Eur J Vasc Endovasc Surg (2009) 38, 1–9

Validation of Video-based Skill Assessment in Carotid Artery Stenting

I. Van Herzeele [a,b,*], R. Aggarwal [b], I. Malik [c], P. Gaines [d], M. Hamady [e], A. Darzi [b], N. Cheshire [b,f], F. Vermassen [a], (European Virtual Reality Endovascular Research Team EVEResT)

[a] *Department of Thoracic and Vascular Surgery, Ghent University Hospital, Gent, Belgium*
[b] *Department of Biosurgery and Surgical Technology, Imperial College Healthcare NHS Trust, London, UK*
[c] *Department of Cardiology, Imperial College Healthcare NHS Trust, London, UK*
[d] *Sheffield Vascular Institute, Northern General Hospital, Sheffield, UK*
[e] *Department of Radiology, Imperial College Healthcare NHS Trust, London, UK*
[f] *Regional Vascular Unit, Imperial College Healthcare NHS Trust, London, UK*

Submitted 19 November 2008; accepted 16 March 2009
Available online 8 April 2009

KEYWORDS
Carotid artery stent procedure;
Virtual reality simulation;
Assessment;
Video-based;
Rating scale;
Proficiency

Abstract *Objectives:* To develop weighted error-based, generic and procedure-specific rating scales, to validate these scales for video-based assessment during virtual carotid artery stent (CAS) procedures and correlate them with simulator-derived metrics.
Methods: A questionnaire was developed to assess the technique during live CAS procedures. Errors were rated from 1 (unimportant) to 5 (life-threatening) by 28 highly experienced CAS (>50 CAS) physicians. Virtual CAS procedure was performed by 21 interventionalists with varied CAS experience. Fluoroscopy screen and hand movements were video-taped, and simulator-derived metrics recorded. Experienced CAS practitioners then rated the video-taped performances using weighted error, generic and procedure-specific rating scales.
Results: Of the 23 errors assessed, 12 were regarded as moderate (score 3), six serious (score 4) and four life-threatening (score 5). The generic rating scale was able to detect significant differences in performance between inexperienced and experienced CAS operators (score 25 *vs.* 32 respectively, $P < 0.01$). All scoring systems demonstrated good inter-rater reliability ($\alpha = 0.61-0.87$). Significant correlations were observed between simulator-derived and video-based scores: weighted error-based score (r: 0.76, $P < 0.01$), generic (r: 0.62, $P < 0.01$) and procedure-specific (r: 0.76, $P < 0.01$) rating scales.
Conclusions: The generic endovascular rating scale differentiated between levels of CAS experience among skilled interventionalists and correlated to simulator-based error scoring.

* Corresponding author. I. Van Herzeele, Department of Thoracic and Vascular Surgery, 2K12IC, University Hospital Ghent, De Pintelaan 185, 9000 Gent, Belgium. Tel.: +32 9 332 5108; fax: +32 9 332 5726.
E-mail address: vhisabelle@gmail.com (I. Van Herzeele).

1078-5884/$36 © 2009 European Society for Vascular Surgery. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejvs.2009.03.008

Introduction

Methods for prevention of stroke and treatment of carotid artery stenosis currently represent a major debate in cardiovascular medicine.[1,2] Carotid artery stenting (CAS) as a treatment for carotid artery stenosis is primarily offered as an alternative to carotid endarterectomy in symptomatic patients to reduce the risk of stroke.[3] The efficacy of this intervention depends upon perioperative mortality and morbidity rates.[4] Operators must have appropriate cognitive and technical skills, proficiency and experience in CAS to maximise patient safety.[5,6]

The literature has demonstrated steep learning curves associated with CAS.[7] New operators in these complex techniques must first acquire generic skills for catheter-based interventions. It has been suggested that physicians need to have performed at least 150 supra-aortic diagnostic and/or therapeutic endovascular procedures and 75 CAS interventions prior to independent CAS practice.[1] This statement highlights that currently proficiency is based primarily upon number of procedures performed. However, case-load is a surrogate rather than an exact measure of proficiency.[4]

In other high-risk arenas such as aviation, virtual reality (VR) simulation has been used extensively to train and assess pilots. In the medical field, VR simulation may improve patient safety through introduction of risk-free training as an adjunct to traditional practice and to ensure proficiency of endovascular physicians prior to performing procedures such as CAS.[8,9]

Prior to assessment of physicians, the construct validity of VR simulators, that is, that one is measuring the trait that one purports to measure, needs to be proven. Construct validity is often affirmed by establishing that the results of performance (assessment parameters) improve with experience. Nonetheless, the current generation of VR simulators are only able to differentiate levels of CAS experience based on parameters such as total procedure and fluoroscopy time.[10–14]

