capturing burnout and strategies to reduce burnout should consider targeting emotional intelligence.

The Effect and Use of Milestones in the Assessment of Neurological Surgery Residents and Residency Programs

OBJECTIVES:
The purpose of this study was to determine the effect of the Accreditation Council for Graduate Medical Education Milestones on the assessment of neurological surgery residents. The authors sought to determine the feasibility, acceptability, and utility of this new framework in making judgments of progressive competence, its implementation within programs, and the influence on curricula. Residents were also surveyed to elicit the effect of Milestones on their educational experience and professional development.

DESIGN, SETTING, AND PARTICIPANTS:
In 2015, program leadership and residents from 21 neurological surgery residency programs participated in an online survey and telephone interview in which they reflected on their experiences with the Milestones. Survey data were analyzed using descriptive statistics. Interview transcripts were analyzed using grounded theory.

RESULTS:
Response themes were categorized into 2 groups: outcomes of the Milestones implementation process, and facilitators and barriers. Because of Milestones implementation, participants reported changes to the quality of the assessment process, including the ability to identify struggling residents earlier and design individualized improvement plans. Some programs revised their curricula based on training gaps identified using the Milestones. Barriers to implementation included limitations to the adoption of a developmental progression model in the context of rotation block schedules and misalignment between progression targets and clinical experience. The shift from time-based to competency-based evaluation presented an ongoing adjustment for many programs. Organized preparation before clinical competency committee meetings and diverse clinical competency committee composition led to more productive meetings and perceived improvement in promotion decisions.

CONCLUSIONS:
The results of this study can be used by program leadership to help guide further implementation of the Milestones and program improvement. These results also help to guide the evolution of Milestones language and their implementation across specialties.

Identifying Gaps in the Performance of Pediatric Trainees Who Receive Marginal/Unsatisfactory Ratings

PURPOSE:
To perform a derivation study to determine in which subcompetencies marginal/unsatisfactory pediatric residents had the greatest deficits compared with their satisfactorily performing peers and which subcompetencies best discriminated between marginal/unsatisfactory and satisfactorily performing residents.

METHOD:
Multi-institutional cohort study of all 21 milestones (rated on four or five levels) reported to the Accreditation Council for Graduate Medical Education, and global marginal/unsatisfactory versus satisfactory performance reported to the American Board of Pediatrics. Data were gathered in 2013- 2014. For each level of training (postgraduate year [PGY] 1, 2, and 3), mean differences between milestone levels of residents with marginal/unsatisfactory and satisfactory performance adjusted for clustering by program and C-statistics (area under receiver operating characteristic curve) were calculated. A Bonferroni-corrected significance threshold of .0007963 was used to account for multiple comparisons.

RESULTS:
Milestone and overall performance evaluations for 1,704 pediatric residents in 41 programs were obtained. For PGY1s, two subcompetencies had almost a one-point difference in milestone levels between marginal/unsatisfactory and satisfactory trainees and outstanding discrimination (≥0.90): organize/prioritize (0.93; C-statistic: 0.91) and transfer of care (0.97; C- statistic: 0.90). The largest difference between marginal/unsatisfactory and satisfactory PGY2s was trustworthiness (0.78). The largest differences between marginal/unsatisfactory and satisfactory PGY3s were ethical behavior (1.17), incorporating feedback (1.03), and professionalization (0.96). For PGY2s and PGY3s, no subcompetencies had outstanding discrimination.

CONCLUSIONS:
Marginal/unsatisfactory pediatric residents had different subcompetency gaps at different training levels. While PGY1s may have global deficits, senior residents may have different performance deficiencies requiring individualized counseling and targeted performance improvement plans.

Comprehensive Assessment of Struggling Learners Referred to a Graduate Medical Education Remediation Program

BACKGROUND:
Implementation of the Next Accreditation System has provided a standardized framework for identifying learners not meeting milestones, but there is as yet no corresponding framework for remediation.

OBJECTIVE:
We developed a comprehensive assessment process that allows correct diagnosis of a struggling learner's deficit(s) to promote successful remediation.

METHODS:
At the University of Pennsylvania, resident learners within the Department of Medicine who are not meeting milestones are referred to the Early Intervention Remediation Committee (EIRC). The EIRC, composed of 14 faculty members with expertise in remediation, uses a standardized process to assess learners' deficits. These faculty members categorize primary deficits as follows: medical knowledge, clinical reasoning, organization and efficiency, professionalism, and communication skills. The standardized process of assessment includes an analysis of the learner's file, direct communication with evaluators, an interview focused on learner perception of the problem, screening for underlying medical or psychosocial issues, and a review of systems for deficits in the 6 core competencies. Participants were surveyed after participating in this process.

RESULTS:
Over a 2-year period, the EIRC assessed and developed remediation plans for 4% of learners (14 of a total 342). Following remediation and reassessment, the identified problems were satisfactorily resolved in all cases with no disciplinary action. While the process was time intensive, an average of 45 hours per learner, the majority of faculty and residents rated it as positive and beneficial.

CONCLUSIONS:
This structured assessment process identifies targeted areas for remediation and adds to the tools available to Clinical Competency Committees.

Impact of an Event Reporting System on Resident Complication Reporting in Plastic Surgery Training: Addressing an ACGME and Plastic Surgery Milestone Project Core Competency

BACKGROUND:
The Accreditation Council for Graduate Medical Education and Plastic Surgery Milestone Project has identified practice-based learning and improvement, which involves systematically analyzing current practices and implementing changes, as a core competency in residency education. In surgical care, complication reporting is an essential component of practice-based learning and improvement as complications are analyzed in morbidity and mortality conference for quality improvement. Unfortunately, current methods for capturing a comprehensive profile of complications may significantly underestimate the true occurrence of complications. Therefore, the objectives of this study are to evaluate an intervention for complication reporting and compare this to current practice, in a plastic surgery training program.

METHODS:
This is a preintervention and postintervention study evaluating resident reporting of complications on a plastic surgery service. The intervention was an online event reporting system developed by department leadership and patient safety experts. The cohorts consisted of all patients undergoing surgery during two separate 3-month blocks bridged by an implementation period. A trained reviewer recorded complications, and this served as the reference standard. Fisher's exact test was used for binary comparisons.

