BMJ Open

Simulation based multi-professional obstetric anaesthesia training conducted in situ versus off site leads to similar individual and team outcomes: a randomised educational trial

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2015-008344
Article Type:	Research
Date Submitted by the Author:	28-Mar-2015
Complete List of Authors:	Sorensen, Jette Led; Rigshospitalet, University of Copenhagen, Juliane Marie Centre for Children, Women and Reproduction Van der Vleuten, Cees; Maastricht University, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences Rosthøj, Susanne; University of Copenhagen, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences Østergaard, Doris; Capital Region of Denmark and University of Copenhagen, Danish Institute for Medical Simulation, Herlev Hospital Leblanc, Vicki; University of Toronto, The Wilson Centre Johansen, Marianne; Rigshospitalet, University of Copenhagen, Department of Obstetrics, Juliane Marie Centre for Children, Women and Reproduction Ekelund, Kim; Rigshospitalet, University of Copenhagen, Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction Starkopf, Liis; University of Copenhagen, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences Lindschou, Jane; Rigshospitalet, University of Copenhagen, Copenhagen Trial Unit, Centre for Clinical Intervention Research Gluud, Christian; Rigshospitalet, Copenhagen University Hospital, The Copenhagen Trial Unit, Centre for Clinical Intervention Research Weikop, Pia; Rigshospitalet, University of Copenhagen, Laboratory of Neuropsychiatry Ottesen, Bent; Rigshospitalet, University of Copenhagen, Juliane Marie Centre for Children, Women and Reproduction
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Obstetrics and gynaecology, Anaesthesia, Emergency medicine
Keywords:	OBSTETRICS, MEDICAL EDUCATION & TRAINING, Anaesthesia in obstetrics < ANAESTHETICS, patient simulation, in situ simulation, interprofessional

TITLE PAGE

Title

Simulation based multi-professional obstetric anaesthesia training conducted in situ versus off site leads to similar individual and team outcomes: a randomised educational trial

Corresponding author

Jette Led Sørensen, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark.

jette.led.soerensen@regionh.dk

Telephone +45 35458901 Mobile phone +45 26711484

Authors name

Jette Led Sørensen, consultant obstetrician, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Cees Van der Vleuten, professor, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, 6200 Maastricht, Netherlands.

Susanne Rosthøj, PhD, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, 1014 Copenhagen, Denmark.

Doris Østergaard, professor, Danish Institute for Medical Simulation, Herlev Hospital, Capital Region of Denmark and University of Copenhagen, 2730 Herlev, Denmark.

Vicki LeBlanc, associate professor, The Wilson Centre, University of Toronto, 1 ES-565 Toronto, Canada.

Marianne Johansen, consultant obstetrician, Department of Obstetrics, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Kim Ekelund, consultant anaesthetist, Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Liis Starkopf, computing assistant, Section of Biostatistics, Institute of Public Health, Faculty of Health S and Medical Sciences, University of Copenhagen, 1014 Copenhagen, Denmark.

Jane Lindschou, clinical research consultant, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Christian Gluud, head of department, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.
Pia Weikop, senior scientist Laboratory of Neuropsychiatry, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.
Bent Ottesen, professor, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark

Keywords Patient simulation, in situ simulation, emergencies, interprofessional, stress

Word count abstract: 297

Word count main text (excluding title page, abstract, references, figures and tables): 4,065
Figures: 2
Tables: 8
Supplementary material (online): 1table

ABSTRACT

Objective

To investigate the effect of in situ simulation (ISS) versus off site simulation (OSS) on knowledge, patient safety attitude, stress, motivation, perceptions of simulation, team performance, and organisational impact.

Design

Investigator-initiated single-centre randomised superiority educational trial.

Setting

Obstetrics and anaesthesiology departments, Rigshospitalet, University of Copenhagen, Denmark.

Participants

One-hundred participants in teams of ten, comprising midwives, specialised midwives, auxiliary nurses, nurse anaesthetists, operating theatre nurses, and consultant doctors and trainees in obstetrics and anaesthesiology.

Interventions

Two multi-professional simulations (clinical management of an emergency caesarean section and a postpartum haemorrhage scenario) were conducted in teams of ten in the ISS versus the OSS settings.

Primary outcome

Knowledge assessed by a multiple choice question test.

