

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

## Authors

Martin G. Tolsgaard, Charlotte Ringsted, Eva Dreisler, Lone N. Nørgaard, Jørgen H. Petersen, Mette E. Madsen, Nina L.C. Freiesleben, Jette L. Sørensen, Ann Tabor

## Abstract

### Affiliations

Martin G. Tolsgaard, MD, PhD, Assistant professor, Department of Obstetrics and Gynecology, Juliane Marie Centre and Centre for Clinical Education, Rigshospitalet, Capital Region and University of Copenhagen, Denmark.

Charlotte Ringsted, MD, MHPE, PhD, Professor and Director, Department of Anesthesia and The Wilson Centre, University of Toronto and University Health Network, Toronto, Canada.

Eva Dreisler, MD, PhD, Consultant, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet.

Lone N. Nørgaard, MD, Consultant, Department of Obstetrics and Gynecology, Nordsjælland Hospital Hillerød, University of Copenhagen, Denmark.

Jørgen H. Petersen, PhD, Associate Professor, Department of Biostatistics, University of Copenhagen, Denmark.

Mette E. Madsen, MD, Research Fellow, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet.

Nina L.C. Freiesleben, MD, PhD, Senior Registrar, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet, and Department of Obstetrics and Gynecology, Naestved Hospital, University of Copenhagen.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.14780

Jette L. Sørensen, MD, MEd, Head of Education, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet.

Ann Tabor, MD, Dr.Sci., Professor, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet.

**Corresponding author**

Martin G. Tolsgaard, MD, PhD, Department of Obstetrics and Gynecology, Juliane Marie Centre, Rigshospitalet. University of Copenhagen, Denmark. Blegdamsvej 9, DK-2100 OE, Copenhagen. Tel: +45 61303072. Fax: +45 35454437 E-mail: [martintolsgaard@gmail.com](mailto:martintolsgaard@gmail.com).

**Key-words**

Simulation-based medical education; simulation-based ultrasound training; ultrasound assessment; ultrasound competence; medical education.

## Abstract

### Objective

The objective was to study the effect of initial simulation-based transvaginal ultrasound training compared to only clinical training on the clinical performances of residents in Obstetrics and Gynecology (OB-GYN) measured at two months into the residency.

### Methods

In a randomized study, new residents in OB-GYN (N=33) without prior ultrasound experience were included from three teaching hospitals. Participants were allocated to simulation-based training and subsequent clinical training (n=18) or only clinical training (n=15). The simulation-based training was performed on a virtual-reality transvaginal ultrasound simulator until an expert performance level was attained followed by training on a pelvic mannequin. After two months of clinical training, one transvaginal ultrasound scan was recorded for assessment of participants' clinical performance. Two blinded ultrasound experts rated the scans using the *Objective Structured Assessment of Ultrasound Skills* (OSAUS) scale.

### Results

During the two months of clinical training, participants in the intervention and control group completed an average of 57 (SD 41) and 63 (SD 47) scans, respectively ( $p = 0.67$ ). On the subsequent clinical performance test the intervention group achieved higher OSAUS-scores than the control group (mean 59.1% vs. 37.6%;  $p < 0.001$ ). A greater proportion of intervention group participants (85.7%) passed a pre-established pass/fail level compared to the controls (8.3%),  $p < 0.001$ .

## Conclusion

Simulation-based ultrasound training leads to substantial improvements in clinical performances that are sustained after two months of clinical training.

