Ultrasound Obstet Gynecol 2002; **19**: 478–483

Teaching invasive perinatal procedures: assessment of a high fidelity simulator-based curriculum

R. PITTINI*, D. OEPKES§, K. MACRURY*, R. REZNICK†, J. BEYENE‡ and R. WINDRIM*

*Departments of *Obstetrics and Gynecology, †Surgery and ‡Paediatrics, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada, §Department of Obstetrics, Leiden University Medical Center, Leiden, Netherlands*

KEYWORDS: Amniocentesis, Medical education, Patient simulation

ABSTRACT

Objective *Learning curves pose a difficult problem in the teaching of technical skills: how do you teach procedural skills without compromising patients' health? A simulator-based curriculum has been designed to minimize the risks to patients undergoing amniocentesis by shifting the learning curve away from patients and into the laboratory. This study evaluated the effectiveness of a high-fidelity simulator-based curriculum in improving the performance of amniocentesis by obstetric trainees.*

Design *Thirty trainees received a course on the practice of amniocentesis. The curriculum consisted of a lecture, a syllabus, and a hands-on training session with the simulator. Pre- and post-training performance were evaluated with two rating scales. Training and performance evaluation were completed using the same simulator. The effectiveness of the simulator-based workshop and the effect of year of training were assessed using a two-way analysis of variance.*

Results *Performance scores improved from a mean score of 55% to 94% using checklist scoring and from 57% to 88% using global ratings. The two-way analysis of variance revealed a significant effect of training ($F_{1,60} = 43.57$; $P < 0.001$) accounting for 45% of the variance in scores, and a significant effect of experience level ($F_{2,60} = 9.16$; $P < 0.001$) accounting for 25% of the variance in scores.*

Conclusions *A comprehensive curriculum based on a high-fidelity simulator was effective at improving skills demonstrated on the simulator. The challenge remains to establish that skills acquired on a simulator are transferable to the clinical setting.*

INTRODUCTION

The safety of technical procedures is dependent on the experience of the operator. Teaching technical skills without compromising the safety of patients is a challenge for educators. Increasingly, educational programs directed at the acquisition of technical skills are studying the benefits of novel approaches to training. While the utility of animal models and cadavers has been limited by ethical considerations, cost and availability, simulators have continued to be developed and are becoming more realistic and pervasive.

Invasive perinatal procedures ranging from simple genetic amniocentesis to complex *in-utero* fetal surgery lend themselves well to simulator-based teaching. The procedures require excellent hand–eye coordination, the ability to conceptualize a three-dimensional space based on two-dimensional ultrasound images and a thorough understanding of fetal anatomy[1]. Many invasive perinatal procedures are performed in only a very few select centers, carry a high risk of complications and are performed on an infrequent basis[1]. Amniocentesis increases the risk of miscarriage by 0.5–1.0%[2]. This risk of complications is related to both patient and physician factors[3]. Patient factors can only be controlled by appropriate patient selection. Physician factors include operator experience and the number of needle insertions[3]. These factors can be modified through education and training.

Several attempts have been made in the last decade to develop simulations for performing invasive fetal procedures. Early attempts described by Timor-Tritsch and Yeh[4] in 1987 and Angel *et al.*[5] in 1989 involved a biological model constructed from human birth byproducts simply immersed in water. Both placentae and umbilical cords were utililized to simulate different aspects of the procedures being taught. A second model described by Ville *et al.* in 1995[6] was a synthetic model that simulated the maternal abdominal wall, uterus, placenta and umbilical cord. The realism of this model was limited to the ultrasound images it produced. Additional models have been described by Smith *et al.*[7] as well as Maher *et al.*[8] in 1998; these models incorporate a number of 'targets' suspended in gelatin. These models achieved some degree of realism in the ultrasound reproduction of invasive procedures and incorporated a didactic component.