Exploratory outcomes

Individual outcomes: scores on the Safety Attitudes Questionnaire, stress measurements (State-Trait Anxiety Inventory, cognitive appraisal, and salivary cortisol), Intrinsic Motivation Inventory, and perceptions of simulations. Team outcome: video assessment of team performance. Organisational impact: suggestions for organisational changes.

Results

The trial was conducted from April to June 2013. No differences between the two groups were found for the multiple choice question test, patient safety attitude, stress measurements, motivation, or the evaluation of the simulations. The participants in the ISS group scored the authenticity of the simulation significantly higher than the participants in the OSS group did. Expert video assessment of team performance showed no differences between the ISS versus the OSS groups. The ISS group provided more ideas and suggestions for changes at the organisational level.

Conclusion

In this randomised trial no significant differences were found regarding knowledge, patient safety attitude, or stress measurements when comparing ISS versus OSS. Although participant perception of the autheticity of ISS versus OSS differed significantly, there were no differences in other outcomes between the groups except that the ISS group generated more suggestions for organisational changes.

Trial registration

ClincialTrials.gov NCT01792674

ARTICLE SUMMARY

Strengths and limitation

- The study is the first randomised trial conducted to assess the effects of two different simulation settings, in situ simulation versus off site simulation, on a broad variety of outcomes.
- Previous non-randomised studies have recommended in situ simulation. But in this randomised trial no significant differences regarding knowledge, patient safety attitude, or stress measurements were found when comparing in situ simulation versus off site simulation. The participants in the in situ group scored the authenticity of the simulation significantly higher than the participants in the off site simulation group did; however, this perception did not influence any other individual and team outcomes. On the outcome on organisational level the in situ group generated more suggestions for organisational changes.
- A strength of this trial is the involvement of authentic teams that mirrored teams in real life, that resembles the real clinical setting in every possible way. This was important for the so-called sociological fidelity.
- A limitation of the study is the fact that the outcome was based only on immediate measurements of knowledge level and of team performance. Only perceptions of simulation were measured after one week (evaluation and motivation) and safety attitudes after one month. No clinical outcome was measured.

INTRODUCTION

Frequently recommended as a learning modality,[1-5] simulation-based medical education is described as "devices, trained persons, lifelike virtual environments, and contrived social situations that mimic problems, events, or conditions that arise in professional encounters".[5] Its key elements, however, remain to be studied in depth in order to improve simulation-based medical education. One potential aspect that may influence the effect of this kind of education is the level of fidelity, or authenticity in more layman's terms. Fidelity is traditionally described to be assessed on two levels: 1) engineering fidelity, i.e. does the simulation look realistic, 2) psychological fidelity i.e. does the simulator contain the critical elements to accurately simulate the behaviours required to complete a task.[6,7]

Simulation-based medical education has traditionally been conducted as off site simulation (OSS), either at a simulation centre or in facilities in the hospital set up for the purpose of simulation. Recently, in situ simulation (ISS) has been introduced and described as "a team-based simulation strategy that occurs on the actual patient care units involving actual healthcare team members within their own working environment".[8-12] A yet unanswered question is whether ISS is superior to OSS. It has been argued that ISS has more fidelity compared to OSS and that ISS can lead to better teaching and a greater organisational impact.[8-14]

We hypothesised that the physical setting could influence fidelity, and hence that ISS could be more effective for educational purposes. To our knowledge, no randomised trials have been conducted comparing the ISS versus the OSS setting. Two recent articles that do use randomisation and compare ISS to OSS focus on frequency of training and not setting,[15] nor did they include a relevant control group.[16] Previous studies have been criticised for having small sample sizes, weak study designs, and a lack of meaningful evaluations of the effectiveness of the programmes.[8]

Human factors such as stress and motivation impact learning,[17-25] which is why we set out to investigate how stress and motivation were affected by ISS versus OSS. We anticipated that the participants would experience ISS as more demanding and as creating higher levels of stress and motivation, which might enhance their learning. Furthermore, we hypothesised that ISS might

provide the investigators with more information on changes needed in the organisation to improve patient safety and quality of care.

In this trial, we wanted to apply simulation-based medical education in the field of obstetrics, as delivery wards are challenging work places, where patient safety is high on the agenda and rare, unexpected emergencies occur.[26-33] Simulation-based medical education is thus argued to be an essential learning strategy for labour wards.[4,34] The objective of this randomised educational trial was to investigate the effect of ISS versus OSS on knowledge, patient safety attitude, stress, motivation, perception of the simulation, team performance, and organisational impact among multi-professional obstetric anaesthesia teams.

