Simulation-based endovascular skills assessment: The future of credentialing?

Maureen M. Tedesco, MD, Jimmy J. Pak, MD, E. John Harris Jr, MD, Thomas M. Krummel, MD, Ronald L. Dalman, MD, and **Jason T. Lee, MD,** *Stanford, Calif*

Objectives: **Simulator-based endovascular skills training measurably improves performance in catheter-based image-guided interventions. The purpose of this study was to determine whether structured global performance assessment during endovascular simulation correlated well with trainee-reported procedural skill and prior experience level.**

Methods: **Fourth-year and fifth-year general surgery residents interviewing for vascular fellowship training provided detailed information regarding prior open vascular and endovascular operative experience. The pretest questionnaire responses were used to separate subjects into low (<20 cases) and moderate (20 to 100) endovascular experience groups. Subjects were then asked to perform a renal angioplasty/stent procedure on the Procedicus Vascular Intervention System Trainer (VIST) endovascular simulator (Mentice Corporation, Gothenburg, Sweden). The subjects' performance was supervised and evaluated by a blinded expert interventionalist using a structured global assessment scale based on angiography setup, target vessel catheterization, and the interventional procedure. Objective measures determined by the simulator were also collected for each subject. A postsimulation questionnaire was administered to determine the subjects' self-assessment of their performance.**

Results: **Seventeen surgical residents from 15 training programs completed questionnaires before and after the exercise and performed a renal angioplasty/stent procedure on the endovascular simulator. The beginner group (n = 8) reported prior experience of a median of eight endovascular cases (interquartile range [IQR], 6.5-17.8; range, 4-20), and intermediate group (n = 9) had previously completed a median of 42 cases (IQR, 31-44; range, 25-89, *P* = .01). The two groups had similar prior open vascular experience (79 cases vs 75, *P* = .60). The mean score on the structured global assessment scale for the low experience group was 2.68 of 5.0 possible compared with 3.60 for the intermediate group (*P* = .03). Scores for subcategories of the global assessment score for target vessel catheterization (*P* = .02) and the interventional procedure (*P* = .05) contributed more to the differentiation between the two experience groups. Total procedure time, fluoroscopy time, average contrast used, percentage of lesion covered by the stent, placement accuracy, residual stenosis rates, and number of cine loops utilized were similar between the two groups (*P* > .05).**

Conclusion: **Structured endovascular skills assessment correlates well with prior procedural experience within a high-fidelity simulation environment. In addition to improving endovascular training, simulators may prove useful in determining procedural competency and credentialing standards for endovascular surgeons. (J Vasc Surg 2008;47:1008-14.)**

The teaching and assessment of technical skills on high-fidelity simulation systems has its roots in the aviation industry and has become an important consideration in surgical education.[1,2] Simulated training environments allow for practice in a realistic setting without the inherent risk of harming others or oneself, or both. During the past decade, surgical simulators have been developed and popularized as a component of skills training for a variety of minimally invasive surgical procedures, including laparoscopic techniques in general surgery, obstetrics/gynecology, and urology.[1] Training on these virtual reality simulation systems has been found to improve the actual operating room performance of surgical residents.[2]

Vascular surgeons have embraced minimally invasive approaches for treatment of aortoiliac, cerebrovascular, and peripheral occlusive diseases. These new techniques have required that training programs adapt and offer the acquisition of advanced endovascular skills to their trainees as well as practicing vascular surgeons. Endovascular training is particularly well suited to the high-fidelity simulator environment, because the basic concept of manipulating a wire or catheter in a three-dimensional field while viewing it on a two-dimensional screen can be modeled within a very realistic user interface. In recognition of its unique utility in this regard, the United States Food and Drug Administration in 2004 accepted a proposal that virtual reality simulation-based training be an important part of a training package for carotid stenting.[3,4] The inherently high-risk nature of this particular procedure, coupled with the lack of uniformly accepted credentialing guidelines for catheter-based procedures, led to this recommendation.

