

Development and construct validation of performance metrics for multivessel coronary interventions on the VIST virtual reality simulator at PCR2005

Anthony G Gallagher¹, PhD; Jean Renkin^{2*}, MD; Hans Buyl³; Hendrik Lambert³; Jean Marco⁴, MD

1. National Surgical Training Centre, Royal College of Surgeons in Ireland, Dublin, Ireland; 2. Cardiology Department, UCL - St Luc University Hospital, Brussels, Belgium; 3. Institute for Therapy Advancement - Europe, Brussels, Belgium; 4. Cardiology Department, Clinique Pasteur, Toulouse, France

This paper also includes accompanying moving images published at the following website: www.europcronline.com/eurointervention

KEYWORDS

Coronary disease, simulation, virtual reality, performance metrics, validation

Abstract

Background: The first aim of this study was the development of performance metrics in a virtual reality simulation for interventional procedure in a real case of multivessel coronary disease. A second aim was to assess the construct validity of these metrics by comparing the performance of interventional cardiologists with different levels of experience.

Methods: Ninety-four practicing interventional cardiologists attending the EuroPCR 2005 in Paris, France participated in this study. Complete data was available on eighty-nine participants (95%). Participants were divided into three categories depending on experience. Group 1 (novices): N = 33, < 1 years experience; Group 2 (intermediate experience): N = 14, >50 cases per year for the last two years and Group 3 (master physicians): N = 42 participants completed > 100 cases per year during the last five years.

Procedure: Over a period of months during 2004-2005 we identified potential performance metrics for cases of coronary artery disease which were then applied to a case of a patient admitted because of stable angina (class 1) with multivessel coronary disease. Patient's coronary anatomy and lesions were then reconstructed and implemented for the VIST virtual reality simulator. All participants were required to perform this case.

Results: Overall, experienced interventional cardiologists performed significantly better than the novices for traditional metrics such as time ($p = 0.03$), contrast fluid ($p = 0.0008$) and Fluoroscopy time ($p = 0.005$). Master physicians performed significantly better than the other two groups on metrics which assessed technical performance e.g., time to ascend the aorta ($p = 0.0004$) and stent placement accuracy ($p = 0.02$). Furthermore, master physicians made fewer handling errors than the intermediated group who in turn made fewer than the novice group ($p = 0.0003$). Performance consistency was also a linear function of experience.

Conclusions: Novel performance metrics developed for the assessment of technical skills for a simulated intervention for multi-vessel coronary disease showed that more experienced interventional cardiologists performed the procedure better than less experienced interventionalists thus demonstrating construct validity of the metrics.

* Corresponding author: Cardiology Department, UCL - St.-Luc University hospital, Hippocrate Avenue 10, 1200 Brussels, Belgium

E-mail: jean.renkin@card.ucl.ac.be

Introduction

In interventional cardiology most current medical training continues in the traditional mentored method where trainees are exposed to patient procedures with the guidance of an experienced teacher.¹ The experience is dictated by the random admission of patients, rather than a timely consistent exposure to the fundamental problems evolving from easy to more complex anatomy, lesion and clinical presentation. During percutaneous coronary interventions, most of the errors may be caused by human factor problems associated with the unique particularity of the 2D image guided techniques. There is difficulty in learning the hand-eye coordination of instruments, catheters and guidewires because of the counter-intuitive movement of instruments due to the fulcrum effect of the body wall² in a 3D environment. This causes a proprioceptive-visual conflict, which takes the operator's brain some time to overcome. Furthermore, the current training paradigm lacks objective feedback on trainee (and teacher) performance.³

The state of the art for training in many high skill professions is virtual reality (VR) defined as a communication interface based on interactive 3D visualization which allows the user to interface, interact with and integrate different types of sensory inputs that simulate important aspects of real world experience.⁴ Within the medical community, the acceptance of the VR training approach is slow partly because of skepticism about the realism of the simulator but also due to the lack of well controlled clinical trials. However, recent studies showed that residents who were trained using VR made significantly fewer intra-operative errors during the performance of a laparoscopic cholecystectomy than a standard trained group,^{5,6} and the American College of Surgeons (ACS)⁷ has outlined its overwhelming support for simulation in the drive to improve patient safety. Cardiology has arrived at the same position relatively quickly. The driving force has been the introduction of a new procedure, i.e. carotid artery stenting.⁸ Physicians who wish to learn this procedure will not train on patients, but virtual patients. VR training will be a requirement by the FDA to credential a physician to perform this procedure.⁹ One of the simulators that physicians will train on is the Vascular Interventional System Training (or VIST), which simulates the human cardiovascular system and provides visual and haptic feedback, very close (similar) to what a physician would see and feel if they were performing the procedure on a patient.¹⁰