Correspondence: Dr R. Pittini, Sunnybrook & Women's College Health Sciences Center, 60 Grosvenor St., Suite 642, Toronto, ON M5S 1B6, Canada (e-mail: richard.pittini@swchsc.on.ca)

Accepted 4-3-02

Previous simulations have focused on ultrasound skills and failed to recreate the entire process of the invasive perinatal procedure they were simulating. In the reported studies, the simulators were used for teaching in the absence of well-planned and comprehensive curricula. With each successive generation of simulator the images produced and the tactile properties obtained improved but none of the models incorporated simulated patients, assistants, or realistic facilities. Each of these limitations could affect the transfer of skills acquired in the simulated environment to the actual patient.

In addition to limited transferability, skills learned on simulators may lack depth if they are learned without the benefit of a structured curriculum. A structured curriculum is one that is designed with a logical sequence of learning objectives and associated activities. In the context of invasive perinatal procedures, this involves learning the theoretical background, developing communication skills, understanding equipment, and learning to utilize assistants optimally. In order to teach trainees to perform an amniocentesis from beginning to end there should be an attempt to address each of these aspects. A curriculum that incorporates all of these features would be considered comprehensive.

This study aimed to demonstrate that a comprehensive simulator-based curriculum would be an effective means of teaching medical trainees at various levels to perform an amniocentesis. The curriculum utilized several teaching methods and incorporated a high-fidelity amniocentesis simulator. 'High fidelity' refers to the close reproduction of the actual clinical environment as distinct from 'low fidelity', which reproduces the clinical environment in an incomplete or rudimentary fashion. The objective was to teach the process of performing an amniocentesis rather than focusing exclusively on the strictly technical aspects of the procedure.

Figure 1 Ultrasound image showing the needle tip in a pocket of fluid immediately above the 'fetal head'.

Figure 2 Simulator development.

METHODS

This study consisted of the development of a comprehensive simulator-based curriculum and the subsequent assessment of its effectiveness. It employed a pretest–post-test research design to evaluate the effect of trainee year, simulator-based training and the interaction between the two. The curriculum consisted of a 4-h workshop that included a baseline evaluation, review of a syllabus, attendance at a brief lecture, a faculty demonstration, supervised hands-on practice on a simulator and a final evaluation. The performance of trainees was compared to a benchmark set by experts in the field of invasive perinatal procedures. Participant performance was evaluated by independent trained examiners with both content and educator expertise. Assessments were made using both a global rating form (GRF) and a detailed checklist (DCL) developed by national consensus (unpublished report by the TIPP Study Group: Canadian Perinatologist Survey on Teaching Amniocentesis, 2000). The reliability and validity of these evaluation scales were assessed. Questionnaires administered prior to and following the workshop were used to collect qualitative data regarding participant perceptions of the value of the simulation.

The development of a simulator with appropriate sonographic properties (Figure 1), tactile properties, anatomical realism (Figure 2) and the ability to be integrated into a simulation that included standardized patients and assistants was commissioned. A panel of perinatology experts developed a syllabus for the training session, which included the program goals and objectives, content outline and training manual. The steps selected for inclusion in the curriculum were initially determined by the panel of perinatology experts; the available literature was then reviewed and where evidence was lacking the opinion of the respondants to a survey of Canadian perinatologists was used (Table 1). Standardized patients, i.e. actors trained to accurately and consistently portray patients, were used in conjunction with the amniocentesis simulator to enhance the realism of the simulation (Figure 3) and to facilitate evaluation of communication skills as they pertain to amniocentesis.

Study participants included undergraduate and postgraduate trainees as well as fellows in maternal fetal medicine from the Department of Obstetrics and Gynecology at the University of Toronto. Trainee participation was on a voluntary basis and served as supplementary training. Trainees were grouped into three cohorts according to their year of

Table 1 Amniocentesis technique used in curriculum

1	Materials verified
2	Appropriate utilization of assistants
3	Initial ultrasound
4	Selection of site
5	Prep and drape
6	Maintain sterile technique
7	Needle insertion under ultrasound guidance
8	Continuous ultrasound monitoring
9	Withdraw fluid
10	Remove needle
11	Post-procedure ultrasound
12	Verify fetal well-being
13	Label specimen
14	Effective communication with patient
15	Post-procedure counseling

Figure 3 Standardized patients were used in conjunction with the amniocentesis simulator to enhance the realism of the simulation.

training. The novice group consisted of medical students and residents in their first and second postgraduate years (PGY 0–2). The intermediate group consisted of residents in their senior years of residency (PGY 3–5). The final, near expert, group consisted of fellows in maternal fetal medicine (PGY 6+). Although the risk to the trainee participants was estimated to be minimal, ethical approval was obtained from the University of Toronto research ethics board to conduct research using human subjects. Written informed consent was obtained from all participants.

