HAL
open science

Virtual Reality, Ultrasound-guided Liver Biopsy Simulator: Development and Performance Discrimination

Sheena Johnson, Carianne Hunt, Helen Woolnough, Martin Crawshaw, Caroline Kilkenny, Derek Gould, Amrita Sinha, Andrew England, Pierre-Frédéric Villard

To cite this version:

Sheena Johnson, Carianne Hunt, Helen Woolnough, Martin Crawshaw, Caroline Kilkenny, et al.. Virtual Reality, Ultrasound-guided Liver Biopsy Simulator: Development and Performance Discrimination. British Journal of Radiology, 2011, 10.1259/bjr/47436030 . inria-00560487

HAL Id: inria-00560487

https://inria.hal.science/inria-00560487

Submitted on 25 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Virtual Reality, Ultrasound-guided Liver Biopsy Simulator: Development and Performance Discrimination[1]

Running head: Virtual Reality Liver Biopsy Simulator

Word count (exc. Figures / tables): 4210

Key Words: Interventional Radiology, Simulation, Virtual Reality, Training, Task Analysis

[1] This article presents independent research commissioned by the UK National Institute for Health Research (NIHR, project reference HTD136). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

ABSTRACT

Purpose: Identify and prospectively investigate simulated ultrasound-guided targeted Liver Biopsy performance metrics as differentiators between levels of expertise in Interventional Radiology.

Materials and Methods: Task analysis produced detailed procedural step documentation allowing identification of critical procedure steps and performance metrics for use in a virtual reality ultrasound-guided targeted Liver Biopsy procedure. Consultant (n=14, male = 11, female = 3) and trainee (n=26, male = 19, female = 7) scores on the performance metrics were compared. Ethical approval was granted by the Liverpool Research Ethics Committee (UK). Independent t-tests and ANOVA investigated differences between groups.

Results: Independent t-tests revealed significant differences between trainees and consultants on 3 performance metrics: *targeting, $p = 0.018$, $t = -2.487$ (-2.040 to -0.207); probe usage time, $p = 0.040$, $t = 2.132$ (11.064 to 427.983); mean needle length in beam, $p = 0.029$, $t = -2.272$ (-0.028 to -0.002)*. ANOVA reported significant differences across years of experience (0-1, 1-2, 3+ years) on seven performance metrics: *no-go area touched, $p = 0.012$; targeting, $p = 0.025$; length of session, $p = 0.024$; probe usage time, $p = 0.025$; total needle distance moved, $p = 0.038$; number of skin contacts, $p < 0.001$; total time in no-go area, $p = 0.008$*. More experienced participants consistently received better performance scores on all 19 performance metrics.

Conclusions: It is possible to measure and monitor performance using simulation, with performance metrics providing feedback on skill level and differentiating levels of expertise. However a transfer of training study is required.

Key Terms

IR	Interventional Radiology
TA	Task Analysis
CPS	Critical Procedure Steps
SME	Subject Matter Expert

INTRODUCTION

Training in Interventional Radiology (IR) uses the traditional apprenticeship model despite recognised drawbacks e.g. difficulty articulating expertise, pressure to train more rapidly [1], reduced training opportunities. Moreover it has been described as inefficient, unpredictable and expensive [2, 3] and its suitability for training questioned due to no mechanism for measuring post-training skill [4]. There is increasing need to develop alternative training methods [5]. Using simulators to train offers numerous benefits including gaining experience free from risk to patients, learning from mistakes, and rehearsal of complex cases [6]. IR is particularly appropriate for simulator training since skills such as interpreting 2D X-ray or ultrasound images, can be reproduced in a simulator in the same way as in real-life procedures.