Some authors have developed alternative scoring systems to assess individual technical skills such as objective structured assessment of technical skill (OSATS)-derived rating scales, which have proven to be construct-valid for both simulated and live procedures in various domains.[15–18] Others preferred to use procedure-specific checklists or rating scales whereby each part of the procedure is rated on a five-point Likert scale.[19] In contrast, error scales have rarely been used for individual assessment. However, if errors are rated according to their severity, because some are more critical than others and may lead to stroke, then this weighted error-based scoring system may prove to be useful to evaluate performances in CAS. [20,21]

The aims of this study were first to develop a weighted error-based score, generic and procedure-specific rating scales to assess technical performances during CAS procedures; second, to determine the construct validity of these rating scales used by experienced CAS interventionalists for video-based assessment of performances in a standardised simulated CAS procedure; and finally to correlate video scores with simulator-derived metrics.

Methods

Questionnaire

An initial questionnaire was created and it included the errors currently recorded by the Vascular Intervention Simulation Trainer (VIST, Mentice, Gothenburg, Sweden) as well as errors that can occur during live CAS procedures, as suggested by five interventionalists. A total of 23 possible errors were identified.[14] This questionnaire was sent to highly experienced CAS interventionalists (>50 CAS procedures),[22,23] who were asked to weight the errors on a Likert scale anchored at points 1 (unimportant), 3 (moderate) and 5 (life-threatening) (Table 1).

Development of rating scales (Fig. 1)

The weighted error-based rating scale was based on the results of the questionnaire. Each error that occurred when watching the video of a simulated CAS procedure was marked on the sheet. The sum of all errors was the total error score. This score was subsequently weighted according to the results of the questionnaire, enabling a weighted error score to be derived. The higher these scores were, the poorer the performance.

The global rating scale used in OSATS as described by Martin et al. is generally used for assessment of open technical surgical skills and consists of seven categories, each rated on a Likert scale from 1 to 5.[24] This scale was modified to enable assessment of generic endovascular skills. Since all participants were routinely assisted by the same experienced vascular individual, one category (usage of the assistant) was dropped and in analogy with the modified Reznick scale created by Hislop et al., two categories were added, that is, judgement of overall performance and final product.[12] This modified OSATS-derived rating scale, named 'generic endovascular rating scale' is meant to test various aspects of basic endovascular skills (Fig. 2). Descriptive comments for each of the eight technical domains were developed at the anchor points for scores of 1 (very poor), 3 and 5 (clearly superior). Thus the maximum score possible was 40; the higher the score, the better the quality of performance.

The procedure-specific rating scale was developed to assess endovascular skills mandatory to carry out a CAS procedure.[14] This assessment tool is a hybrid of a procedure-specific checklist and a global rating scale. Five experienced interventionalists in CAS with various medical backgrounds were involved in the development and final approval of this instrument. A total of seven categories were defined entailing the five key tasks of a CAS procedure, quality of final product and overall performance, each rated on a Likert scale from 1 to 5. Thus the maximum score possible was 35; the higher the score, the better the quality of performance.

Subjects

Twenty-one experienced endovascular physicians (>100 endovascular therapeutic interventions) were recruited at various international meetings and the simulation laboratory of Imperial College, London. Five interventional cardiologists, eight interventional radiologists and five

Table 1 The questionnaire results (median) based on the scores of 28 highly experienced interventionalists in CAS (>50).

Moving diagnostic catheter without support of guide wire	1	2	***3***	4	5
Moving guiding catheter, balloon, stent without support of guide wire	1	2	3	***4***	5
Pressing diagnostic catheter against the wall	1	***2***	3	4	5
Pressing guiding catheter against the wall	1	2	***3***	4	5
Moving 0.035 guide wire or catheter near the lesion	1	2	***3***	4	5
Losing position once CCA, ECA is cannulated	1	2	***3***	4	5
Moving embolic protection device during deployment	1	2	***3***	4	5
Moving embolic protection device after deployment	1	2	3	***4***	5
Not keeping the tip of guide wire or catheters in view	1	2	***3***	4	5
Not keeping the embolic protection device in view	1	2	3	***4***	5
No accurate pre- and post-dilation	1	2	3	***4***	5
No accurate choice and deployment of stent	1	2	3	***4***	5
Residual stenosis > 30%	1	2	***3***	4	5
Engagement of stent while removing the protection device	1	2	3	***4***	5
High fluoroscopy time	1	2	***3***	4	5
High amount of contrast used	1	2	***3***	4	5
High number of catheters used	1	2	***3***	4	5
Long time taken to complete the procedure	1	2	***3***	4	5
Poor flow of the intervention	1	2	***3***	4	5
Not monitoring heart rate during PTA + stent of carotid artery	1	2	3	4	***5***
No administration of heparin	1	2	3	4	***5***
No measures to prevent air emboli	1	2	3	4	***5***
Improper patient selection	1	2	3	4	***5***

1, unimportant error; 2, minor error; 3, moderate error; 4, serious error; 5, life-threatening error.

vascular surgeons participated. Over 70% had performed at least 500 endovascular interventions as primary operator.