Early adoption of endovascular simulation as an education tool has allowed several groups to preliminarily validate training courses or to document procedural improvement on either objective or "expert" graded global rating

From the Division of Vascular Surgery and the Department of Surgery, Stanford University Medical Center.

Competition of interest: Dr Lee has received an educational grant from Cordis Endovascular to study simulation-based learning. Drs Lee, Dalman, and Krummel received technical grants from Cordis Endovascular in the form of two endovascular simulators; however, these two simulators received in the grant were not used in this project.

Presented at the Twenty-second Annual Meeting of the Western Vascular Society, Kona, Hawaii, Sep 8-11, 2007.

Reprint requests: Jason T. Lee, MD, Division of Vascular Surgery, Stanford University Medical Center, 300 Pasteur Dr, Ste H3600, Stanford, CA 94305 (e-mail: jtlee@stanford.edu).

0741-5214/$34.00

doi:10.1016/j.jvs.2008.01.007

Fig 1. Vascular Interventional Surgical Trainer (VIST), Mentice Corporation.

scales after simulation-based training.[5,6] Multiple studies demonstrate improved performance in medical students, residents, fellows, and attendings of multiple specialties after simulation-based training, with generally positive feedback from learners after focused educational sessions.[1,5-7] More recent applications of simulation technology have focused on skills assessment. The concept of construct validity, that is, that the simulator can actually distinguish real-world skills assessment, can only be applied when evidence confirms that skilled individuals achieve superior simulator-assessed outcomes.[8] Skills assessment remains an important theoretic concern, because in the United States, no formal structured assessment of the technical competency exists for residents, fellows, or attending vascular surgeons.

The purpose of this study was to investigate the potential utility of simulation-based skills assessment by analyzing whether a structured global performance assessment and objective computerized measurements during endovascular simulation correlated well with self-reported procedural skill and prior experience level.

METHODS

Participants. This study was approved by Stanford University's Institutional Review Board for studies involving human subjects as an exempt protocol that did not require written consent. Participation in the simulation exercise is part of the standard educational curriculum within the residency program. The study population comprised 17 fourth- and fifth-year general surgery residents interviewing for vascular fellowship training. This cohort recorded information regarding prior open vascular and endovascular operative experience. Data collected consisted of months spent on a vascular surgery service, estimated numbers of major index vascular procedures, and estimated numbers of specific endovascular procedures, including diagnostic arteriograms, aortic stent grafts, lower extremity angioplasty/stenting, renal stenting, and carotid stenting. Subjects were not asked specifically about prior simulator experience.

Subjects were then asked to perform a renal angioplasty and stenting module on the Procedicus Vascular Intervention System Trainer (VIST) simulator (Mentice Corporation, Gothenburg, Sweden). Throughout the simulation exercise, subject proficiency was evaluated on a standardized global rating scale with specific checkpoints of key portions of the procedure by an experienced interventionalist who was blinded to the subjects' experience level.

Simulator. The Goodman Simulation Center is a multidisciplinary effort within the Department of Surgery at Stanford University Medical Center and is accredited by the American College of Surgeons as a Level I Educational Institute. The endovascular curriculum uses the Mentice VIST system, which consists of a mechanical interface device, a high-performance desktop computer, and two display screens (Fig 1). The simulation software allows the selection of dozens of different carotid, renal, iliac, and superficial femoral artery scenarios. The interface device is designed as the virtual patient with simulated right femoral arterial access. Three separate haptic units allow the introduction of real guidewires, guiding catheters, and balloons/stents into simulated images. Separate controls are provided for the stent deployment, contrast injection, fluoroscopic C-arm and table movement, cine video runs, roadmapping capabilities, and measurements.

Various objective measures of the procedure are recorded and tabulated for each participant by the system software. For the purposes of this study, a technician trained in using and servicing the simulator was present for each participant's exercise and provided a standardized introduction to the simulator. Subjects were not, however, allowed time to practice before testing. There were no technical issues with the simulator during the study.