RESULTS:
There were 32 complications detected in 219 patients from June to August of 2015 and 35 complications in 202 patients from October to December of 2015. The proportion of complications reported in the preintervention group was nine of 32 (28.1 percent). After the intervention, this significantly increased to 32 of 35 (91.4 percent) (p < 0.001).

CONCLUSION:
An intervention using an event reporting system, supported by departmental leadership, led to significant improvements in complication reporting by plastic surgery residents.

Using the Teaching Perspectives Inventory as an Introduction to a Residents-as-Teachers Curriculum

OBJECTIVE:
The Anesthesiology Milestone Project includes a milestone for assessing the teaching attributes of residents within the competency of Practice-based Learning and Improvement. We intend to develop a Residents-as-Teachers educational curriculum to assist our residents in successfully achieving this milestone. The goal of this study is to identify the specific teaching perspectives and intentions of our residents and to promote residents' comprehension of their own teaching philosophy.

METHODS:
We invited our residents to complete the Teaching Perspective Inventory (TPI) and a follow-up survey to gather information regarding dominant and recessive teaching perspectives, their intended career pathway, and their view of the importance of understanding teaching perspectives.

RESULTS:
The two most common dominant teaching perspectives are apprenticeship and nurturing for residents who are planning a career in both academic medicine and private practice. A greater percentage of residents planning an academic career agree that identifying their teaching perspective is beneficial to their role as a clinical educator, compared to those anticipating a career in private practice.

CONCLUSIONS:
Based on this pilot data, our Residents-as-Teachers curriculum will include instruction of educational strategies specifically designed towards the apprenticeship and nurturing perspectives.

Towards Consensus: Training in Procedural Skills for Diagnostic Radiology Residents – Current Opinions of Residents and Faculty at a Large Academic Center

RATIONALE AND OBJECTIVES:
The Diagnostic Radiology Milestones Project provides a framework for measuring resident competence in radiologic procedures, but there are limited data available to assist in developing these guidelines.
We performed a survey of current radiology residents and faculty at our institution as a first step toward obtaining data for this purpose. The survey addressed attitudes toward procedural standardization and procedures that trainees should be competent by the end of residency.

MATERIALS AND METHODS:
Current residents and faculty members were surveyed about whether or not there should be standardization of procedural training, in which procedures residents should achieve competency, and the number of times a procedure needs to be performed to achieve competency.

RESULTS:
Survey data were received from 60 study participants with an overall response rate of 32%. Sixty-five percent of respondents thought that procedural training should be standardized. Standardization of procedural training would include both the list of procedures that trainees should be competent in at the end of residency and the standard minimum number of procedures to achieve competency. Procedures that both residents and faculty agreed are important in which to achieve competency included central line/port procedures; CT-guided abdominal, thoracic, and musculoskeletal procedures; minor fluoroscopic-guided procedures; general fluoroscopy; peripheral line placements; and US-guided abdominal procedures. For most of these categories, most respondents believed that these procedures needed to be performed 6-20 times to achieve competency.

CONCLUSION:
Both resident and faculty respondents agreed that procedural training should be standardized during residency, and competence in specific procedures should be achieved at the completion of residency. Although this study is limited to a single institution, our data may provide assistance in developing future guidelines for standardizing image-guided procedure training. Future studies could be expanded to create a national consensus regarding the implementation of the Diagnostic Radiology Milestones Project.

How Do Clinical Competency Committees Use Different Sources of Data to Assess Residents' Performance on the Internal Medicine Milestones? A Mixed Methods Pilot Study

PURPOSE:
This study examines how Clinical Competency Committees (CCCs) synthesize assessment data to make judgments about residents' clinical performances.

METHODS:
Between 2014 and 2015, after four six-month reporting periods to the Accreditation Council for Graduate Medical Education (ACGME), 7 of 16 CCC faculty at Rush University Medical Center completed questionnaires focused on their perspectives about rating residents on their achievement of the milestones and participated in a focus group. Qualitative data were analyzed using grounded theory. Milestones ratings for two six-month ACGME reporting cycles (n = 100 categorical residents) were also analyzed.

RESULTS:
CCC members weighted resident rotation ratings highest (weight = 37%), followed by faculty rotation comments (weight = 27%) and personal experience with residents (weight = 14%) for making judgments about learner's milestone levels. Three assessment issues were identified from qualitative analyses: (1) "design issues" (e.g. problems with available data or lack thereof); (2) "synthesis issues" (e.g. factors influencing ratings and decision-making processes) and (3) "impact issues" (e.g., how CCC generated milestones ratings are used).

CONCLUSIONS:
Identifying factors that affect assessment at all stages of the CCC process can contribute to improving assessment systems, including support for faculty development for CCCs. Recognizing challenges in synthesizing first and second-hand assessment data is an important step in understanding the CCC decision-making process.

Mindfulness, Burnout, and Effects on Performance Evaluations in Internal Medicine Residents

PURPOSE:
Burnout has been documented at high levels in medical residents with negative effects on performance. Some dispositional qualities, like mindfulness, may protect against burnout. The purpose of the present study was to assess burnout prevalence among internal medicine residents at a single institution, examine the relationship between mindfulness and burnout, and provide preliminary findings on the relation between burnout and performance evaluations in internal medicine residents.

METHODS:
Residents (n = 38) completed validated measures of burnout at three time points separated by 2 months and a validated measure of dispositional mindfulness at baseline. Program director end- of-year performance evaluations were also obtained on 22 milestones used to evaluate internal medicine resident performance; notably, these milestones have not yet been validated for research purposes; therefore, the investigation here is exploratory.

RESULTS:
Overall, 71.1% (n = 27) of the residents met criteria for burnout during the study. Lower scores on the "acting with awareness" facet of dispositional mindfulness significantly predicted meeting burnout criteria $\chi^2(5) = 11.88$, $p = 0.04$. Lastly, meeting burnout criteria significantly predicted performance on three of the performance milestones, with positive effects on milestones from the "system-based practices" and "professionalism" domains and negative effects on a milestone from the "patient care" domain.