They were subdivided into four categories according to CAS experience: six inexperienced (0 CAS), three minor (1—20 CAS), five moderately (21—50 CAS) and seven highly experienced (>50 CAS). The purpose of the study was explained to all the physicians. The group with minor experience in CAS only included three participants and was therefore excluded from further analysis. Thus, only the performances of inexperienced, moderately and highly experienced groups were evaluated in this article.

Simulation and task performed

Initial didactic teaching regarding the VIST simulator was delivered, followed by familiarisation with the VR simulator

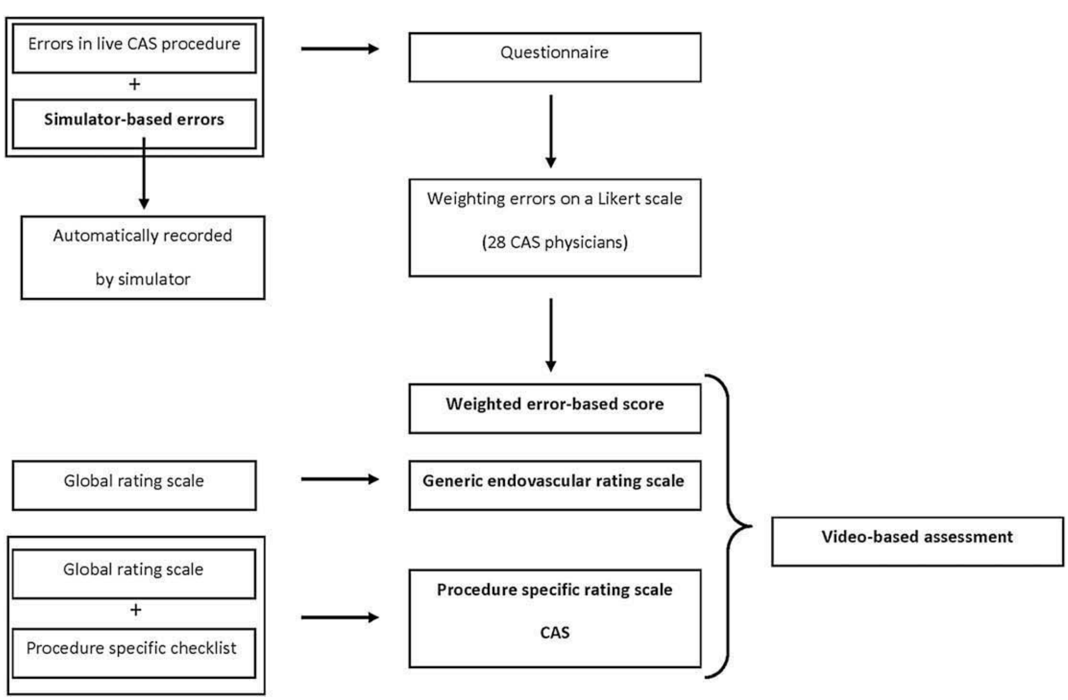

Figure 1 Scheme of the creation of the various scoring methods.

Objective structured assessment of technical skills
Modified Global rating scale: generic endovascular skills

Interventionalist code: Procedure: Date:

Please circle the physician's performance on the following scale:

Respect for tissue (stenosis or occlusion)	1 Frequently used unnecessary force on tissue or caused damage by inappropriate use of material	2	3 Careful handing of tissue but occasionally caused inadvertent damage.	4	5 Consistently approached tissues appropriately with minimal damage.
Time and motion	1 Make unnecessary moves.	2	3 Efficient time/motion but some unnecessary moves.	4	5 Clear economy of movement and maximum efficiency.
Knowledge of endovascular material	1 Frequently asked for the wrong tool or used an inappropriate material	2	3 Knew names of most endovascular toolsl and used appropriate material	4	5 Obviously familiar with endovascular material and their names.
Handling of endovascular material	1 Repeatedly awkward moves and unsure with loss of access, poor stability of the tools and inaccurate positioning of balloon/stent	2	3 Competent use with hardly any loss of access, moderate stability of tools and good positioning of balloon/stent but appeared stiff and awkward occasionally	4	5 Fluid movements with stability of the tools, maintenance of access and perfect positioning of balloon/stent
Flow of intervention	1 Frequently stopped intervention or needed to discuss the next move.	2	3 Demonstrated some forward planning and reasonable progression of procedure.	4	5 Obviously planned course of intervention with efficiency from one move to another
Knowledge of procedure	1 Insufficient knowledge. Looked unsure and hesitant.	2	3 Knew all important steps of the intervention.	4	5 Demonstrated familiarity with all steps of the intervention.
Overall performance	1 Very poor	2	3 Competent	4	5 Clearly superior
Quality of final product	1 Very poor	2	3 Competent	4	5 Clearly superior