Renal artery stenosis model. Subjects performed a right renal angioplasty and stenting module on a high-grade ostial lesion on the VIST simulator (Fig 2). Throughout the simulated procedure, subject performance was evaluated on a global rating scale by an experienced interventionalist who was blinded to the subjects' background or prior experience. The examiner was given strict instructions not to assist the subject with the procedure unless solicited. If the subject asked for advice from the examiner, this was reflected in the score. The examiner used a structured global rating scale focusing on the subject's ability to perform an angiogram, gain access to the target artery, and perform the angioplasty/stent procedure (Table I). This global ratings checklist was modified from examiner checklists used in several other studies of endovascular simulation[7,9] in which the scales were modified from

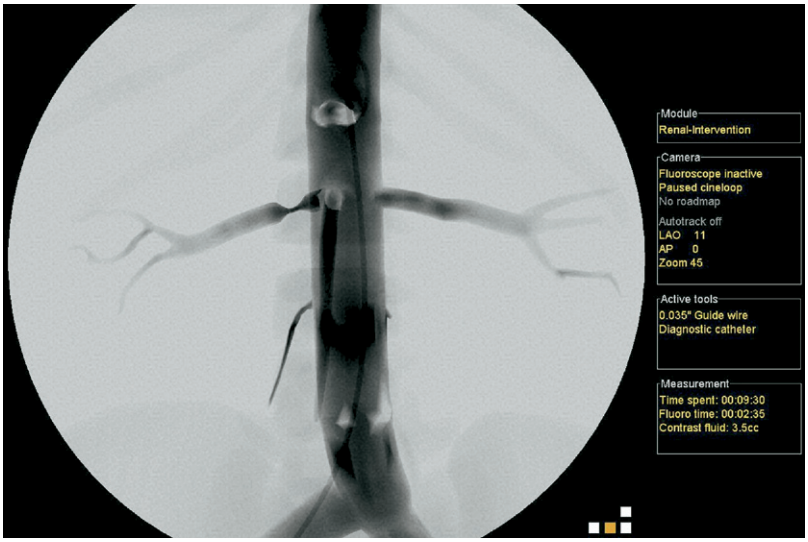

Fig 2. Ostial lesion of right renal artery treated during simulation exercise.

Table I. Score sheet used to assess individual categories in the renal artery angioplasty/stenting global rating scale[a]

Category	*Subcategory*	*Grade*				
Angiogram	• advance wire into suprarenal aorta without forming a J or pushing against obstruction	1	2	3	4	5
	• place pigtail catheter into renal angiogram position/wire manipulation	1	2	3	4	5
	• knowledge of renal anatomy/ perform angiogram	1	2	3	4	5
Wire access	• select proper catheter/wire for renal canalization	1	2	3	4	5
	• safely traverse lesion	1	2	3	4	5
Intervention	• select guiding catheter	1	2	3	4	5
	• select appropriate renal stent	1	2	3	4	5
	• deploy renal stent	1	2	3	4	5
	• select proper balloon for renal angioplasty post-stent	1	2	3	4	5
	• perform completion angiogram	1	2	3	4	5

[a]For each task the expert observer scored the subject on a 1-5 scale: 1 = *fail*, 2 = *poor*, 3 = *satisfactory*, 4 = *good*, 5 = *superior*. See the text for descriptions of types of techniques that led to certain grades. The score for each category was averaged, and the average of the three categories (weighted equally) was calculated to determine the subjects' global rating score.

previously validated skills assessment scales popularized by Reznick et al,[8] which consists of seven categories used to rate operative performance.