CONCLUSION:
Burnout rates were high in this sample of internal medicine residents and rates were consistent with other reports of burnout during medical residency. Dispositional mindfulness was supported as a protective factor against burnout. Importantly, results from the exploratory investigation of the relationship between burnout and resident evaluations suggested that burnout may improve performance on some domains of resident evaluations while compromising performance on other domains. Implications and directions for future research are discussed.

Critical Deficiency Ratings in Milestone Assessment: A Review and Case Study

PURPOSE:
The Accreditation Council for Graduate Medical Education (ACGME) requires programs to report learner progress using specialty-specific milestones. It is unclear how milestones can best identify critical deficiencies (CDs) in trainee performance. Specialties developed milestones independently of one another; not every specialty included CDs within milestones ratings. This study examined the proportion of ACGME milestone sets that include CD ratings, and describes one residency program's experiences using CD ratings in assessment.

METHOD:
The authors reviewed ACGME milestones for all 99 specialties in November 2015, determining which rating scales contained CDs. The authors also reviewed three years of data (July 2012-June 2015) from the University of Cincinnati Medical Center (UCMC) internal medicine residency assessment system based on observable practice activities mapped to ACGME milestones. Data were analyzed by postgraduate year, assessor type, rotation, academic year, and core competency. The Mantel-Haenszel chi-square test was used to test for changes over time.

RESULTS:
Specialties demonstrated heterogeneity in accounting for CDs in ACGME milestones, with 22% (22/99) of specialties having no language describing CDs in milestones assessment. Thirty-three percent (63/189) of UCMC internal medicine residents received at least one CD rating, with CDs accounting for 0.18% (668/364,728) of all assessment ratings. The authors identified CDs across multiple core competencies and rotations.

CONCLUSIONS:
Despite some specialties not accounting for CDs in milestone assessment, UCMC's experience demonstrates that a significant proportion of residents may be rated as having a CD during training. Identification of CDs may allow programs to develop remediation and improvement plans.

Evaluating Surgical Residents Quickly and Easily Against the Milestones Using Electronic Formative Feedback

Hartranft TH, Yandle K, Graham T, Holden C, Chambers LW. J Surg Educ. 2017 Mar- Apr;74(2):237-242. doi: 10.1016/j.jsurg.2016.09.006. Epub 2016 Oct 13.

OBJECTIVE:
This study was conducted to assess the effectiveness of a newly implemented electronic web- based review system created at our institution for evaluating resident performance relative to established milestones.

DESIGN:
Retrospective review of data collected from a survey of general surgery faculty and residents.

SETTING:
Tertiary care teaching hospital system and independent academic medical center.

PARTICIPANTS:
A total of 12 general surgery faculty and 17 general surgery residents participated in this study. The survey queried the level of satisfaction before and after the adoption of QuickNotes using several statements scored on a 5-point scale, with 1 being the lowest rating as "not satisfied," and 5 being the highest rating as "completely satisfied."

RESULTS:
The weighted average improvements from pre- to post-QuickNotes implementation for the faculty responding to the survey ranged from 10% to 40%; weighted average improvements for the residents responding to the survey ranged from 5% to 73%. For the survey of faculty, both sets of weighted averages tended to be higher than the weighted average for the resident's survey responses. The highest rated topic was the faculty's level of satisfaction with the "frequency to provide feedback" with a post-QuickNotes implementation weighted average of 4.25, closely followed by the residents' level of satisfaction with the "evaluation includes positive feedback" with a post-QuickNotes implementation weighted average of 4.24. The most notable increases in weighted averages from preimplementation to postimplementation were noted for "overall satisfaction" (20% increase for faculty, 37% for residents), "reflects actual criteria that matter" (36% increase for faculty, 73% for residents), faculty "opportunity for follow- up" (increase of 40%), resident "reflects overall trends" (increase of 37%), and resident "provides new information about my performance" (increase of 37%).

CONCLUSIONS:
Our institutional adoption of QuickNotes into the resident evaluation process has been associated with an overall increased level of satisfaction in the evaluation process by both faculty and residents. The design of QuickNotes facilitates its integration into the resident training environment, as it is web based, easy to use, and has no additional cost over the standard New Innovations subscription. Although it is designed to capture snapshots of trainee behavior and performance, monthly reports through QuickNotes can be used effectively in conjunction with the more traditional end-of-rotation evaluations to show trends, identify areas of strength that should be reinforced, demonstrate areas needing improvement, allow for a more tailored individual education plan to be developed, and permit a more accurate determination of milestone progression.

Competent for Unsupervised Practice: Use of Pediatric Residency Training Milestones to Assess Readiness

Li ST, Tancredi DJ, Schwartz A, Guillot AP, Burke AE, Trimm RF, Guralnick S, Mahan JD, Gifford KA; Association of Pediatric Program Directors (APPD) Longitudinal Educational Assessment Research Network (LEARN) Validity of Resident Self-Assessment Group. Acad Med. 2017 Mar;92(3):385-393. doi: 10.1097/ACM.0000000000001322.

PURPOSE:
To describe clinical skills progression during pediatric residency using the distribution of pediatric milestone assessments by subcompetency and year of training and to determine reasonable milestone expectations at time of graduation.

METHOD:
Multi-institutional cohort study of the milestones reported to the Accreditation Council for Graduate Medical Education for all 21 pediatric subcompetencies. Most subcompetencies were measured using five milestone levels (1 = novice, 2 = advanced beginner, 3 = competent, 4 = proficient, 5 = master); 3 subcompetencies had only four levels defined.

RESULTS:
Milestone assessments for 2,030 pediatric residents in 47 programs during academic year 2013-2014 were obtained. There was significant variation in end-of-year milestone ratings for residents within each level of training, which decreased as training level increased. Most (78.9%; 434/550) graduating third-year pediatric residents received a milestone rating of ≥ 3 in all 21 subcompetencies; fewer (21.1%; 116/550) received a rating of ≥ 4 in all subcompetencies. Across all training levels, professionalism and interpersonal communication skills were rated highest; quality improvement was rated lowest.