One of the major advantages of VR for training technical skills is that the opportunity can be used to replace the early part of the learning curve, which would otherwise be achieved in the clinical situation by practicing on live patients. Another advantage is that the trainee can make mistakes without exposing the patient (and himself/herself to some extent) to risk unlike *in vivo* procedures.¹¹ However, the potential of VR for patient safety, improved training and the development and market roll out of new procedures require controlled clinical trials. An optimal approach to using simulation for training would be to first establish an objective benchmark on the simulator, based on the performance of experienced operators with defined quantitative metrics and then require trainees to train until they reach the benchmark, consistently (e.g., for two consecutive trials).

However, this assumes that the performance metrics that are used in the simulator measure meaningful aspects of performance reli-

ably.¹² The VIST simulator is a full physics simulator and allows for the precise assessment of the catheter and wire handling skills of the operator. However, the exact performance metrics must be developed and defined by experienced operators. They need to identify aspects of performance that denote the skill (or the lack of it) of the operator on a continuous or interval/ratio scale. Once these have been identified and defined these metrics must then be validated in a series of empirical studies.

One of the goals of the study reported here is to investigate the construct validity of new performance metrics developed for the VIST simulator for one of the most commonly performed interventional cardiology procedures, i.e., coronary angioplasty and stenting. Over a series of meetings we developed a wide range of intra-operative performance metrics for coronary angioplasty and stenting and at PCR2005 we set about assessing the validity of them. In particular we sought to establish their construct validity. This would be demonstrated if the metrics distinguished between the performance of very experienced or master physicians and more junior physicians with less experience in coronary angioplasty and stenting on the simulator. We hypothesized that the performance metrics we had developed for coronary angioplasty and stenting would be able to distinguish between the intra-operative performance of master physicians and less experienced.

Methods

Study population

Ninety four practicing interventional cardiologists attending the EuroPCR 2005 in Paris, France participated in this study. Complete data were available on eighty nine participants (95%). According to their own declaration participants were divided into three categories depending on experience. Group 1 (novices): 33 participants (35.1%) had < 1 years experience; Group 2 (intermediate experience): 14 (14.9%) participants completed >50 cases per year for the last two years and Group 3 (master physicians): 42 (44.7%) participants completed > 100 cases per year during the last five years.

Material

The Procedicus VISTTM simulator (Mentice AB, Gothenburg, Sweden) is based on a dual processor (2 x 2.8 Ghz processor), Pentium IV PC running Windows Microsoft XP Professional with 1 GB RAM, a 40 GB hard disk drive, a GeForce FX5200 128MB graphics card, and two 17" flat-panel monitors (Figure 1). The simulation interface device is designed to sense the simultaneous translation and rotation of three co-axial tools (clinical tools), the flow of air from a syringe, pressure in fluid compressed by an inflator, and on/off foot switch. Output of the device to the user is the application of force and torque on each of the tools based on the calculations of the simulator.

The forces applied to the clinical tools are sensed by strain gage sensors, fitted between a cart base and a suspended mechanism, which is locked on the tool. The resolution of the force measurement system (which includes a chain with the sensor, preamplifier and A/D-converter) is 0.025 Newtons. The calibration of the sensor is performed dynamically (in run-time) and the offset error is lower