Figure 2 OSATS-derived rating scale, generic rating scale.

by endovascular stent placement of an ipsilateral common iliac artery stenosis (Fig. 3). The subsequent CAS procedure entailed stenting of a right-sided CAS with type I aortic arch using a wire-mounted filter. Passive assistance was provided by an experienced endovascular team, and a protocol was available detailing the steps of the CAS procedure.

Performance evaluation

Simulator-based assessment

The VR simulator consists of a desktop personal computer and two monitors coupled to an interface device that allows the user to insert and manipulate endovascular tools.[22] The interface device functions as a virtual patient with a simulated groin. This simulator can assess performance during CAS procedures by objectively and instantly recording quantitative metrics (total procedure and fluoroscopy time) and measurements of quality of performance (clinical parameters and errors) (Table 2).

Video-based assessment

Video recording

In order to record the fluoroscopy screen and hand movements of the subjects, two video cameras were mounted on tripods and focussed upon the LCD monitor and the operator's hands. Recording commenced upon entry of the guide wire into the femoral artery and was completed upon removal of endovascular tools from the groin. Complete, unedited videos of each procedure were recorded into Microsoft Windows .avi format (Microsoft Corporation,

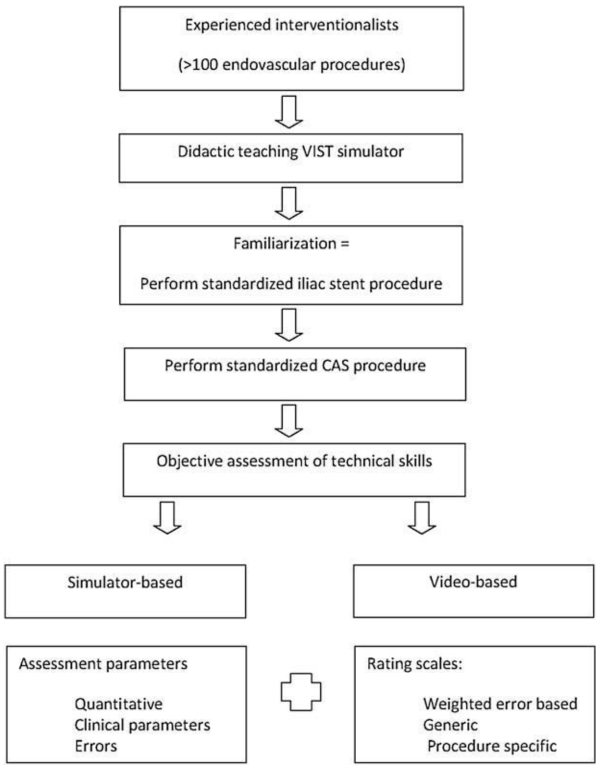

Figure 3 Methods used for the evaluation of endovascular performances of experienced interventionalists.

Redmond, WA, USA). All data files were coded by an alphanumeric code to ensure subject identity was blinded.

Video-based rating
Three interventionalists who had significant experience in techniques of CAS evaluated performances during virtual CAS procedures by observing the recorded videos. These assessors, not involved in the data collection, separately evaluated and graded performances using a weighted error-based score, generic and procedure-specific rating scales. Time required per review was not recorded.

Statistics

The non-parametric distribution of the data entailed the use of non-parametric tests of significance. Derivation of the weighted error-based rating scale was performed by frequency analysis of the scores from the 28 experienced CAS interventionalists. Construct validity of each scale was based upon a between-group comparison using the Kruskal–Wallis test of significance. The Spearman's rank test was used to correlate video-based scores with simulator-derived metrics. Finally, inter-rater reliability between the three video assessors was calculated with the Cronbach's alpha test statistic. A P value of <0.05 was considered statistically significant.

Results

Questionnaire

Four interventional cardiologists, 11 interventional radiologists and 13 vascular surgeons completed the questionnaire. No errors were regarded by these 28 experienced CAS physicians as unimportant (i.e., 1) and only one was rated as a minor mistake (i.e., 2). The majority of the errors were considered to be moderate ($n = 12$) or serious ($n = 6$) and focussed mainly on technical errors. Four errors were rated as life-threatening (i.e., 5). These comprised absence of physiological monitoring of the patient, incorrect drug administration, failure to prevent air emboli and inappropriate patient selection (Table 1).

Performance assessment

Video-based assessment
All videos were complete and of adequate quality for video-based analysis.

