


# Can ACGME Milestones predict surgical specialty board passage: an example in Obstetrics and Gynecology

Sarah Ottum<sup>1</sup>, Conrad Chao<sup>2,§</sup>, Sejal Tamakuwala<sup>3</sup>, Joshua Dean<sup>4</sup>, Adib Shafi<sup>5</sup>, Katherine Jennifer Kramer<sup>6</sup>, Satinder Kaur<sup>7,§</sup>, Maurice-Andre Recanati<sup>8,\*</sup> 

<sup>1</sup> Department of Obstetrics and Gynecology, University of Cincinnati, Cincinnati, OH 45267, USA

<sup>2</sup> Department of Maternal-Fetal Medicine, University of New Mexico School of Medicine, Albuquerque, NM 87131, USA

<sup>3</sup> Department of Obstetrics and Gynecology, Emory University, Atlanta, GA 30322, USA

<sup>4</sup> Department of Obstetrics and Gynecology, Beaumont, Royal Oak, MI 48073, USA

<sup>5</sup> Department of Computer Sciences, Wayne State University, Detroit, MI 48201, USA

<sup>6</sup> Department of Obstetrics and Gynecology, St. Vincent's Catholic Medical Centers, New York, NY 10011, USA

<sup>7</sup> Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI 48201, USA

<sup>8</sup> Department of Obstetrics and Gynecology and NIH-Women's Reproductive Health Research Scholar, Wayne State University, Detroit, MI 48201, USA

\*Correspondence: [marecanati@wayne.edu](mailto:marecanati@wayne.edu) (Maurice-Andre Recanati)

§ Current or former program director or chair.

DOI: [10.31083/j.ceog4805168](https://doi.org/10.31083/j.ceog4805168)

This is an open access article under the CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>).

Submitted: 23 May 2021 Revised: 13 July 2021 Accepted: 20 July 2021 Published: 15 October 2021

**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.

## Keywords

CREOG; ABOG; Board passage; Milestones; USMLE

## 1. Introduction

The success of a residency program relies upon ensuring that residents who are in the program are learning the necessary clinical skills and academic knowledge to successfully

practice and pass the specialty board examination. To ensure that residents are developing appropriately over the course of their training and developing competence appropriately, several tools have been implemented over the years.

The American medical system has evolved tremendously over the last century. The practice of medical education initially consisted of unstructured apprenticeships with virtually no requirements for medical licensure [1]. In the early 20th century William Halstead established the foundation of the American surgical residency at the Johns Hopkins School of Medicine [2]. Abraham Flexner uncovered that many training programs were of substandard quality [3], which led to accreditation as the method of establishing uniform standards. The Accreditation Council for Graduate Medical Education (ACGME), which accredits post-graduate training programs, is tasked with “improving healthcare and population health by assessing and advancing the quality of resident physicians’ education through accreditation” [4].

Up until recently, residencies needed to show that trainees were exposed to adequate clinical volumes and a didactic program [5]. Focus then turned to producing physicians that are “competent” and possess a skill set that will “improve patient outcomes” [6, 7]. In 1998, with the support of the Robert Wood Johnson Foundation, the ACGME developed the Outcomes Project. It required physicians’ development of competencies in six broad core domains [8] (education, medical knowledge, patient care, interpersonal and communication skills, systems-based practice, and practice-based learning and professionalism), each measured by individual metrics [9, 10].

The Next Accreditation System (NAS), created out of concerns that graduates lack the readiness to practice independently [11, 12], led to a system in which resident skills are tracked through individual milestones within the six core competencies [13]. Milestones, introduced in 2014, are fine-grained developmental levels that correspond to and expand on the core competencies [14]. Numerous contributions, such as attending feedback, technical skills, clinical skills, nursing feedback, peer-review, end-of-rotation evaluations, patient feedback, and performance in simulation are used to derive individual milestone scores [15]. These 28 scores are then aggregated by an institutional clinical competency committee (CCC) and each resident's performance is reported to the ACGME on a semiannual basis [16]. Milestone evaluations require a great deal of time and detailed effort on the part of faculty and CCC committees, but very little has been published validating the contributions of the Milestone evaluation process to quantifiable outcomes of educational efficacy. This validation is essential not only to determine the value of these processes but also to create a framework for improvements in the milestones themselves.

