

Improvement of the predictive probability value reporting through machine learning approach: a case study

Kenji Yamazaki¹

Stanley Hamstra² and Eric Holmboe³

¹ Accreditation Council For Graduate Medical Education

² University of Toronto

Abstract

ACGME uses Milestone data to estimate residents' probability of reaching recommended milestones graduation goals. These predictive probabilities are now supported by validity evidence, including lower interpersonal communication skills (ICS) ratings associated with higher post-training patient complaints. However, ACGME's current predictive probability estimation lacks consideration of rating variability among programs. In this proof-of-concept study, we addressed these limitations using machine learning (ML) techniques to predict the penultimate milestones ratings.

We analyzed the ICS ratings in emergency medicine milestones. The training set included 5808 residents from 144 programs across four academic cohorts (2013-2016); the test set comprised 1639 residents from 141 programs in the 2017 cohort. Using 11 algorithms, we predicted which residents would be rated 3.5 or lower (5-point scale) for the ICS subcompetency during the penultimate assessment period. Predictors consisted of each resident's rating progression over the initial three semi-annual assessments, including program identifiers. Model evaluation utilized the f1-score (combination of sensitivity and positive predictive value(ppv)) and accuracy metrics on the test set.

The support vector machine algorithm with program identifiers was the best-fitting model in the training dataset, outperforming any other models without program identifiers. The test set included 959 residents (59%) from 129 programs, who were rated 3.5 or lower. Model evaluation yielded 0.78, 0.82, 0.75, and 0.74 for f-1, sensitivity, ppv, and accuracy, respectively.

Forecasting the performance status of residents at the penultimate assessment period based on the initial three assessment periods depends on program-level rating behaviors, suggesting ML approaches can produce tailored probability estimation for each resident by program.

Our approach enables program directors to identify struggling trainees earlier and support their competency development under program-specific educational conditions. Further research should extend this method to other competencies for other specialties and include trainee's and training site's characteristics relevant to trainee's early career performance.

References (maximum three)

Holmboe ES, Yamazaki K, Nasca TJ, Hamstra SJ. Using longitudinal milestones data and learning analytics to facilitate the professional development of residents: early lessons from three specialties. *Acad Med.* 2020;95(1):97–103.

Han M, Hamstra SJ, Hogan SO, Holmboe E, Harris K, Wallen E, Hickson G, Terhune KP, Brady DW, Trock B, Yamazaki K, Bienstock JL, Domenico HJ, Cooper WO. Trainee Physician Milestone Ratings and Patient Complaints in Early Posttraining Practice. *JAMA Netw Open.* 2023 Apr 3;6(4):e237588.

Hamstra SJ, Yamazaki K. A Validity Framework for Effective Analysis and Interpretation of Milestones Data. *J Grad Med Educ.* 2021 Apr;13(2 Suppl):75-80.