Construct validity of rating scales
Comparison between the three different groups during simulated CAS procedures revealed statistically significant differences for only the generic OSATS-derived rating scale (Table 3).

Inter-rater reliability of rating scales
The inter-rater reliability coefficient alpha for each scale is shown in Table 3, revealing both the OSATS-derived global rating scale and procedure-specific rating scale possess excellent inter-rater reliability (>0.80).

Simulator-based assessment
None of the simulator metrics were able to differentiate level of CAS experience of the participants: total procedure time (medians 17 min vs. 19 min vs. 14 min, $P = 0.14$ for inexperienced, moderately experienced and highly experienced operators, respectively) and fluoroscopy time (8 min vs. 10 min vs. 7 min, $P = 0.14$). Simulator-based error scoring was also not a valid mode of assessment (16 min vs. 23 min vs. 20 min, $P = 0.85$).

Correlation between simulator-based metrics and video-based ratings
Errors recorded by the simulator such as movement of the embolic protection device after deployment showed positive correlations with video-based weighted error scores and inverse correlations with generic and procedure-specific rating scales. There were statistically significant correlations between the number of errors recorded by the simulator and video-based rating scores (Table 4, Fig. 4). There were no significant correlations between quantitative metrics and *post hoc* video-based rating scores.

Discussion

Traditional modes of characterising satisfactory operative or endovascular performance have relied upon the numbers of procedures performed, senior evaluation[20,25] or completion of a prescribed course.[26] These recommendations rely upon crude data that are recognised to be unreliable and indirect measures of technical skill.[27] Winckel et al. have shown evaluation of performance in the operating room to be difficult, leading to efforts focussed upon

Table 2 Metrics recorded during a CAS procedure by the VIST simulator.

Quantitative metrics	Qualitative metrics	
	Errors	Clinical parameters
Total time (min)	Catheter vessel errors — Pressing diagnostic catheter against the wall — Pressing guiding catheter against the wall	Tools used during procedure — Diagnostic catheter — Guiding catheter or sheath — EPD size/angle tip — Predilation balloon size/length — Stent size/length — Postdilation balloon size/length
Total amount of contrast fluid used (cm^3)	Catheter movement errors — Moving diagnostic catheter without support of a guide wire — Moving guiding catheter without support of a guide wire	Placement accuracy — Predilation balloon — Stent — Postdilation balloon
Number of cineloops	Moving near lesion — Moving guide wire near lesion — Moving diagnostic catheter near lesion — Moving guiding catheter near lesion	% Lesion covered with — Balloon — Stent
Total fluoroscope time (min)	Moving EPD — During deployment — After deployment	Balloon-vessel ratio: Stent-vessel ratio
	Deployment of stent — Moving stent during deployment — In guiding catheter or sheath Balloon inflated in guiding catheter or sheath	Max pressure reached during deployment — Balloon — Stent
		Residual stenosis after — Predilation — Stent — Postdilation

techniques to standardise assessment outside the interventional suite.[28]

More recently, VR simulation has been suggested as a mode of objective assessment of performance, prior to intervention on patients.[10–13,22] However, the current generation of simulators is only able to differentiate level of experience in CAS based on the quantitative assessment parameters.[11,13,22] These surrogate markers for skill (total procedure and fluoroscopy time, amount of contrast given during the intervention) do not measure the quality of performance during a CAS intervention.[29]

What are the ultimate measures of proficiency to assess the performance of an interventionalist who carries out a CAS procedure? The most frequently measured and reported end points to evaluate surgical performances are patient outcomes and complications.[30] However, the use of morbidity and mortality data as the sole indicators of performance has its limitations. They do not differentiate the role of technical skills of the surgeon from patient pathophysiological risk factors and factors related to teamwork and dynamics.[31] Furthermore, these morbidity data do not provide prescriptive information of how errors

Table 3 Validity and inter-rater reliability of the video-based scores by experienced interventionalists and simulator-based metric.

		0 CAS $n = 6$	21–50 CAS $n = 5$	>50 CAS $n = 7$	P value	Inter-rater reliability (α)
Simulator errors	Median IQR	16 11–18	23 13–26	20 11–29	0.85	NA
Weighted error score	Median IQR	12 9–17	7 4–21	12 4–22	0.55	0.61
Generic rating scale	Median IQR	25 23–26	32 25–39	32 24–38	<0.01*	0.84
CAS rating scale	Median IQR	23 20–26	26 13–32	26 19–32	0.52	0.87

CAS, carotid artery stent; IQR, interquartile range.
*$P < 0.05$.

Table 4 Correlation between simulator-based error scoring during a CAS procedure and the blinded video-based scorings by experienced interventionalists (Spearman's rank test *r*).