There is increased use of medical virtual reality simulators with some validated to show improved clinical skills, e.g. laparascopic surgery [7], colonoscopy [8] and anaesthetics [9]. However, within IR no simulator has met this standard [5, 6, 10] with validation studies typically failing to discriminate accurately between experts and novices [11], although differences have been observed [12]. Length of time to complete procedures on simulators is a frequently reported expertise discriminator [6] but there is a worrying lack of emphasis on the number of errors made or other clinically relevant parameters. A recent review [6] reported ‘fundamental inconsistencies’ and ‘wide variability in results’ in validation studies, concluding that the analysis of errors and quality of the end product should be the focus of assessment. The authors proposed that, to fully develop and validate simulators, there is a need for task analysis (TA) to deconstruct individual procedural tasks followed by

metric definition and critical performance indicator identification. This echoes previous calls for expert involvement in simulator design [13].

To the best of our knowledge no IR simulators have been developed through the use of TA of real world tasks despite the critical role of such techniques in training development and system design for the past 100 years [14]. TA identifies knowledge and thought processes supporting task performance, and the structure and order of individual steps, with particular relevance in deconstructing tasks conducted by experts [15, 16]. TA techniques are increasingly being used as a medical educational resource, e.g. the development of surgical training [17], teaching of technical skills within surgical skills laboratories [18].

Using task analysis, this research identified and prospectively investigated simulated ultrasound-guided targeted Liver Biopsy performance metrics as differentiators between levels of expertise in Interventional Radiology.

MATERIALS AND METHOD

The research method incorporated three objectives:

1) Produce a detailed hierarchical and cognitive TA of an ultrasound-guided, targeted liver biopsy procedure, to inform simulator design.

2) Identify the critical performance steps (CPS) key to safe completion of the procedure, to inform measurement of performance on the simulator.

3) Perform construct validation comparing novice and expert performance on the simulator.

These objectives are described in more detail below, followed by technical information into simulator design.

1 Task analysis (TA): data was collected through: discussions with clinician collaborators; observation of procedures; video-recording procedures; interviews.

Discussions with clinical collaborators informed on the nature and aim of the procedure. Observing a small number of IR procedures allowed understanding of the complexity of the tasks, equipment and environment of the IR suite prior to the acquisition and observation of video data.

Ethical and Research Governance approval was granted for video data collection from the Liverpool Research Ethics Committee. Data was collected at two UK hospitals and a total of four liver biopsy videos obtained. Patient written informed consent was obtained.

Twelve interviews were conducted. A wide selection of experts was assured through interviews at clinical sites in the UK and at three high profile IR conferences: CIRSE (Copenhagen, Sept 2008); BSIR (Manchester, Nov 2008); SIR (Washington, Mar 2008; San Diego, Mar 2009). Interviewees were identified as subject matter experts (SME) by their professional bodies. SMEs were asked to describe the liver biopsy procedure and questioned about the following: aims and stages of the procedure; tools and techniques; 'decision points'; potential complications; the amount of risk individual steps in a procedure pose to a patient.

Interviews were semi-structured to allow exploration of points raised by SMEs and lasted approximately one hour. Salient points and issues were explored in more detail with subsequent SMEs. SMEs were asked to justify the use of certain techniques and decisions, and the information/cues used to guide this decision making.

Whilst data collection was ongoing, a first draft of the TA for the liver biopsy procedure was produced. Observable actions were identified from video data and included in task descriptions. Further information was taken from the interviews including: detailed descriptions of the procedure; potential errors/complications, including their identification and prevention; cues that indicate procedure success, safety and completion; decisions made and the information that guided these decisions. Areas needing clarification were identified and explored in subsequent interviews. Throughout this process the TA was redrafted and refined. Drafting the document in this manner provided a fully developed tool.

2 **Critical Performance Steps (CPS)**: the TA was used in the design of a questionnaire to provide data to identify the CPS of the liver biopsy procedure. Respondents rated each step of the TA on seven-point Likert scales ranging from *very strongly disagree* to *very strongly agree* in two areas: (a) whether the step is of critical importance to a successful procedure; (b) whether the step poses a high risk to a patient. Eight subject matter experts completed the CPS questionnaire. Interview transcripts were reviewed to identify the critical parts of the procedure as discussed by SMEs. Performance metrics relating to the CPS were identified and incorporated in the liver biopsy simulator to allow measurement of participant performance.

