Research Article

iMedPub Journals
www.imedpub.com

Journal of Emergency and Internal Medicine

2017
Vol.1 No.2:12

Inter-Rater Reliability of Select Emergency Medicine Milestones in Simulation

Kathleen Wittels[1], Michael E Abboud[2], Yuchiao Chang[3], Alexander Sheng[4], and James K Takayesu[5*]

[1]Department of Emergency Medicine, Brigham and Women's Hospital, Boston MA, USA

[2]Harvard-Affiliated Emergency Medicine Residency at Massachusetts General Hospital and Brigham and Women's Hospital, Boston MA, USA

[3]Department of Medicine, Massachusetts General Hospital, Boston MA, USA

[4]Department of Emergency Medicine, Boston Medical Center, Boston MA, USA

[5]Department of Emergency Medicine, Massachusetts General Hospital, Boston MA, USA

***Corresponding author:** James K. Takayesu, Massachusetts General Hospital, 5 Emerson Place, Room 108, Boston, MA 02114, USA, Tel: 617-816-6112; E-mail: jtakayesu@partners.org

Rec Date: September 24, 2017, **Acc Date:** September 29, 2017, **Pub Date:** September 30, 2017

Citation: Wittels K, Abboud ME, Chang Y, Sheng A, Takayesu JK (2017) Inter-Rater Relaiability of Select Emergency Medicine Milestones in Simulation. J Emerg Intern Med Vol.1:No.2:12.

Abstract

Objectives: In 2012, the ACGME established the Milestones in emergency medicine (EM) training to provide competency-based benchmarks for residency training. Small observational studies have shown variable correlation between faculty assessment and resident self-assessment. Using simulation clinical scenarios, we sought to determine (1) the correlation between resident self-assessment and faculty assessment of clinical competency using selected Milestones; and (2) the inter-rater reliability between EM faculty using both Milestone scoring and a critical actions checklist.

Methods: This is an observational study in which second-year EM residents at an urban academic medical center were assessed with two simulation cases focusing on management of cardiogenic shock and sepsis. Twenty-three residents completed both cases; they were assessed by two EM faculty in eight select Milestones (scored 1-5, increments of 0.5) and with a checklist of critical actions to perform (scored 0 or 1). Intra-class correlation coefficients (ICC) were used to compare Milestone scoring between faculty and to assess correlation between resident self-assessment and faculty scoring. Faculty checklist inter-observer agreement was assessed using kappa statistics. Correlation between Milestone achievement and checklist performance were assessed using Spearman and Pearson correlation coefficients.

Results: The ICCs for inter-rater agreement between faculty for Milestone level were 0.12 and 0.15 for the cardiogenic shock and sepsis cases, respectively. The ICC comparing resident self-assessment with the average of faculty Milestone level scoring for each case was 0.00. The inter-rater agreement on checklist items for the cardiogenic shock and sepsis cases had kappa coefficients of 0.83 and 0.78, respectively. Pearson and Spearman correlation coefficients comparing Milestone scoring and checklist items in the cardiogenic shock case were 0.27 and 0.29; in the sepsis case, 0.085 and -0.021.

Conclusion: When compared to critical action checklists, use of Milestones lacks consistency between faculty raters for simulation-based competency assessment. Resident self-assessment shows no correlation with faculty assessments.

Keywords: Cardiogenic shock; Spearman and Pearson correlation coefficients; Accreditation Council for Graduate Medical Education (ACGME); Emergency Medicine (EM)

Introduction

The Emergency Medicine (EM) Milestones are the current standard that all residencies must use to measure progress in resident learning and clinical growth along the pathway from novice to expert. Residencies are required to provide the Accreditation Council for Graduate Medical Education (ACGME) with milestones rating every six months through reviews conducted by Core Competency Committees to document resident progress in training.

