Ultrasound Obstet Gynecol 2015; **46**: 312–318
Published online 6 August 2015 in Wiley Online Library (wileyonlinelibrary.com). **DOI:** 10.1002/uog.14780.

Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial

M. G. TOLSGAARD*†, C. RINGSTED‡, E. DREISLER†, L. N. NØRGAARD§, J. H. PETERSEN¶, M. E. MADSEN*†, N. L. C. FREIESLEBEN†**, J. L. SØRENSEN† and A. TABOR†

**Centre for Clinical Education, Rigshospitalet, Capital Region and University of Copenhagen, Copenhagen, Denmark; †Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet, Capital Region and University of Copenhagen, Copenhagen, Denmark; ‡Faculty of Health, Aarhus University, Aarhus, Denmark; §Department of Obstetrics and Gynecology, Nordsjælland Hospital Hillerød, University of Copenhagen, Hillerød, Denmark; ¶Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark; **Department of Obstetrics and Gynecology, Næstved Hospital, University of Copenhagen, Næstved, Denmark*

KEYWORDS: medical education; simulation-based medical education; simulation-based ultrasound training; transvaginal; ultrasound assessment; ultrasound competence

ABSTRACT

Objective *To study the effect of initial simulation-based transvaginal sonography (TVS) training compared with clinical training only, on the clinical performance of residents in obstetrics and gynecology (Ob-Gyn), assessed 2 months into their residency.*

Methods *In a randomized study, new Ob-Gyn residents (*n = *33) with no prior ultrasound experience were recruited from three teaching hospitals. Participants were allocated to either simulation-based training followed by clinical training (intervention group;* n = *18) or clinical training only (control group;* n = *15). The simulation-based training was performed using a virtual-reality TVS simulator until an expert performance level was attained, and was followed by training on a pelvic mannequin. After 2 months of clinical training, one TVS examination was recorded for assessment of each resident's clinical performance (*n = *26). Two ultrasound experts blinded to group allocation rated the scans using the Objective Structured Assessment of Ultrasound Skills (OSAUS) scale.*

Results *During the 2 months of clinical training, participants in the intervention and control groups completed an average ± SD of 58 ± 41 and 63 ± 47 scans, respectively (*P = *0.67). In the subsequent clinical performance test, the intervention group achieved higher OSAUS scores than did the control group (mean score, 59.1%* vs *37.6%, respectively;* P < *0.001). A greater proportion of the intervention group passed a pre-established pass/fail level than did controls (85.7%* vs *8.3%, respectively;* P < *0.001).*

Conclusion *Simulation-based ultrasound training leads to substantial improvement in clinical performance that is sustained after 2 months of clinical training.* © *2015 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of the International Society of Ultrasound in Obstetrics and Gynecology.*

INTRODUCTION

Ultrasonography is being used increasingly in the field of obstetrics and gynecology (Ob-Gyn). Although ultrasound imaging is traditionally considered safe, its use is highly operator-dependent[1]. The lack of sufficient operator skills can lead to diagnostic errors that may compromise patient safety due to unnecessary tests or interventions[2]. However, ultrasound training is associated with long learning curves and is therefore time-consuming and requires extensive teaching resources[3,4]. Consequently, some residents may never acquire the basic skills and knowledge needed for independent practice[5]. Simulation-based medical education (SBME) has been suggested as an adjunct to early ultrasonography training[5–11] but there is limited evidence of skill transfer from simulation to clinical performance[12,13]. Existing studies on SBME that involve technical or interventional procedures have focused predominantly on the initial effects of training[14–16] and only a few studies have documented the sustained effects on clinical performance[17,18]. Many resources are currently being allocated to SBME within multiple disciplines, but its effectiveness may be

Correspondence to: Dr M. G. Tolsgaard, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, DK-2100 Ø, Copenhagen, Denmark (e-mail: martintolsgaard@gmail.com)

Accepted: 3 January 2015

overestimated if only immediate outcomes are evaluated. For ultrasonography, it could be argued that the effects of SBME should extend beyond initial training to justify financial and time expenditure[19] as there is little harm associated with supervised clinical training alone. Hence, the aim of this study was to explore the sustained effects of simulation-based transvaginal sonography (TVS) training, measured after 2 months of clinical training. The aim of this study was to investigate, in a group of new Ob-Gyn residents, the effect of initial simulation-based TVS training followed by clinical training, compared with clinical training alone, on the quality of scans performed on patients at 2 months into their residency.