3 **Construct validation:** investigated the use of CPS and related performance metrics to measure performance and to differentiate across expertise. Ethical and research governance approval was granted by the Liverpool Research Ethics Committee and the study conducted in three UK hospitals. SMEs and IR trainees participated in the study. A total of 41 respondents (SMEs, n=14; trainees, n= 26; unreported status, n=1) participated and provided full informed consent. All participants completed the simulated liver biopsy procedure and their performance assessed with the performance metrics. Independent t-tests investigated whether SMEs / consultants produced better performance metrics on the simulator than trainees. ANOVA investigated performance differences across number of years experience in IR.

Simulator design: The simulator is composed of 1) a mannequin to initially scan a patient and perform a skin nick with a scalpel; 2) a virtual environment to display the patient while inserting a needle; 3) two haptic devices to replicate the medical tools and 4) a simulated ultrasound (see figure 1). Each component is detailed below: 1)

the mannequin has a human torso shape and motors reproduce the breathing behavior. It is linked with two sensors to track the position of the faked ultrasound probe and the scalpel respectively. 2) The virtual environment displays a three-dimensional patient, with 3D glasses on a semi-transparent mirror. It includes organ deformations and the visualization of both the needle and the ultrasound probe. 3) The first haptic device simulates the ultrasound probe action with a stiff contact feedback when touching the external skin and a simulated ultrasound rendered according to the position and orientation of the device. The second haptic device simulates the needle with force feedback based on the insertion behavior while penetrating through soft tissue. Its position is used to compute the metrics. 4) The ultrasound is displayed on a monitor and is a computer-based simulation based on the organ geometry associated with their deformation.

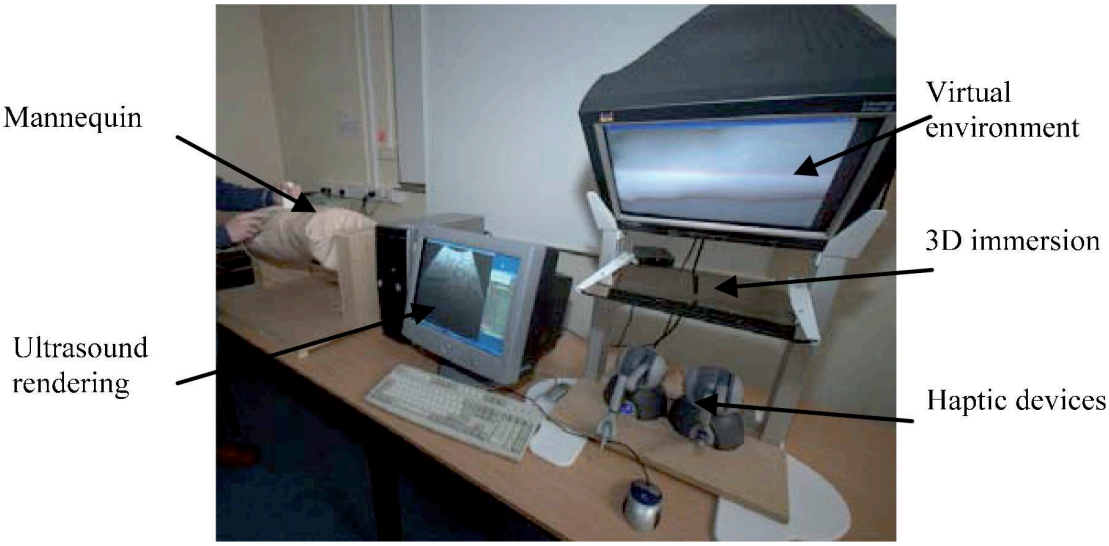

Figure 1. Ultrasound-guided liver biopsy simulator components.

RESULTS

1 Task analysis (TA): the first column of Table 1 shows a short segment of the ultrasound-guided targeted liver biopsy TA; the complete document involves 174 procedural steps and is therefore not reproduced in full.