ClinicalTrials.gov Identifier NCT01895868

## Introduction

Ultrasonography is increasingly used in Obstetrics and Gynecology (OB-GYN). Although ultrasound imaging is traditionally considered safe, its use is highly operator dependent.<sup>1</sup> The lack of sufficient operator skills can lead to diagnostic errors that eventually compromise patient safety due to unnecessary tests or interventions.<sup>2</sup> However, ultrasound training is associated with long learning curves and is therefore time consuming and requires extensive teacher resources.<sup>3,4</sup> Consequently, some residents may never acquire the basic skills and knowledge needed for independent practice.<sup>5</sup> Simulation-based medical education (SBME) has been suggested as an adjunct to early ultrasonography training<sup>5-11</sup> but there is limited evidence of skills transfer from simulation to clinical performances.<sup>12,13</sup> Existing studies on SBME that involve technical or interventional procedures have predominantly focused on the initial effects of training<sup>14-16</sup> and only few studies document sustained effects on clinical performance.<sup>17,18</sup> Large resources are currently being allocated to SBME within multiple disciplines but its effectiveness may be over-estimated if only immediate outcomes are evaluated. For ultrasonography, it could be argued that the effects of SBME should extend beyond initial training in order to justify financial and time expenditure<sup>19</sup> 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 ultrasound training measured after two months of clinical training. The research question was: In a group of new residents in OB-GYN, what is the effect of initial simulation-based transvaginal ultrasound training followed by clinical training compared to only clinical training on the quality of scans performed on patients measured at two months into the residency?

## Methods

### Study design and setting

The study was a multi-centre randomized observer-blind superiority trial conducted between May 1<sup>st</sup> 2013 and April 4<sup>th</sup> 2014 and reported according to the CONSORT statement.<sup>20</sup> The study was carried out in the departments of OB-GYN of three teaching hospitals in Eastern Denmark affiliated with the University of Copenhagen; Rigshospitalet, Nordsjaellands Hospital Hilleroed, and Naestved Hospital. Ethical approval was obtained from the Regional Ethical Committee of the Capital Region, Denmark, (Protocol No H-3-2012-154) and the Danish Data Protection Agency approved the storing of patient relevant information (Protocol No 2007-58-0015). The trial was reported to clinicaltrials.gov prior to inclusion of participants (ClinicalTrials.gov Identifier NCT01895868).

### Participants

Participants included new residents in OB-GYN at the three gynecological departments. Inclusion criterion was proficiency in written and oral Danish. The exclusion criteria were 1) prior employment at a department of OB-GYN, 2) any formal ultrasound training with or without hands-on practice or 3) prior virtual reality simulation experience. The participants were recruited by e-mail two to four weeks before their first day in the OB-GYN residency. The primary investigator (MGT) was responsible for inclusion of participants. A research fellow (TT) at The Center for Clinical Education, Rigshospitalet, randomized participants stratified by hospital to either simulation-based transvaginal ultrasound training and clinical training (intervention) or clinical training alone (control). The randomization was performed by computer using an allocation ratio of 1:1.

## Interventions

Intervention group participants were trained in the initial phase of the residency on two different types of simulators. The first was a virtual-reality simulator (Scantrainer, Medaphor™, UK) designed for transvaginal ultrasound. 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 pictures obtained from real patients as well as 3D animated illustration of the probe's anatomical scan position. 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 minutes introduction to the simulated environment and equipment where a systematic examination of the normal female pelvis was demonstrated. The participants trained 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 (MGT or MEM) and was limited to 10-minutes of feedback each time participants 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 pre-defined expert level of performance corresponding to 88.4% of maximum total score.<sup>12</sup> All virtual-reality simulator training was dispersed in sessions of maximum two hours duration and took place in the Juliane Marie Center, 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 transvaginal ultrasound (BluePhantom, USA). This mannequin allowed participants to train handling the ultrasound equipment and available functions (i.e. knobology training) using their local ultrasound equipment. The mannequin training continued until proficiency was reached and took place in the OB-GYN department where the participants undertook their residency. Proficiency on the

mannequin was determined using pass/fail-levels on the *Objective Structured Assessment of Ultrasound Skills* (OSAUS) scale.<sup>21-23</sup> The training was discontinued if the participants had not completed both types of simulation-based training within the first four weeks of their residency. None of the intervention group participants were 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 participants. During the first week of their residency, all participants received a one-hour introductory lecture locally at the three teaching hospitals on basic ultrasound physics, knobology, female pelvic anatomy, and steps included in the systematic examination. Clinical training was 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 where supervision was requested, a clinical supervisor would supervise 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 demanded a second-opinion from a senior supervisor according to national guidelines.