METHODS

The study was a multicenter, randomized observer-blind superiority trial conducted between 1 May 2013 and 4 April 2014 and reported following the CONSORT statement[20]. The study was carried out in the Ob-Gyn departments of three teaching hospitals in Eastern Denmark affiliated with the University of Copenhagen: Rigshospitalet, Nordsjaellands Hospital Hillerød and Næstved Hospital. Ethical approval was obtained from the Regional Ethical Committee of the Capital Region, Denmark (Protocol No. H-3-2012-154). The Danish Data Protection Agency approved the storage of relevant patient information (Protocol No. 2007-58-0015). The trial was reported to clinicaltrials.gov prior to the inclusion of participants (ClinicalTrials.gov Identifier NCT01895868).

The study cohort of participants comprised new residents in Ob-Gyn who were training at the three participating gynecological departments. The inclusion criterion was proficiency in written and oral Danish. The exclusion criteria were (1) prior employment at an Ob-Gyn department, (2) any formal ultrasound training with or without hands-on practice and (3) prior virtual-reality simulation experience. The participants were recruited by e-mail 2–4 weeks before the first day of their Ob-Gyn residency.

The primary investigator (M.G.T.) was responsible for the selection of participants. A research fellow (T.T.) at The Centre for Clinical Education, Rigshospitalet, randomized participants, stratified by hospital, to either simulation-based TVS training and subsequent clinical training (intervention) or clinical training alone (control). The randomization was performed by computer using an allocation ratio of 1 : 1.

Participants of the intervention group were trained on two different types of simulators in the initial phase of their residency. The first was a virtual-reality simulator (Scantrainer, Medaphor^{TM}, Cardiff, UK) designed for TVS. This system consists of a monitor and a transvaginal probe docked into a haptic device that provides realistic force-feedback when moving the probe. The monitor provides B-mode ultrasound images obtained from real patients and three-dimensional (3D) animated illustrations of the anatomical scan position of the probe. The system includes various training modules ranging from basic to advanced gynecological and early pregnancy modules. After completing a module, the simulator provides automated feedback using dichotomous metrics in a number of task-specific areas (e.g. scanning through the entire uterus), as well as general performance aspects (e.g. sufficiently optimizing the image). The participants were provided with a 30-min introduction to the simulated environment and equipment, during which a systematic examination of a normal female pelvis was demonstrated. The participants underwent training alone but were able to request verbal feedback on the metrics that indicated a fail. The verbal feedback was provided by one of two simulator instructors (M.G.T. or M.E.M.) and was limited to 10 min of feedback after a participant had completed all training modules. Instructors were present at the simulation center during all training sessions in case participants needed technical assistance. The participants were required to train on seven selected modules until they passed a predefined expert level of performance corresponding to 88.4% of the maximum total score[12]. All virtual-reality simulator training was dispersed in sessions of maximum 2-h duration and took place in the Juliane Marie Centre, University of Copenhagen.

Once the participants attained the expert level of performance on the virtual-reality simulator, their training was continued using a pelvic mannequin designed for TVS (BluePhantom, CAE Healthcare, Sarasota, FL, USA). This mannequin allowed participants to practice handling the ultrasound equipment and using available functions (i.e. knobology training) with their local ultrasound equipment. The mannequin training took place in the Ob-Gyn department at which the participants undertook their residency and was continued until proficiency was reached. Proficiency on the mannequin was determined using pass/fail levels on the Objective Structured Assessment of Ultrasound Skills (OSAUS) scale[21–23]. Training was discontinued if the participant had not completed both types of simulation-based training within the first 4 weeks of their residency. None of the intervention group participants was informed about their test scores during training.

Participants in both groups underwent clinical training but this was the only type of training provided to the control group. During the first week of their residency, all participants received a 1-h introductory lecture at one of the three teaching hospitals on basic ultrasound physics, knobology, female pelvic anatomy and the stages included in the systematic examination. Clinical training comprised a traditional apprenticeship model of learning under supervision. The protocol for all residents was to call for assistance whenever needed during the ultrasound examinations. In cases that requested supervision, a clinical supervisor would oversee the trainee's performance or repeat the scan. There was no specified minimum number of supervised scans required before independent practice was allowed at any of the three participating hospitals. However, certain diagnoses such as suspected fetal demise or ectopic pregnancy always

required a second opinion from a senior supervisor, according to national guidelines.