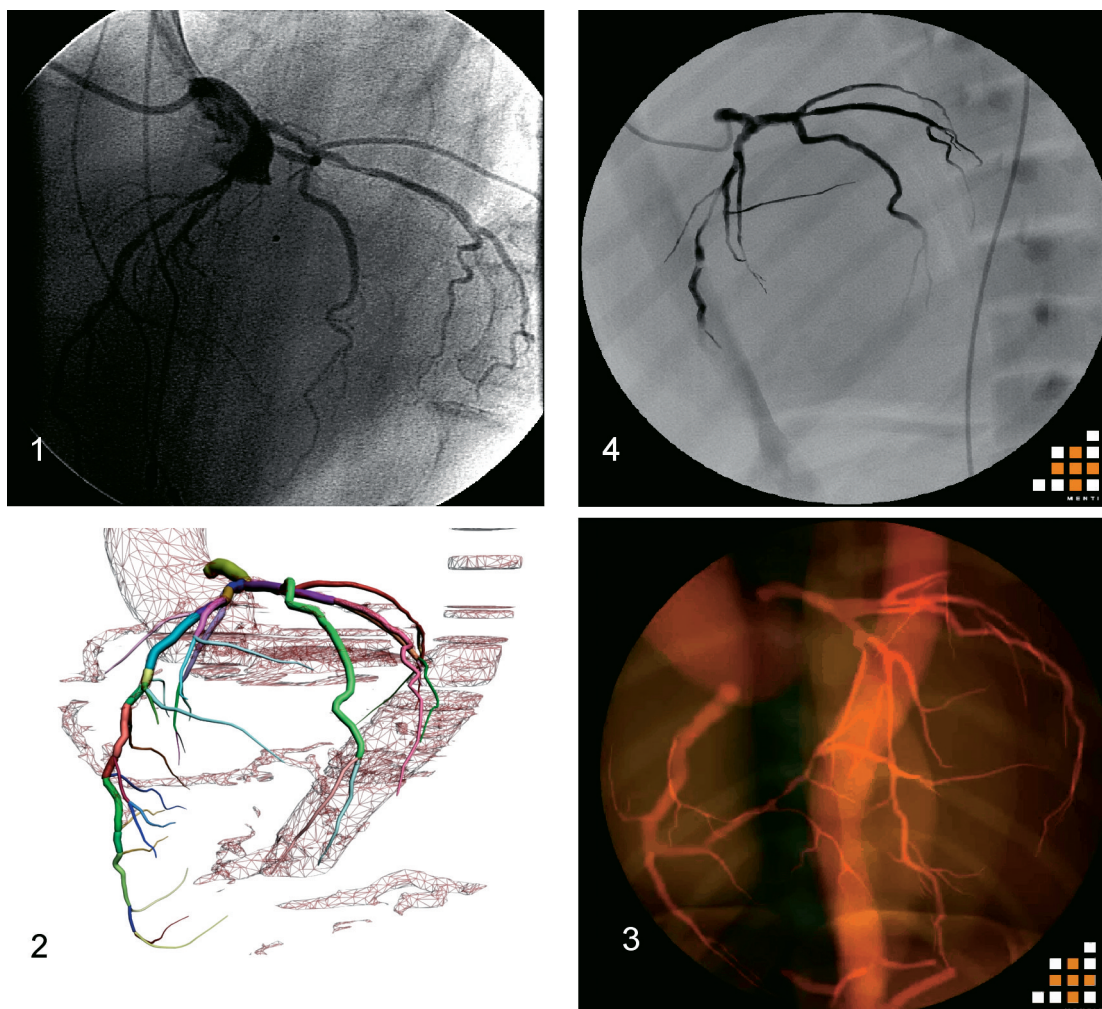


Figure 1. The different steps from diagnostic angiography to simulation images. 1. Diagnostic angiography in left anterior oblique, cranial projection. 2. 3D reconstruction of coronary artery tree from CT data. 3. 3D coronary artery tree implemented in the VIST simulator. 4. 2D angiography displayed in the VIST simulator.

than 0.025 Newtons. The span of the force measurement is ± 2.5 Newtons. Within this range, the forces in the force feedback loop are controlled in a closed loop. The force feedback range is (theoretically) ± 30 Newton, and after 2.5 N the forces are controlled in an open loop.

The translational position is measured with an optical encoder, which in combination with the transmission system gives a resolution of 0.11 mm. The rotational angle is measured with an optical encoder, which in combination with the gear ratio to the locking device gives a resolution of 7.9 - 31.4 milliradians (depending on which cart). The tool diameters are measured with an infrared optical sensor, which in the existing construction gives a resolution of 0.02mm and has a precision of about $\pm 15\%$. The algorithms that calculate the diameter calibrates the parameter settings in run-time, to avoid drifting. The measurement span is between 0.1 - 3.0 mm.