METHODS

Design

An investigator-initiated, single-centre randomised superiority educational trial previously described in a design article.[35]

Setting and participants

The setting was the Department of Obstetrics and the Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, which has approximately 6,300 deliveries per year. Participants were healthcare professionals who worked in shifts on the labour ward: consultant and trainee doctors in obstetrics and anaesthesiology, midwives, specialised midwives, auxiliary nurses, nurse anaesthetists, and operating theatre nurses. Participants gave written informed consent. Exclusion criteria were lack of informed consent, employees with managerial and staff responsibilities, staff members involved in the design of the trial, and employees who did not work in shifts.[35]

Recruitment of participants

Eligible participants were provided with information via meetings, a website, and personal letters but additional verbal and written information could also be obtained from the principal investigator (JLS). Informed written consent was obtained if people decided to participate in the study.[35]

Interventions

The experimental intervention was a pre-announced ISS,[8,9] i.e. simulation- based medical education in the delivery room and operating theatre. The control intervention was an OSS, which took place in hospital rooms set up for the occasion but away from the patient care unit.

An appointed working committee consisting of representatives from all the healthcare professionals participating in the trial developed its aims and objectives, and they designed simulated scenarios for the ISS and the OSS. [35] The two simulation scenarios were: 1) management of an emergency caesarean section; and 2) a postpartum haemorrhage. Focusing mainly on inter-professional skills and communication, the scenarios gave each healthcare profession a significant role to play.[36]

In the first part of the simulation in the delivery room, someone who has been instructed in role playing acted as the patient in both the ISS and the OSS setting. In both the real and the simulated operating theatre a full-body birthing simulator, a SimMom, was used for parts of the simulation scenario.[37] Recruited from the working committee, the instructors conducting the simulations were trained in facilitating simulations and doing debriefings. The working committee were trained in local organised courses and attended a British National train the trainers course: PROMPT (PRactical Obstetric Multi-Professional Training).[38] They worked in groups of two comprising either a consultant obstetrician with a nurse anaesthetist or a consultant anaesthetist with a midwife. The debriefings lasted 50-60 minutes and comprised three phases: description, analysis, and application.[39] In addition to the simulation-based medical education, the training day also included video-based, case-based,[40] and lecture-based teaching sessions.

Primary outcome

The primary outcome was the results from a knowledge test based on a 40-item multiple choice question (MCQ) test developed specifically for this trial.[41] The participants completed the MCQ test at the beginning and at the end of the training day. They were asked not to discuss the MCQ test with other participants or instructors during the training day.

Exploratory outcomes

The Safety Attitudes Questionnaire (SAQ) is validated in a Danish context.[42] It included 33 items covering five dimensions: 1)team work climate; 2)safety climate; 3)job satisfaction; 4)stress recognition; and 5)work conditions.[43,44] The participants did the Safety Attitudes Questionnaire one month prior to and one month after participating in the training day.

Stress: Salivary cortisol levels were used as an objective measure of physiological stress. [35] The salivary cortisol samples were obtained as a baseline before the first and the second simulation and three additional times for the two simulations (figure 1). The subjective stress level was measured using the State-Trait Anxiety Inventory (STAI) and cognitive appraisal (CA) (figure 1).[20,22,45,46]

Intrinsic Motivation Inventory (IMI) included 22 items with four dimensions: 1)interest / enjoyment; 2)perceived competence; 3)perceived choice; and 4)pressure or tension (reversed scale).[47] The IMI was given to the participants at the end of the training day.

Evaluation questionnaire: Together with the IMI, each participant received an evaluation questionnaire at the end of the training day, and they were asked to be return it within a week.[35]

Team performance was video recorded and assessed by experts using a Team Emergency Assessment Measure (TEAM).[35,48,49] The TEAM scale was used in the original version in English and supplemented with a translated Danish version. The scoring of team performance was done by two consultant anaesthetists and two consultant obstetricians from outside the hospital. All four video assessors jointly attended two three-hour training sessions on video rating but assessment of the videos was done individually. Each video-assessor received an external hard disc with 20 simulated scenarios in random order of teams and scenarios of respectively management of an emergency caesarean section and a postpartum haemorrhage.