CONCLUSIONS:
Trainees entered residency with a wide range of skills. As they advanced, skill variability within a training level decreased. Most graduating pediatric residents were still advancing on the milestone continuum toward proficiency and mastery, and an expectation of milestone ratings ≥ 4 in all categories upon graduation is unrealistic; milestone ratings ≥ 3 upon graduation may be more realistic. Understanding current pediatric residents' and graduates' skills can help to identify key areas that should be specifically targeted during training.

Milestones: A Rapid Assessment Method for the Clinical Competency Committee

INTRODUCTION:
Educational milestones are now used to assess the developmental progress of all U.S. graduate medical residents during training. Twice annually, each program's Clinical Competency Committee (CCC) makes these determinations and reports its findings to the Accreditation Council for Graduate Medical Education (ACGME). The ideal way to conduct the CCC is not known. After finding that deliberations reliant upon the new milestones were time intensive, our internal medicine residency program tested an approach designed to produce rapid but accurate assessments.

MATERIAL AND METHODS:
For this study, we modified our usual CCC process to include pre-meeting faculty ratings of resident milestones progress with in-meeting reconciliation of their ratings. Data were considered largely via standard report and presented in a pre-arranged pattern. Participants were surveyed regarding their perceptions of data management strategies and use of milestones. Reliability of competence assessments was estimated by comparing pre-/post-intervention class rank lists produced by individual committee members with a *master* class rank list produced by the collective CCC after full deliberation.

RESULTS:
Use of the study CCC approach reduced committee deliberation time from 25 min to 9 min per resident ($p < 0.001$). Committee members believed milestones improved their ability to identify and assess expected elements of competency development ($p = 0.026$). Individual committee member assessments of trainee progress agreed well with collective CCC assessments.

CONCLUSIONS:
Modification of the clinical competency process to include pre-meeting competence ratings with in-meeting reconciliation of these ratings led to shorter deliberation times, improved evaluator satisfaction and resulted in reliable milestone assessments.

Do Attending Surgeons and Residents See Eye to Eye? An Evaluation of the Accreditation Council for Graduate Medical Education Milestones in General Surgery Residency

OBJECTIVE:
The Accreditation Council for Graduate Medical Education requires accredited general surgery residencies to implement competency-based developmental outcomes in resident evaluations. Overall, 16 milestones are evaluated by a clinical competency committee (CCC). The milestones span 8 domains of surgical practice and 6 Accreditation Council for Graduate Medical Education clinical competencies. The highest level suggests preparedness for independent practice. Our objective was to compare self-assessments and committee evaluations within the milestone framework.

STUDY DESIGN:
All residents underwent semiannual evaluations from 2013 to 2015. Residents independently completed a self-assessment using the milestones. The CCC completed the milestones document using resident evaluations and consensus opinion of committee members. Assessment differences were calculated for each evaluation. A negative value indicated that the residents evaluated themselves at a lower level than the committee. Major assessment disparities were defined as >0.5 on a 4-point scale.

SETTING:
An independent academic medical center.

PARTICIPANTS:
General surgery residents.

RESULTS:
Overall, 20 residents participated; 7 were female. In total, 5 (7%) evaluations had a mean overall assessment difference >0.5, whereas 6 (8%) had a difference <-0.5. Residents evaluated themselves lower than the committee with a median assessment difference of -0.06 [-0.25 to 0.16] (p = 0.041). Evaluations were similar across surgical domains. Negative self-evaluations were more common for medical knowledge (-0.25 [-0.25 to 0.25], p = 0.025). Female residents had 2% positive and 13% negative major assessment disparity rates versus 10% positive and 9% negative rates among male residents. Postgraduate year III residents had 12% positive and 4% negative major disparity rates; all other years had higher negative than positive rates.

CONCLUSIONS:
Surgery residents within our program demonstrated adequate self-awareness, with most self- evaluations falling within a half level of the CCC report. This self-awareness was consistent across surgical domains and most clinical competencies. Residents perceived a lower level of medical knowledge than the CCC. Subgroup analysis revealed interesting trends in the effects of sex, postgraduate year level, and academic year timing, which will take additional study to fully delineate.

Using Milestones as Evaluation Metrics during an Emergency Medicine Clerkship

Quinn SM, Worrilow CC, Jayant DA, Bailey B, Eustice E, Kohlhepp J, Rogers R, Kane BG. J Emerg Med. 2016 Oct;51(4):426-431.

BACKGROUND:
The Accreditation Council for Graduate Medical Education's (ACGME) Milestones presumes graduating medical students will enter residency proficient at Milestone level 1 for 23 skills. The Next Accreditation System now includes Milestones for each postgraduate specialty, and it is unlikely that schools will document every emergency medicine (EM) applicant's EM-specific skills in their performance evaluation.

OBJECTIVES:
The goals of this research were to determine if assessment of the Milestones was feasible during a medical student clerkship and examine the proportion of medical students performing at Milestone Level 1.

METHODS:
This study was conducted at a center with Liaison Committee on Medical Education-approved medical training and a 4-year EM residency. Using traditional clerkship, we studied the feasibility of an ACGME EM Milestones-based clerkship assessment. Data led to redesign of the clerkship and its evaluation process, including all level 1 anchor(s) to add "occasionally" (>60%), "usually" (>80%), and "always" (100%) on a Likert scale to on-shift assessment forms.

RESULTS:
During the feasibility phase (2013-14), 75 students rotated though the clerkship; 55 evaluations were issued and 50 contained the Milestone summary. Eight deficiencies were noted in Milestone 12 and three in Milestone 14. After changes, 49 students rotated under the new evaluation rubric. Of 575 completed on-shift evaluations, 16 Milestone deficiencies were noted. Of 41 institutional evaluations issued, only one student had deficiencies noted, all of which pertained to patient care. All evaluations in this second cohort contained each student's Milestone proficiency.

CONCLUSIONS:
Assessment of the Milestones is feasible. Communication of ACGME EM Milestone proficiency may identify students who require early observation or remediation. The majority of students meet the anchors for the Milestones, suggesting that clerkship assessment with the ACGME EM Milestones does not adequately differentiate students.

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

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.