Simulator errors	Post hoc ratings	Weighted error score	Generic score	CAS score
Catheter movement	*r*	0.52	−0.56	−0.53
error[a]	*P* value	0.03	0.02	0.02
Moving protection	*r*	0.73	−0.57	−0.60
device after deployment	*P* value	<0.01	0.01	<0.01
Catheter vessel	*r*	0.54	NS	−0.62
error[b]	*P* value	0.02		<0.01
VIST total errors	*r*	0.76	− 0.62	−0.76
	P value	<0.01	<0.01	<0.01

VIST, vascular intervention simulation trainer.
[a] Movement of a guiding catheter too close to the lesion.
[b] Selective catheter is scraping the vessel wall.

are made and are unable to specify how the performance of a procedure can be improved.[32]

Subsequently, most investigators have begun to evaluate the performance itself instead of evaluating the outcomes of the intervention.[31,33] The definition, identification and measurement of errors is one method by which proficiency can be assessed.[34] The more errors an endovascular physician causes, the more likely it is that the procedure will have a poor outcome. Therefore, in this study, a weighted error-based rating scale has been developed based on 28 questionnaires completed by highly experienced interventionalists in CAS.[14] Similar to open and laparoscopic surgery, generic and procedure-specific rating scales were created to allow assessment of endovascular technical skills, which have been shown to be valid and reliable measures of technical skill in other domains.[24,35]

In this study, the three different rating scales were compared on a simulated carotid module as a preliminary to do the same in real carotid cases. The weighted error-based score was unable to differentiate level of CAS experience probably since the simulator did not record life-threatening errors such as lack of monitoring of physiological measures, not administering crucial drugs and failure to prevent air emboli. Therefore these mistakes which are part of the weighted error-based score could not be picked up during video assessments. Thus this score primarily took notes of errors that were recorded by the simulator.

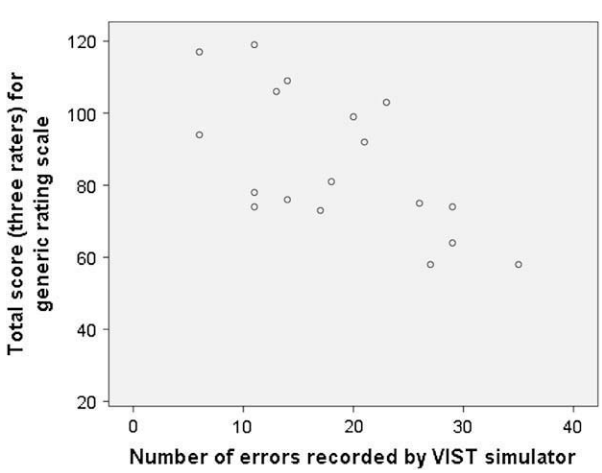

Figure 4 Scatter plot demonstrating the correlation between the simulator-based error scoring and the *post hoc* video-based total scores by expert raters using the generic rating scale. (Spearman's rank test $r = -0.62$, $P < 0.01$).

Only the generic (OSATS-derived) rating scale, employed by the three experienced raters to assess endovascular performance, demonstrated construct validity. Interventionalists, who scored high on the generic rating scale, were more likely to be experienced in CAS. A wide variability within these groups was noted (Table 2), which underlines the importance of assessing a performance objectively rather than relying on the number of procedures to demonstrate proficiency prior to allowing independent practice in actual patients.

In the endovascular literature, Hislop et al. have also proven the construct validity of an OSATS-derived modified Reznick scale used for *post hoc* video-based rating by two blinded observers during a virtual selective carotid angiography.[12] Tedesco et al. demonstrated that a single blinded expert observer using a structured global rating scale was able to discern differences in endovascular experience during a virtual renal artery stent procedure.[36] However, these studies included large groups of inexperienced endovascular doctors. Although our study included solely experienced interventionalists, it confirms these findings.

The procedure-specific rating scale was not valid possibly because the carotid lesion was not complex enough to differentiate the performances in experienced interventionalists. Within our department, recent work in the laparoscopic field came to a similar conclusion, namely that there was no benefit in the use of a procedure-specific scale to rate performance in complex procedures.[15,37] However, since it deconstructs the CAS procedure into its key components, it may have an important role to provide formative feedback to the physician by identifying areas of weakness in CAS while a generic scale may be more beneficial for summative assessment.[4,15] Both global and procedure-specific rating scales can be used to demonstrate the improvement in technical performance after VR endovascular training of individuals on computer-based or bench models and in real life.[38,39]

Our group and others have demonstrated that quantitative metrics are able to differentiate CAS experience in experienced interventionalists during the same CAS module.[11,13,22] In contrast, the simulator-based metrics in this study were unable to differentiate level of CAS experience possibly because of the small number of participants in this study. Nevertheless, a strong correlation was perceived between the blinded video-based generic and procedure-specific rating scores and simulator-based errors during a CAS procedure suggesting concurrent validity of the latter mode of assessment. The simulator metrics that demonstrated moderate-to-good inverse correlations with video-based scores assessed both generic endovascular skills (catheter movement and vessel error) and procedure-specific skills such as keeping the protection device stable during the CAS procedure (Table 4).

