P301: Validation of Task-Specific Rating Scale for Open Balloon Catheter Arterial Embolectomy: An Assessor-Blinded Quasi-Experimental Pilot Study

Nadeem Ahmed Siddiqui, Amna Riaz Khan, Shiraz Hashmi, Iram Naz and Ziad Sophie
Aga Khan University Hospital, Karachi, Pakistan

INTRODUCTION

The technical Ability of a clinician is one of the most important components of surgical competency. Global Rating Scale (GRS) is used to objectively assess a trainee's surgical skill (1). However, GRS is generic and lacks the ability to assess the specific aspects of diverse surgical procedures performed (2), particularly vascular procedures. Combining the GRS and procedure-specific checklists by including additional safety and efficacy items has been found to be a more effective and reliable assessment for surgical dexterity (3).

AIM

Primary: To develop and validate a task-specific rating scale (TSRS) by comparing it with GRS to evaluate the procedural steps of Balloon arterial embolectomy (BAE).

Secondary: To estimate criterion cut-off points of the TSRS against overall GRS binary scores to declare trainees as successful candidates.

METHODS

Participants were divided into expert and novice groups who were oriented and were allowed to practice on the locally developed simulator model. The following day, arterial embolectomy procedure was performed independently by the participants on same simulator and graded by two independent assessors using the GRS and TSRS, which was self developed checklist designed specifically for BAE.

Fig 1

A locally developed simulation model of brachial artery embolectomy: (A) Arteriotomy, (B) Thrombectomy, (C) balloon catheter embolectomy, (D) closure of arteriotomy
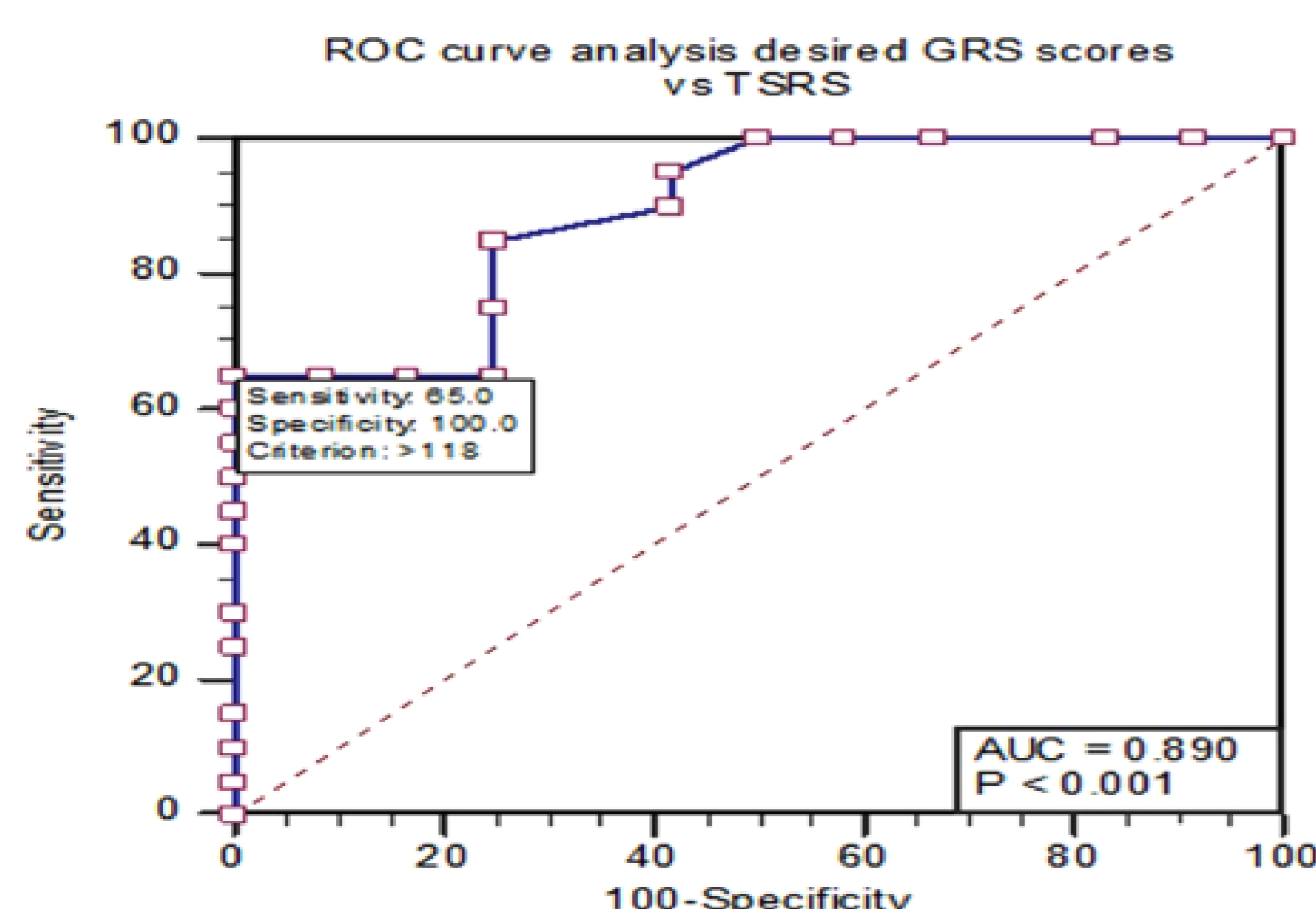

Fig. 2 Receiver operating characteristic (ROC) curve analysis showing the task-specific rating score cut points against 65% of the Global Rating Scale scores.