In parallel to the Milestones scores, the Council on Resident Education in Obstetrics and Gynecology (CREOG) has administered a multiple choice in-training examination since 1970. This test objectively assesses a resident's cognitive knowledge in the specialty [17] while also ensuring that training programs provide adequate teaching to residents [18]. Several studies have attempted to show a correlation between CREOG scores and outcomes on the ABOG written examination [5, 19, 20]. Other studies have shown a correlation between CREOG scores and United States Medical Licensing Examination Step scores [21, 22], suggesting that these grades can be used to predict residents who will do well academically and have a high likelihood of passing written boards.

In 2016, the Obstetrics and Gynecology program at Wayne State University/Detroit Medical Center had a 50% failure rate on the American Board of Obstetrics and Gynecology (ABOG) written examination. In 2018, the failure rate was 30% in an academic program accommodating ten residents per year. These outcomes prompted this study. This study sought to determine metrics that could identify residents at risk for failure on the specialty board examination. We hypothesized that Milestone and CREOG scores would predict performance on the written Obstetrics and Gynecology (OB/GYN) board examination. To our knowledge, we are one of the few studies to date that conducted a regression analysis and built a predictive model for ABOG written board passage.

## 2. Materials and methods

Under IRB approval (042317MP2X), data for this retrospective study were collected from electronic and paper documents on residents who graduated from the Wayne State University's Obstetrics and Gynecology Residency program

between July 2013 and July 2018 and attempted to pass the ABOG written examination. Data collected during residency, and used in the analysis included the following:

*United States Medical Licensing Examination (USMLE) Scores:* The USMLE is a three-step examination for medical licensure in the United States and assesses a physician's ability to apply the knowledge, concepts and principles and demonstrate patient-centered skills that constitute the basis of safe and effective care. Step 1 tests sciences basic to medical practice, Step 2 assesses clinical knowledge, while Step 3 assesses the application of medical and biomedical knowledge essential for the unsupervised practice of medicine. Scores range from 1 to 300, with mean (and standard deviation) typically 230 (19), 244 (16), 226 (15) for each step respectively [23]. USMLE step scores were analyzed individually as well as in aggregate, having defined "average USMLE score" as the average of Step 1, 2, and 3 scores.

*CREOG scores:* The CREOG in-training examination is a national subspecialty multiple-choice test given to all OB/GYN residents annually and offers a formative assessment of a resident's developing medical knowledge. The overall reported standardized score are calculated with a mean of 200 and a standard deviation of 20 for all test-takers, thus 99.7% of scores range between 140–260. Internally, the raw score, which is not reported, theoretically ranges from zero to  $n$ , where  $n$  is the number of scored items on a given exam, and where  $n$  is usually around 318. CREOGs were analyzed individually and the "average CREOG" was defined as the average score obtained over the four yearly exams.

*ACGME Milestones:* Milestones provide a framework for assessment of the development of residents in key dimensions of the elements of physician competency in a specialty. While many ACGME accredited specialties use Milestones, in Obstetrics and Gynecology attending-level clinicians evaluate residents with scores from 1 to 5. Level 1 corresponds to the level of an incoming intern, level 2 is below that expected at mid-residency, while level 4 is the graduation target and level 5 corresponds to an attending with some years of practice [24]. We compiled data from all 28 milestones grades from all years for all candidates, as available, at the time the study was done. To facilitate analysis, the 28 individual milestones, recorded semi-annually since the program began recording these scores, were grouped into seven categories (Fig. 1) and analyzed using matrix analysis.

*ABOG Qualifying Exam Passage:* Passage of the ABOG written exam was determined by the ABOG candidate status change to "active candidate". The passing rate for first time test takers varied between 91–93%. Successful board performance was defined as passing the ABOG written boards on first attempt.

Data was de-identified and tabulated in an Excel spreadsheet and statistical analysis was performed using both ANOVA and nonparametric Mann-Whitney tests (SPSS version 26, IBM, Chicago, IL, USA). Analysis of the subgroups were performed using the F-test. Binomial logistic regres-

#### Obstetric Milestones

1. PC 1: Antepartum Care and Complications of Pregnancy
2. PC2: Care of Patients in the Intrapartum Period
3. PC3: Care of Patients in the Postpartum Period
4. PC 4: Obstetrical Technical Skills
5. PC 5: Immediate Care of the Newborn