### **Outcomes**

The primary outcome was clinical performances on patients measured at two months into the residency. For each resident one independently performed transvaginal ultrasound scan was recorded using a hard-disc recorder (MediCapture-200™). Eligible patients included emergency patients, who were referred to a gynecological department for a transvaginal ultrasound examination. The recordings were made on-call (04:00PM to 08:00AM) between seven and 11 weeks from their first day of 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 assessments by two blinded raters. The raters were consultant gynecologists, who were experts in transvaginal ultrasound. Performance assessments were made using the OSAUS scale.<sup>21-23</sup> The number of completed scans at the time of assessment and the proportion of these 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 the assessments. For the participants, who completed simulation-based training, time used on the simulator, simulator scores for each attempt on the simulator test, and number of attempted modules was recorded.

### **Performance assessment**

The OSAUS scale is used to rate ultrasound competence and consists of six items pertaining to equipment knowledge, image optimization, systematic examination, image interpretation, documentation of findings, and medical decision making; Appendix 1. 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 based on video-performances and patient records. The patient records were used to assess the interpretation and documentation of the scan as well as the medical decision making following the scan. Each OSAUS item is rated on a five-point Likert scales (1=poor performance, 5=excellent performance). The OSAUS scale has demonstrated content validity,<sup>21</sup> construct validity,<sup>23</sup> high inter-rater reliability and internal consistency,<sup>22,23</sup> as well as evidence of structural validity.<sup>5</sup> Finally, credible pass/fail standards have been established for the OSAUS scale in a previous study.<sup>23</sup> The raters completed comprehensive training in assessing pre-recorded ultrasound performances until ratings consensus was reached, which occurred after two videos. The raters were instructed to assess performances according to what can be 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,<sup>12</sup> in which the validity and reliability of simulator metrics were

determined. Only metrics that demonstrated construct validity (that is, they significantly differed between novice and expert performances) were included in the analysis of simulated performances in the present study.

### **Sample size calculations**

Sample size calculations were based on data from previous studies on clinical performances of ultrasound novices with and without simulation-based ultrasound training.<sup>23,24</sup> 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 two months of clinical practice,<sup>25</sup> an alpha-level of 0.05 and a power of 0.80, the total number of participants needed was 26.<sup>26</sup> Participants were recruited consecutively until the required number of participants had completed the performance test.

### **Statistics**

Data was analyzed using SPSS 20 (Chicago, IL) by the primary investigator (MGT) and the trial statistician (JHP). All scores were calculated into percentages of maximum score and OSAUS-scores were calculated into mean scores. A two-way ANOVA was performed with hospital and group (intervention vs. control) as the independent variables and OSAUS-scores as the dependent variable. Assumptions of the model (homogeneity of variance and normally distributed residuals) were assessed for the OSAUS-scores. The proportion of residents that achieved an OSAUS-score above a pre-established pass/fail level of 50.0%<sup>23</sup> 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 on the individual six 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 Intra-class 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 as well as 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 test, and Chi-square test when appropriate. Two-sided significance levels of 0.05 were used for all analyses.

## Results

Participant enrolment, randomization, and follow-up are illustrated in Figure 1. Participant baseline and follow-up characteristics are shown in Table 1. The mean time for participants in the intervention group to attain the expert performance level on the virtual-reality simulator was 3 hours and 16 minutes (95% CI, 2h:56m to 3h:36m), and the mean number of attempted modules was 30.3 (95% CI, 27.6-32.9). Learning curves on the virtual-reality transvaginal simulator for intervention group participants' first four training rounds on the simulator are shown in Figure 2. Two participants needed 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 group had spent an average of 60.4 (95% CI, 55.3-65.7) and 62.9 (95% CI, 56.6-69.3) days of clinical training, respectively ( $p = 0.46$ ). There were no differences 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,  $p = 1.00$ ) for the intervention and control group, respectively.

Ultrasound examinations of 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,  $p = 0.77$ ). The clinical performance test OSAUS scores of the intervention group were significantly higher than the control group; mean 59.1% (SD 9.3) vs. 37.6% (SD 11.8),  $p <$

0.001). The adjusted absolute difference in OSAUS scores between the two groups was 20.1 percentage points; 95% CI, 11.1-29.1. There was no main effect of hospital allocation ( $p = 0.34$ ) or interaction between hospital and group allocation ( $p = 0.84$ ). A significantly higher number of participants from the intervention group (85.7%) passed a pre-established pass/fail level of 50.0% in OSAUS-scores compared to the control group (8.3%),  $p < 0.001$ . Only 25.0% of the control group participants attained scores above the worst performing participant in the intervention group. There were statistically significant differences between the two groups' scores 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 not on knowledge of equipment ( $p = 0.095$ ), Figure 3.