Performance was assessed using two evaluation scales, a GRF and a DCL. The content of both of these scales was determined by the survey responses from a national survey of perinatologists. The scales were designed as independent evaluation tools and each covered all pertinent content areas. The GRF was a twenty-item scale. Each item was rated using a Likert scale from 1 to 5, with descriptive anchors for ratings 1,3, and 5. The items used were adapted from the Objective Structured Assessment of Technical Skills (OSATS) GRF used in the evaluation of general surgery residents at the University of Toronto[9]. The GRF assessed knowledge of procedure, sterile technique, and technical skills. Two items were included to assess communication skills. These two items were adapted from a GRF developed by Dr Cleo Boyd, educator at the University of Toronto, specifically for the assessment of communication skills. A rating for overall performance was included in the twenty-item scale. The DCL consisted of 25 items highlighting the steps felt to be important in the performance of an amniocentesis. The items were rated 'done correctly' or 'not done/done incorrectly'.

Evaluators completed these forms during and immediately following the completion of the procedure by the trainee. Items not observed were not assigned a score. Crude total scores were converted to proportion scores by dividing the score assigned by the maximum score possible based on the number of items observed. An average score for each scale was calculated for each of the three cohorts: PGY 0–2, PGY 3–5, and Fellows. These scores were calculated both pre- and post-test.

The improvement observed following training and the effect of postgraduate year was evaluated using a two-way analysis of variance (ANOVA). The significance level of the observed improvement was determined using the *F*-value for training, PGY level and the interaction between these two factors. This analysis was repeated using both the DCL and the GRF. Improvements within each cohort were analyzed using a two-tailed paired *t*-test.

The reliability of each scale was assessed in terms of inter-rater reliability and interform reliability. Inter-rater reliability was estimated using the intraclass correlation coefficient. Individual intraclass correlation coefficients were calculated for each of four observer pairings. An overall intraclass correlation coefficient was based on the average of these four observer pairings. The coefficient was based on a two-way random effects model and used average measures. The interform reliability was based on a comparison of the performance scores obtained using both the DCL and the GRF. A correlation coefficient was calculated based on the average of pretest scores assigned by four raters for all subjects for each scale. The 95% confidence interval (CI) was calculated for all correlation coefficients.

The ability to discriminate individuals with various levels of prior training was used to test the construct validity of each of the rating scales. The construct validity of the performance scores was evaluated by an ANOVA on pretest scores based on the level of postgraduate training. Postgraduate trainees were grouped as follows: PGY 0–2, PGY 3–5, Fellows, and Faculty. The faculty members were experts in the field of amniocentesis as well as experienced medical teachers. Their performance was taken as gold standard. Significance levels were based on the *F*-value, effect size was estimated using eta^2 and a post hoc analysis was used to determine which groups differed significantly. The authenticity of the performance scales was evaluated by surveying participants with the following question: 'Please rate the realism of the amniocentesis simulator'; possible responses ranged from 1 (not at all realistic) to 5 (extremely realistic). Feedback was also sought from the participants regarding the helpfulness of the workshop. The descriptor for a score of 1 was 'not at all helpful' and for 5 was 'extremely helpful'.

RESULTS

Thirty trainees participated in the amniocentesis workshop. The first cohort included 12 participants in their PGY 0 (undergraduate medical students) and PGY 1. These participants had minimal prior experience with amniocentesis, only one subject having ever attempted an amniocentesis. The second cohort consisted of 15 senior residents in their PGY 3–5. Their level of prior experience and exposure to amniocentesis varied. The average number of prior amniocenteses was two; however, eight of the subjects had not performed an amniocentesis. The final cohort consisted of three maternal fetal medicine Fellows. Each Fellow had prior experience of performing an amniocentesis. The average number of prior amniocenteses performed was 50.