The primary outcome of the study was the clinical performance of the participants on real patients, assessed 2 months into their residency. For each resident, one independently performed TVS examination was recorded using a hard-disk recorder (MediCapture-200^{TM}, Philadelphia, PA, USA). Eligible patients for the clinical performance test were emergency patients who were referred to a gynecological department for a TVS examination. The recordings were made while on call (from 16:00 to 08:00 h) between 7 and 11 weeks from the first day of the participant's residency. The first eligible patient to consent was selected for the assessment. The ultrasound recordings were matched with a copy of patient record transcripts made by the participants. The identity of the participants was masked on the ultrasound videos and the corresponding patient record transcripts for subsequent assessment by two blinded raters. The raters were consultant gynecologists who were experts in TVS. Performance assessments were made using the OSAUS scale[21–23]. The number of completed ultrasound scans at the time of assessment and the proportion that had been supervised by a senior gynecologist were recorded for all participants, to account for any differences in clinical training between groups at the time of assessment. For participants who completed simulation-based training, the time used on the simulator, simulator scores for each attempt on the simulator test, and number of attempted modules was recorded.

The OSAUS scale was used to rate ultrasound competence; the scale consists of six items pertaining to equipment knowledge, image optimization, systematic examination, image interpretation, documentation of findings and medical decision-making (Appendix). The original OSAUS scale also contains the optional item 'indication for the examination', which was not included in the performance assessment in the present study. The OSAUS items were rated for each participant based on video performance and patient records. The patient records were used to assess the interpretation and documentation of the TVS examination as well as the medical decision made following the scan. Each OSAUS item was rated on a 5-point Likert scale (scale of 1–5, where 1 represents poor performance and 5 represents excellent performance). The OSAUS scale has demonstrated content validity[21], construct validity[23], high inter-rater reliability and internal consistency[22,23], as well as evidence of structural validity[5]. Credible pass/fail standards were established for the OSAUS scale in a previous study[23]. The raters completed comprehensive training in assessing prerecorded ultrasound performances until rating consensus was reached, which occurred after assessing two videos. The raters were instructed to assess performance according to that expected from a recently certified consultant gynecologist.

The selection of the seven simulator modules and performance standards were based on results from a previous study[12], in which the validity and reliability of simulator metrics were determined. Only metrics that demonstrated construct validity (that is, they differed significantly between novice and expert performances) were included in the analysis of simulated performances in the present study.

Sample size calculations were based on data from previous studies on clinical performance of ultrasound novices, with and without simulation-based ultrasound training[23,24]. From these studies, the expected difference in OSAUS scores between groups was 17.0% (pooled SD 9.0%). Assuming a dilution of initial training effects of 40% after 2 months of clinical practice[25], an alpha-level of 0.05 and a power of 0.80, a total of 26 participants were needed in the study[26]. Participants were recruited consecutively until the required number had completed the performance test.

Statistical analysis

Data were analyzed by the primary investigator (M.G.T.) and the trial statistician (J.H.P.) using SPSS 20 (IBM Corp., Chicago, IL, USA). All scores were calculated as percentages of maximum score and OSAUS scores were calculated as mean scores. A two-way ANOVA was performed with hospital and group (intervention *vs* control) as independent variables and OSAUS scores as a dependent variable. Assumptions of the model (homogeneity of variance and normally distributed residuals) were assessed for OSAUS scores. The proportion of residents that achieved an OSAUS score above a pre-established pass/fail level of 50.0%[23] was calculated and compared between the two groups using logistic regression, adjusting for effect of the different hospitals and interaction between hospital and group. Scores of the six individual OSAUS items were compared between groups using Mann–Whitney *U*-tests. Internal consistency for the OSAUS items was calculated using Cronbach's alpha; inter-rater reliability for the pre- and post-test assessments was calculated using intraclass correlation coefficients (ICC).

The simulator scores were calculated as the sum of metrics with established validity evidence. Simulator scores on the first and final attempt, time spent on the simulator, and number of attempted modules were correlated to OSAUS scores for the intervention group using multiple linear regression. Finally, differences in baseline characteristics between groups were assessed using independent-samples *t*-test, Mann–Whitney *U*-test, and chi-square test when appropriate. Two-sided significance levels of $P < 0.05$ were used for all analyses.

RESULTS

Participant enrollment, randomization and follow-up are illustrated in Figure 1. Participant baseline and follow-up characteristics are shown in Table 1. The mean length of time that participants in the intervention group took to attain the expert performance level on the virtual-reality simulator was 3 h and 16 min (95% CI, 2 h 56 m to 3 h 36 m) and the mean number of attempted modules