The performances of the intervention group participants during the simulation-based training did not predict their subsequent clinical performances, 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, Intraclass Correlation Coefficient = 0.63.

## Discussion

Whereas the efficacy of technical skills training using simulation has been well-documented,<sup>27,28</sup> there has until now been limited evidence of the effectiveness in terms of transfer of skills to clinical settings.<sup>5,13,29</sup> This study adds to this evidence by demonstrating that simulation-based ultrasound training during the initial part of residency followed by clinical training compared to only clinical training of new residents in OB-GYN had a sustained impact on clinical performance on patients measured at two 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 participants were able to pass a pre-established pass/fail level compared to the large majority of the intervention group participants.

Previous studies in other areas of medicine have consistently shown large immediate effects of simulation-based training when compared to no training.<sup>30</sup> However, these studies carry the risk of over-estimating the clinical importance of simulation-based training when only evaluating immediate effects. To assess dilution of training effects over time, we chose to evaluate participants' performances at two months into their residency. The concept of the intervention in our study was 'proficiency-based training' in accordance with current recommendations.<sup>27</sup> 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 including technical aspects of performance, image perception and interpretation as well as medical decision-making.<sup>5,31-33</sup> 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.<sup>23</sup> However, our results demonstrate that large effects were not only observed for participants' technical skills but also in other areas of performance including image interpretation, documentation, and medical decision-making. It is conceivable that mastering basic technical aspects reduced cognitive load<sup>34</sup> during clinical training. This may have enabled the intervention group participants 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 the fact that residents were trained systematically in a safe environment, which allowed them to commit errors and practice until proficiency.<sup>35, 36</sup>

Despite having completed an average 60 scans of which more than 70% were supervised, only a small proportion of the control group participants passed the clinical performance test.

Consequently, two months of clinical training in itself was insufficient to ensure competence at a pre-defined basic level, which is consistent with previous findings.<sup>5,23</sup> 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 two 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 this study's context. Although we did not investigate details on 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 of their learning curves. However, external factors rather than individual training needs may also determine the level of supervised practice according to recent studies.<sup>5</sup>

Although the intervention group participants varied in simulator scores and amount of time 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 simulated and clinical setting because of dilution of individual performance differences after two months of clinical training.<sup>37</sup> However, the lack of any 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.<sup>38</sup>

Strengths of this study include the use of a randomized single-blinded 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<sup>12,13,28</sup> and

is among few studies that have examined the sustained effects of simulation on clinical performance.<sup>16-18</sup>

We acknowledge some limitations to this study. One is the degree of variance in cases used for assessment. However, only a limited number of diagnoses were included 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 participants. Although the effects of training cannot be attributed to either one of these types of simulators, 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 transvaginal ultrasound as the intimate nature of this exam makes it particularly suitable for simulation-based training. However, we cannot rule out that the type and intimacy of the transvaginal ultrasound 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 studies. 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.

Although performance improvements have been 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 and time costs associated with simulation-based training as well as its long-term effects should be explored to assess how simulation-based practice compares to other training strategies.<sup>19</sup>

**Conflicts of Interest and Source of Funding**

This study was funded by a grant from the Tryg Foundation. No conflicts of interest were declared by any of the authors.

**Acknowledgements**

The authors would like to thank Dr. Tobias Todsen, Rigshospitalet, Denmark, for helping with generating the randomization key for this trial.

**Author contributions**

MGT was the principal investigator and was responsible for the integrity of the data. A statistician, JHP, was responsible for the statistical analyses. MGT, CR, and AT designed and planned the study and wrote the first draft of the manuscript. ED and LNN were responsible for assessment of all ultrasound performances. NLCF, LNN, and MGT were responsible for the data collection at the three study sites. MEM contributed during the simulation-based training of intervention group participants. All authors took part in data analysis, critically revised and approved the final manuscript. All authors listed have contributed sufficiently to the project to be included as authors, and all those who are qualified to be authors are listed in the author byline. To the best of our knowledge, no conflict of interest, financial or other, exists.