The Procedicus VISTTM simulates the entire coronary procedure exactly as it would be performed in the cath lab. All testing was performed in a quiet room with the table height of approximately 100cm with the monitor position at, or slightly below, eye level. The

devices used were real functional devices. A 0.035" Supra Core 35 guide wire (Guidant Europe NV/SA, Diegem, Belgium) was used to introduce the Viking guiding catheter (Guidant Europe NV/SA, Diegem, Belgium). The floppy tip of the 0.035" guide wire as well as the tip (before the shape) of the guiding catheter was removed. This allows the VISTTM to grab and control the devices in an efficient way. A balloon Crossail (Guidant Europe NV/SA, Diegem, Belgium) less or equal to 3.0 mm was used to inflate the lesion. The same balloon was used to simulate the stent. The balloon or stent diameter and length could have been selected independent in function of the anatomy that needed to be treated. A 0.014" guide wire (Guidant Europe NV/SA, Diegem, Belgium) was used in both balloon and stent. The floppy tip of the 0.014" guide wire was removed similar to the 0.035" guide wire. The selection of the devices came from a virtual stock that was programmed in the system.

Procedure

A case of a patient admitted because stable angina (class 2) with multivessel coronary disease was selected for this particular study. Diagnostic coronary angiography used for the study showed a dif-

fuse three vessel disease with a long severe lesion located on the mid segment of the Left Anterior Descending artery, a short severe lesion on the proximal Circumflex artery and a moderate lesion on the distal postero-lateral branch of the Right coronary artery. The process to get simulation is summarized hereunder.

Following diagnostic angiography the patient was restudied using a 64 slice computer tomography (Siemens GmbH Erlangen, Germany). From CT data and Dicom angio images a 3D coronary anatomy was reconstructed using specific software and implemented in the VIST system.

Physicians first viewed the procedure performed by a very experienced interventional cardiologist (JR) on, and familiar with, the VIST simulator in a 300 seat auditorium with the images from the simulator projected onto two large screens. As the procedure was performed the master trainer explained each stage of the procedure and the rationale for material and strategy selection. Before carrying out the case each participant viewed a short video summarizing the main points of the case. They were assisted at the simulator by support staff experienced in the use of the VIST simulator and who had been briefed and trained in the experimental protocol. Support staff could operate the simulator controls but only on the instruction of the operating participant or give advice on request. Support staff also corrected deviations from procedure protocol but only after the participant had given clear direction as to their next step. Procedure deviations that required support staff interventions were recorded for later analysis. For practical reasons subjects were given 22 minutes to complete the procedure.

Statistical analysis

The independent variable in this study was operative experience (Groups 1-3) and dependant variables were in two groups the first being summative metrics which included time to complete the procedure, amount of contrast agent used, and fluoroscopy exposure time. We also assessed a number of other variables that directly assessed technical skills associated with performance of the procedure, reported on Table 1. Technical skills metrics provided formative feedback on a second-by-second basis.

Table 1. Performance metrics assessed during the PCR2005 study.

1. Time to ascend the aorta
2. The number of guide catheter rotations (i.e., >1800 rotation)
3. Frequency of entering the wrong vessel
4. Stent placement accuracy
5. Catheter scraping against the vessel wall
6. Advancing the guide catheter ahead of the guide wire
7. Frequency of inserting the guide catheter too far into the vessel (i.e., >3mm)
8. Frequency of lost access to the vessel once the ostium had been engaged
9. Catheter manipulation errors
10. Guide wire manipulation errors

The scores of six of these variables were summed to form a new dependant variable which we called 'handling errors' because they were dynamic measurements of performance in an ongoing fashion. The six variables that made up this category were variables 5, 6, 7, 8, 9 & 10.

Data was analyzed with SPSS 10 for windows (Chicago, IL, USA). Differences between the groups were compared for significance with one-factor analysis of variance (ANOVA) and Scheffe post-hoc contrasts were used to identify specific statistically significant differences between the groups.