Organisational impact was registered using: 1)two open-ended questions included in the evaluation questionnaire on suggestions for organisational changes; and 2)debriefing and evaluation at the end of the training day, where participants reported ideas for organisational changes. The principal investigator (JLS) took notes during these sessions, where were then discussed in previously mentioned working committee, which included authors MJ and KE.

Sample size calculation

Unable to identify data on the effectiveness of ISS training to do sample size calculations, we chose data from knowledge tests from previous studies instead.[50,51] We assumed the distribution of the primary outcome (the percentage of correct MCQs answers) to be normally distributed with a standard deviation of 24%. If a difference in the percentage of correct MCQ answers between the two groups (ISS and OSS) was 17%, then 64 participants had to be studied to be able to reject the null hypothesis with a power of 80%. As the interventions were delivered in teams (clusters), observations from the same team were likely to be correlated.[52,53] The reduction in effective sample size depends on the cluster correlation coefficient, which is why the crude sample size had to be multiplied by a design effect. With a design effect of 0.05 the minimum sample size was increased to 92.8 participants.[53] We decided to include a total of 100 participants.

Randomisation and blinding

Randomisation was performed by the Copenhagen Trial Unit using a computer-generated allocation sequence concealed to the investigators. The randomisation was conducted in two steps. First, the participants were individually randomised 1:1 to the ISS or the OSS group. The allocation sequence consisted of nine strata, one for each healthcare professional group. Each stratum was composed of one or two permuted blocks with the size of ten. Second, the participants in each group were randomised into five teams for the ISS and OSS settings using simple randomisation that took into account the days they were available for training.

Questionnaire data were transferred from the paper versions and coded by independent data managers. The intervention was not blinded for the participants, facilitators providing the educational intervention, the video assessors, or the principal investigator drawing the conclusions. The data managers and statisticians were blinded to the allocated intervention groups.

Data analysis and statistical methods

Due to the low number of missing values, no missing data techniques were applied. Single missing items in the MCQ test or more than one answer to an MCQ item were treated as incorrect answers. Single missing items in inventories (Safety Attitudes Questionnaire, Intrinsic Motivation Inventory, Stress-Trait Anxiety Inventory) were excluded from the calculation of the mean summary scores.

Comparisons of means were performed by the Welch Two Sample t-test and comparisons of location were performed by the Kruskal-Wallis rank sum test. Calculation of 95% confidence intervals (CI) and comparison of means of outcomes obtained after the simulation intervention were based on generalised estimating equations,[54] which were used since observations from individuals on the same team were potentially correlated.

The team data, i.e. the ratings from the four assessors, were analysed using linear mixed models to take into account the repeated measurements on the teams by the same assessors. A random effect for each team nested in the randomisation group and in each assessor was included. A model including the interaction between randomisation group and simulation was used to estimate means, whereas an additive model was used to determine the overall difference in mean between the ISS and OSS intervention and the first (emergency caesarean section) and the second (postpartum haemorrhage) simulation (no interaction between randomisation and simulations was found).

For each outcome, assessments of means and mean differences in subgroup analyses were adjusted for multiple testing using the Benjamini-Hochberg method.[55] Furthermore, the comparisons of items in the evaluation questionnaire were adjusted for multiple testing.

Ideas for organisational changes were registered by participants and the reported suggestions were categorised as qualitative data and analysed using part of the framework Systems Engineering Initiative for Patient Safety model.[56]

Post hoc analyses were performed to evaluate whether the simulation intervention had an effect on knowledge, patient safety attitudes, stress measurements, and team performance. Outcomes obtained before and after the training day and outcomes obtained in the first simulation (emergency caesarean section) and in the second simulation (postpartum haemorrhage) were compared.

SAS version 9.2, R version 3.0.2, and IBM SPSS Statistics 20 were used for statistical analysis. Two-sided p-values <0.05 were considered significant.

RESULTS

Recruitment, basic characteristics, and follow-up on participants

Informed written consent for participation in the trial was provided by 116 healthcare professionals. See figure 2 and table 1 for the flow of participants. The two intervention groups were comparable (table 2).

Table 1. Reasons for lost to follow-up.