Achievement of specific Milestones is fundamentally based on the successful demonstration of defined entrustable professional activities. These activities are complex, relying on the integration of knowledge and skills to demonstrate competence. Evidence for attainment of specific Milestone levels can be garnered from a variety of assessments, including direct observation in both the clinical and simulation learning environments.

Research on the reliability of Milestone scoring in direct observational assessment and resident self-evaluation has had variable results. While small observational studies in non-EM residency programs have shown strong correlation between

resident self-assessment and faculty assessment of those residents [1,2] a small observational study performed in an EM residency found that residents consistently scored themselves higher than faculty assessments [3].

It is unclear whether incorporating Milestone-based assessments in simulation-based competency assessments provides reliable data and added value over checklist-based rubrics. This pilot study sought to determine the inter-rater reliability of Milestones relative to checklists for faculty observers when used to measure performance in two simulation-based competency assessments. In addition, we sought examine the correlation between EM resident self-assessment and faculty assessment.

Methods

Study design

This is an observational study in which EM residents were assessed with two simulation cases focusing on management of cardiogenic shock and sepsis.

Study setting and population

Second-year EM residents at a four-year, academic medical center underwent hands-on training modules in a state-of-the-art medical simulation center using mannequins, simulation software, and ample medical equipment. Residents were familiar with the equipment and setup through multiple other simulation modules performed on site. These simulations were performed over two years (2013 and 2014), so two classes of residents took part in this study. Each resident independently ran two simulation cases, one involving a patient in septic shock from pneumonia ("sepsis case") and another focusing on a patient in cardiogenic shock from ST-elevation MI ("cardiogenic shock case"). Cases were obtained from the Council of Emergency Medicine Residency Directors teaching case bank [4].

Measurements

Two faculty members familiar with the Milestones and resident evaluation observed the case (one inside the room with the resident and one observing through a one-way mirror). They evaluated residents using a checklist of critical actions (expanded from the checklists included with the cases) as well as selected EM Milestones most relevant to the individual cases. Checklist items included important actions necessary for the diagnosis, stabilization and treatment of the simulation patients (**Table 1**), and residents were scored as either having performed or not performed each checklist item. Specific Milestones were selected prior to the simulation, and residents were scored from Level 1 (lowest level) to Level 5 (highest level) in increments of 0.5 (**Table 2**). After completion of the simulation cases, residents were asked to score themselves using the same Milestone scoring system without having seen the faculty assessments of their cases.

Sepsis Case	Cardiogenic Shock Case
Measure temperature	Obtain EKG
Obtain history from EMS/family	Recognize STEMI
Appropriate medications for intubation	Obtain chest x-ray
Place endotracheal tube (ETT)	Obtain cardiac enzymes
Confirm placement of ETT	Consult cardiology
Identify pneumonia on chest x-ray	Screen for contraindications for lytics
Give bolus of intravenous fluids	Administer lytic medications
Place central venous catheter	Give bolus of intravenous fluids
Initiate pressors	Initiate pressors
Give appropriate antibiotics	Transfer patient to higher level of care
Explain medical situation to mother	Explain need for transfer to patient
Act calm and professional	--
Disposition patient to ICU	--

Table 1 Checklist items.

Emergency Medicine Milestones
Emergency stabilization
Focused history and physical exam
Diagnostic studies
Diagnosis
Pharmacotherapy
Observation and reassessment
Disposition
Patient-centered communication

Table 2 Selected emergency medicine milestones.

Data analysis

The data was analyzed to assess the reliability and reproducibility of faculty assessment and resident self-assessment. The Milestone scoring performed by each faculty member was compared to the other faculty member using intra-class correlation coefficients. The faculty-assessed Milestone score for each Milestone was then averaged between the faculty, and this result was compared to resident self-assessment of Milestone level using intra-class correlation coefficients. Inter-rater agreement for faculty assessment of critical actions performed was determined using kappa statistics. Finally, Pearson and Spearman correlation coefficients were calculated to determine the correlation between the number of critical actions performed and the average Milestone score given to each resident by the faculty. This study was determined by the IRB to be exempt.