Table 1: Segment of Liver Biopsy TA with quantitative and qualitative critical performance steps (CPS)

Procedure Steps Quantitative CPS[1]	Qualitative CPS	Performance Metric
Inject local (this produces a bleb on the skin which looks like a nettle sting and gives near (0-2) mins instantaneous anaesthesia). 1-2 cc initially.	Identify access point and inject local: in correct position for access given trajectory and approach	Tick box.
Approx 1cm deep skin nick at planned incision site. The blade can be pushed directly down as planned biopsy site has been identified so there is no danger of damaging anything underneath the skin. A deeper cut reduces any resistance when needle inserted. *NOTE: Skin nick made to facilitate needle insertion. More difficult to insert biopsy needle without skin nick.*	Identify access point and nick skin: in correct position for access given trajectory and approach	Tick box.
Looking at the ultrasound screen roll the probe back and forth on the incision site until the planned biopsy site is identified.		Target visible as a percentage of total probe usage time.
Note the depth of the planned biopsy site using the markers on the ultrasound screen.		Simulated US screen shows measurement markers.
Pick up the biopsy needle from the green sheet - make note of the measurement markers on the needle to identify how deep this needle is going to be inserted.		Not measured.
Holding the needle shaft begin to insert the needle into incision.		Time on target (i.e. to lesion with needle).
Position needle and probe and planned biopsy site in the same plane (to ensure all necessary information can be viewed on the ultrasound screen).	Triangulation correct, i.e. angle of probe and angle of needle correct to reach target	Time on target (i.e. to lesion with needle). Average needle length in beam. Needle in beam as a percentage of total probe usage time.

Slowly advance the needle towards the planned biopsy site, constantly viewing the ultrasound image to guide orientation and progress. NOTE: If utilizing an intercostal approach, aim the needle over the top of the lower rib as opposed to the bottom of the upper rib to avoid hitting nerves.	Coordinate hands and images: keep length of needle on screen at all times). NB. In large patients need to insert needle before seeing it due to angle	Time on target (i.e. to lesion with needle). Average needle length in beam. Needle in beam as a percentage of total probe usage time. Needle tip in beam as a percentage of total probe usage time. No-go area touched. Total time in no-go areas. Number of tissues encountered. Target visible as a percentage of total probe usage time.
Continue advancing the needle in this way until the planned biopsy site is reached (the needle should be positioned just short of the planned sample site, if a tumour is being biopsied the needle needs to be at the edge of the tumour), be guided both by the image and by your knowledge of the depth of the planned biopsy site.	Watch needle in real time on US as introduced into the liver	Time on target (i.e. to lesion with needle). Average needle length in beam. Needle in beam as a percentage of total probe usage time. Needle tip in beam as a percentage of total probe usage time. No-go area touched. Total time in no-go areas. Number of tissues encountered. Target visible as a percentage of total probe usage time. Targeting quality. Number of biopsy samples taken.

[1] Critical procedure steps in bold were rated as important and carrying significant risk if performed incorrectly; steps in italics were rated as important but were not considered a significant risk.

Many procedural details derived from the TA were included in the simulator to increase the immersion and procedure realism. Two examples are detailed here. Firstly, the influence of respiration on the procedure was incorporated through development of a breathing model to allow participants to see organ motion on the ultrasound and to have needle oscillation in the 3D environment. Secondly, the haptic feedback was developed to fully reproduce the needle insertion steps: skin perforation, various soft-tissue behaviour and liver capsule penetration.

2 Critical Performance Steps (CPS): The CPS questionnaire responses provided frequency data to assist identification of those aspects of the procedure that should be measured to inform on performance in the simulator. More detailed statistical analysis of the quantitative data was not appropriate due to the restricted sample size (n=8). If a step was rated as either *important to successful completion of the procedure* or *carrying significant risk to the patient*, or both, by the majority of experts (i.e. by at least five experts) it was highlighted in the TA documentation. A total of 30 CPS steps were identified for the liver biopsy procedure. The first column of Table 1 indicates the quantitative CPS for this segment of the liver biopsy procedure.