#### Gynecology Milestones

6. PC 6: Gynecology Technical Skills: Laparotomy (e.g. Hysterectomy, Myomectomy, Adnexectomy)
7. PC 7: Gynecology Technical Skills: Vaginal Surgery (e.g. Vaginal Hysterectomy, Colporrhaphy, Mid-urethral Sling)
8. PC 8: Gynecology Technically Skills: Endoscopy (e.g. Laparoscopy, Hysteroscopy, Cystoscopy)
9. MK 1: Perioperative Care
10. MK 2: Abdominal/Pelvic Pain (Acute and Chronic)
11. MK 3: Abnormal Uterine Bleeding (Acute and Chronic)
12. MK 4: Pelvic Mass
13. MK 5: Pelvic Floor Disorders (Urinary Incontinence, Pelvic Prolapse, Anal Incontinence)
14. MK 6: First Trimester Bleeding
15. MK 7: Health Care Maintenance and Disease Prevention

#### Office Practice Milestones

16. PC 9: Family Planning
17. PC 10: Ambulatory Gynecology
18. PC 11: Care of the Patient with Non reproductive Medical Disorders

#### Systems-Based Practice Milestones

19. SBP 1: Patient Safety and Systems Approach to Medical Errors: Participate in Identifying System Errors and Implementing Potential Systems Solutions
20. SBP 2: Cost-Effective Care and Patient Advocacy

#### Practice-Based Learning And Improvement Milestones

21. PBL 1: Self-Directed Learning/Critical Appraisal of Medical Literature
22. PBL 2: Quality Improvement Process: Systematically Analyze Practice Using Quality Improvement Methods and Implement Changes With the Goal of Practice Improvement

#### Professionalism Milestones

23. Pro 1: Compassion, Integrity, and Respect for Others
24. Pro 2: Accountability and Responsiveness to the Needs of Patients, Society, and the Profession
25. Pro 3: Respect for Patient Privacy, Autonomy, Patient-Physician Relationship

#### Interpersonal And Communication Skills Milestones

26. ICS 1: Communication With Patients and Families
27. ICS 2: Communication With Physicians and Other Health Professionals and Teamwork
28. ICS 3: Informed Consent and Shared Decision Making

**Fig. 1. Grouping of ACGME milestone into categories.** All 28 individual milestones were grouped into seven categories: obstetrics, gynecology, office, systems-based practice (SBP), practice-based learning (PBL), professionalism (PRO), and interpersonal and communication (ICS).

sion, using the criterion of  $p = 0.05$  to enter and  $p = 0.01$  to remove variables, was used to create a prediction equation for board passage as a binary outcome (pass or fail). Positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity were computed using standard methods.

### 3. Results

#### 3.1 Descriptive statistics on cohort of residents

The dataset, which includes residents who were enrolled from July 2013 to 2018, is comprised of 45 residents, 16 males (35.5%) and 29 females (64.5%). Out of 45 graduates, 9 failed their ABOG written boards on first attempt.

#### 3.2 Predictor of board passage

We looked for predictors of successful board performance and compared metrics from passing and failing groups (Fig. 2). Average USMLE score were not significantly associated with board passage (at  $p < 0.05$ ), however USMLE1 came closest ( $p = 0.055$ ). Using a cutoff of  $\geq 217$  on USMLE1, we can predict 8 of the 9 board failures and 21 of the 36 passes (PPV = 95.8%, NPV = 38%, Sensitivity = 63.9%, Specificity = 88.9%) (Fig. 3A).

Academic knowledge, such as average CREOG score, is associated with board passage ( $p = 0.015$ ). Using a cutoff of  $\geq 198$  we are able to predict all failures and 21 out of the 36 passes (PPV = 100%, NPV = 37.5%, Sensitivity = 58.3%, Specificity = 100%) (Fig. 3B).

PREDICTOR	sample size fail/pass	FAILED BOARDS		PASSED BOARDS		F-test
		mean	s.d.	mean	s.d.	Significance
USMLE1	9/36	211.6	8.7	223.4	17.4	0.055
USMLE2	9/34	219.3	16.1	228.9	18.9	0.170
USMLE3	8/33	203.6	8.1	211.8	14.2	0.125
USMLE Average	9/36	211.8	8.3	221.1	14.8	0.079
CREOG1	8/35	172.4	18.2	188.0	19.3	0.043*
CREOG2	8/36	182.0	13.6	199.1	21.5	0.038*
CREOG3	9/35	186.1	12.4	206.7	23.2	0.013*
CREOG4	8/36	193.3	15.6	211.3	18.3	0.014*
CREOG Average	9/36	184.6	11.1	201.5	19.1	0.015*
Mean ICS	9/30	3.3	0.6	3.7	0.5	0.074
Mean Ob/Gyn/Office	9/30	2.9	0.3	3.4	0.5	0.011*
Mean PRO	9/30	3.3	0.6	3.7	0.5	0.030*
Mean SBP&PBL	9/31	2.9	0.3	3.4	0.5	0.002*
Milestones Average	9/31	3.1	0.4	3.6	0.5	0.008*