Potential limitations introduced in this study include a relatively small number of participants that may have caused a type II error. Furthermore, these participants were divided into arbitrary groups, who may not have been a representative sample of interventionalists in the real world.

Assessment occurred during a single CAS procedure. A competent interventionalist may perform better when being assessed; others do less well owing to anxiety.[33] The chosen virtual CAS module was a non-complex lesion and may have been too straightforward to differentiate the two more experienced groups. Moreover, only technical skills were evaluated since performances were assessed by observing videos of the fluoroscopy screen and hand movements; any differences in use of assistants, non-verbal communication and distractions may not have been noted.[15,40]

In addition, participants may not have treated the simulator seriously because they realised that it was not a real patient. Ultimately the transferability of objective assessment using the valid generic rating scale of endovascular CAS procedures to real live cases needs to be proven as well as the correlation between skills and patient outcomes (stroke, survival and restenosis).

In the future, optimal (summative) assessment of trainees and interventionalists may be a combination of a valid structured global rating scale along with objective computerised measurements that will best reflect the subject's skill level. Standardised feedback during training may be delivered using procedure-specific rating scales (formative feedback).

Acknowledgements

The authors are extremely grateful to Mr Robert Brightwell, Mr Andrew Choong and Mr Simon Neequaye for their valuable assistance through the data collections. This study was supported by funding from the Imperial College Healthcare Biomedical Research Centre, UK; a Clinical Doctoral Grant from the Fund for Scientific Research — Flanders, Belgium; grants from Mentice, Gothenburg, Sweden and from Boston Scientific International, France. However they did not interfere in the study design, data collection, analysis or interpretation of the data or in the decision to submit the article for publication.

Conflict of Interest

None.

References

1 Cremonesi A, Setacci C, Bignamini A, Bolognese L, Briganti F, Di Sciascio G, et al. Carotid artery stenting: first consensus document of the ICCS-SPREAD Joint Committee. *Stroke* 2006;**37**(9): 2400—9.

2 Naylor AR. Where next after SPACE and EVA-3S: 'the good, the bad and the ugly!'. *Eur J Vasc Endovasc Surg* 2007;**33**(1):44—7.

3 Hobson 2nd RW, Mackey WC, Ascher E, Murad MH, Calligaro KD, Comerota AJ, et al. Management of atherosclerotic carotid artery disease: clinical practice guidelines of the Society for Vascular Surgery. *J Vasc Surg* 2008;**48**(2):480—6.

4 Beard JD, Choksy S, Khan S. Assessment of operative competence during carotid endarterectomy. *Br J Surg* 2007;**94**(6): 726—30.

5 Hobson 2nd RW, Howard VJ, Roubin GS, Ferguson RD, Brott TG, Howard G, et al. Credentialing of surgeons as interventionalists for carotid artery stenting: experience from the lead-in phase of CREST. *J Vasc Surg* 2004;**40**(5):952—7.

6 Lin PH, Bush RL, Peden E, Zhou W, Kougias P, Henao E, et al. What is the learning curve for carotid artery stenting with neuroprotection? Analysis of 200 consecutive cases at an academic institution. *Perspect Vasc Surg Endovasc Ther* 2005; **17**(2):113—23. discussion 123—15.

7 Fiehler J, Jansen O, Berger J, Eckstein HH, Ringleb PA, Stingele R. Differences in complication rates among the centres in the SPACE study. *Neuroradiology* 2008;**50**(12):1049—53.

8 Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg* 2007;**246**(5): 771—9.

9 Dawson S, Gould DA. Procedural simulation's developing role in medicine. *Lancet* 2007;**369**(9574):1671—3.

10 Aggarwal R, Black SA, Hance JR, Darzi A, Cheshire NJ. Virtual reality simulation training can improve inexperienced surgeons' endovascular skills. *Eur J Vasc Endovasc Surg* 2006;**31**(6): 588—93.

11 Dayal R, Faries PL, Lin SC, Bernheim J, Hollenbeck S, DeRubertis B, et al. Computer simulation as a component of catheter-based training. *J Vasc Surg* 2004;**40**(6):1112—7.

12 Hislop SJ, Hsu JH, Narins CR, Gillespie BT, Jain RA, Schippert DW, et al. Simulator assessment of innate endovascular aptitude versus empirically correct performance. *J Vasc Surg* 2006;**43**(1):47—55.