Results

Table 2 shows the mean and standard deviation scores for the three groups for seven of the thirteen variables studied. The upper plate shows the ANOVA results of the more traditional measures of Time, Contrast fluid used and Fluoroscopy exposure time. Significant differences were observed for all three measures. Intermediate experienced cardiologists performed the procedure faster than the Master physicians and the Novices however Scheffe F-test contrasts showed that no two groups differed significantly. The Master physicians used the least contrast fluid during the procedure but this was not significantly less than the Intermediated experienced group ($p = 0.09$). However, Scheffe F-test comparisons showed that only for the

Table 2. Mean and standard deviation performance metrics of the three groups (df = 2, 86).

Groups	Novices		Intermediate experience		Master physicians		F	Sig.
	Mean	sd	Mean	sd	Mean	sd		
Traditional metrics								
Metric variables	Mean	sd	Mean	sd	Mean	sd	F	Sig.
Total time (secs.)	1455.0	604	1169.0	259.0	1207.0	332	3.51	0.03
Contrast fluid	192.5	107	140.1	50.3	121.5	56	7.76	0.001
Fluoroscopy	676.5	264	498.8	166.0	514.0	207	5.66	0.005
Technical Performance metrics								
Metric variables	Mean	sd	Mean	sd	Mean	sd	F	Sig.
Time to ascend the aorta (secs.)	62.20	77.0	27.1	24.0	16.0	19.0	8.2	0.001
No. guide catheter rotations (>180°)	3.90	4.7	2.5	2.2	1.9	1.6	3.6	0.03
Frequency of entering wrong vessel	1.42	2.0	0.6	0.9	0.6	0.7	3.9	0.02
Stent placement accuracy	2.73	2.4	3.4	2.1	1.7	2.0	4.1	0.02

Table 3. Mean and standard deviation 'handling error' metric scores for the three groups.

Groups	Novices		Intermediate experience		Master physicians	
	Handling errors					
Metric variables	Mean	sd	Mean	sd	Mean	sd
Catheter scraping against vessel wall	0.73	1.31	0.57	0.85	0.36	0.58
Advancing cath. ahead of guide-wire	0.33	0.82	0.50	0.86	0.24	0.49
Inserting cath. too far into vessel	14.55	25.70	9.93	14.50	3.19	3.50
Loss access to vessel after ostium has been engaged	4.85	4.10	3.64	6.60	1.86	1.60
Catheter manipulation errors	0.06	0.24	0.07	0.27	0.14	0.42
Guide-wire manipulation errors	0.24	1.20	0.21	0.58	0.05	0.81

Master physicians differed statistically significant from the novices ($p = 0.0004$). There also was a significant difference in the fluoroscopy exposure times for the three groups. Although the Intermediate experienced group used less fluoroscopy that the other two groups this difference was not statistically significant compared to the Master physicians ($p = 0.79$). Both of these did use significantly less fluoroscopy than the novices (Intermediate experienced Vs Novices $p = 0.02$; Master physicians Vs Novices $p = 0.004$).

The same type of statistical analysis was applied to the technical performance metrics and the means and standard deviation scores for these variables for the three groups are shown in the lower plate of Table 2. The Master physicians demonstrated the best technical performance scores for three of the metrics (i.e., time to ascend the aorta, number of guide catheter rotations and stent placement accuracy). The Intermediate experienced group scored marginally better than the Master physicians for frequency of entering the wrong vessel. However, across all four metric measurements the Master physicians performed the most consistently as demonstrated by the lowest standard deviations.

Scheffe F-test contrasts showed that only the Master physician performance differed significantly from the Novices in the time taken to ascend the aorta with the catheter and wire ($p = 0.0003$). The differences between the Novices and the Intermediate experienced group were observed to be approaching statistical significance but did not reach the acceptable level ($p = 0.09$). Master physicians also made significantly fewer guide catheter rotations than the novices ($p = 0.03$) but not the Intermediate experienced group ($p = 0.07$). The Intermediate experienced group were the least likely to enter the wrong vessel however, only the master physicians performance differed significantly from the Novices ($p = 0.04$; Master Vs Intermediate, $p = 0.84$). Indeed the Novices were two and a half time more likely to enter the wrong vessel than the other two groups. The Master physicians were the most accurate at deploying the stent followed by the novice group and the difference between their performance and the Intermediate experienced group was found to be statistically significant (Intermediate Vs Master Physicians $p = 0.009$; Novice Vs Master, $p = 0.04$).