**Fig. 2. Predictors of board passage.** The cohort of residents who passed ABOG written exam on first attempt was compared with those that failed the exam. Mean scores and standard deviation for each predictor variable are shown, along with *p*-values (\* denotes  $p < 0.05$ ).

Clinical metrics, such as mean milestones were also significantly correlated with board passage ( $p = 0.0085$ ). In particular, the average SBP ( $p = 0.01$ ) and PBL ( $p = 0.002$ ) milestone score were most strongly associated with board passage. Combining both SBP and PBL together (SBP&PBL) provided the highest association ( $p = 0.002$ ) of any single or combination of metrics. Using a cutoff for the average SBP&PBL of  $\geq 3.25$  we are able to predict all fails and 20 of 31 passes (PPV = 100%, NPV = 45%, Sensitivity = 64.5%, Specificity = 100%) with 73% accuracy (Fig. 3C). Board passage was also associated with other milestones including office-based practice ( $p = 0.01$ ) and professionalism ( $p = 0.03$ ) (not shown). Rate of improvement, year-to-year in milestones was not correlated with board passage.

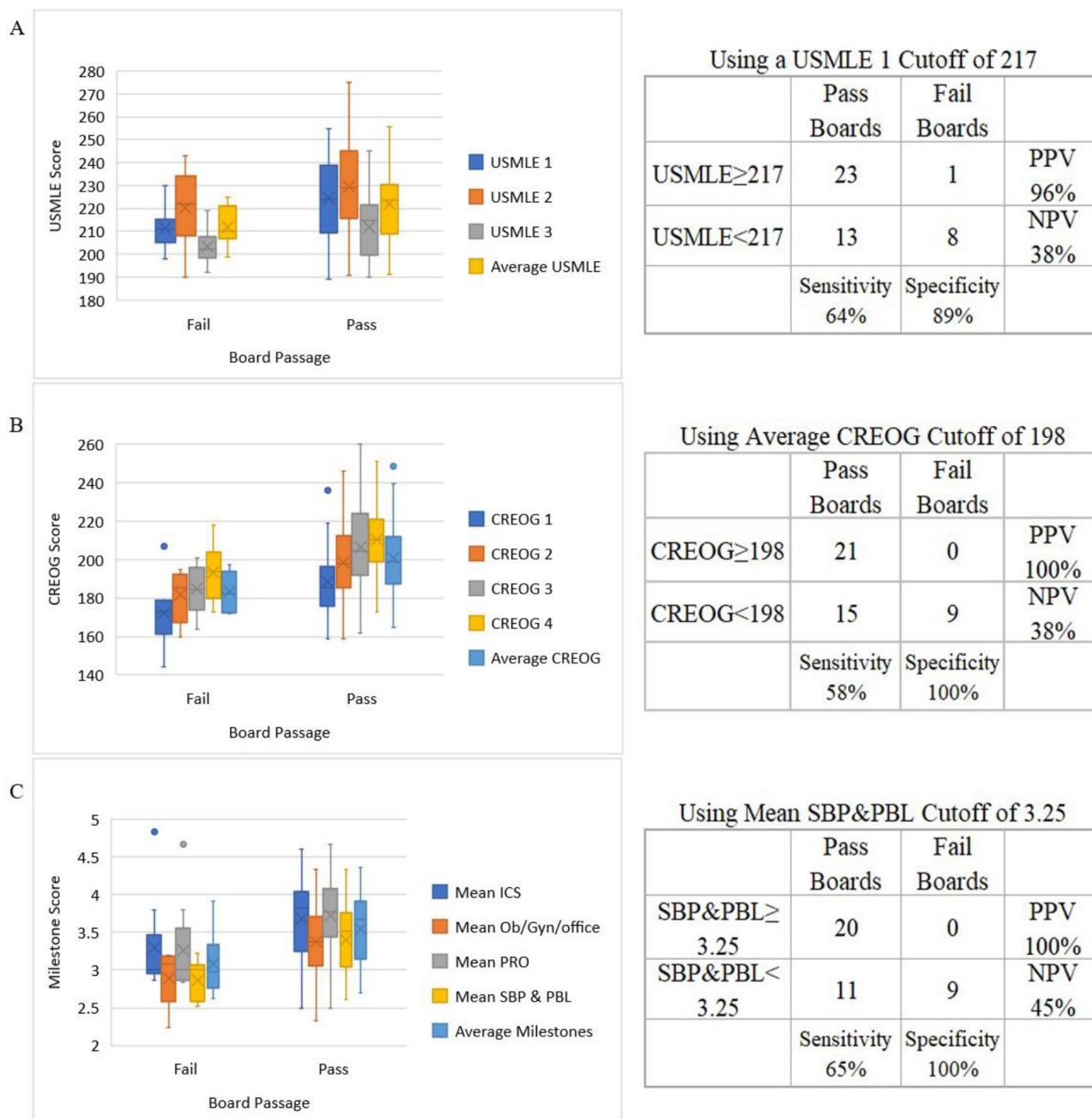
A stepwise binomial logistic regression confirms that SBP & PBL are the best predictors of board passage and were the only metrics which entered the regression. The prediction equation has a Nagelkerke  $R^2 = 0.425$  ( $p = 0.014$ ) and is 82%

