

ACS 2024 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

P-C-04

Research Abstracts

Correcting for Rater Effects in Surgical Skills Assessment in the Operating Room

Ryan Chou; Hajira Naz; Kofi D. O. Boahene; Jessica H. Maxwell; John R. Wanamaker, MD; Patrick J. Bryne, MD; Ira D. Papel; Theda C. Kontis; Gregory D. Hager, PhD; Lisa E. Ishii; Sonya Malekzadeh; S. Swaroop Vedula; and Masaru Ishii

Whiting School of Engineering, Johns Hopkins University, Baltimore, MD; Johns Hopkins University School of Medicine, Baltimore, MD; MedStar Georgetown University Hospital, Washington, DC; Cleveland Clinic, Cleveland, OH

Introduction: Rater effects are a source of inter-rater variation in surgical skills assessments, and can obscure the true skill of trainees. This affects the validity and utility of skill data collected for use in the training, certifying, credentialing, and privileging surgeons, the establishment of national benchmarks and surgical curricula, and machine learning methods. Our objective is to estimate and adjust for rater effects in surgical skill assessment using a structured rating scale for nasal septoplasty in the operating room.

Methods: In a prospective cohort study, attending surgeons (raters) rated residents and fellows (trainees) performing nasal septoplasty using seven items from the Septoplasty Global Assessment Tool. We fit a structural equation model with the item scores regressed on a latent component of skill, then fit a second model including the rating surgeon as a random effect to model a rater effect adjusted latent surgical skill. We validated the second model against conventional measures including the level of expertise and post graduation year (PGY) commensurate with the trainee's performance, the trainee's actual PGY, and whether the surgical goals were achieved.

Results: Our dataset included 188 assessments by 7 raters and 41 trainees. The model with one latent construct for surgical skill and the rater as a random effect was best. Some items in the rating scale had rater coefficients almost as high as their skill coefficients (Table). Rater-adjusted latent skill scores increased with attending-estimated skill levels and PGY of trainees, increased with the actual PGY, and appeared constant over different levels of achievement of surgical goals.

Conclusions: Our work provides a method to obtain rater effect adjusted surgical skill assessments in the operating room using structured rating scales. Our method allows for the creation of standardized (i.e., adjusted for rater effects) quantitative surgical skill benchmarks using national-level databases on trainee assessments.

Estimated Regression Coefficients from the Structural Equation Model

	Skill Coefficient	Skill Standard Error	Rater Coefficient	Rater Standard Error
SGAT I	1.8512	0.2868	1.5818	0.5248
SGAT II	3.2670	0.4641	1.4513	0.5050
SGAT III	3.7284	0.5576	0.5225	0.3381
SGAT IV	5.4008	1.0089	0.7813	0.4836
SGAT V	2.9306	0.4278	1.3346	0.4710
SGAT VI	5.2512	1.0216	1.7701	0.6919
SGAT VII	2.8854	0.4041	1.6775	0.5519