The modified global assessment scorecard we used separated the procedure into angiography, wire access, and intervention subcategories. Within each subcategory, the examiner used a traditional Likert scale score of 1 = *fail*, 2 = *poor*, 3 = *satisfactory*, 4 = *good*, and 5 = *superior*. Examples of criteria for a fail grade (score, 1) would be frequently stopping the procedure, clearly being unsure of the next move, awkward or inappropriate movements that would have potential to injure the vessels, and sizing the target lesion that might lead to vessel rupture. Examples of criteria for a superior grade (score, 5) would be consistent handling the wires and catheters with minimal damage to the vessels, clear economy of motion and efficiency, a well-thought-out plan of procedure with effortless flow, and demonstration of sound knowledge of the appropriate wires/catheters for this particular procedure.

The angiography grade consisted of the ability to advance the wire, place the catheter, and perform an angiogram. The wire access grade consisted of selecting the appropriate catheter and traversing the lesion. The intervention grade consisted of selecting the appropriate catheter, selecting the appropriate stent, deployment of the stent, and performance of a completion angiogram. The total score was averaged within the three categories, and the final global assessment score was average of the three scores (possible range, 1.00-5.00).

In addition to the global assessment score determined by the examiner, the simulation software provided objective output for each subject, including total time, fluoroscopy time, volume of contrast used (mL), percentage of lesion covered, placement accuracy, presence of residual stenosis, and number of cine loops used.

Once the task was completed, or in some cases when the subject had reached the time limit of 15 minutes, the subject was asked to complete a questionnaire regarding his or her experience on the simulator. Each subject was asked to grade his or her own performance on a scale of 1 to 5 (self-assessment score).

Statistical analysis. Subjects were divided into experience groups according to their previous experience of the total number of endovascular cases performed. Subjects who reported a total number of endovascular cases of ≤20 were identified as "low experience," and those with 20 to

100 endovascular cases were considered "moderate experience." These group designations were somewhat arbitrary but based on previous studies' designation of novice interventionalists.[3,9]

Statistical analyses were performed using Excel 2003 (Microsoft Corp, Redmond, Wash) and StatView software (SAS Institute, Cary, NC). A Student *t* test was used to determine differences between the two experience groups. Analysis of variance was used to determine differences between multiple groups. Values of $P < .05$ were considered significant.

RESULTS

The study was completed by 17 surgical residents (16 fourth-year residents, 1 fifth-year resident) from 15 general surgery residency programs. The subjects reported a median of 70 open vascular cases in the pretest survey (interquartile range [IQR], 50-100; overall range, 20-150 cases). The mean number of endovascular cases for the entire cohort was 25 cases (IQR, 9-42; range, 4-89 cases). Eight subjects in the low experience (LE) group (<20 cases) performed a median of eight endovascular cases (IQR, 6.5-17.8, range, 4-20 cases). Nine subjects in the moderate experience (ME) group (20-100 cases) performed a median of 42 endovascular cases (IQR, 31-44; range, 25-89 cases). The LE group and the ME group had done a similar mean number of open cases (78.8 vs 75.0, $P = .60$).

The average overall global assessment score for the LE group (n = 8) was 2.68 of a possible 5.0, whereas the global assessment score for the ME group (n = 9) was 3.61 ($P = .03$). When the groups were analyzed by the subcategories of the overall score, the wire access and intervention grades contributed more than the angiography grades to the overall discrimination of the two experience levels (Table II). The simulator measured objective scores for the groups, and these were not found to be significantly different. Total procedural time, fluoroscopy time, volume of contrast used (mL), percentage of lesion covered, placement accuracy, presence of residual stenosis, and number of cine loops used were similar between the LE and ME groups (Table II).

Analysis of the post-test questionnaire demonstrated a relatively poor correlation between the examiner's global assessment score and subjects' self-assessment score ($R = 0.4$, Fig 3). Of the 17 participants, 11 gave themselves a satisfactory score of 3 on a scale of 1 to 5, even though the range of the structured global assessment score from the expert observer in these 11 ranged from 1.1 to 4.2.