correct in predicting pass/fail (PPV = 86%, NPV = 60%, Sensitivity = 92.6%, Specificity = 42.9%) (Fig. 4).

#### 4. Discussion

The shared academic responsibility between a residency program and a resident is critical for the combined and individual success of each entity. Early markers identifying residents at risk of failure on the written examination or having trouble acquiring the clinical skills necessary for independent practice are essential to enable program directors to identify and assist struggling residents.

Milestones were the best predictor of passing the ABOG written exam. Milestones measure clinical knowledge of a topic as a resident progress over five levels, from uncomplicated management, to more complicated cases, and ultimately to a level expected for independent practice [25]. While Milestone scores are somewhat subjective, at our institution, a clinical competency committee, attended by nu-



**Fig. 3. Predictors of board passage.** The cohort of residents who passed ABOG written exam on first attempt (n = 36) was compared with those that failed the exam (n = 9). **Left:** Distribution showing mean (X), median (line), first & third quartile and range are depicted graphically, dots represent outliers. **Right:** Prediction accuracy. Contingency tables at specified cutoff values for pass and fail groups. Panels include (A) USMLE scores, (B) CREOGs, and (C) Milestones.

merous attendings from various subspecialties as well as private (non-faculty) physicians offer differing perspectives and a balanced consensus. This approach helps regulate and moderate grades assigned by faculty members who are uncomfortable giving low scores [26] or who are influenced by personal biases or opinions of residents [27]. Often, milestone scores from the CCC committee were very much in-line with resident self-assessment [15, 16] and were as objective as possible in measuring clinical performance. The bulletin for the

ABOG basic written examination [28] suggests that the exam spans mostly clinical topics in obstetrics, gynecology and office practice including topics on postoperative care, patient evaluation, surgical management and possible complications. It is likely that one of the reasons that Milestones most closely predict a resident's performance on this exam is because the focus of the exam is clinical and mirrors the practical experience as well as practical knowledge needed to take care of patients. As residents progress through the steepest part of the

$$Probability(Pass) = 1/(1 + e^{11.52 - 4.129 * Mean\ SBP\ \&\ PBL})$$

	Pass Boards	Fail Boards	
Model Pass	25	4	PPV 86%
Model Fail	2	3	NPV 60%
	Sensitivity 93%	Specificity 43%	

**Fig. 4. Binomial Logistic Regression Model.** Mean SBP&PBL can be used to calculate the probability of passing. Insert depicts 2 × 2 contingency table for the model.

learning curve over the first year, significant gains in clinical knowledge are made that are reflected in improved milestone scores [29]. The rate of milestone improvement, however, did not correlate with board passage. This may be because different residents progress at different rates; however the absolute score is more reflective of performance.