	Lost to follow-up (n=100) n (%)
Pre MCQ test	3 (3%)[1]
Post MCQ test	3 (3%)[1]
Salivary cortisol level at emergency caesarean section simulation	5 (5%)[1,2]
Salivary cortisol level at postpartum haemorrhage simulation	4 (4%)[1,3]
STAI at emergency caesarean section simulation	3 (3%)[1]
STAI at postpartum haemorrhage simulation	4 (4%)[1,3]
CA at caesarean section simulation	3 (3%)[1]
CA at postpartum haemorrhage simulation	4 (4%)[1,3]
Evaluation questionnaire	4 (4%)[1,4]
IMI	5 (5%)[1,5]
Pre-SAQ	5 (5%)[6,7]
Post-SAQ	9 (9%)[1,8]

CA: cognitive appraisal; IMI: Intrinsic Motivation Inventory; MCQ: multiple choice question; SAQ: Safety Attitudes Questionnaire; STAI: Stress-Trait Anxiety Inventory

[1)] Participants ill and did not participate (n=3).
[2)] Two measurements were clear outliers. A re-evaluation of the data collection indicated that the two samples had most likely been swapped between two participants, which is why these measurements were excluded from all analyses (n=2).
[3)] Because one participant was temporarily called away for clinical work, the cortisol measurement after the simulation in postpartum haemorrhage is lacking and he was unable to answer parts of the questionnaires (n=1).
[4)] Questionnaires not returned (n=1).
[5)] Questionnaires not returned (n=2).
[6)] For three participants pre-SAQ data were excluded because these participants were employed in other departments prior to participating in the training days, hence their responses did not refer to the department in question (n=3).
[7)] Of the individuals who did not participate due to illness (n=3), one filled out the pre SAQ anyhow.
[8)] Questionnaires not returned (n=6).

Table 2. Baseline characteristics of participants in ISS and OSS groups

	ISS group	OSS group
Number of participants	48[1]	49[2]
Number of females / males	42 / 6	43 / 6
Median age (range)	44.5 (26-63)	42 (27-65)
Median year of obstetric work experiences (range)	7 (0.6-38)	7 (0.6-39)
Previous simulation experiences:		
No experience	8	10
Simple simulation	25	24
Full-scale simulation	15	15
Pregnant participants	2	2
Participants on any kind of medication	19	20
Participants on medication with no expected influence on cortisol measurement[3]	12	9
Participants on medication with potential influence on cortisol measurement	7	11
Intranasal and inhaled corticosteroids (mometasone furoate, budesonide/formoterol, budesonide, fluticasone/salmeterol)	2	3
Levothyroxine	1	2
Metformin	1	1
Norethisterone/estradiolacetate	0	1
Oral contraceptives	1	3
Beta blockers (metoprolol)	0	1
Antidepressants (nortriptyline, fluoxetine)	2	0

ISS: in situ simulation to OSS: off site simulation.

[1] Not included due to illness: A consultant obstetrician and an operating room nurse (n=2).
[2] Not included due to illness: An auxiliary nurse (n=1).
[3] Intrauterine contraceptive devices, angiotensin II receptor antagonists, angiotensin-converting-enzyme inhibitor, simvastatin, alendronate, pantoprazole, antihistamine, tinzaparine.

Intervention delivery

The trial was conducted from April to June 2013. Out of 100 participants included, 97 participated (figure 2 and table 1). The ten simulations were conducted as planned, although one had to be postponed for 15 minutes due to an ongoing, real emergency caesarean section. The mean number of minutes spent on the caesarean section simulation in ISS and OSS was 18 and 15 minutes, respectively (P=0.70), while the mean for the postpartum haemorrhage simulation was 26 and 24 minutes, respectively (P=0.40).

Primary outcome

Multiple Choice Question (MCQ) test: There were no differences in the post-MCQ scores between the ISS versus the OSS group (table 3). Additional analyses based on the MCQ test, including 33 or 29 of the 40 items, gave similar results (data not shown). These additional analyses were performed

because validation of the MCQ test revealed that seven to eleven of 40 MCQ items were disputable.[41]

MCQ post hoc analysis: The increase in percentage of correct answers in the MCQ test following training was the same in the two groups, respectively 13.1% (95% CI, 11.0% to 15.3%) and 12.7% (95% CI, 10.3% to 15.2%). This increase was statistically significant (P<0.0001) from pre-training to post-training in both the ISS and the OSS groups (table 3).