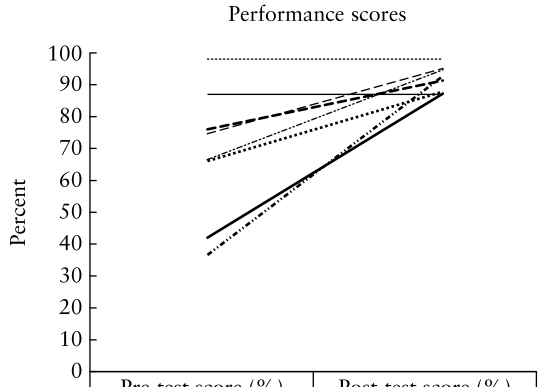
An improvement in the performance of a simulated amniocentesis by the study subjects was observed across all cohorts using both the DCL and the GRF. The DCL performance score improved from 54% to 94% while the GRF performance score improved from 57% to 88% following participation in the amniocentesis workshop.

In comparing the improvement of group scores on the GRF, the two-way ANOVA revealed a significant effect of training ($F_{1,60} = 43.57$; $P < 0.001$), accounting for 45% of the variance in scores, a significant effect of experience level ($F_{2,60} = 9.16$; $P < 0.001$), accounting for 25% of the variance in scores, and a significant interaction of training by experience ($F_{2,60} = 7.25$; $P < 0.002$), such that the junior trainees showed greater improvement than the more senior trainees, with experience, accounting for 21% of the variance in improvement.

For the detailed checklist score, the two-way ANOVA also revealed a significant effect of training ($F_{1,60} = 34.82$; $P < 0.001$), accounting for 39% of the variance in scores, a significant effect of experience level ($F_{2,60} = 8.05$; $P = 0.001$), accounting for 23% of the variance in scores, and a significant interaction of training by experience ($F_{2,60} = 6.26$; $P = 0.004$), such that the junior trainees showed greater improvement than the more senior trainees, with experience, accounting for 19% of the variance in improvement. The improvement observed within each cohort of trainees, was statistically significant on the basis of the two-tailed paired *t*-test (Figure 4, Table 2).

The intraclass correlation coefficients for each rating scale were based on seven or eight observations for each of four observer pairings. The GRF was found to be reliable, with an average intraclass correlation coefficient of 0.86 (95% CI, 0.44–0.97; alpha = 0.92). The DCL was also found to be a reliable measure of performance. The average intraclass correlation coefficient for the detailed checklist was 0.93 (95% CI, 0.68–0.99; alpha = 0.96). Agreement between rating scales was based on pretest comparisons only and compared the average ratings of four raters. Each rater evaluated 15 candidates. The correlation coefficient between GRF and DCL was 0.81.

The one-way ANOVA of pretest scores by level of postgraduate training was based on evaluation of 34 subjects. The subjects evaluated included the original 30 subjects plus an additional four subjects who were practicing perinatologists.



	Pre-test score (%)	Post-test score (%)
DCL:PGY0–2	42	87
DCL:PGY3–5	66	88
DCL:Fellows	76	91
DCL:Faculty	98	98
GRF:PGY0–2	37	93
GRF:PGY3–5	67	95
GRF:Fellows	75	95
GRF:Faculty	87	87

Figure 4 Amniocentesis performance scores.

Table 2 Performance scores: *t*-test significance levels

Postgraduate level	*GRF*			*DCL*		
	d.f.	t	P	*d.f.*	t	P
Years 0–2	11	9.04	< 0.0001	11	7.27	< 0.00002
Years 3–5	14	5.48	< 0.0001	14	5.03	< 0.0002
Fellows	2	3.48	< 0.008	2	43.58	< 0.0005

GRF, global rating form; DCL, detailed checklist; d.f., degrees of freedom.