Our data suggests that two milestones were most significant. PBL reflects self-directed learning and improvements that allow a doctor to become a more masterful clinician. This is a measure of internal motivation, desire to improve, acceptance of responsibility for one's own learning combined with the initiative and self-discipline to complete goal-oriented self-directed learning. Individuals who score high in this area are typically able to identify their own deficiencies and limits in knowledge, analyze their own practice, incorporate feedback from faculty and assimilate evidence from scientific publications into their practice [30]. These life-long basic study skills translate into residents who enjoy learning, as well as teaching their junior colleagues, and typically have the knowledge and self-confidence to do well clinically as well as on board exams, as demonstrated in our results. The other highly predictive Milestone, SBP, reflects on a physician's awareness of (and responsiveness to) the health-care universe in which they deliver care. Clinicians who score high generally exhibit strong interpersonal and communications skills and professionalism enabling them to interface well with other members of the team. They emphasize shared-decision-making, advocate for quality patient care, actively seek to improve the quality of care and find creative solutions to limitations inherent to the system [31]. Importantly, they take steps to enhance patient safety while embracing personal responsibility and quality improvement. While the ABOG qualifying exam is a written exam, patient safety is one of the important areas tested and this foundational Milestone touches on all cross-content areas examined, including ethics, communication and health literacy [32]. This may help explain the importance of this Milestone in our results.

Performance on the ABOG written examination was correlated with CREOG and Milestone scores. Our analysis showed that average CREOGs greater than 198 were associated with board passage. This finding was similar to other

studies which determined that a score of 200 either as a PGY-4 or two times throughout residency correlated with a 100% probability of passage [19], or a score over 202 on the third year CREOG was predictive of board passage [20]. While other studies found that CREOGs were not useful in identifying residents at risk of board failure [5], our analysis validated the utility of Milestones. Milestones were more predictive than other metrics such as CREOGs. One possible explanation is that the ABOG exam essentially tests a candidate's ability to interpret clinical data and derive a treatment plan [28], while CREOG questions tend to test knowledge differently. Prior publications have demonstrated that in-training exam scores are expected to increase with each year of training as more knowledge is accumulated [18, 33]. In our study, year-to-year improvement on CREOGs did not significantly correlate with board passage but score value did [17]. Unlike other studies which found that USMLE Step 1 correlated with board passage and that Step 1 >200 was suggestive of successful outcomes on boards [21], our study did not echo these findings. In our cohort, Step 1 was not significantly correlated to board passage at a  $p < 0.05$  and mean Step 1 scores in those who failed written boards was 211. This may be due to factors specific to our program or to our propensity to highly rank medical students with high USMLEs.

One potential weakness in this study is the unusually high failure rate in our program during the study period, compared to the nationwide failure rate of 3–9%. Possible confounding factors, which may affect the applicability of our study, may include particulars of our residency program at the time, such as service to teaching balance, structure of the curriculum and composition of our medical residents. However, data collected from our single site offers consistency in faculty evaluators (the same individual faculty members) and clinical rotations (no changes in the number of weeks on the various rotations), making the residents one of the only variables. Without a large number of individuals failing, it would be statistically difficult to characterize the unique characteristics of residents who have failed. The limitation, however, is the smaller sample size. Strengths of this study include the predictive regression analysis. Because of the diversity in our program, including residents who transferred in from other programs, our results may not be applicable to other institutions.

## 5. Conclusions

We have validated the utility of the OB/GYN residency Milestones by demonstrating their ability to predict OB/GYN written board passage, a critical accomplishment that constitutes an objective external assessment of educational efficacy and attainment. The encouraging results from this pilot study suggest that a larger study encompassing Milestone data from all residency programs across all surgical specialties would confirm these preliminary findings.

## Author contributions

SO—statistical analysis, wrote parts of manuscript, created figures. CC—revising the manuscript, important intellectual contributions, interpreting statistics. ST—conception, data extraction. JD—data extraction, wrote parts of manuscript. AS—statistical analysis and interpretation. KK—conception, design, drafting manuscript, created figures. SK—data extraction, interpretation of data, rewrote parts of manuscript. MR—conception of study, wrote part of manuscript, analyzed statistics and provided guidance, responsible for the accuracy and integrity of the work presented here. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

This work was conducted under the approval of the Wayne State Institutional Review Board (042317MP2X) as well as the Detroit Medical Center's contract research organization (CRO) (DMC14380). Informed consent was waived for this study.

## Acknowledgment

This research was made possible by the NIH-Women's Reproductive Health Research Career Development Award at Wayne State University. We wish to thank the residents at Wayne State University and the faculty and staff at the Detroit Medical Center.