13 Hsu JH, Younan D, Pandalai S, Gillespie BT, Jain RA, Schippert DW, et al. Use of computer simulation for determining endovascular skill levels in a carotid stenting model. *J Vasc Surg* 2004;**40**(6):1118—25.

14 Van Herzeele I, Aggarwal R, Neequaye S, Hamady M, Cleveland T, Darzi A, et al. Experienced endovascular interventionalists objectively improve their skills by attending carotid artery stent training courses. *Eur J Vasc Endovasc Surg* 2008;**35**(5):541—50.

15 Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Ann Surg* 2008; **247**(2):372—9.

16 Friedman Z, Katznelson R, Devito I, Siddiqui M, Chan V. Objective assessment of manual skills and proficiency in performing epidural anesthesia-video-assisted validation. *Reg Anesth Pain Med* 2006;**31**(4):304—10.

17 Swift SE, Carter JF. Institution and validation of an observed structured assessment of technical skills (OSATS) for obstetrics and gynecology residents and faculty. *Am J Obstet Gynecol* 2006;**195**(2):617—21. discussion 621—623.

18 Watterson JD, Beiko DT, Kuan JK, Denstedt JD. Randomized prospective blinded study validating acquistion of ureteroscopy

skills using computer based virtual reality endourological simulator. *J Urol* 2002;**168**(5):1928–32.

19 Dath D, Regehr G, Birch D, Schlachta C, Poulin E, Mamazza J, et al. Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. *Surg Endosc* 2004;**18**(12):1800–4.

20 Hanna GB, Shimi SM, Cuschieri A. Randomised study of influence of two-dimensional versus three-dimensional imaging on performance of laparoscopic cholecystectomy. *Lancet* 1998; **351**(9098):248–51.

21 Tang B, Hanna GB, Joice P, Cuschieri A. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg* 2004;**139**(11):1215–20.

22 Van Herzeele I, Aggarwal R, Choong A, Brightwell R, Vermassen FE, Cheshire NJ. Virtual reality simulation objectively differentiates level of carotid stent experience in experienced interventionalists. *J Vasc Surg* 2007;**46**(5):855–63.

23 Wholey MH, Wholey M, Mathias K, Roubin GS, Diethrich EB, Henry M, et al. Global experience in cervical carotid artery stent placement. *Catheter Cardiovasc Interv* 2000;**50**(2): 160–7.

24 Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997;**84**(2): 273–8.

25 Elliot DL, Hickam DH. Evaluation of physical examination skills. Reliability of faculty observers and patient instructors. *JAMA* 1987;**258**(23):3405–8.

26 Gaines P, Nicholson T. A suggested training programme for carotid artery stenting (CAS). *Eur J Radiol* 2006;**60**(1):37–9.

27 Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *BMJ* 2003;**327**(7422):1032–7.

28 Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. *Am J Surg* 1994;**167**(4):423–7.

29 Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. *BMJ* 1999;**318**(7188):887–8.

30 Campbell B. Surgical morbidity and mortality meetings. *Ann R Coll Surg Engl* 1988;**70**(6):363–5.

31 Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;**239**(4):475–82.

32 Joice P, Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery – a human reliability analysis. *Appl Ergon* 1998; **29**(6):409–14.

33 Beard JD. Assessment of surgical competence. *Br J Surg* 2007; **94**(11):1315–6.

34 Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Andersen DK, Satava RM. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc* 2004;**18**(4):592–5.

35 Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative ''bench station'' examination. *Am J Surg* 1997;**173**(3):226–30.

36 Tedesco MM, Pak JJ, Harris Jr EJ, Krummel TM, Dalman RL, Lee JT. Simulation-based endovascular skills assessment: the future of credentialing? *J Vasc Surg* 2008;**47**(5):1008–14.

37 Aggarwal R, Boza C, Hance J, Leong J, Lacy A, Darzi A. Skills acquisition for laparoscopic gastric bypass in the training laboratory: an innovative approach. *Obes Surg* 2007;**17**(1):19–27.

38 Berry M, Lystig T, Beard J, Klingestierna H, Reznick R, Lonn L. Porcine transfer study: virtual reality simulator training compared with porcine training in endovascular novices. *Cardiovasc Intervent Radiol* 2007;**30**:455–61.

39 Chaer RA, Derubertis BG, Lin SC, Bush HL, Karwowski JK, Birk D, et al. Simulation improves resident performance in catheter-based intervention: results of a randomized, controlled study. *Ann Surg* 2006;**244**(3):343–52.

40 Scott DJ, Rege RV, Bergen PC, Guo WA, Laycock R, Tesfay ST, et al. Measuring operative performance after laparoscopic skills training: edited videotape versus direct observation. *J Laparoendosc Adv Surg Tech A* 2000;**10**(4):183–190.