## 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**:1247–52.
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**:1024–7.
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**:444–51.
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**:213–7.
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**:525–9.
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**:595–6.

9. Tutschek B, Pilu G. Virtual reality ultrasound imaging of the normal and abnormal fetal central nervous system. *Ultrasound Obstet Gynecol* 2009;**34**:259-67.
10. Heer IM1, 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**:581-5.
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. DOI: 10.1002/uog.13400.
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**:30-7.
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:b1802-2.
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**:146-50.
17. Barsuk JH, McGaghie WC, Cohen ER, O'Leary KJ, Wayne DB. Simulation-based mastery

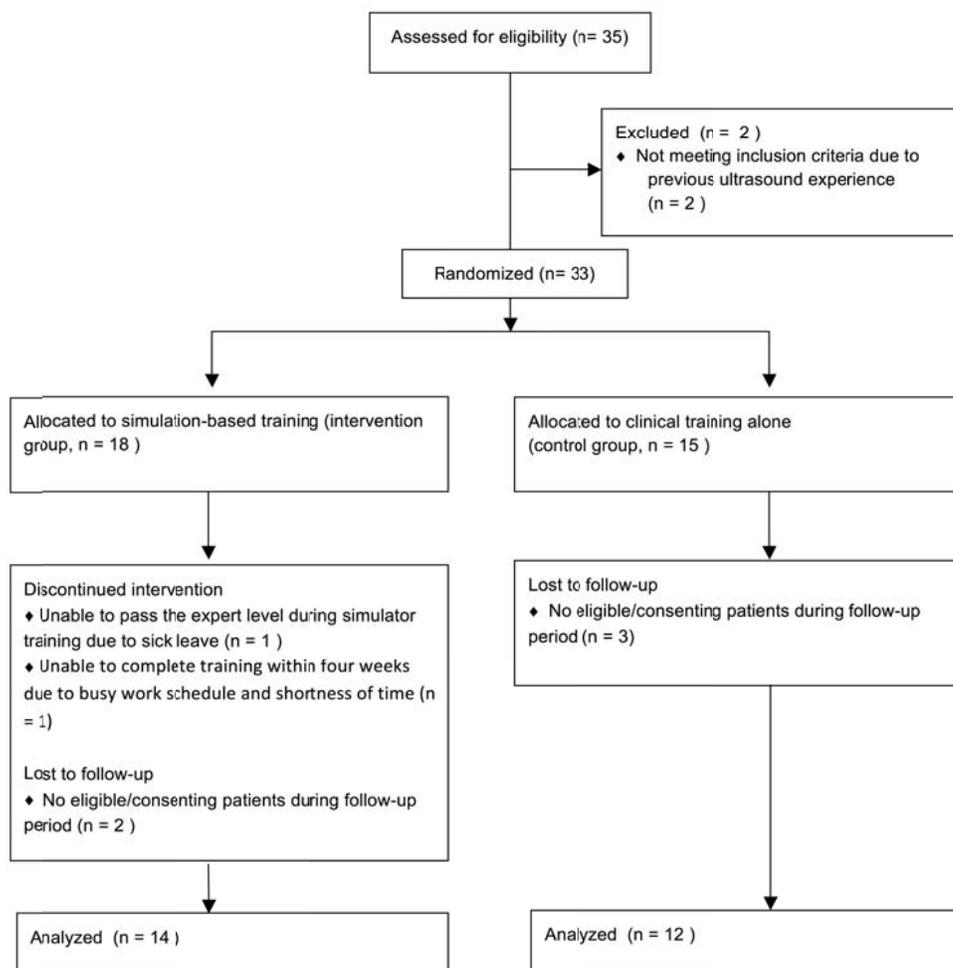
- learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med* 2009;**37**:2697–701.
18. Barsuk JH, Cohen ER, McGaghie WC, Wayne DB. Long-Term Retention of Central Venous Catheter Insertion Skills After Simulation-Based Mastery Learning. *Acad 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**:160–76.
  20. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials. *Ann Intern Med* 2010;**152**:726–32.
  21. Tolsgaard MG, Todsén T, Sørensen 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(2):e57687.
  22. Todsén 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*. 2014. DOI: 10.1097/SLA.0000000000000552.
  23. Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sørensen JL, Ottesen B, Tabor A.. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound Obstet Gynecol* 2014;**43**:437–43.
  24. Tolsgaard MG, Madsen ME, Ringsted C, Ringsted C, Oxlund B, Oldenburg A, Sørensen JL, Ottesen B, Tabor A. The effect of dyad versus individual simulation-based training on skills transfer. *Med Educ* 2014. In press.
  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**:146–51.
26. Whitley E, Ball J. Statistics review 4: sample size calculations. *Crit Care* 2002;**6**:335-41.
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 in Health Sci Educ* 2011;**17**:137–44.
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**:978-88.
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 in Health Sci Educ* 2014. DOI: 10.1007/s10459-013-9488-y.
32. Krupinski EA. Current perspectives in medical image perception. *Attention, Perception & Psychophysics* 2010;**72**:1205–17.
33. Kundel HL, Nodine CF. A visual concept shapes image perception. *Radiology* 1983;**146**:363–8.
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.