## Funding

This research was supported under NIH-Women's Reproductive Health Research Career Development Award (K-12HD001254), awarded to MAR.

## Conflict of interest

The authors declare no conflict of interest. MAR is on our Reviewer Board. Given the role on the Reviewer Board, has no involvement in the peer-review of this article and has no access to information regarding its peer-review.

## Previous presentation

This work was presented in abstract form for poster publication at the Council on Resident Education in Obstetrics and Gynecology (CEOG) & the Association of Professors of Gynecology and Obstetrics (APGO) in New Orleans, LA, February 2019.

## Data availability

The datasets that support the findings of this study are available from the corresponding author (MAR) upon reasonable written request.

## References

- [1] Joyner BD. An historical review of graduate medical education and a protocol of accreditation council for graduate medical education compliance. *Journal of Urology*. 2004; 172: 34–39.
- [2] Niederee MJ, Knudtson JL, Byrnes MC, Helmer SD, Smith RS. A survey of residents and faculty regarding work hour limitations in

- surgical training programs. *Archives of Surgery*. 2003; 138: 663–671.
- [3] Ludmerer KM. Time to heal: American medical education from the turn of the century to the era of managed care (pp. xxvi, 514 pages). New York: Oxford University Press, Oxford.1999.
- [4] Accreditation Council for Graduate Medical Education. Mission, Vision and Values. June 2021. Available at: <https://www.acgme.org/About-Us/Overview/Mission-Vision-and-Values/> (Accessed: 7 June 2021).
- [5] Withiam-Leitch M, Olawaiye A. Resident Performance on the in-Training and Board Examinations in Obstetrics and Gynecology: Implications for the ACGME Outcome Project. *Teaching and Learning in Medicine*. 2008; 20: 136–142.
- [6] Carraccio C, Wolfsthal SD, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Academic Medicine*. 2002; 77: 361–367.
- [7] Leach DC. A model for GME: shifting from process to outcomes. a progress report from the Accreditation Council for Graduate Medical Education. *Medical Education*. 2004; 38: 12–14.
- [8] Rose SH, Long TR, Elliott BA, Brown MJ. A Historical Perspective on Resident Evaluation, the Accreditation Council for Graduate Medical Education Outcome Project and Accreditation Council for Graduate Medical Education Duty Hour Requirement. *Anesthesia & Analgesia*. 2009; 109: 190–193.
- [9] Swing SR. Assessing the ACGME general competencies: general considerations and assessment methods. *Academic Emergency Medicine*. 2002; 9: 1278–1288.
- [10] Swing SR. The ACGME outcome project: retrospective and prospective. *Medical Teacher*. 2007; 29: 648–654.
- [11] Park YS, Zar FA, Norcini JJ, Tekian A. Competency Evaluations in the next Accreditation System: Contributing to Guidelines and Implications. *Teaching and Learning in Medicine*. 2016; 28: 135–145.
- [12] Damewood RB, Blair PG, Park YS, Lupi LK, Newman RW, Sachdeva AK. "Taking Training to the next Level": the American College of Surgeons Committee on Residency Training Survey. *Journal of Surgical Education*. 2017; 74: e95–e105.
- [13] Cogbill TH, Malangoni MA, Potts JR, Valentine RJ. The general surgery milestones project. *Journal of the American College of Surgeons*. 2014; 218: 1056–1062.
- [14] Carraccio CL, Benson BJ, Nixon LJ, Derstine PL. From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. *Academic Medicine*. 2008; 83: 761–767.
- [15] Watson RS, Borgert AJ, O Heron CT, Kallies KJ, Sidwell RA, Mellinger JD, *et al*. A Multicenter Prospective Comparison of the Accreditation Council for Graduate Medical Education Milestones: Clinical Competency Committee vs. Resident Self-Assessment. *Journal of Surgical Education*. 2017; 74: e8–e14.
- [16] Lyle B, Borgert AJ, Kallies KJ, Jarman BT. Do Attending Surgeons and Residents See Eye to Eye? An Evaluation of the Accreditation Council for Graduate Medical Education Milestones in General Surgery Residency. *Journal of Surgical Education*. 2016; 73: e54–e58.
- [17] Visscher HC. Validity and purpose of in-training examination in obstetrics and gynecology, 1979–1982. *Obstetrics and Gynecology*. 1984; 63: 253–259.
- [18] Holzman GB, Downing SM, Power ML, Williams SB, Carpentieri A, Schulkin J. Resident performance on the Council on Resident Education in Obstetrics and Gynecology (CREOG) in-Training Examination: years 1996 through 2002. *American Journal of Obstetrics and Gynecology*. 2004; 191: 359–363.
- [19] Lingenfelter BM, Jiang X, Schnatz PF, O'Sullivan DM, Minassian SS, Forstein DA. CREOG in-Training Examination Results: Contemporary Use to Predict ABOG Written Examination Outcomes. *Journal of Graduate Medical Education*. 2016; 8: 353–357.
- [20] Spellacy WN, Carlan SJ, McCarthy JM, Tsibris JCM. Prediction of ABOG written examination performance from the third-year