A Survey of Ultrasound Milestone Incorporation into Emergency Medicine Training Programs

OBJECTIVES:
With the introduction of the Emergency Medicine Milestone Project in 2013, residencies now assess emergency ultrasound (US) skills at regular intervals. However, it is unclear how programs are implementing the emergency US milestones and assessing competency. With the use of the milestone tool, a survey was distributed to emergency US educators to determine when programs are providing emergency US education, when residents are expected to attain competency, and whether the milestones reflect their expectations of trainees.

METHODS:
We conducted a prospective cross-sectional survey study distributed electronically to designated emergency US experts at 169 programs. Participants were queried on education and competency evaluation within the context of the milestones by designating a postgraduate year when the 5 milestone levels were taught and competency was expected. Survey findings were reported as percentages of total respondents from descriptive statistics.

RESULTS:
Responses were received from 53% of programs, and 99% were familiar with the milestones. Most programs provide level 1 (88%) and 2 (85%) instruction during postgraduate year 1. Most programs expect level 1 competency before residency (61%) and expect mastery of level 2 by the end of postgraduate year 1 (60%). Sixty-two percent believe the milestones do not accurately reflect their expectations, citing insufficient minimum scan numbers, lack of specificity, and unattainable level 5 requirements.

CONCLUSIONS:
There is substantial variability in the frequency and methods of competency evaluation using the emergency US milestones. However, most responders agree that residents should obtain level 2 competency by postgraduate year 1. Variation exists regarding what year and what skills define level 3 or greater competency.

The Problem Resident Behavior Guide: Strategies for Remediation

ABSTRACT:
In 2012, the ACGME supplemented the core competencies with outcomes-based milestones for resident performance within the six competency domains. These milestones address the knowledge, skills, abilities, attitudes, and experiences that a resident is expected to progress through during the course of training. Even prior to the initiation of the milestones, there was a paucity of EM literature addressing the remediation of problem resident behaviors and there remain few readily accessible tools to aid in the implementation of a remediation plan. The goal of the "Problem Resident Behavior Guide" is to provide specific strategies for resident remediation based on deficiencies identified within the framework of the EM milestones. The "Problem Resident Behavior Guide" is a written instructional manual that provides concrete examples of remediation strategies to address specific milestone deficiencies. The more than 200 strategies stem from the experiences of the authors who have professional experience at three different academic hospitals and emergency medicine residency programs, supplemented by recommendations from educational leaders as well as utilization of valuable education adjuncts, such as focused simulation exercises, lecture preparation, and themed ED shifts. Most recommendations require active participation by the resident with guidance by faculty to achieve the remediation expectations. The ACGME outcomes-based milestones aid in the identification of deficiencies with regards to resident performance without providing recommendations on remediation. The Problem Resident Behavior Guide can therefore have a significant impact by filling in this gap.

Driving Care Quality: Aligning Trainee Assessment and Supervision through Practical Application of Entrustable Professional Activities, Competencies, and Milestones

ABSTRACT:
To address the long-standing challenge of meaningful trainee assessment, the authors reviewed and expanded on the Accountable Assessment for Quality Care and Supervision (AAQCS) equation. The equation proposes that care quality is the product of the interaction between trainee performance (measured by workplace assessment) and supervision (required level of intervention to ensure care quality) in the context of the environment where the care occurs: Trainee performance × Appropriate supervision = Safe, effective patient-centered care. Assessing trainee performance and matching that performance to "appropriate" supervision, however, is fraught with challenges. The authors suggest a unifying framework that integrates entrustable professional activities (EPAs), competencies, and milestones to inform trainee assessment and supervision, thereby enabling the practical application of the AAQCS equation in the workplace. Because the unit of measure for an EPA is the outcome of whether the trainee can safely and effectively perform the professional activity without supervision, the proposed unifying framework directly aligns with the dependent variable in the AAQCS equation: care quality. The value of applying a unifying framework that integrates EPAs, competencies, and milestones to the AAQCS equation in the clinical learning environment lies in its ability to provide supervisors with a shared mental model of performance expectations for trainees, reducing unwanted variability and improving assessment accuracy; guidance for aligning performance milestones of trainees with the needed level of supervisor intervention to ensure care quality; and substrate for specific feedback to improve the trainee's professional development as a way to ensure future care quality.

A Milestone-Based Evaluation System-The Cure for Grade Inflation?

Kuo LE, Hoffman RL, Morris JB, Williams NN, Malachesky M, Huth LE, Kelz RR. J Surg Educ. 2015 Nov-Dec;72(6):e218-25.

PURPOSE:
Controversy exists over the optimal use of the Milestones in the process of resident Evaluation and feedback. We sought to evaluate the performance of a Milestones-based feedback system in comparison to a traditional model.

METHODS:
The traditional evaluation system (TES) consisted of a generic 16-item survey using a 5-point Likert scale ranging from 1 to 5, and a free-text comments section. The Milestones-based evaluation system (MBES) was launched in July 2014, ranging from 0 to 4. Individual milestones were mapped to rotations based on resident educational goals by postgraduate year (PGY). The MBES consisted of a survey with a maximum of 7 items, followed by a free-text comment section. Within each evaluation system, an overall composite score was calculated for each categorical general surgical resident. To scale the 2 systems for comparison, TES scores were adjusted downward by 1 point. Descriptive statistics were performed. Univariate analysis was performed with the Wilcoxon signed-rank test. A test for trend across PGY was used for the MBES only.

RESULTS:
In the traditional system, the median score was 3.66 (range: 3.2-4.0). There was no meaningful difference in the median score by PGY. In the new system, the median score was 2.69 (range: 1.5-3.7, p < 0.01). The median score differed across PGY and increased by PGY of training (p < 0.01). There was an increase in differences between median scores by PGY.

CONCLUSIONS:
On using the milestones to facilitate faculty evaluation of resident knowledge and skill, there was a trend in increasing score by PGY of training. In the MBES, scores could be used to better discriminate resident skill and knowledge levels and resulted in improved differentiation in scoring by PGY. The use of the milestones as a basis for evaluation enabled the program to provide more meaningful feedback to residents and represents an improvement in surgical education.