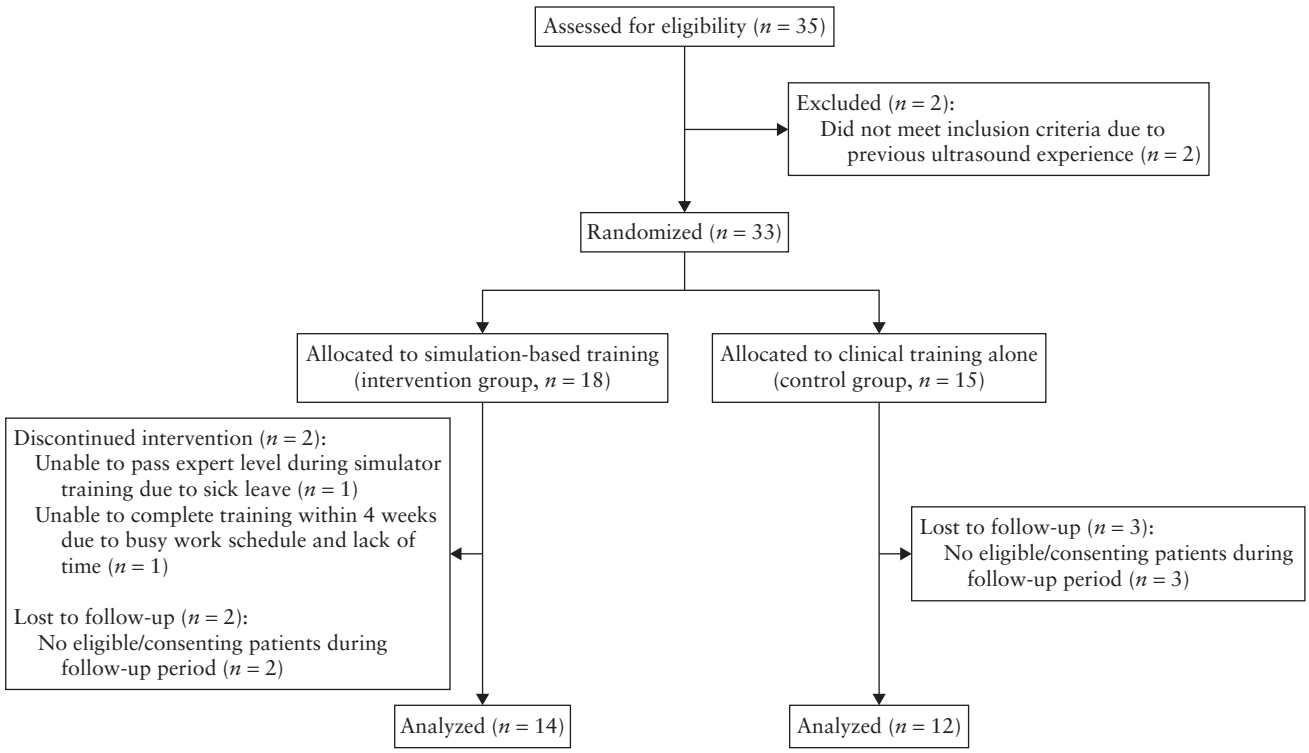

Figure 1 Flowchart of the study showing participant enrollment, randomization, allocation of interventions and follow-up.

Table 1 Baseline and follow-up characteristics of participants who completed simulation-based ultrasound (US) training followed by clinical training and those who underwent clinical training only

Characteristic	*Simulation-based US training* (n = *14*)	*Clinical training only* (n = *12*)	P
Gender (*n* (%))			1.00
Male	4 (28.6)	3 (25.0)	
Female	10 (71.4)	9 (75.0)	
Mean age (years)	34.1	33.5	0.71
Independently performed US scans (mean ± SD)	57.6 ± 40.5	62.5 ± 46.9	0.67
Supervised US scans (mean ± SD (%*))	43.9 ± 38.1 (76.2)	45.0 ± 38.1 (72.0)	1.00
Allocation of participants (*n* (%))			0.23
Copenhagen University Hospital Righospitalet (*n* = 8)	5 (62.5)	3 (37.5)	
Nordsjaellands University Hospital Hillerød (*n* = 5)	4 (80.0)	1 (20.0)	
Næstved University Hospital (*n* = 13)	5 (38.5)	8 (61.5)	
US diagnoses in performance test (*n*)			0.94
Normal pelvic US with or without intrauterine pregnancy	8	6	
PUL or ectopic pregnancy	3	3	
Complete/incomplete spontaneous miscarriage, missed miscarriage or blighted ovum	3	3	

*Percentage of total number of scans completed. PUL, pregnancy of unknown location.

was 30.3 (95% CI, 27.6–32.9). Learning curves of the first four rounds of training on the virtual-reality transvaginal simulator for the intervention group are shown in Figure 2. Two participants required more than four rounds of training to attain the expert level.