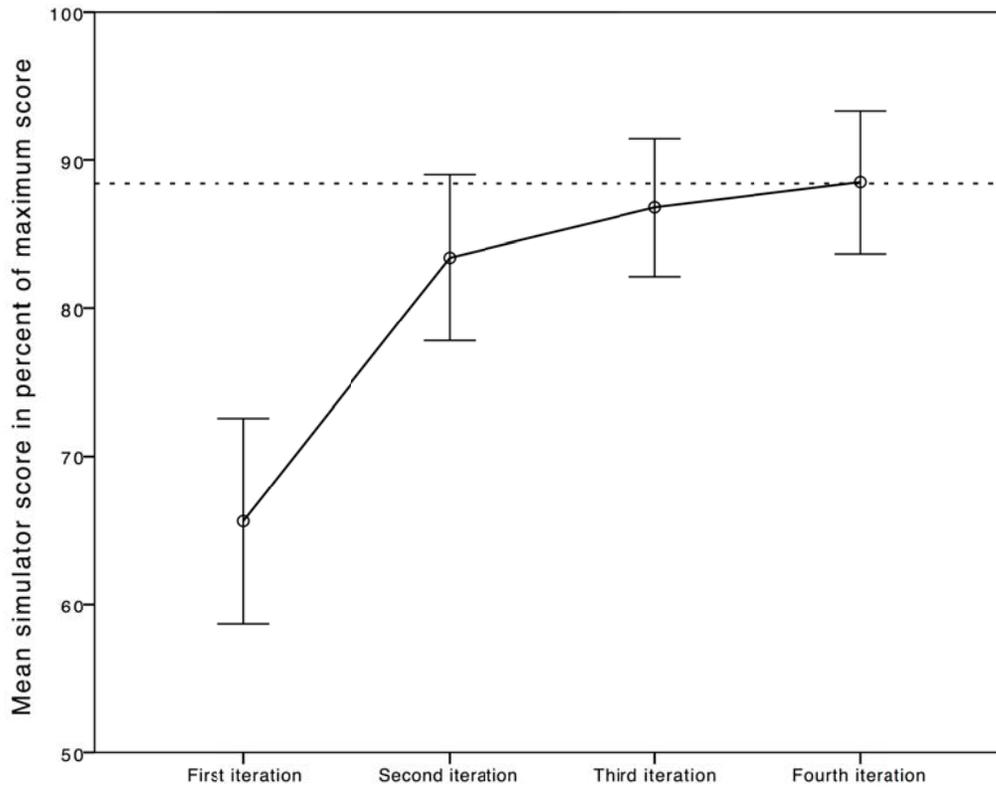
*Acad Med* 2014;**89**:1063-8.

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**:162-7.
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.