Emergency Medicine Residents Consistently Rate Themselves Higher than Attending Assessments on ACGME Milestones

INTRODUCTION:
In 2012 the Accreditation Council for Graduate Medical Education (ACGME) introduced the Next Accreditation System (NAS), which implemented milestones to assess the competency of residents and fellows. While attending evaluation and feedback is crucial for resident development, perhaps equally important is a resident's self-assessment. If a resident does not accurately self-assess, clinical and professional progress may be compromised. The objective of our study was to compare emergency medicine (EM) resident milestone evaluation by EM faculty with the same resident's self-assessment.

METHODS:
This is an observational, cross-sectional study that was performed at an academic, four-year EM residency program. Twenty-five randomly chosen residents completed milestone self- assessment using eight ACGME sub-competencies deemed by residency leadership as representative of core EM principles. These residents were also evaluated by 20 faculty members. The milestone levels were evaluated on a nine-point scale. We calculated the average difference between resident self-ratings and faculty ratings, and used sample t-tests to determine statistical significance of the difference in scores.

RESULTS:
Eighteen residents evaluated themselves. Each resident was assessed by an average of 16 attendings (min=10, max=20). Residents gave themselves statistically significant higher milestone ratings than attendings did for each sub-competency examined (p<0.0001).

CONCLUSION:
Residents over-estimated their abilities in every sub-competency assessed. This underscores the importance of feedback and assessment transparency. More attention needs to be paid to methods by which residency leadership can make residents' self-perception of their clinical ability more congruent with that of their teachers and evaluators. The major limitation of our study is small sample size of both residents and attendings.

Development and Evaluation of Standardized Narrative Cases Depicting the General Surgery Professionalism Milestones

PURPOSE:
Residency programs now are required to use educational milestones, which has led to the need for new methods of assessment. The literature suggests that narrative cases are a promising tool to track residents' progress. This study demonstrates the process for developing and evaluating narrative cases representing the five levels of the professionalism milestones.

METHOD:
In 2013, the authors identified 28 behaviors in the Accreditation Council for Graduate Medical Education general surgery professionalism milestones. They modified previously published narrative cases to fit these behaviors. To evaluate the quality of these cases, the authors developed a 28-item, five-point scale instrument, which 29 interdisciplinary faculty completed. The authors compared the faculty ratings by narrative case and specialty with the authors' initial rankings of the cases by milestone level. They used t tests and analysis of variance to compare mean scores across specialties.

RESULTS:
The authors developed 10 narrative cases, 2 for each of the 5 milestone levels. Each case contained at least 20 of the 28 behaviors identified in the milestones. Mean faculty ratings matched the milestone levels. Reliability was good (G coefficient = 0.86, phi coefficient = 0.85), indicating consistency in raters' ability to determine the proper milestone level for each case.

CONCLUSIONS:
The authors demonstrate a process for using specialty-specific milestones to develop narrative cases that map to a spectrum of professionalism behaviors. This process can be applied to other competencies and specialties to facilitate faculty awareness of resident performance descriptors and provide a frame of reference for milestones assessment.

Development and Validation of an Assessment of Regional Anesthesia Ultrasound Interpretation Skills

Woodworth GE, Carney PA, Cohen JM, Kopp SL, Vokach-Brodsky LE, Horn JL, Missair A, Banks SE, Dieckmann NF, Maniker RB.

BACKGROUND:
Interpretation of ultrasound images and knowledge of anatomy are essential skills for ultrasound-guided peripheral nerve blocks. Competency-based educational models promoted by the Accreditation Council for Graduate Medical Education require the development of assessment tools for the achievement of different competency milestones to demonstrate the longitudinal development of skills that occur during training.

METHODS:
A rigorous study guided by psychometric principles was undertaken to identify and validate the domains and items in an assessment of ultrasound interpretation skills for regional anesthesia. A survey of residents, academic faculty, and community anesthesiologists, as well as video recordings of experts teaching ultrasound-guided peripheral nerve blocks, was used to develop short video clips with accompanying multiple choice-style questions. Four rounds of pilot testing produced a 50-question assessment that was subsequently administered online to residents, fellows, and faculty from multiple institutions.

RESULTS:
Test results from 90 participants were analyzed with Item Response Theory model fitting indicating that a 47-item subset of the test fits the model well (P = 0.11). There was a significant linear relation between expected and predicted item difficulty (P < 0.001). Overall test scores increased linearly with higher levels of formal anesthesia training, regional anesthesia training, number of ultrasound-guided blocks performed per year, and a self-rating of regional anesthesia skill (all P < 0.001).

CONCLUSIONS:
This study provides evidence for the reliability, content validity, and construct validity of a 47- item multiple choice-style online test of ultrasound interpretation skills for regional anesthesia, which can be used as an assessment of competency milestone achievement in anesthesiology training.

Use of Emergency Medicine Milestones as Items on End-of-Shift Evaluations Results in Overestimates of Residents' Proficiency Level

BACKGROUND:
The emergency medicine milestones were developed to provide more objective resident assessment than current methods. However, little is known about the best practices for applying the milestones in resident assessment.

OBJECTIVE:
We examined the utility of end-of-shift evaluations (ESEs) constructed using the milestones in resident assessment.

METHODS:
We developed 14 daily ESEs, each of which included 9 or 10 emergency medicine milestones. Postgraduate year (PGY)-1 and PGY-2 residents were assessed on milestone levels 1 through 3; PGY-3 and PGY-4 residents were assessed on levels 3 through 5. Each milestone was rated on a nominal scale (yes, no, or not applicable). The Clinical Competency Committee combined the ESE data with data from other assessments to determine each resident's proficiency level for the emergency medicine subcompetencies. We used descriptive statistics to summarize resident ESEs and milestone levels. We analyzed differences in ESE score across PGY levels using t tests and analyses of variance.

RESULTS:
Faculty completed 763 ESEs on 33 residents with a range of 2 to 54 (median=22) ESEs per resident. Faculty rarely (8%, 372 of 4633) rated a resident as not achieving a milestone on the ESEs. Analyses of variance revealed that ESE scores on level 3 milestones did not differ significantly by PGY level. There was poor agreement between ESE scores and Clinical Competency Committee ratings.