At the time of the clinical performance test, participants in the intervention and control groups had undergone an average of 60.4 (95% CI, 55.3–65.7) days and 62.9 (95% CI, 56.6–69.3) days of clinical training, respectively ($P = 0.46$). There were no differences observed in the reported number of completed scans (mean, 57.6 *vs* 62.5; $P = 0.67$) or supervised scans (mean, 43.9 *vs* 45.0; $P = 1.00$) performed by the intervention and control groups, respectively.

Ultrasound examinations for the clinical performance test were recorded for a total of 26 participants, thereby reaching the estimated sample size. Assumptions for the two-way ANOVA model were fulfilled (normally distributed residuals and homogeneity of variance; Levene's test, $P = 0.77$). OSAUS scores of the clinical performance test were significantly higher in the intervention than in the control group (mean, 59.1 ± 9.3% *vs* 37.6 ± 11.8%; $P < 0.001$). The adjusted absolute difference in OSAUS scores between the two groups was 20.1 (95% CI, 11.1–29.1) percentage points. There was no main effect of hospital allocation ($P = 0.34$) or interaction between hospital and group allocation ($P = 0.84$) on clinical performance. A significantly higher number of participants

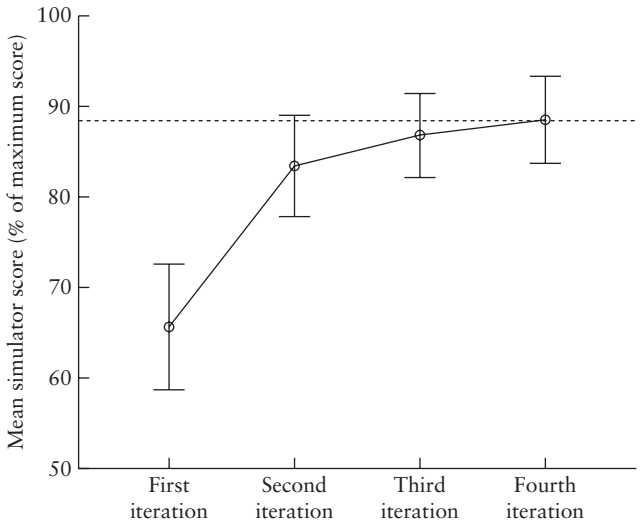

Figure 2 Learning curve of participants in first four training rounds on virtual-reality transvaginal simulator. Two participants required more than four rounds of training to attain expert level (dotted line). Error bars indicate ± 2 standard errors.

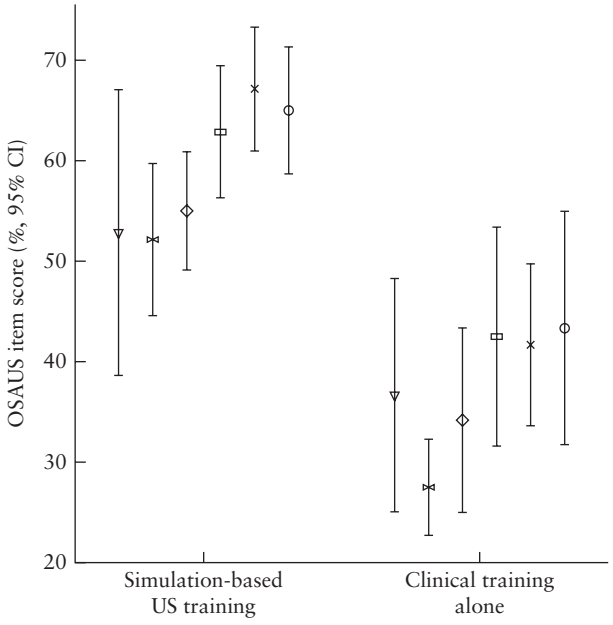

Figure 3 Objective Structured Assessment of Ultrasound Skills (OSAUS) scores of participants who underwent simulation-based ultrasound (US) training followed by clinical training and those who underwent clinical training only, measured after 2 months into residency. ▽, Knowledge of equipment; ⋈, image optimization; ◇, systematic examination; ▭, interpretation of images; ×, documentation of images; ○, medical decision-making.

from the intervention group passed a pre-established pass/fail level of 50.0% in OSAUS score compared with the control group (85.7% *vs* 8.3%, respectively; $P < 0.001$). Only 25.0% of the control group attained scores greater than the lowest performing participant in the intervention group. There were statistically significant differences between the scores of the two groups (Figure 3) on image optimization ($P < 0.001$), systematic examination ($P = 0.001$), interpretation of images ($P < 0.001$), documentation of examination ($P < 0.001$), and medical decision-making ($P = 0.005$) but no difference was observed for knowledge of equipment ($P = 0.095$).