Table II. Structured global assessment scores and objective measurements provided by the simulator for the low-experience and moderate-experience groups

Scores and Measurements	*Low experience (n = 8)*	*Moderate experience (n = 9)*	P
Structured global assessment			
Angiography	3.08	3.63	.14
Wire Access	2.44	3.67	.02
Intervention	2.53	3.53	.05
Overall global assessment	2.68	3.61	.03
Objective simulator measurements			
Total procedure time, s	896	947	.58
Fluoroscopy time, s	460	412	.62
Contrast used, mL	15.6	19.2	.48
Lesion covered, %	96.8	94.9	.70
Placement accuracy,[a] mm	4.85	6.64	.31
No residual stenosis[b]	75%	89%	.93
Cine loops, No.	5.5	4.7	.50

[a]Placement accuracy is distance in millimeters from the center of the stent to the center of the lesion.
[b]No residual stenosis is percentage of cases where there was complete resolution of the stenosis.

DISCUSSION

In this study, we demonstrate a statistically significant difference in the determination of a structured global assessment score between two groups of surgical residents with varying levels of self-reported procedural experience on a high-fidelity endovascular simulator. Although neither group was very experienced with renal angioplasty/stenting, the simulator environment allowed a blinded expert observer using a structured global rating scale to discern even minimal differences in experience. No statistically significant difference was found between the two study groups in the simulator software's objective measurements with respect to total time of procedure, fluoroscopy time, volume of contrast used, percent of lesion covered, placement accuracy, residual stenosis, or number of cine loops performed.

Determining proficiency in an objective manner for surgical procedures remains a challenge for educators, licensing boards, and hospital credentialing bodies. Global assessment scores that are "expert" opinions of observed tasks are quite subjective and optimally would consist of structured checklists so that maximal reliability, validity, usability, comprehensiveness, and discrimination can be achieved. A more formal assessment that adds objectivity has been suggested as a way to more accurately determine improvement of the trainee over time, such as using the validated Objective Structured Assessment of Technical Skills (OSATS).[8] The focus of simulator-based education research on endovascular skills assessment should be on validation of a standardized global score that combines the most appropriate subjective and objective criteria. In this study, the overall global assessment score had components that were more discriminating between the two experience groups, namely the wire access and intervention grades over the angiography grade.

The utility of the endovascular simulator to provide such a standardized exercise and framework for objective skills assessment is attractive. Being able to repeatedly simulate the same task within the same environment will allow comparison of the same subject over time, a group a subjects at the same level of training, or a group of subjects of different backgrounds or specialties. The modified global

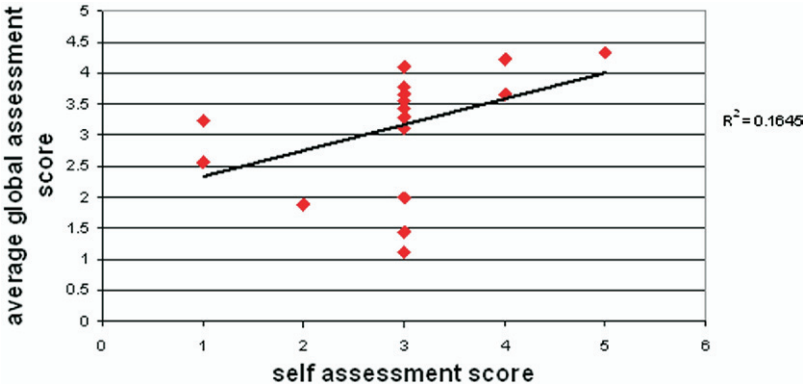

Fig 3. Subjects' average global assessment scores vs self-assessment scores from the post-test survey. The calculated correlation coefficient ($R2$) is 0.16.

assessment score would be considered a reliable test if studies showed that multiple blinded observers reached the same conclusion. We were not able to have a second observer in this particular study to determined the reliability of the assessment score we created, which has been a limitation of other studies.[7,10,11] In fact, only the study from Hislop et al[9] used two blinded expert observers, and in 49 of 61 cases, there was agreement on the overall pass/fail of the endovascular task. We believe that all future studies should include at least two observers so that the reliability of the assessment tool can be shown.