Results

Twenty-three out of possible thirty second-year residents took part in this study due to scheduling constraints, with each running identical cases on cardiogenic shock and septic shock. Fifteen of the twenty-three residents who participated were female.

Faculty observers found that residents completed an average of 91.7% of critical actions in the cardiogenic shock case and 82.2% of critical actions in the septic shock case. Between faculty members, a kappa correlation coefficient of 0.83 (95% confidence interval 0.69-0.96) indicated high inter-rater agreement for completion of critical actions performed during the cardiogenic shock case. For the sepsis case, the kappa correlation coefficient was similarly high at 0.78 (95% CI 0.68-0.88).

The intra-class correlation coefficients (ICC) for agreement between faculty observers for Milestone level were 0.12 and 0.15 for the cardiogenic shock and sepsis cases, respectively. The ICC comparing resident self-assessment with the average of faculty Milestone scoring for each case was 0.00. On average, residents scored themselves 0.70 points higher (standard deviation 0.49) than the average faculty Milestone score for the cardiogenic shock case, and 0.35 points higher (standard deviation 0.62) in the sepsis case.

Figures 1 and 2 compare the average faculty Milestone score to average percentage of checklist items completed. The Pearson and Spearman correlation coefficients comparing average faculty Milestone scoring and checklist items in the cardiogenic shock case were 0.27 and 0.29; in the sepsis case, 0.085 and -0.021.

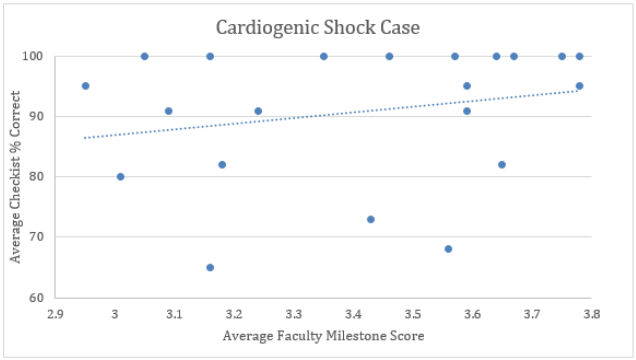

Figure 1 Comparison of faculty milestone score to checklist percentage correct for cardiogenic shock case.

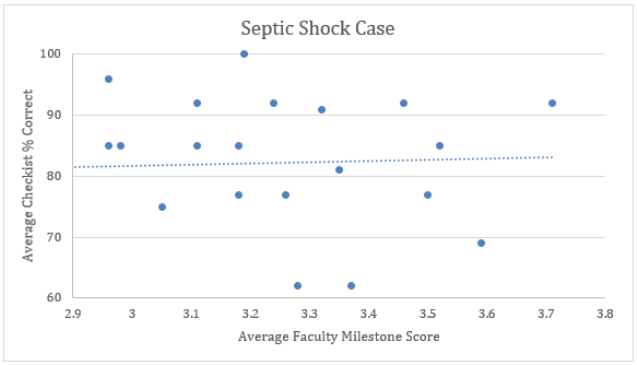

Figure 2 Comparison of faculty milestone score to checklist percentage correct for septic shock case.

Discussion

Since the ACGME introduced the six core competencies in 1999, the development of assessment methods with proven reliability and validity in measuring physician competence has proven challenging [5]. In 2008, the ACGME EM Milestones Project sought to provide standardized competency-based benchmarks to measure a resident's individual developmental progression in 23 specific domains during residency and beyond [6]. Evidence for attainment of specific Milestone levels relies on multiple direct observations of an individual resident demonstrating the effective integration of knowledge and skills over time.

While many observations occur in the clinical environment, extended in depth assessment of patient care skills in a busy ED is challenging. Simulation-based assessments provide a controlled environment for benchmarking and deliberate practice of patient care skills free of these constraints [7,8]. Checklist-based assessments, based on specific observable actions and behaviors, can be designed to assess successful execution of sequenced critical actions to determine successful achievement of a specific skillset or knowledge level. The "yes or no" nature of checklist assessments removes much of the subjectivity of resident evaluation. Our data show that using checklists for simulation-based assessment provides reliable measurements between faculty observers, consistent with other studies [9].