Reviewing the interview transcripts allowed the identification of 13 aspects of the procedure that were described by experts as being critical to successfully and safely completing the procedure. These corresponded with the important and risky steps identified through the CPS questionnaires (the second column of Table 1 reports qualitative CPS). It was not unexpected to identify fewer CPS points in the qualitative data collection as verbal descriptions of procedural actions are less detailed than a documented TA. One qualitative CPS step therefore incorporates a number of quantitative CPS steps.

A final TA document was produced that detailed the procedure steps, the quantitative and qualitative CPS. This document informed simulator development. Decisions on the inclusion of performance metrics was made in relation to (1) the CPS identified by the procedure detailed above, and (2) the capability of the simulator to measure performance on these steps.

Where possible each CPS was measured directly. However, some aspects of the procedure were not represented in the simulator and yet had been identified by experts as CPS. Care was therefore taken to ensure these steps were represented in the simulator despite not being fully simulated. For example *ensure there are no air bubbles in a syringe of local anaesthetic* was brought to the attention of the participant through the use of a 'check-box'. This ensured key procedural steps were not fully omitted.

A total of 19 quantitative performance metrics were included in the simulator, see Table 2. To provide further context, the third column of Table 1 indicates the performance metric for each step of the procedure. It was therefore possible to produce accurate measurement of participant performance on each of these 19 performance metrics. Participants with differing levels of expertise in IR were then compared, allowing discriminant validation of the simulator and the selected performance measures.

Table 2: Liver Biopsy Performance Metrics Incorporated into Simulator

Liver Biopsy Performance Metrics
Number of no-go[1] areas touched
Targeting quality (no sample, no pathology, some cells, perfect)
Number of attempts
Number of biopsy samples taken
Length of session [s]
Probe usage time [s]
Time on target (needle accurately directed towards lesion) [s]
Probe usage as a percentage of total procedure time [%]
Average needle length in beam [mm]
Target visible as a percentage of total probe usage time [%]
Needle in beam as a percentage of total probe usage time [%]
Needle tip in beam as a percentage of total probe usage time [%]
Total probe distance moved [mm]
Average probe distance moved [mm]

Total needle distance moved [mm]
Average needle distance moved [mm]
Number of skin contacts
Number of tissues encountered
Total time in no-go[1] areas [%]

[1]*no-go area: an anatomical region where needle penetration would represent a risk for complication or patient discomfort.*

It was predicted that performance on the simulator would relate to IR competency. Specifically, participants more experienced in real life IR procedures would perform better on the simulator.

Detailing each technical implementation of the metrics would be too tedious for the reader, instead only the first two are explained here: “Number of no-go areas touched” and “Targeting quality”. First, a list of anatomical no-go areas was obtained from the SME interviews and a detection collision algorithm used to check if the virtual position of the needle tip touches any of these areas during the process. Summarizing the results on the whole body gives the total number of no-go areas touched. Second, the simulator records the distance *d* between the center of the sphere representing the tumour and the position of the virtual needle when the haptic device button is pushed (representing the firing of the biopsy needle). The targeting quality is computed as follows:

$$\text{targeting quality} = \begin{cases} \text{"necrosis debris"} \; if \; d < 5 \\ \text{"perfect"} \; if \; d < 15 \\ \text{"some cells"} \; if \; d < 20 \\ \text{"no pathology"} \; \text{if d} = else \end{cases}$$

<u>3</u> Construct validation studied performance of IR consultants (n=14) and trainees (n=26). Table 3 reports the significant differences revealed for three of the performance metrics: *targeting; probe usage time; mean needle length in beam*. These

were in the expected direction with consultants achieving better performance metrics than trainees. Specifically, consultants: obtained significantly better 'biopsy samples'; used the probe significantly less (suggesting they were more proficient and better able to locate the target); had significantly more needle length in the beam (indicating greater proficiency in needle identification and tracking). While significant differences were shown in just 3 metrics, the mean scores for the remaining 16 performance metrics showed a consistent trend of better performance for consultants than trainees.