The performance of the intervention-group participants during the simulation-based training did not predict their subsequent clinical performance, as there were low correlations between OSAUS scores and simulator metrics: number of attempted simulator modules ($P = 0.58$), first-attempt simulator scores ($P = 0.43$), final-attempt simulator scores ($P = 0.38$), time spent on the simulator to achieve expert level ($P = 0.09$). The internal consistency of the OSAUS items was high (Cronbach's alpha, 0.91) and the inter-rater reliability was acceptable (ICC, 0.63).

DISCUSSION

Although the efficacy of technical skills training using simulation has been well-documented[27,28], there has been limited evidence of the effectiveness of diagnostic simulation, in terms of transfer to clinical settings[5,13,29]. Our study adds to this evidence by demonstrating that, compared with clinical training only, simulation-based ultrasound training during the initial part of residency followed by clinical training of new residents in Ob-Gyn had a sustained impact on clinical performance on patients, measured at 2 months into the residency. The absolute difference in clinical performance between our intervention and control groups was large and only a small fraction of the control group were able to pass a pre-established pass/fail level as compared with the majority of the intervention group.

Previous studies in other areas of medicine have consistently shown large immediate effects of simulation-based training when compared with no training[30]. However, these studies carry the risk of overestimating the clinical importance of simulation-based training when evaluating only the immediate effects. To assess the dilution of training effects over time, we chose to evaluate participants' performance 2 months into their residency. The concept of the intervention in our study was 'proficiency-based training' in accordance with current recommendations[27]. This included continuous performance assessment until a certain competence level was attained, and the effect of the intervention can therefore be attributed to a combination of training and testing.

Existing literature has identified three major components of ultrasound competence – technical aspects of performance, image perception and interpretation – as well as medical decision-making[5,31–33]. Of these, the simulation-based training in our study primarily involved technical aspects of performance, as there is evidence that even advanced residents lack basic technical management skills and image optimization skills[23]. However, our results demonstrate that large effects were observed not only for participants' technical skills but also in other areas of performance: image interpretation, documentation and medical decision-making. It is conceivable that mastering the basic technical aspects reduced cognitive load[34] during clinical training. This may have enabled participants of the intervention group to allocate cognitive

resources more effectively to higher-order tasks such as image interpretation and medical decision-making. In other words, providing residents with systematic basic hands-on training may be beneficial to subsequent clinical training. Thus, the effective component in our study may be that residents were trained systematically in a safe environment, which allowed them to commit errors and to practice until proficient[35,36].

Despite having completed an average of 60 ultrasound examinations, of which more than 70% were supervised, only a small proportion of the control group passed the clinical performance test. Consequently, 2 months of clinical training in itself was insufficient to ensure competence at a predefined basic level, which is consistent with previous findings[5,23]. This raises concerns regarding patient safety and the efficiency of the apprenticeship model for clinical training. Interestingly, participants in both groups reported the same amount of supervision despite substantial performance differences after 2 months of training. This suggests that competence in itself was not a strong predictor for supervision but other factors probably influenced the amount of supervision provided in the context of this study. Although we did not investigate details of the reasons for requesting supervision and the content of the feedback provided, these may have differed between groups as a result of being at different levels in their learning curves. However, according to recent studies, external factors rather than individual training needs may also determine the level of supervised practice[5].

Although participants of the intervention group varied in simulator scores and amount of time they required to achieve an expert performance level on the simulator, there were no significant correlations between performance measures in the simulated setting and the clinical setting. The low predictive validity of simulator metrics may indicate that the sample size was inadequate to establish a correlation between performance in a simulated and clinical setting due to dilution of differences in individual performance after 2 months of clinical training[37]. However, the lack of correlation between performance measures used in the simulated and clinical settings may also reflect the limited predictive value of in-training assessment for subsequent clinical performances[38].

Strengths of this study include the use of a randomized single-blind design involving several institutions, well-defined intervention and control circumstances, outcome measures with established validity evidence, and the use of a clinical performance test on real patients. This study is the first to examine skills transfer after simulation-based ultrasound training[12,13,28] and is among the few studies that have examined the sustained effects of simulation on clinical performance[16–18].