Validity of the structured global assessment score for endovascular procedures has been suggested by most of the publications[3,5,10] because novices and experts can be distinguished at baseline in both subjective and objective measures. This suggests that these types of checklist grading systems truly measure endovascular skills. Hsu et al[3] compared trained and untrained surgeons in a carotid stenting model, evaluating simulator scores at baseline and after the training session. They demonstrated an improvement after teaching on the vascular simulator in the novice group that was greater than in the advanced group. Aggarwal et al[5] similarly evaluated experienced and novice vascular surgeons with a renal artery angioplasty and stenting on an endovascular simulator. Each group received up to six training sessions on the simulator before re-evaluation of the skill. Minimally experienced surgeons were found to improve more than experienced surgeons at this task. Before instituting a formal type of board exam or credentialing test, however, further studies on the validity of these modified Reznick-like global scales will have to be done for endovascular skills assessment.

For a skills assessment tool to be widely adopted, the test also has to be usable and easy to administer. Within the environment of the simulation exercise in this study, it was relatively easy to train the expert observer to use the structured global assessment score and accurately record data. It is likely that the optimal assessment of trainees in the future will be some combination of the structured global rating scale along with objective computerized measurements that will best reflect the subjects' skill level. New metrics will need to be defined that the simulator might be able to objectify.

Of interest in this study, the objective measures provided by the software alone did not distinguish the low-experience and intermediate-experience groups. Other studies have confirmed that these surrogate markers for skill, including total procedure time, fluoroscopy time, and amount of contrast used are typically only different when the novice and the established expert are compared.[12,13] Because we had no expert (>100 cases) interventionalists, the simulator's objective measures may not have been discriminatory enough to distinguish novice and intermediate subjects. It may also be a limitation of the particular brand of simulator we used, since Dawson et al[14] found that the SimSuite (Medical Simulation Corp, Denver, Colo) endovascular simulator objective measures were a very useful marker of improvement after comprehensive interactive training in vascular surgery fellows.

Although the concept of endovascular simulators to enrich the educational environment is not debated, questions remain about its utility as a skills assessment tool. This concept will be important because simulator-based testing may become a component of licensing exams. Because practitioners from multiple specialties with varying training paradigms are typically allowed to perform similar endovascular procedures at most centers, standardized credentialing may be required. A study from Europe has documented the use of surgical simulators as part of the board examination.[15] Eight examiners and 20 examinees were evaluated for technical competence using several models, including suturing on a low-fidelity simulator. An interesting finding was that there was no correlation between technical competence on the simulator and oral examination scores. Examiners scored better on the suturing task compared with examinees, as expected.

The American Board of Surgery has been interested in standardizing an educational curriculum that can be competency and skill-based. Certainly, the assessment of train-

ees in such a standardized fashion would be facilitated by the use of simulators. The American Board of Vascular Medicine currently has an endovascular medicine certifying examination that uses simulators for endovascular skills assessment. Further research will be necessary to understand how the measured outcomes from these simulation exercises will translate into improved patient care, safety, and quality.

The post-test questionnaire in this study revealed that subjects who scored well were aware of their competency. Some subjects who scored poorly were also aware of their lack of ability. Most of the subjects, however, assessed themselves as average, irrespective of the global assessment score, which led to the poor correlation coefficient seen in Fig 3. This likely indicates that individuals cannot accurately assess their own competency. This analysis is limited, however, because the subjects were in the stressful environment of being interviewed for a vascular fellowship and may have not been as accurate in their own assessment owing to concerns about how it would affect their overall evaluation.