Table 3: Consultant and trainee performance metric comparison

Performance Metric	Participant	Mean	Standard Deviation	t	p	95% confidence interval
Targeting*	Trainee	0.96	1.27	-2.487	0.018	-2.040 to -0.207
	Consultant	2.08	1.31			
Probe usage time	Trainee	699.32	335.86	2.132	0.040	11.064 to 427.983
	Consultant	479.80	255.20			
Mean needle length in beam	Trainee	0.03	0.02	-2.272	0.029	-0.028 to -0.002
	Consultant	0.04	0.02			

* 0 = no sample, 1 = no pathology, 2 = some cells, 3 = perfect

Performance was also compared across number of years of experience: 0 to 1 years (first year trainees), n = 8; 1 to 2 years, n = 13; 3 or more years (including final year trainees and consultants), n = 10. Participants who did not report their years of experience were removed from the dataset prior to analysis.

Table 4 reports the significant differences revealed across years of experience for seven of the performance metrics: *number of no-go areas touched; targeting; length of session; probe usage time; total needle distance moved; number of skin contacts; total time in no-go area*. Each of these was in the expected direction. Inspection of the

mean scores for all performance metrics revealed that, although only these 7 were significant, the performance metrics again followed a predictable pattern across experience. That is, more experienced participants consistently received better scores on each of the performance metrics.

Table 4: Comparison of performance metrics across years of experience in IR

Performance Metric	Experience (in years)	Mean	Standard Deviation	F	p
Number of no-go areas touched	**0-1**	3.75	2.05	5.218	0.012
	1-2	2.00	1.35		
	3+	1.33	1.41		
Targeting*	**0-1**	0.38	0.74	4.258	0.025
	1-2	1.25	1.42		
	3+	2.11	1.27		
Length of session	**0-1**	866.18	201.60	4.277	0.024
	1-2	690.25	396.20		
	3+	447.29	227.09		
Probe usage time	**0-1**	809.82	255.88	4.209	0.025
	1-2	645.65	377.43		
	3+	400.96	211.60		
Total needle distance moved	**0-1**	24.15	10.16	3.691	0.038
	1-2	16.10	8.71		
	3+	12.80	8.34		
Number of skin contacts	**0-1**	58.00	14.57	10.792	<0.001
	1-2	33.69	22.26		
	3+	19.40	11.74		
Total time in no-go area	**0-1**	50.32	27.22	5.812	0.008
	1-2	19.81	34.38		
	3+	7.13	12.34		

* 0 = no sample, 1 = no pathology, 2 = some cells, 3 = perfect

Tukey's post hoc test showed that there were differences between the *0-1 years experience* and *3 or more years experience* on all seven significant performance metrics. Significant differences were also found between *0-1 years experience* and *1-2 years experience* on two of the metrics, *number of skin contacts* and *total time in no-go area.*

The acceptability and utility of the simulator to participants is indicated with 84% of participants agreeing it would be useful in learning the steps of a procedure, and 81% believing it would be useful for procedure rehearsal. The quality of the feedback provided by the simulator was also indicated with 80% of participants reporting the feedback as accurate.

DISCUSSION

This paper has described the deconstruction of the ultrasound-guided targeted liver biopsy procedure using TA techniques, resulting in detailed documentation of the steps required for successful completion of the procedure. Additionally, those steps of most importance to the success of the procedure and in managing risk to the patient were identified (i.e. the CPS). Performance metrics that allowed the measurement of these CPS were incorporated into the simulated procedure, thus providing a means by which to measure and monitor participant performance. Comparing performance on the simulated liver biopsy, both between trainees and consultants, and between differing lengths of experience in IR, revealed significant differences on a total of eight performance metrics. Each of these was in the expected direction with more experienced participants consistently performing better than less experienced participants. Furthermore, a consistent trend was revealed across all performance metrics, i.e. more experienced participants received better scores on all 19 performance metrics. The results indicate that the identified performance metrics can accurately discriminate for expertise, and that number of years of experience in IR is a more informative level of comparison than a simple trainee / consultant split.