We acknowledge some limitations to this study, including the degree of variance in the patients used for assessment. However, only a limited number of diagnoses were included in the assessment and there was no difference in the distribution of case presentations between the two groups. In the present study, a virtual-reality simulator and physical mannequin were used for training the intervention group. Although the effects of training cannot be attributed to either one of these types of simulator, the aim of this study was to examine the efficacy of simulation as a training method and not to explore the relative effectiveness of different simulators. We chose to focus on TVS, as the intimate nature of this examination makes it particularly suitable for simulation-based training. However, we cannot rule out that the type and intimacy of the TVS examination affects the amount and quality of the supervision provided during clinical training, and therefore the generalizability of the results to other types of examinations, such as abdominal ultrasound, requires further study. Finally, the quality of clinical training may differ between institutions and countries with regard to the level of supervised practice and amount of feedback provided, which may affect the value of adding simulation-based ultrasound training.

In conclusion, despite the performance improvements demonstrated in the present study, the effects on diagnostic error, patient satisfaction, need for re-examination and supervision from a senior colleague are among the factors that need to be explored in future studies involving ultrasound simulation. Furthermore, the monetary costs and time expenditure associated with simulation-based training, as well as its long-term effects, should be explored to assess how simulation-based practice compares with other training strategies[19].

ACKNOWLEDGMENTS

This study was funded by a grant from the Tryg Foundation. We would like to thank Dr Tobias Todsen, Righospitalet, Denmark, for helping to generate the randomization key for this trial.

REFERENCES

1. European Federation of Societies for Ultrasound in Medicine. Minimum training requirements for the practice of Medical Ultrasound in Europe. *Ultraschall Med* 2010; **31**: 426–427.
2. Moore CL, Copel JA. Point-of-care ultrasonography. *N Engl J Med* 2011; **24**: 749–757.
3. Jang TB, Ruggeri W, Dyne P, Kaji AH. The Learning curve of resident physicians using emergency ultrasonography for cholelithiasis and cholecystitis. *Acad Emerg Med* 2010; **17**: 1247–1252.
4. Jang TB, Jack Casey R, Dyne P, Kaji A. The learning curve of resident physicians using emergency ultrasonography for obstructive uropathy. *Acad Emerg Med* 2010; **17**: 1024–1027.
5. Tolsgaard MG, Rasmussen MB, Tappert C, Sundler M, Sorensen JL, Ottesen B, Ringsted C, Tabor A. Which factors are associated with trainees' confidence in performing obstetric and gynecological ultrasound examinations? *Ultrasound Obstet Gynecol* 2014; **43**: 444–451.
6. Burden C, Preshaw J, White P, Draycott TJ, Grant S, Fox R. Usability of virtual-reality simulation training in obstetric ultrasonography: a prospective cohort study. *Ultrasound Obstet Gynecol* 2013; **42**: 213–217.
7. Salvesen KÅ, Lees C, Tutschek B. Basic European ultrasound training in obstetrics and gynecology: where are we and where do we go from here? *Ultrasound Obstet Gynecol* 2010; **36**: 525–529.
8. Tutschek B, Tercanli S, Chantraine F. Teaching and learning normal gynecological ultrasonography using simple virtual reality objects: a proposal for a standardized approach. *Ultrasound Obstet Gynecol* 2012; **39**: 595–596.
9. Tutschek B, Pilu G. Virtual reality ultrasound imaging of the normal and abnormal fetal central nervous system. *Ultrasound Obstet Gynecol* 2009; **34**: 259–267.
10. Heer IM, Middendorf K, Müller-Egloff S, Dugas M, Strauss A. Ultrasound training: the virtual patient. *Ultrasound Obstet Gynecol* 2004; **24**: 440–444.
11. Maul H, Scharf A, Baier P, Wüstemann M, Günter HH, Gebauer G, Sohn C. Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. *Ultrasound Obstet Gynecol* 2004; **24**: 581–585.