Figure 2 shows the mean and standard deviation scores for the handling errors variable. There appears to be a linear trend of increasing handling error scores with decrease in clinical experience. Intermediate experience participants on average made more handling errors than the Master physicians and likewise Novices made more than the Master group. This also appears to have been the case for

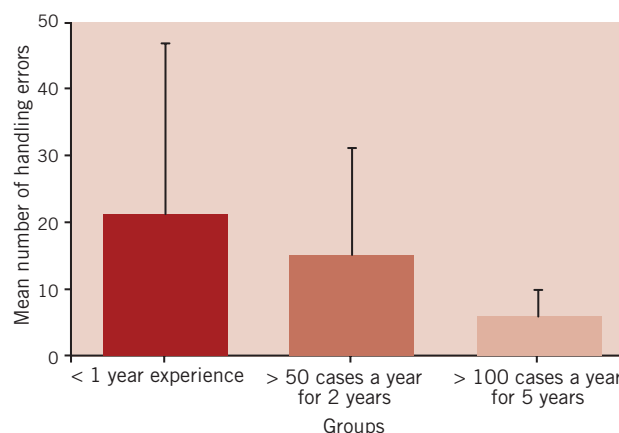


Figure 2. Mean and standard deviation number of handling errors made by the three groups.

performance variability with Master physicians showing the greatest performance consistency and Novices showing the least. Indeed novice standard deviation scores were observed to be six and a half times greater than master physicians. This was also the case for Intermediate experienced participants who showed a greater score variability than the Master physicians. Differences between the groups were examined for statistical significance with Kruskal-Wallis. Overall a significant effect for handling errors was found [$df = 2,3$; $H = 15.96$, $p < 0.0003$ (17 tied groups)]. Specific contrasts between performances by the different groups were performed with Mann-Whitney U tests. The Master physicians made significantly fewer handling errors than the Novice group ($Z = -3.99$, $p < 0.0001$) and although they made fewer errors than the Intermediate experienced group this did not reach statistical significance ($Z = -1.84$, $p < 0.07$). The difference between the Intermediate experienced group and Novices was not found to be statistically significant ($Z = -0.92$, $p < 0.36$).

Discussion

The development of virtual reality simulation in medicine is facilitating a change in how doctors' skills are trained and assessed. These changes originally occurred in laparoscopic surgery¹³ but more recently have impacted heavily on endovascular medicine. Furthermore, the simulators that are available in endovascular medicine are orders of magnitude superior to those available in surgery. One of the most important aspects of virtual reality is to develop valid metrics for each learning step that are sensitive to inter-operative performance. These metrics need to be developed by experi-

enced physicians along with behavioural scientists who have expertise in operational definition and measurement of performance and computer scientists. One of the most crucial aspects of this enterprise is the ability to operationally define these performance characteristics which is important for their accurate reliable measurement. Performance characteristics that do not meet these criteria must be excluded before progressing further in the validation process. One of the sterner challenges for the metrics that have survived this process is the empirical demonstration of construct validity, i.e., the metrics should be able to empirically distinguish between the intra-operative performance of experienced physicians and junior less experienced physicians.

Being able to complete the procedure without major complications is too crude an indicator of skill level particularly during training. Metrics such as time to complete the procedure, contrast fluid used and fluoroscopy exposure time are probably also too crude for assessment purposes as they only provide summative measures of performance. A more complete picture of intra-operative performance may be provided by metrics that measure deviations from optimal performance. The metrics that we have reported on appear to provide that insight. These new technical performance metrics tended to demonstrate the superior performance of the master physicians across almost all of the variables. As well as superior performance the master physicians also showed a lower variability in performance in comparison to the other groups. This is demonstrated clearest in the data on handling errors in Figure 2. Consistency as a performance metric is emerging as one of the most important indicator of 'skill' across a range of disciplines. However, there has been little systematic investigation of this metric in procedural based medicine. This methodological and measurement issue is important because we envisage that score consistency appears to be emerging as an important element in operationally defining exactly what we mean by 'proficiency'.