Table 3. Means (95% CI) of percentages of correct answers in the MCQ test before (pre-MCQ) and after (post-MCQ) in the ISS and OSS groups. The primary analysis comprised a comparison of mean post-MCQ of ISS versus OSS group

MCQ test % correct	Simulation intervention	Pre-MCQ mean (start of training day)	Post-MCQ mean[1] (end of training day)	Mean change between pre-MCQ and post-MCQ[1]	P[1,2]
	ISS	69.4 (65.4 to 73.4)	82.6 (79.3 to 85.8)	ΔISS: 13.1 (11.0 to 15.2)	<0.0001
	OSS	70.6 (66.0 to 75.2)	83.3 (80.4 to 86.1)	ΔOSS:12.7 (10.3 to 15.1)	<0.0001
			P[1] = 0.74		

CI: confidence interval; ISS: in situ simulation; OSS: off site simulation; MCQ: multiple choice question (range: 0-100%)

[1)] CI and P-values based on generalised estimating equations to account for potential correlation within teams.
[2)] Adjusted for multiple testing.

Exploratory outcomes

Safety Attitudes Questionnaire (SAQ): No differences were found between the ISS and OSS groups with respect to all the dimensions of post-SAQ (table 4).

Safety Attitudes Questionnaire (SAQ) post hoc analysis: The post hoc analysis showed no differences in any of the dimensions from pre-SAQ to post-SAQ (table 4).

Table 4. Means (95% CI) of SAQ (converted to percentages) for five dimensions one month before (pre-SAQ) and one month after (post-SAQ) the simulation training day with ISS and OSS. Exploratory analysis comprised a comparison of the mean post-SAQ of the ISS versus the OSS group

	Simulation intervention	Pre-SAQ mean (1 month before)	Post-SAQ mean (1 month after)	Mean (95% CI change between pre-SAQ and post-SAQ	P[1,2]
SAQ Team work	ISS	80.5 (76.8 to 84.2)	81.1 (76.7 to 85.5)	ΔISS: 0.0 (-2.9 to 2.8)	0.98
	OSS	78.4 (74.2 to 82.6)	81.2 (77.5 to 85.0)	ΔOSS: 1.9 (-1.9 to 5.7)	0.60
			P[1] = 0.97		
SAQ Safety Climate	ISS	66.7 (61.9 to 71.5)	70.6 (65.9 to 75.2)	ΔISS: 3.4 (0.5 to 6.3)	0.20
	OSS	69.3 (65.5 to 72.9)	70.8 (66.8 to 74.8)	ΔOSS: 1.3 (-1.3 to 3.9)	0.61
			P[1] = 0.93		
SAQ Job Satisfaction	ISS	86.4 (83.0 to 89.7)	87.5 (83.3 to 91.7)	ΔISS: 0.4 (-1.9 to 2.6)	0.95
	OSS	85.6 (81.7 to 89.5)	85.7 (81.9 to 89.5)	ΔOSS: -0.0 (-2.8 to 2.7)	0.98
			P[1] = 0.54		
SAQ Stress recognition	ISS	69.7 (63.5 to 75.8)	68.8 (62.4 to 75.1)	ΔISS: -1.4 (-6.5 to 3.7)	0.84
	OSS	67.3 (61.4 to 73.1)	69.2 (64.0 to 74.4)	ΔOSS: 3.2 (-3.2 to 9.6)	0.60
			P[1] = 0.92		
SAQ Work condition	ISS	66.4 (60.9 to 71.9)	64.9 (59.0 to 70.8)	ΔISS: -2.5 (-7.0 to 2.1)	0.60
	OSS	65.8 (60.1 to 71.6)	64.0 (58.1 to 69.8)	ΔOSS: -1.7 (-5.4 to 2.0)	0.60
			P[1] = 0.82		

CI: confidence intervals; ISS: in situ simulation; OSS: off site simulation; SAQ: Safety Attitudes Questionnaire (range: 0-100%).

[1)] CI and p-values based on generalised estimating equations to account for potential correlation within teams. [2)] Adjusted for multiple testing.

Salivary cortisol: No significant differences were found in the mean changes between baseline and peak for salivary cortisol in the ISS and OSS groups (table 5).

Salivary cortisol post hoc analysis: The salivary cortisol level increased significantly from baseline to peak in the ISS and OSS groups following both the first (caesarean section) and the second (postpartum haemorrhage) simulation (table 5).

Stress-Trait Anxiety Inventory (STAI) and Cognitive Appraisal (CA): No differences were found in the mean changes between baseline and peak for STAI and CA between the ISS and OSS groups (table 5).