- CREOG in-training examination results. *Journal of Reproductive Medicine*. 2006; 51: 621–622.
- [21] Armstrong A, Alvero R, Nielsen P, Deering S, Robinson R, Frattarelli J, *et al*. Do U.S. medical licensure examination step 1 scores correlate with council on resident education in obstetrics and gynecology in-training examination scores and American board of obstetrics and gynecology written examination performance? *Military Medicine*. 2007; 172: 640–643.
- [22] Bell JG, Kanellitsas I, Shaffer L. Selection of obstetrics and gynecology residents on the basis of medical school performance. *American Journal of Obstetrics and Gynecology*. 2002; 186: 1091–1094.
- [23] United States Medical Licensing Examination. USMLE Score Interpretation Guidelines. 2019. Available at: [https://www.usmle.org/pdfs/transcripts/USMLE\\_Step\\_Examination\\_Score\\_Interpretation\\_Guidelines.pdf](https://www.usmle.org/pdfs/transcripts/USMLE_Step_Examination_Score_Interpretation_Guidelines.pdf) (Accessed: 7 July 2021).
- [24] The Accreditation Council for Graduate Medical Education. The Obstetrics and Gynecology Milestone Project. September 2015. Available at: <https://www.acgme.org/Portals/0/PDFs/Milestones/ObstetricsandGynecologyMilestones.pdf?ver=2016-04-04-151917-043> (Accessed: 7 July 2021).
- [25] Bienstock JL, Edgar L, McAlister R. Obstetrics and gynecology milestones. *Journal of Graduate Medical Education*. 2014; 6: 126–128.
- [26] Kinnear B, Bensman R, Held J, O'Toole J, Schauer D, Warm E. Critical Deficiency Ratings in Milestone Assessment: a Review and Case Study. *Academic Medicine*. 2017; 92: 820–826.
- [27] Meier AH, Gruessner A, Cooney RN. Using the ACGME Milestones for Resident Self-Evaluation and Faculty Engagement. *Journal of Surgical Education*. 2016; 73: e150–e157.
- [28] The American Board of Obstetrics and Gynecology. 2019 Bulletin for the Written Examination for Basic Certification in Obstetrics and Gynecology. June 2019. Available at: <https://www.abog.org/docs/default-source/bulletins/2019/2019-qualifying-examination-in-obstetrics-and-gynecology.pdf> Available at: (Accessed: 7 July 2021).
- [29] Goldman RH, Tuomala RE, Bengtson JM, Stagg AR. How Effective are New Milestones Assessments at Demonstrating Resident Growth? 1 Year of Data. *Journal of Surgical Education*. 2017; 74: 68–73.
- [30] Burke AE, Benson B, Englander R, Carraccio C, Hicks PJ. Domain of competence: Practice-based learning and improvement. *Academic Pediatrics*. 2014; 14: S38–S54.
- [31] Guralnick S, Ludwig S, Englander R. Domain of competence: Systems-based practice. *Academic Pediatrics*. 2014; 14: S70–S79.
- [32] The American Board of Obstetrics and Gynecology. 2014 Bulletin for the Written Examination for Basic Certification in Obstetrics and Gynecology. June 2014. Available at: <https://www.abog.org/docs/default-source/bulletins/2014/basic-written-bulletin-2014.pdf> (Accessed: 7 July 2021).
- [33] McKenna BJ. The American Society for Clinical Pathology Resident in-Service Examination: does Resident Performance Provide Insight into the Effectiveness of Clinical Pathology Education? *Clinics in Laboratory Medicine*. 2007; 27: 283–291.