Longitudinal Performance Assessment of a Cohort of Anesthesia Resident Trainees

Richard H. Blum [1] , Sharon Muret-Wagstaff [2] , John R. Boulet [3] , Jeffrey B. Cooper [4] , Emil R. Petrusa [5]

1. Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital, Harvard Medical School 2. Department of Surgery, Emory University School of Medicine 3. Research and Data Resources, Foundation for Advancement of International Medical Education and Research 4. Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School 5. Department of Surgery, Massachusetts General Hospital, Harvard Medical School

✉ **Corresponding author:** Richard H. Blum, richard.blum@childrens.harvard.edu

How to cite this poster
Blum R H, Muret-Wagstaff S, Boulet J R, et al. (2017) Longitudinal Performance Assessment of a Cohort of Anesthesia Resident Trainees. Cureus 9(6): e.

Abstract

Background: Obtaining valid resident performance measures over time is critical to both patient safety and a comprehensive learning system. Earlier Harvard Assessment of Anesthesia Resident Performance (HARP) assessment work found it was feasible to assess resident performance on domains felt to be critical to safe independent practice with high reliability and initial evidence of validity.[1]

Little is known about the development of performance of anesthesiology residents from the 1st to 3rd year of training; the HARP assessment was used to characterize longitudinal performance of a group of residents in their 1st and 3rd year of training.

Methods: 20 residents independently engaged 4 scenarios designed for 1st year residents. 20 Trained anesthesiologists rated videos using the HARP rubric comprised 5 domains: 1) Synthesizes information to formulate a clear anesthetic plan, 2) Implements a plan based on changing conditions, 3) Demonstrates effective interpersonal and communication skills, 4) Identifies ways to improve performance, 5) Recognizes own limits. 10% of scenarios were rated twice for determination of inter-rater reliability.

Scenario score is the average of 5 domain ratings, domain score was the average of those domain ratings across 4 scenarios and total score was the average of all scenario scores. Descriptive statistics were used to summarize performance and provide evidence to support validity. Differences in performance by training year were analyzed using RM-ANOVA.

Results: Inter-rater reliability was 0.76. Over the 2 years, residents improved their performance by about 1/3 of a standard deviation. They improved most on scenario 3 (difference=0.9 points, ES=0.78) and did not improve on scenario 2 (difference=-0.2, ES=0.11). 12 residents had higher total scores; 8 had lower total scores.

RM-ANOVA yielded a significant time effect ($F_{1,73}=5.8$, $p<0.025$; ES=0.33). 1st year average total score was 4.4 and 3rd year was 4.8. 4 of 5 domain scores were significantly higher in the 3rd year. Domain 4 (Improve performance) was not significantly higher.

Conclusions: The HARP assessment had good rater agreement and further evidence of validity as residents' overall performance improved over the 2 years of training.

8 residents had overall lower performance and scored lower for Scenario 2 (hemorrhage). Possible explanations include a ceiling effect as the scenarios were designed for 1st year residents, expert behavior with the use of short-cuts, less motivation by the 3rd year group that was close to graduation and results were confidential; this require further study. Finally, little improvement was seen in domain 4 (identifies ways to improve performance) was consistent with our other study comparing a large cohort of 1st and 3rd year residents.[2] This domain reflects aspects of reflective practice which may not be well addressed in their training.

65