Milestone assessments rely on the demonstration of much broader clinical skillsets using novice-to-expert format. Observations of clinical skills may be subject to inter-observer variance based on a variety of factors, including prior experience with the learner in other settings (e.g. halo or millstone effects), expertise of the observer, perceived fluidity of the performance, and understanding of how to interpret the Milestone assessment language. This study found a poor degree of inter-rater reliability for the EM milestones related to direct patient care skills (EM Milestones 1-8) between faculty observers as well as for resident self-assessments and faculty ratings. These findings underscore the importance of anchoring Milestones assessments to specific performance

criteria to ensure that rating is more consistent and reliable [10].

Unlike the procedure-oriented Milestones, Milestones related to direct patient care skills may require a series of observed performances in order to reliably determine competence. Simulation-based assessments in which faculty observe a single patient care encounter may not provide enough opportunity for faculty to adequately assess Milestone achievement. When used in conjunction with checklists, Milestone ratings may be used to provide formative feedback to residents. However, our data indicate that there are significant limitations to using Milestones for the purposes of generating a summative assessment for single-encounter patient care simulations.

Limitations

Our study was conducted in a single, four-year residency training program; therefore, the results may have limited external validity. While faculty observers were trained on the interpretation of both checklist items and Milestones levels prior to the assessment, it is possible that misinterpretation of specific criteria persisted. Lastly, all faculty observers had previous real-world patient care experiences with the PGY-2 residents who participated in the assessment, raising the possibility of the halo and millstone effects influencing certain Milestone ratings.

Conclusion

Compared to critical action checklists, use of Milestones ratings in simulation-based assessments lacks consistency between faculty raters for simulation-based competency assessment. Resident self-assessment shows no correlation with faculty assessments. While Milestone ratings may be used to provide formative feedback to residents, there are significant limitations in using Milestones for summative assessment for single-encounter patient care simulations.

References

1. Lyle B, Borgert AJ, Kallies KJ, Jarman BT (2016) Do attending surgeons and residents see eye to eye? An evaluation of the Accreditation Council for Graduate Medical Education Milestones in General Surgery Residency. J Surg Educ 73: e54-e58.
2. Ross FJ, Metro DG, Beaman ST, Cain JG, Dowdy MM, et al. (2016) A first look at the Accreditation Council for Graduate Medical Education anesthesiology milestones: implementation of self-evaluation in a large residency program. J Clin Anesth 32: 17-24.
3. Goldflam K, Bod J, Della-Giustina D, Tsyrulnik A (2015) Emergency medicine residents consistently rate themselves higher than attending assessments on ACGME milestones. West J Emerg Med 16: 931-935.
4. www.CORDEM.org
5. Ling LJ, Beeson MS (2012) Milestones in emergency medicine. J Acute Med 2: 65-69.
6. Carraccio C, Burke AE (2010) Beyond competencies and milestones: adding meaning through context. J Grad Med Educ 2: 419-422.
7. Takayesu JK, Kulstad C, Wallenstein J, Gallahue F, Gordon D, et al. (2012) Assessing patient care: Summary of the breakout group on assessment of observable learner performance. Acad Emerg Med 19: 1379-1389.
8. McGaghie WC, Siddall VJ, Mazmanian PE, Myers J (2009) Lessons for continuing medical education from simulation research in undergraduate and graduate medical education: Effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. Chest 135: 62S-68S.
9. Ilgen JS, Ma IWY, Hatala R, Cook DA (2015) A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ 49: 161-173.
10. Boateng BA, Bass LD, Blaszak RT, Farrar HC (2009) The development of a competency-based assessment rubric to measure resident milestones. J Grad Med Educ 1: 45-48.