The ANOVA for the GRF by PGY level revealed a significant difference between cohorts ($F_{3,33} = 15.36$; $P < 0.001$), with PGY level accounting for 61% of the variance according to the eta[3] test. Post hoc analysis using the Tukey method revealed that at pretesting, only the novices and near experts groups differed significantly at baseline testing. The ANOVA for the DCL by PGY level was also significant ($F_{3,33} = 6.771$; $P = 0.001$), with PGY level accounting for 40% of the variance. The post hoc analysis revealed significant differences only between the PGY 0–2 group and the other groups. There were no significant differences noted among the remaining groups.

Feedback provided by the study participants indicated that the workshop was found to be helpful. When asked to rate the helpfulness of the workshop using a five-point Likert scale, participants rated the workshop 4.67. The median and mode ratings were 5. The simulator realism was assigned an average score of 4.13 of a possible 5; the median and modal scores were 4. Participant confidence increased following the amniocentesis workshop. The rating on the five-point Likert score for comfort level in performing a genetic amniocentesis given on the preworkshop questionnaire was 1.67. This comfort level increased to 3.7 with a median and modal value of 4 on the postworkshop questionnaire (Figure 5).

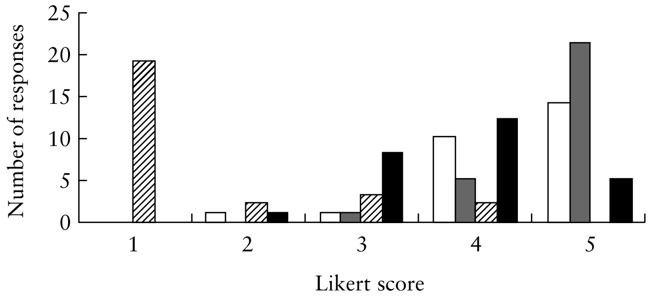


Figure 5 Participant feedback: questionnaire responses. □, simulator realism; ■, simulator helpfulness; ▨, comfort level preworkshop; ■, comfort level postworkshop.

DISCUSSION

This study was able to establish that the performance of a genetic amniocentesis can be effectively taught to trainees at various levels of training using a high fidelity simulator-based curriculum. The effectiveness of the training was evaluated using performance scales designed specifically for the evaluation of the performance of an amniocentesis. The psychometric attributes of these performance scales indicate that the performance scores recorded were reliable and valid measures. Feedback provided by participants indicated that the workshop was perceived as a useful educational activity.

We have shown that medical students and residents can be trained to almost the same level of skill as Fellows in maternal fetal medicine. The most novice participants demonstrated the greatest improvement which is consistent with the theory that simulator-based training can effectively shift the steep portion of the learning curve away from patients and into the teaching laboratory. It also clearly indicated that participants at all levels of training could benefit from such an educational activity.

While the gold standard score set by faculty members was high, it was less than 100%. This was primarily due to suboptimal communication. This finding was helpful in identifying an area in which teachers could improve their teaching. The positive feedback received regarding the incorporation of standardized patients into a simulation suggests that other similar areas of training may also benefit from combining simulators with standardized patients.

Simulators allow teaching to occur in a controlled safe environment where skills can be practiced repetitively. Procedures that are infrequently performed or are very difficult can be effectively simulated. Feedback can be provided unhampered by patient anxiety and medicolegal concerns. Stress is reduced when learning by simulation and this reduction in stress has been associated with a beneficial effect on subsequent performance[10]. The assessment of the skills obtained by this method of teaching is facilitated by standardization and the ability to allow trainees to make mistakes.

The performance scores obtained during evaluations on the amniocentesis simulator could be used for comparative purposes. The evaluation tools developed for this workshop can now be applied in the clinical environment. Follow-up evaluation of trainee performance in the clinical setting can be correlated with their performance in the simulated environment. Once a relationship between simulator scores and clinical scores is established the performance score could be used to determine when adequate training in the safe simulated environment has been achieved prior to allowing training to occur on patients.