METHODS (Cont..)

Validity was evaluated using Pearson's correlation coefficient (r), reliability by the interclass correlation coefficient (ICC), and agreement by Bland–Altman plots. A p-value <0.05 was considered significant.

RESULTS

Thirty-two participants were enrolled in this study. The overall TSRS was found to be a valid assessment tool (r=0.82; 95% confidence interval [CI]: 0.66, 0.91; p<0.001). Domain-specific analyses showed a moderate positive association between all domains (p<0.05), except for instrument handling (r=0.09, 95% CI: -.027, 0.42: p=0.642). The ICC for overall scores showed excellent reliability for both GRS and TSRS, with values of 0.97 and 0.92, respectively. The TSRS cut point estimated on the ROC was 118, which corresponds to 65% of the overall GRS scores that can be used to discriminate participants who performed well from those who need further improvement (Fig. 2). Both the GRS and mean TSRS scores could discriminate the expert group from the novice group in overall scores (GRS: 27.8±3 .13 and 22.16± 6.24; p= 0.011; and TSRS: 32.45±1 .08 and 30.34± 1.73; p= 0.001). Instrument handling and the use of an assistant could not be differentiated by either the GRS or the TSRS.

CONCLUSIONS

The TSRS was found to be a valid and reliable assessment tool for BAE; however, for some domains, such as instrument handling and time and motion, it has limited reliability.

BIBLIOGRAPHY

1. Ahmed K, Miskovic D, Darzi A, et al. Observational tools for assessment of procedural skills: a systematic review. Am J Surg 2011; 202: 469-80.

2. Mitchell EL, Sheahan MG, Schwiesow M. Simulation in vascular surgery. In: Stefanidis D, Korndorffer Jr J, Sweet R eds. Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties. Indianapolis: Springer, Cham, 2019; 327-47.

3. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005; 190: 107-13.

Table 1 Mean Global Rating Scale (GRS) and Task- Specific Rating scale (TSRS) scores and construct validity of both instruments assessed by Pearson's correlation coefficient and 95% CI, by task: n+32

Domains	Mean GRS ±SD	Mean* TSRS ±SD	r† (95%CI)	p-Value
Respect for tissue	3.5±0.81	3.96±0.36	0.49 (0.18, 0.72)	0.004
Time and motion	3.47±1.03	4.34±0.25	0.43 (0.10, 0.68)	0.014
Instrument handling	3.44±1.05	4.67±0.23	0.09 (−0.27, 0.42)	0.642
Knotting and suturing	3.19±1.20	4.64±0.34	0.58 (0.28, 0.77)	0.001
Use of assistant	3.19±1.12	4.36±0.67	0.59 (0.31, 0.78)	<0.001
Procedural flow	3.64±0.92	4.69±0.27	0.45 (0.12, 0.69)	0.009
Quality of final product	3.5±1.00	4.35±0.60	0.60 (0.32, 0.79)	<0.001
Overall	23.92±6.02	31.0±1.83	0.82 (0.66, 0.91)	<0.001

Table 2 Reliability of the Global Rating Scale (GRS) and Task-Specific Rating Scale (TSRS); n=32

Task	Interclass correlation coefficients* (95%CI)			
	GRS	p-Value	TSRS	p-Value
Respect for tissue	0.93 (0.86, 0.97)	<0.001	0.71 (0.42, 0.86)	<0.001
Time and motion	0.94 (0.88, 0.97)	<0.001	0.23 (−0.57, 0.63)	0.233
Instrument handling	0.92 (0.84, 0.96)	<0.001	0.38 (−0.27, 0.70)	0.095
Knotting and suturing	0.92 (0.83, 0.96)	<0.001	0.94 (0.87, 0.97)	<0.001
Use of assistant	0.93 (0.85, 0.96)	<0.001	0.85 (0.70, 0.93)	<0.001
Procedural flow	0.81 (0.62, 0.91)	<0.001	0.91 (0.82, 0.96)	<0.001
Quality of final product	0.94 (0.87, 0.97)	<0.001	0.97 (0.95, 0.99)	<0.001
Overall	0.97 (0.94, 0.98)	<0.001	0.92 (0.83, 0.96)	<0.001

*The average of the two raters was analyzed.

Table 3 Comparison of domain-specific means of GRS and TSRS scores among expert and novice groups, by task; n=32

Task	GRS‡			TSRS†		
	Expert n=10	Novice n=22	p-Value	Expert n=10	Novice n=22	p-Value
Respect for tissue	4.1±0.32	3.23±0.83	0.003	4.09±0.26	3.9±0.38	0.166
Time and motion	4.1±0.74	3.18±1.03	0.017	4.43±0.17	4.3±0.27	0.180
Instrument handling	3.6±0.7	3.36±1.19	0.565	4.76±0.32	4.63±0.18	0.144
Knotting and suturing	3.9±0.94	2.86±1.19	0.021	4.87±0.23	4.53±0.33	0.007
Use of assistant	3.7±0.63	2.95±1.22	0.081	4.6±0.57	4.25±0.7	0.178
Procedural flow	4.3±0.42	3.34±0.93	0.004	4.86±0.23	4.61±0.26	0.016
Quality of final product	4.1±0.57	3.23±1.04	0.019	4.85±0.23	4.12±0.57	0.001
Overall	27.8±3.13	22.16±6.24	0.011	32.45±1.08	30.34±1.73	0.001

‡Global Rating Scale. †Average task-specific rating scale. *The p-value is derived from an independent t-test or the Mann–Whitney U test. Values are means±standard deviations.