CONCLUSIONS:
The ESEs constructed using the milestones resulted in grade or milestone inflation. Our results do not support using milestones as a stand-alone assessment tool.

Realizing the Promise of Competency-Based Medical Education

ABSTRACT
Competency-based medical education (CBME) places a premium on both educational and clinical outcomes. The Milestones component of the Next Accreditation System represents a fundamental change in medical education in the United States and is part of the drive to realize the full promise of CBME. The Milestones framework provides a descriptive blueprint in each specialty to guide curriculum development and assessment practices. From the beginning of the Outcomes project in 1999, the Accreditation Council for Graduate Medical Education and the larger medical education community recognized the importance of improving their approach to assessment. Work-based assessments, which rely heavily on the observations and judgments of clinical faculty, are central to a competency-based approach. The direct observation of learners and the provision of robust feedback have always been recognized as critical components of medical education, but CBME systems further elevate their importance. Without effective and frequent direct observation, coaching, and feedback, the full potential of CBME and the Milestones cannot be achieved. Furthermore, simply using the Milestones as end-of-rotation evaluations to "check the box" to meet requirements undermines the intent of an outcomes-based accreditation system. In this Commentary, the author explores these challenges, addressing the concerns raised by Williams and colleagues in their Commentary. Meeting the assessment challenges of the Milestones will require a renewed commitment from institutions to meet the profession's "special obligations" to patients and learners. All stakeholders in graduate medical education must commit to a professional system of self- regulation to prepare highly competent physicians to fulfill this social contract.

Milestone-Based Assessments Are Superior to Likert-Type Assessments in Illustrating Trainee Progression

BACKGROUND:
The Pediatrics Milestone Project uses behavioral anchors, narrative descriptions of observable behaviors, to describe learner progression through the Accreditation Council for Graduate Medical Education competencies. Starting June 2014, pediatrics programs were required to submit milestone reports for their trainees semiannually. Likert-type scale assessment tools were not designed to inform milestone reporting, creating a challenge for Clinical Competency Committees.

OBJECTIVE:
To determine if milestone-based assessments better stratify trainees by training level compared to Likert-type assessments.

METHODS:
We compared assessment results for 3 subcompetencies after changing from a 5-point Likert scale to milestone-based behavioral anchors in July 2013. Program leadership evaluated the new system by (1) comparing PGY-1 mean scores on Likert-type versus milestone-based assessments; and (2) comparing mean scores on the Likert-type versus milestone-based assessments across PGY levels.

RESULTS:
Mean scores for PGY-1 residents were significantly higher on the prior year's Likert-type assessments than milestone-based assessments for all 3 subcompetencies (P, .01). Stratification by PGY level was not observed with Likert-type assessments (eg, interpersonal and communication skills 1 [ICS1] mean score for PGY-1, 3.99 versus PGY-3, 3.98; P 5 .98). In contrast, milestone-based assessments demonstrated stratification by PGY level (eg, the ICS1 mean score was 3.06 for PGY-1, 3.83 for PGY- 2, and 3.99 for PGY-3; P, .01 for PGY-1 versus PGY-3). Significantly different means by trainee level were noted across 21 subcompetencies on milestone-based assessments (P, .01 for PGY-1 versus PGY-3).

CONCLUSIONS:
Initial results indicate milestone-based assessments stratify trainee performance by level better than Likert-type assessments. Average PGY-level scores from milestone-based assessments may ultimately provide guidance for determining whether trainees are progressing at the expected pace.

Impact on House Staff Evaluation Scores When Changing from a Dreyfus- to a Milestone-Based Evaluation Model: One Internal Medicine Residency Program's Findings

PURPOSE:
As graduate medical education (GME) moves into the Next Accreditation System (NAS), programs must take a critical look at their current models of evaluation and assess how well they align with reporting outcomes. Our objective was to assess the impact on house staff evaluation scores when transitioning from a Dreyfus-based model of evaluation to a Milestone- based model of evaluation. Milestones are a key component of the NAS.

METHOD:
We analyzed all end of rotation evaluations of house staff completed by faculty for academic years 2010- 2011 (pre-Dreyfus model) and 2011-2012 (post-Milestone model) in one large university-based internal medicine residency training program. Main measures included change in PGY-level average score; slope, range, and separation of average scores across all six Accreditation Council for Graduate Medical Education (ACGME) competencies.

RESULTS:
Transitioning from a Dreyfus-based model to a Milestone-based model resulted in a larger separation in the scores between our three post-graduate year classes, a steeper progression of scores in the PGY-1 class, a wider use of the 5-point scale on our global end of rotation evaluation form, and a downward shift in the PGY-1 scores and an upward shift in the PGY-3 scores.

CONCLUSIONS:
For faculty trained in both models of assessment, the Milestone-based model had greater discriminatory ability as evidenced by the larger separation in the scores for all the classes, in particular the PGY-1 class.

The Milestones Passport: A Learner-Centered Application of the Milestone Framework to Prompt Real-Time Feedback in the Emergency Department

BACKGROUND:
In July 2013, emergency medicine residency programs implemented the Milestone assessment as part of the Next Accreditation System.

OBJECTIVE:
We hypothesized that applying the Milestone framework to real-time feedback in the emergency department (ED) could affect current feedback processes and culture. We describe the development and implementation of a Milestone-based, learner-centered intervention designed to prompt real-time feedback in the ED.

METHODS:
We developed and implemented the Milestones Passport, a feedback intervention incorporating subcompetencies, in our residency program in July 2013. Our primary outcomes were feasibility, including faculty and staff time and costs, number of documented feedback encounters in the first 2 months of implementation, and user-reported time required to complete the intervention. We also assessed learner and faculty acceptability.

RESULTS:
Development and implementation of the Milestones Passport required 10 hours of program coordinator time, 120 hours of software developer time, and 20 hours of faculty time. Twenty- eight residents and 34 faculty members generated 257 Milestones Passport feedback encounters. Most residents and faculty reported that the encounters required fewer than 5 minutes to complete, and 48% (12 of 25) of the residents and 68% (19 of 28) of faculty reported satisfaction with the Milestones Passport intervention. Faculty satisfaction with overall feedback in the ED improved after the intervention (93% versus 54%, *P* = .003), whereas resident satisfaction with feedback did not change significantly.