Intermediate experienced physicians performed the procedure faster than the Master physicians and also used less fluoroscopy. Their score variability was also slightly lower than the experienced physicians. There are two possible explanations for this finding. The first is the small number of subjects in the intermediate experienced group and the second is the large amount of variability observed in this group in comparison to the experienced physicians. Both of these issues are related as increasing numbers within groups tends to decrease variability.

There was considerable surprise from some members of this writing group at the performance variability of some of the participants, i.e., > 10 standard deviations from the mean. However, this finding may be new in interventional cardiology but it is not new in other procedural based medical disciplines such as laparoscopic surgery. In this sense the findings from the study reported here are reassuring in that we are finding in cardiology what the surgical community has already observed with a lower fidelity simulator. The VIST simulator looks, feels and behaves similar to endovascular instruments inside real patient. Furthermore it has granularized performance metrics. The MIST VR tasks are analogs of the tasks to be performed during a laparoscopic cholecystectomy. However, MIST VR was the first simulator in medicine to demonstrate in a prospective,

randomized double-blinded study that training significantly improved objectively assessed intra-operative performance. Surgical residents who trained to the empirically established proficiency level performed the procedure 29% faster and made six times fewer operative errors than standard trained surgical residents⁷ and have been replicated⁶.

This novel metrics developed by this group and applied to patient specific data distinguished intra-operative performance between interventional cardiologists with different levels of experience. Predicted performance differences were observed. Experienced interventional cardiologists performed better and with greater consistency than less experienced physicians. These operative metrics give value to the use of virtual reality simulation as a training tool to evaluate skill acquisition and gives immediate feedback, which facilitates optimal learning.

Acknowledgments

The authors wish to thank the Guidant Institute for Therapy Advancement, (Brussels, Belgium) and Mentice AB, (Gothenburg, Sweden) for their co-operation, their technical support and data collection.

References

1. Gallagher AG, Cates CU. Virtual reality training for the operating room and cardiac catheterisation laboratory. *Lancet*. 364:1538-40, 2004.
2. Gallagher AG, McClure N, McGuigan J, Ritchie K, Sheehy NP. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endosc*. 30: 617-20, 1998.
3. Darzi A, Smith S, Taffinder N. Assessing operative skill needs to become more objective. *BMJ*; 318: 887-8, 1999.
4. Riva G. Applications of Virtual Environments in Medicine. *Methods Inf. Med.* 42(5): 524 - 534, 2003. Satava RM. Virtual reality surgical simulator; The first steps. *Surg Endosc*: 7: 203-205, 1993.
5. Grantcharov TP, Kristianson VB, Bendix J, Bardram L, Rosenerg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg*. 91: 146-50, 2004.
6. Seymour N, Gallagher AG, O'Brien M, Roman S, Andersen D, Satava RM. Virtual Reality Training Improves Operating Room Performance: Results of a Randomized, Double-Blinded Study. *Ann of Surg*; 236: 458-464, 2002.
7. Healy GB. The College should be instrumental in adapting simulators to education. *Bull of the Am Coll of Surg*; 8: 10 - 12, 2002.
8. Yadav JS. Carotid stenting in high-risk patients: design and rationale of the SAPHIRE trial. *Cleveland Clinic Journal of Medicine*. 71 Suppl 1:S45-6, 2004.
9. Gallagher AG, Cates CU. Approval of Virtual Reality Training for Carotid Stenting; What This Means for Procedural Based Medicine. *JAMA*. 292: 3024-3026, 2004.
10. Cotin S, Dawson SL, Meglan D, et al. ICTS, An interventional cardiology training system. In JD Westwood, HM Hoffman, GT Mogel, RA Robb and D Stredney (Eds.) *Medicine Meets Virtual Reality2000*. Washington DC.: IOS Press, 2000:59-65.
11. Rogers DA, Regehr G, MacDonald J. A role for error training in surgical technical skill instruction and evaluation. *Am J of Surg*. 183: 242-5, 2002.
12. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. *Surg Endosc*. 10: 1525-1529, 2003.
13. Satava RM. Virtual Reality surgical simulator: The first steps. *Surg Endosc*: 7, 203-205, 1993.