12. Madsen ME, Konge L, Nørgaard LN, Tabor A, Ringsted C, Klemmensen AK, Ottesen B, Tolsgaard MG. Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. *Ultrasound Obstet Gynecol* 2014; **44**: 693–699.
13. Blum T, Rieger A, Navab N, Friess H, Martignoni M. A review of computer-based simulators for ultrasound training. *Simul Healthc* 2013; **8**: 98–108.
14. Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. *Ann Surg* 2012; **255**: 30–37.
15. Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS. Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. *BMJ* 2009; **14**; 338: b1802.
16. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 2004; **91**: 146–150.
17. Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB. Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med* 2009; **37**: 2697–2701.
18. Barsuk JH, Cohen ER, McGaghie WC, Wayne DB. Long-term retention of central venous catheter insertion skills after simulation-based mastery learning. *Acad Med* 2010; **85**: 9–12.
19. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: the missing outcome in simulation-based medical education research: a systematic review. *Surgery* 2013; **153**: 160–176.
20. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. *Ann Intern Med* 2010; **152**: 726–732.
21. Tolsgaard MG, Todsen T, Sorensen JL, Ringsted C, Lorentzen T, Ottesen B, Tabor A. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi Consensus Survey. *PLoS One* 2013; **28**; 8: e57687.
22. Todsen T, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, Konge L, Jensen ML, Ringsted C. Reliable and valid assessment of point-of-care ultrasonography. *Ann Surg* 2015; **261**: 309–315.
23. Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sorensen JL, Ottesen B, Tabor A. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2014; **43**: 437–443.
24. Tolsgaard MG, Madsen ME, Ringsted C, Ringsted C, Oxlund B, Oldenburg A, Sorensen JL, Ottesen B, Tabor A. The effect of dyad versus individual simulation-based training on skills transfer. *Med Educ* 2015; **49**: 286–295.
25. Smith CC, Huang GC, Newman LR, Clardy PF, Feller-Kopman D, Cho M, Ennacheril T, Schwartzstein RM. Simulation training and its effect on long-term resident performance in central venous catheterization. *Simul Healthc* 2010; **5**: 146–151.
26. Whitley E, Ball J. Statistics review 4: sample size calculations. *Crit Care* 2002; **6**: 335–341.
27. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. *Med Educ* 2010; **44**: 50–63.
28. Teteris E, Fraser K, Wright B, McLaughlin K. Does training learners on simulators benefit real patients? *Adv Health Sci Educ* 2011; **17**: 137–144.
29. Nitsche JF, Brost BC. Obstetric ultrasound simulation. *Semin Perinatol* 2013; **37**: 199–204.
30. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA* 2011; **306**: 978–988.
31. van der Gijp A, van der Schaaf MF, van der Schaaf IC, Huige JCBM, Ravesloot CJ, van Schaik JPJ, Ten Cate T. Interpretation of radiological images: towards a framework of knowledge and skills. *Adv Health Sci Educ* 2014; **19**: 565–580.
32. Krupinski EA. Current perspectives in medical image perception. *Atten Percept Psychophys* 2010; **72**: 1205–1217.
33. Kundel HL, Nodine CF. A visual concept shapes image perception. *Radiology* 1983; **146**: 363–368.
34. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010; **44**: 85–93.
35. Moak JH1, Larese SR, Riordan JP, Sudhir A, Yan G. Training in transvaginal sonography using pelvic ultrasound simulators versus live models: a randomized controlled trial. *Acad Med* 2014; **89**: 1063–1068.
36. Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005; **27**: 10–28.
37. Cook DA, West CP. Perspective: Reconsidering the focus on 'outcomes research' in medical education: a cautionary note. *Acad Med* 2013; **88**: 162–167.
38. Mitchell C1, Bhat S, Herbert A, Baker P. Workplace-based assessments of junior doctors: do scores predict training difficulties? *Med Educ* 2011; **45**: 1190–1198.

APPENDIX

The Objective Structured Assessment of Ultrasound Skills (OSAUS) scale

	Likert scale		
Item	*1*	*3*	*5*
1. Indication for the examination*: if applicable. Reviewing patient history and knowing why the examination is indicated	Displays poor knowledge of the indication for the examination	Displays some knowledge of the indication for the examination	Displays ample knowledge of the indication for the examination
2. Applied knowledge of ultrasound equipment: familiarity with the equipment and its functions, i.e. selecting probe, using buttons and application of gel	Unable to operate equipment	Operates the equipment with some experience	Familiar with operating the equipment
3. Image optimization: consistently ensuring optimal image quality by adjusting gain, depth, focus, frequency, etc.	Fails to optimize images	Competent image optimization but not done consistently	Consistent optimization of images
4. Systematic examination: consistently displaying systematic approach to the examination and presentation of relevant structures according to guidelines	Unsystematic approach	Displays some systematic approach	Consistently displays systematic approach
5. Interpretation of images: recognition of image pattern and interpretation of findings	Unable to interpret any findings	Does not consistently interpret findings correctly	Consistently interprets findings correctly
6. Documentation of examination: image recording and focused verbal/written documentation	Does not document any images	Documents most relevant images	Consistently documents relevant images
7. Medical decision-making: if applicable. Ability to integrate scan results into the care of the patient and medical decision making	Unable to integrate findings into medical decision making	Able to integrate findings into a clinical context	Excellent integration of findings into medical decision making

Likert is a five-point scale with 1 representing very poor and 5 representing excellent. In the OSAUS rating scale, only three points have descriptive anchors. *Item 1 was not included in the assessment of performances because only cases for which an ultrasound examination was indicated were included.