CONCLUSIONS:
The Milestones Passport feedback intervention was feasible and acceptable to users; however, learner satisfaction with the Milestone assessment in the ED was modest.

Programmatic Assessment of Level 1 Milestones in Incoming Interns

OBJECTIVES:
With the Accreditation Council for Graduate Medical Education (ACGME) Next Accreditation System, emergency medicine (EM) residency programs will be required to report residents' progress through the EM milestones. The milestones include five progressively advancing skill levels, with Level 1 defining the skill set of a medical school graduate and Level 5, that of an attending physician. The ACGME stresses that multiple forms of assessment should be used to ensure capture of the multifaceted competencies. The objective of this study was to determine the feasibility and results of programmatic assessment of Level 1 milestones using multisource assessments for incoming EM interns in July.

METHODS:
The study population was interns starting in 2012 and 2013. Interns' Level 1 milestone assessment was done with four distinct methods: 1) the postgraduate orientation assessment (POA) by the Graduate Medical Education Office for all incoming interns (this multistation examination covers nine of the EM milestones and includes standardized patient cases, task completion, and computer-based stations); 2) direct observation of patient encounters by core faculty using a milestones-based clinical skills competency checklist; 3) the global monthly assessment at the end of the intern orientation month that was updated to reflect the EM milestones; and 4) faculty assessment during procedural labs. These occurred during the July orientation month that included the POA, clinical shifts, didactic sessions, and procedure labs.

RESULTS:
In the POA, interns were competent in 48% to 93% of the milestones assessed. Overall, competency was 70% to 80%, with low scores noted in aseptic technique (patient care Milestone 13 [PC13]) and written and verbal hand-off (interpersonal communications skills [ICS]2). In overall communication, 70% of interns demonstrated competency. In excess of 80% demonstrated competency in critical values interpretation (PC3), informed consent (PC9), pain assessment (PC11), and geriatric functional assessment (PC3). On direct observation, almost all Level 1 milestones were achieved (93% to 100%); however, only 78% of interns achieved competency in pharmacotherapy (PC5). On global monthly evaluations, all interns met Level 1 milestones.

CONCLUSIONS:
A multisource assessment of EM milestones is feasible and useful to determine Level 1 milestones achievement for incoming interns. A structured assessment program, used in conjunction with more traditional forms of evaluation such as global monthly evaluations and direct observation, is useful for identifying deficits in new trainees and may be able inform the creation of early intervention programs.

Operationalizing the Internal Medicine Milestones-An Early Status Report

Nabors C, Peterson SJ, Forman L, Stallings GW, Mumtaz A, Sule S, Shah T, Aronow W, Delorenzo L, Chandy D, Lehrman SG, Frishman WH, Holmboe E. J Grad Med Educ. 2013 Mar;5(1):130-7. doi: 10.4300/JGME-D-12-00130.1.

BACKGROUND:
The internal medicine milestones were developed to advance outcomes-based residency training and will play an important role in the next accreditation system.

INNOVATION:
As an element of our program's participation in the internal medicine educational innovations project, we implemented a milestones-based evaluation process in our general medicine and pulmonary-critical care rotations on July 1, 2010.

MEASURES:
Outcomes assessed included survey-rated acceptability to participating faculty, residents, and clinical competency committee members.

RESULTS:
Faculty and residents agreed that the milestones promoted a common understanding of what knowledge, skills, and attitudes should be displayed at particular points in residents' professional development and enhanced evaluators' ability to provide specific performance feedback. Most residents and faculty members agreed that the milestones promoted fairness and uniformity in the evaluation process. Clinical competency committee members agreed the milestones improved the quality of information available for deliberations and resulted in more uniform promotion standards. Faculty rated the use of too many milestones per form/tool at a mean of 7.3 (where 1 was minimally problematic, and 10 was maximally problematic) and the potential for evaluator fatigue (mean, 8.2) as the most significant challenges to the use of milestones. Eight of 12 faculty members would recommend milestones in other programs; 4 were uncertain.

CONCLUSIONS:
Despite logistical challenges, educators and trainees found that milestones promoted a common understanding of what knowledge, skills and attitudes should be displayed at particular stages of training; permitted greater specificity in performance feedback; and enhanced uniformity and fairness in promotion decisions.

The Pediatrics Milestones: Initial Evidence for their Use as Learning Road Maps for Residents

Schumacher DJ, Lewis KO, Burke AE, Smith ML, Schumacher JB, Pitman MA, Ludwig S, Hicks PJ, Guralnick S, Englander R, Benson B, Carraccio C. Acad Pediatr. 2013 Jan-Feb;13(1):40-7. doi: 10.1016/j.acap.2012.09.003. Epub 2012 Nov 17.

OBJECTIVE:
As the next step in competency-based medical education, the Pediatrics Milestone Project seeks to provide a learner-centered approach to training and assessment. To help accomplish this goal, this study sought to determine how pediatric residents understand, interpret, and respond to the Pediatrics Milestones.

METHODS:
Cognitive interviews with 48 pediatric residents from all training levels at 2 training programs were conducted. Each participant reviewed one Pediatrics Milestone document (PMD). Eight total Pediatrics Milestones, chosen for their range of complexity, length, competency domain, and primary author, were included in this study. Six residents, 2 from each year of residency training, reviewed each PMD. Interviews were transcribed and coded using inductive methods, and codes were grouped into themes that emerged.

RESULTS:
Four major themes emerged through coding and analysis: 1) the participants' degree of understanding of the PMDs is sufficient, often deep; 2) the etiology of participants' understanding is rooted in their experiences; 3) there are qualities of the PMD that may contribute to or detract from understanding; and 4) participants apply their understanding by noting the PMD describes a developmental progression that can provide a road map for learning. Additionally, we learned that residents are generally comfortable being placed in the middle of a series of developmental milestones. Two minor themes focusing on interest and practicality were also identified.

CONCLUSIONS:
This study provides initial evidence for the Pediatrics Milestones as learner-centered documents that can be used for orientation, education, formative feedback, and, ultimately, assessment.