The effectiveness of this simulator-based curriculum in amniocentesis provides a basis for developing simulations for more complex perinatal procedures. The use of an amniocentesis simulator can be rationalized as a safe means of practicing skills where there are limited opportunities for training. This rationale applies even more so to progressively advanced skills. More complex invasive perinatal procedures pose higher risks to patients. The opportunities to learn and practice these skills are even scarcer. The same curriculum used for the amniocentesis simulation can be expanded to teach advanced skills. The performance scores can be modified for each new procedure. Given the general nature of the GRF only minimal modifications would be required.

The transferability of the skills obtained during the simulator-based curriculum cannot be verified by this study. This could only be proven by correlating the performance scores obtained following the simulation-based curriculum with scores obtained by the same trainees in the clinical setting. There are several reasons why this type of study remains to be performed in any field of medicine. While evaluation in the clinical setting could be completed using the same performance scores as used in the simulations, the need to protect patients and the inability to standardize test circumstances constrain such evaluations. It may be feasible to test trainees in the clinical environment once they have achieved a simulator-based score that meets minimum criteria to ensure patient safety. The problem of standardization in the clinical setting remains a difficult challenge to be overcome.

While there are limitations to this study, it lays the groundwork for future studies. It has provided the necessary evidence to establish that objective observable improvements can be achieved through participation in a high-fidelity simulator-based curriculum. The performance measures developed for the study provide an established means of evaluating the outcomes of future studies. Future studies may be able to resolve important issues such as the transferability of skills. Serial evaluations using the amniocentesis performance scores could assess the retention rate for skills obtained during intensive workshops.

The results presented here are encouraging to anyone interested in developing similar simulator-based curricula. We have shown that it is possible to shift medical training from the clinic to the teaching laboratory. This shift provides a safe, standardized setting that fosters important learning concepts such as experiential learning, contextualization and immediate directive feedback. It encourages collaboration between peers and capitalizes on the expertise of faculty members as demonstrators and evaluators. It allows for reliable and valid measurement of skills and does so without subjecting trainees to the stress of learning on actual patients.

ACKNOWLEDGMENTS

This study was generously funded by a grant from the physicians of Ontario through the Physician Services Incorporated

Foundation, Toronto, Ontario, Canada. We would like to thank Mr Francis LeBouthillier for his contribution to the design and manufacturing of the simulator, Dr Glenn Regehr for his guidance and statistical expertise, and all of the nurses, standardized patients and participants who attended the amniocentesis workshop.

REFERENCES

1 Ludomirski A. Training in invasive fetal procedures. *Ultrasound Obstet Gynecol* 1995; 5: 150
2 Tabor A, Phillip J, Madsen M, Bang J, Obel EB. Randomized control trial of genetic amniocentesis in 4606 low-risk women. *Lancet* 1986; 1: 1287–93
3 Silver RK, Russell TL, Kambich MP, Leeth EA, MacGregor SN, Sholl JS. Midtrimester amniocentesis. Influence of operator caseload on sampling efficiency. *J Reprod Med* 1998; 43: 191–5
4 Timor-Tritsch IE, Yeh MN. In vitro training model for diagnostic and therapeutic fetal intravascular needle puncture. *Am J Obstet Gynecol* 1987; 157: 858–9
5 Angel JL, O'Brien WF, Michelson JA, Knuppel RA, Morales WJ. Instructional model for percutaneous fetal umbilical blood sampling. *Obstet Gynecol* 1989; 73: 669
6 Ville Y, Cooper M, Revel A, Frydman R, Nicolaides KH. Development of a training model for ultrasound-guided invasive procedures in fetal medicine. *Ultrasound Obstet Gynecol* 1995; 5: 180–3
7 Smith JF, Bergmann M, Gildersleeve R, Allen R. A simple model for learning stereotactic skills in ultrasound-guided amniocentesis. *Obstet Gynecol* 1998; 92: 303–5
8 Maher JE, Kleinman GE, Lile W, Tolaymat L, Steele D, Bernard J. The construction and utility of an amniocentesis trainer. *Am J Obstet Gynecol* 1998; 179: 1225–7
9 Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; 84: 273–8
10 Linn BS, Zeppa R. Stress in junior medical students: relationship to personality and performance. *Acad Med* 1984; 59: 7–12