A strength of the study is the care taken to ensure the simulated procedure was based upon real-life procedures, with the TA and CPS identified through subject matter expert input. Simulator content and development was guided and underpinned by expert knowledge thus allowing theoretically based performance evaluation using evidence based components. The need for expert input into simulator design is recognized [13] and identified as lacking in existing IR simulation research [6]. Using this approach ensured that the simulated procedure was comprehensive and realistic,

with the monitoring of performance focused upon the most important aspects of the procedure in terms of successful completion and management of risk to the patient. The significant differences that were reported in the performance metrics indicate that it is possible to assess expertise through performance on the simulator. This further suggests that training on the simulator would allow experience to be gained on critical elements of the procedure, with a subsequent improvement in performance to be anticipated following practice.

This is the first study of IR simulation that has accurately discriminated across expertise on a number of performance metrics. It is therefore an important addition to existing research into IR simulation which has to date failed to discriminate expertise, although differences across expertise have been reported [11, 12]. Length of time to complete a simulated procedure has previously been found to discriminate expertise [6, 19] and was confirmed here. The main contribution of the present study though is in the identification of clinically relevant parameters and errors that can discriminate expertise (e.g. total needle distance moved, number of no-go areas touched). Trainees thus receive performance feedback on those areas that allow an assessment of procedural quality, i.e. analysis of errors and the quality of the end product [6].

The research reveals that it is possible to measure and monitor performance on a liver biopsy procedure using simulation, with performance metrics providing feedback on various aspects of the procedure. This should enable participants to identify those elements of a procedure they have either mastered or that require further skill development. This is likely to be useful during training, particularly given the known difficulties in gaining significant practice of core skills in patients. Simulators have

the potential to facilitate training in areas where traditional training is failing to keep up with demand and have proven to be useful in the medical field in other areas (e.g. laparoscopy, colonoscopy). The results from the current study suggest that simulation may also be useful in training key skills in IR. It was also reported that the majority of participants believed the simulator to be an acceptable and useful training tool. It is not proposed that simulation could, or indeed should, replace existing training in IR which is a profession that encompasses far more skills than simulators can currently inform on (e.g. decision making, complexity, and interactions in the real environment). Nevertheless this study, alongside the increasing interest and research into simulation in IR and the existing problems with current training such as lack of opportunity to train and pressure to train more rapidly [6], strongly indicates that if developed and used appropriately IR simulation could prove to be an invaluable and highly informative training tool. We agree with other authors [e.g. 6, 13] that simulation should augment rather than replace existing methods.

There are many advantages to simulator training based on task-specific information that has been appropriately validated. Trainees will: have the opportunity to practice a standardized task with a high degree of realism; receive feedback on their performance through the use of validated assessment metrics derived from experts; gain basic skills that can become automated before procedures are conducted with real patients. Further advantages include an absence of the ethical issues of training in patients and animals, the opportunity for a simulator to cover a number of anatomically different procedures, and the use of patient-specific simulations that allow practice of a complex procedure prior to actual performance in the patient. In addition to these benefits it has been proposed [13] that simulators have the potential

to assist with: aptitude testing; advanced skills training; career-long training; board examination; and credentialing. There are of course a number of disadvantages, the most obvious of which is the cost. Simulator development is time consuming and expensive. Buying a simulator for use by trainees is very expensive and in 2007 was estimated at between $100,000 to $200,000 with additional annual service costs of between $10,000 and $16,000 [5]. Day to day costs of using simulators can also be significant, for example transport and maintenance costs. Some of these costs may, however, be reduced by cost-effective technologies from the games industry.

A limitation of the study is the small numbers of participants and further work with larger sample sizes is desirable to confirm the most useful performance metrics and to develop 'norm' scores to provide performance benchmarks against which trainee scores can be compared. Furthermore, although the results of the current study provide evidence of the discrimination of expertise by use of the simulator performance metrics it does not prove skills can be acquired in the simulator. A transfer of training study of any simulator is needed to provide evidence of skills acquisition on the simulator, and of the transferability of those skills to real world procedures in patients. Evidence of transfer of training in simulation should be a prerequisite for adoption as a training tool.