The study has several limitations. We used only one expert observer. As a result, we cannot determine interobserver variability nor can we average the global rating score between two experts for each subject. Each subject underwent only one session, without the opportunity to practice or learn the equipment. There are many variations with respect to instrumentation. Their results may simply be due to a lack of familiarity with the instrumentation used and not the technical aspects of the procedure. Low-experience subjects might have performed better if they had been more familiar with the tools. For this reason, others have considered the second attempt at the procedure as the baseline assessment to correct for this lack of familiarity with the simulator.[5] In addition, there is a risk of a type II error given the relatively small number of subjects in each group. Plans for a multicenter trial to evaluate residents from multiple institutions are being considered to validate the utility of an endovascular simulator.

Another study limitation was the stress of this particular situation because the study was performed on the day of vascular fellowship interviews. Future trials include repeating this pilot study in different training levels of surgical residents, other types of residents including interventional cardiology and radiology residents, medical students, and also to have expert observers of different specialties.

CONCLUSION

Our preliminary results support a correlation between procedural experience measured by self-reported case completion and a structured global rating score by a blinded observer on an endovascular simulator. The environment created by the simulator allows for the assessment of endovascular skills in a controlled environment and may be an important additional role of simulation besides skills training. Future research is required to determine if simulator-based testing should be incorporated into the credentialing of vascular specialists.

AUTHOR CONTRIBUTIONS

Conception and design: JL, MT, JP
Analysis and interpretation: JL, MT
Data collection: JL, MT, JP
Writing the article: JL, MT
Critical revision of the article: JL, MT, RD, EH
Final approval of the article: JL
Statistical analysis: JL, MT
Obtained funding: JL, RD, TK
Overall responsibility: JL

REFERENCES

1. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg 2006;244:310-4.
2. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002;236:458-63; discussion 463.
3. 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;40:1118-25.
4. Gallagher AG, Cates CU. Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. JAMA 2004; 292:3024-6.
5. Aggarwal R, Black SA, Hance JR, Darzi A, Cheshire NJW. Virtual reality simulation training can improve inexperienced surgeons' endovascular skills. Eur J Vasc Endovasc Surg 2006;31:588-93.
6. Aggarwal R, Tully A, Grantcharov T, Larsen CR, Miskry T, Farthing A, et al. Virtual reality simulation training can improve technical skills during laparoscopic salpingectomy for ectopic pregnancy. BJOG 2006; 113:1382-7.
7. 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: 343-52.
8. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg 1997;173:226-30.
9. 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;43:47-55.
10. 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;40:1112-7.
11. Neequaye SK, Aggarwal R, Brightwell R, Van Herzeele I, Darzi A, Cheshire NJW. Identification of skills common to renal and iliac endovascular procedures performed on a virtual reality simulator. Eur J Vasc Endovasc Surg 2007;33:525-32.
12. Rhodes RS, Biester TW, Bell RH Jr, Lewis FR Jr. Assessing surgical knowledge: a primer on the examination policies of the American Board of Surgery. J Surg Educ 2007;64:138-42.
13. Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, et al. Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. Surgery 2000;128:613-22.
14. Dawson DL, Meyer J, Lee ES, Pevec WC. Training with simulation improves residents' endovascular procedure skills. J Vasc Surg 2007;45: 149-54.
15. Pandey VA, Wolfe JHN, Liapis CD, Bergqvist D. The examination assessment of technical competence in vascular surgery. Br J Surg 2006;93:1132-1138.

Submitted Sep 17, 2007; accepted Jan 6, 2008.

DISCUSSION

Dr David L. Dawson (Sacramento, Calif). Skeptics would say that simulation is a credentialing tool of the future—and it always will be. I suspect, however, that we will see procedural simulation become a more substantial part of the certification process in the future. To date, procedural training has been the primary application of simulation technology. Dr Tedesco and her Stanford surgical colleagues have addressed another important role for simulation—the potential use of the simulator as an assessment tool. Further, their observation that the machine may not work well as a stand-alone assessment tool is an important one.