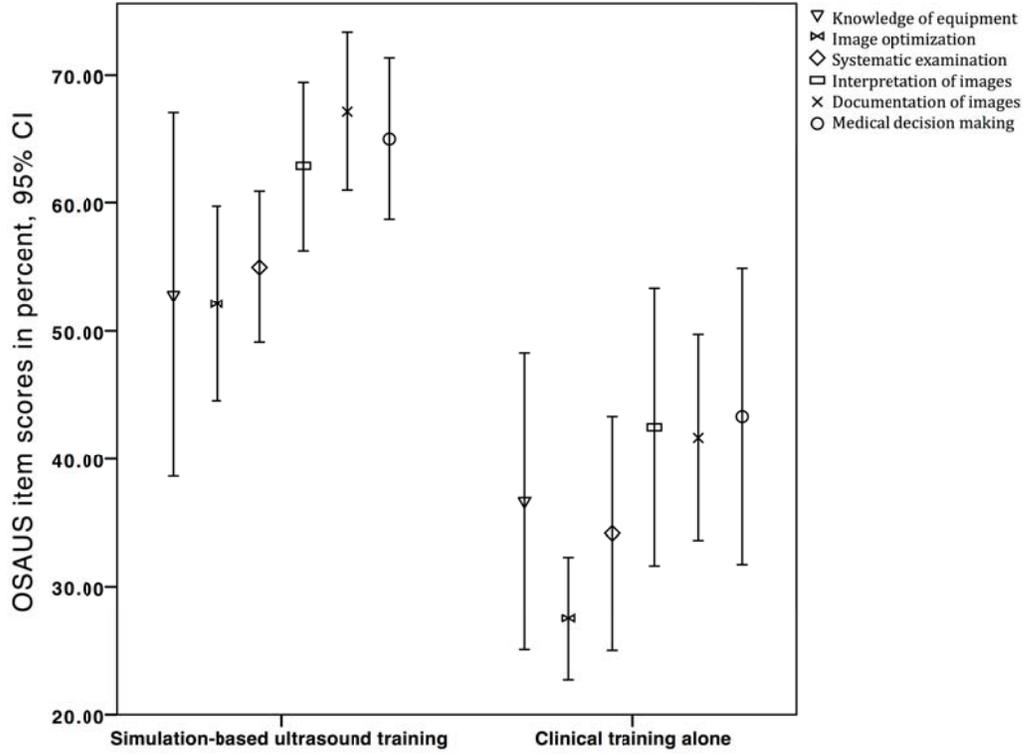
## Figure legends



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



**Figure 2.** Learning curves on the virtual-reality transvaginal simulator for the first four training rounds on the simulator. Two participants needed more than four rounds of training to attain the expert level, which is indicated by the dotted line. The error bars indicate  $\pm 2$  SE.



**Figure 3.** OSAUS-scores of participants that underwent simulation-based ultrasound training followed by clinical training (intervention) or only clinical training (control). Performances were recorded after two months of clinical training.

**Table 1.** Baseline and follow-up characteristics of participants, who completed simulation-based ultrasound training followed by clinical training, and of participants, who only underwent clinical training.

	Intervention group (simulation-based ultrasound training)	Control group (clinical training alone)	p- values
Gender, % (No.)			1.00
- male	28.6% (4)	25% (3)	
- female	71.4% (10)	75.0 (9)	
Age, mean	34.1 years	33.5 years	0.71
Number of independently performed scans (No.)	57.6 scans	62.5 scans	0.67
Proportion of scans that were supervised, % (No.)	76.2% (43.9 scans)	72.0% (45.0 scans)	1.00
Allocation of participants % (No. of participants)			0.23
- Copenhagen University Hospital Rigshospitalet	63% (5)	37% (3)	
- Nordsjaellands University Hospital Hillerød	80% (4)	20% (1)	
- Naestved University Hospital	38.5% (5)	61.5% (8)	
Performance test ultrasound diagnoses			0.94
- Normal pelvic ultrasound scan with or without intrauterine pregnancy	8	6	
- Pregnancy of unknown location (PUL) or ectopic pregnancy	3	3	
- Complete/incomplete spontaneous abortion, missed abortion or blighted ovum	3	3	