REFERENCES

1 BERR: Department for Business Enterprise and Regulatory Reform: Your Guide to the Working Time Regulations http://www.berr.gov.uk/employment/employment-legislation/employment-guidance/page30342.html. Accessed 01/05/10.

2 Bridges M, Diamond DL. The financial impact of training surgical residents in the operating room. *American Journal of Surgery* 1999;177:28-32.

3 Crofts TJ, Griffiths JM, Sharma S, Wygrala J, Aitken RJ. Surgical training: an objective assessment of recent changes for a single health board. *British Medical Journal* 1997;314:814.

4 Gallagher AG, Cates CU. Approval of virtual reality training for carotid stenting: What this means for procedural-based medicine. *Journal of the American Medical Association* 2004;292:3024-3026.

5 Neequaye SK, Aggarwal R, Van Herzeele IV, Darzi A, Cheshire NJ. Endovascular skills training and assessment. *Journal of Vascular Surgery* 2007;46:1055-1064.

6 Ahmed K, Keeling, AN, Fakhry M, Ashrafian H, Aggarwal R, Naughton PA et al. Role of Virtual Reality Simulation in Teaching and Assessing Technical Skills in Endovascular Intervention. *Journal of Vascular Interventional Radiology*, 2010;21:55-66.

7 Larsen CR, Sorensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, et al. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *British Medical Journal*, Online First, BMJ 2009;338:b1802.

8 Sedlack R, Kolars J. Computer simulation training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. *American Journal of Gastroenterology* 2004;99:33-37.

9 Rowe R, Cohen R. An evaluation of a virtual reality airway simulator. *Anaesthesia and Analgesia* 2002;95:62-66.

10 Gould DA, Reekers JA, Kessel DO, Chalmers N, Sapoval M, Patel A, et al. Simulation devices in interventional radiology: validation pending. *Journal of Vascular and Interventional Radiology* 2006;17:215-216.

11 Berry M, Lystig T, Reznick R, Lonn L. Assessment of a virtual interventional simulator trainer. *Journal of Endovascular Therapy*, 2006;13:237-243.

12 Duncan JR, Kline B, Glaiberman CB. Analysis of simulated angiographic procedures: Part 2 – extracting efficiency data from audio and video recordings. *Journal of Vascular and Interventional Radiology*, 2007;18:535-544.

13 Dawson S. Procedural Simulation: A Primer. *Radiology*, 2006;241:17-25.

14 Militello LG, Hutton RJB. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, 1998;41:1618-1641.

15 Reynolds R, Brannick M. Thinking about work/thinking at work: Cognitive Task Analysis. In Tett, R. P., & Hogan, J. C. Recent developments in cognitive and personality approaches to job analysis. *Symposium presented at the 17th annual conference of the Society for Industrial and Organizational Psychology* 2002;Toronto:Canada http://luna.cas.usf.edu/~mbrannic/files/biog/CTASIOP_02.htm. Accessed 01/05/10.

16 Schraagen JM, Chipman SF, Shalin VL. Introduction to cognitive task analysis. In J.M. Schraagen, S.F. Chipman, & V.L. Shalin, (Eds.). *Cognitive Task Analysis.* Mahwah, NJ: Lawrence Erlbaum Associates; 2000.

17 Grunwald T, Clark D, Fisher SS, McLaughlin M, Narayanan S, Piepol D. Using cognitive task analysis to facilitate collaboration in development of simulator to accelerate surgical training. *Proceedings of the 12th Medicine Meets Virtual Reality Conference* 2004:114-120.

18 Velmahos GC, Toutouzas MD, Sillin LF, Clark RE, Theodorou D, Maupin F. Cognitive task analysis for teaching technical skills in an inanimate surgical skills laboratory. *The American Journal of Surgery*, 2004;187:114-119.

19 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. *Journal of Vascular Surgery*, 2007;46:855-863.