Test developers recognize five characteristics of a good assessment tool: reliability, validity, usability, comprehensiveness, and discrimination, the last being the focus of this report. In this small study, there were no measures of test reliability. The validity of the testing methodology is yet to be established, but I think it certainly could be with more studies. Simulation, I believe, is a usable test method, but one associated with substantial costs. As reported, this particular test was not comprehensive, as this pilot project was of limited scope.

Thus, the authors focus primarily on the ability of simulation-based endovascular skills assessment to discriminate between low and high endovascular case experience in residents, assuming that the volume of prior experience is a reasonable surrogate for either clinical competence or proficiency. Their observations suggest that an expert observer can evaluate performance with a structured assessment tool and using this evaluation can distinguish between low and high performance.

They found that self-assessments were not particularly useful, with most subjects grading themselves as midlevel in performance. These self-assessment scores do not appear to effectively discriminate between those who actually did perform well or not, an observation that provides additional support for the argument that other, better, and perhaps more independent assessment tools are needed.

I have three questions. First, this evaluation model is an example of a criterion-based test. How were the criteria established for expected or desired performance? That is, how were the definitions of "poor performance" and "flawless performance" defined and how were the gradations between delineated?

Second, in contrast to the reports of others working with this type of technology, you were not able to use the objective performance metrics of the stimulator to distinguish between the levels of performance of the test subjects. Do you think that this might be attributable to the limited time spent on the single case simulation that each subject performed?

And finally, it is generally accepted that expert observers can meaningfully evaluate professional competence. I think this is a concept that the American Board of Surgery has bought into, as "expert observers" in small hotel rooms often do assess professional competence of people seeking certification. Time and resource constraints, however, limit the practicality of always depending on evaluations by senior subject matter experts. Do you think it would be valid to have subjective assessments made by trained educational testing specialists who are not physicians or endovascular specialists?

I appreciate the opportunity to discuss this interesting and timely paper and I expect that we will be hearing more about this topic as the technologies mature and experience grows.

Dr Tedesco. Thank you, Dr Dawson, for your discussion and questions and especially thank for asking three, not 17, questions. With respect to the first question, our senior author and second-year endovascular fellow developed the questionnaire based on reports by Reznick and others. Specifically addressing how we scaled the Likert scale, we graded the applicants from 1 to 5, and where 1 on the scale indicated a poor performance and a 5, flawless performance. Criteria for a fail grade or poor performance were frequently stopping the procedure, clearly being unsure of the next move, awkward or inappropriate movements that would result in potential injury to the vessels, sizing the target lesion that might result in rupture. Those would be indications for a fail or a poor performance. We defined flawless or a superior performance with a score of 5 as consistently handling the wires and catheters with minimal damage to the vessels, clear economy of motion and efficiency, a well-thought-out plan of procedure with effortless flow or a demonstration of sound knowledge of the appropriate wires and catheters for the renal angioplasty and stenting procedure. That is how we measured the scale that was developed by our senior author.

To address your second question: yes, you are correct. Objective criteria have been used to differentiate novice and expert subjects. However, the expert subjects that were used in prior studies had performed over 300, sometimes over 1000 endovascular cases. In our study, the objective criteria was not able to discern the small difference between the experienced groups which were between—the low experience group, with less than 20 endovascular cases, and the moderate experience group with between 20 and 100. Perhaps the limitation of this study and this simulator is that it is unable to detect small differences, which also highlights the importance and need for the expert observer. In addition, there was limited time for each subject, which created a definite limitation for this study. Perhaps if the low experience group had more time to practice they would have performed better and that is certainly a limitation of our study.

With respect to your last question, the self-assessment was not an accurate assessment of skill level in this particular study and perhaps testing specialists could perform the assessment. However, I think it is potentially dangerous to remove physicians from the testing scenario, as physicians are the people performing the actual live endovascular skill on a day-to-day basis. There are nuances and style points that perhaps only physicians can understand and would be able to score better than trained specialists. This of course has implications for the application of simulation-based skill assessment with respect to physician time and cost.