Stress-Trait Anxiety Inventory (STAI) and Cognitive Appraisal (CA) post hoc analysis: STAI increased significantly from baseline to peak in the ISS and OSS setting in the caesarean section simulation, but no increase was registered in the second simulation on postpartum haemorrhage (table 5). CA decreased significantly from baseline to peak in the ISS and OSS settings in both the caesarean section and in the postpartum haemorrhage simulations (table 5).

Table 5. Mean (95% CI) of salivary cortisol (nmol/L), STAI, CA during simulation in management of an emergency caesarean section and postpartum haemorrhage conducted as ISS and OSS. Exploratory analysis is difference in mean of baseline to peak in the ΔISS versus ΔOSS group

		Baseline	Post-simulation 0 min Mean[1]	Post-simulation 5 min Mean[1]	Post-simulation 10 min Mean	Peak-level mean[1,3]	Change in mean[1] baseline to peak	P[1,2]	Difference in mean[1] baseline to peak of ΔOSS vs. ΔISS
1st simulation: Emergency caesarean section									
Corti-sol	ISS	7.0 (6.3 to 7.8)	8.9 (7.2 to 10.6)	8.1 (6.6 to 9.6)	8.1 (6.6 to 9.5)	9.3 (7.6 to 11.0)	ΔISS: 2.3 (0.8 to 3.7)	0.004	-0.5 (-2.6 to 1.5) P[1] = 0.62
	OSS	7.3 (5.3 to 9.2)	8.2 (6.3 to 10.2)	7.8 (6.1 to 9.6)	8.0 (6.2 to 9.8)	9.0 (6.9 to 11.1)	ΔOSS: 1.7 (0.2 to 3.2)	0.03	
STAI	ISS	32.2 (30.4 to 34.0)	34.8 (32.7 to 37.0)		31.3 (29.5 to 33.1)	36.5 (34.3 to 38.7)	ΔISS: 4.3 (2.5 to 6.1)	<0.0001	-0.4 (-2.7 to 2.0) P[1] = 0.76
	OSS	33.1 (31.1 to 35.0)	34.8 (32.2 to 37.3)		30.7 (29.0 to 32.4)	37.0 (34.7 to 39.3)	ΔOSS: 3.9 (2.3 to 5.5)	<0.0001	
CA	ISS	1.0 (0.9 to 1.1)	0.8 (0.7 to 1.0)		0.8 (0.7 to 0.9)	0.8 (0.7 to 1.0.)	ΔISS: -0.2 (-0.4 to -0.1)	0.01	0.0 (-0.2 to 0.2) P[1] = 0.92
	OSS	1.0 (1.0 to 1.1)	0.8 (0.7 to 0.9)		0.8 (0.6 to 0.9)	0.9 (0.7 to 0.9)	ΔISS: -0.2 (-0.3 to -0.1)	0.002	
2nd simulation: Postpartum haemorrhage									
Corti-sol	ISS	7.4 (6.5 to 8.3)	9.2 (7.7 to 10.7)	7.7 (6.6 to 8.8)	7.4 (6.3 to 8.5)	9.4 (7.9 to 10.9)	ΔISS: 2.0 (1.0. to 3.0)	0.0004	-1.2 (-2.5 to 0.1) P[1] = 0.08
	OSS	6.9 (5.9 to 7.9)	7.5 (6.6 to 8.4)	6.7 (5.8 to 7.7)	6.8 (6.0 to 7.6)	7.7 (6.7 to 8.7)	ΔOSS: 0.8 (0.0 to 1.6)	0.05	
STAI	ISS	31.8 (30.0 to 33.6)	31.8 (30.1 to 33.6)		28.5 (27.3 to 29.7)	32.2 (30.5 to 33.9)	ΔISS: 0.4 (-1.2 to 2.0)	0.6	-0.4 (-1.7 to 2.4) P[1] = 0.74
	OSS	32.1 (29.9 to 34.2)	32.4 (30.5 to 34.3)		30.1 (28.5 to 31.8)	32.8 (31.0 to 34.7)	ΔOSS: 0.8 (-0.5. to 2.0)	0.29	
CA	ISS	1.0 (0.9 to 1.1)	0.8 (0.7 to 0.9)		0.8 (0.7 to 0.9)	0.8 (0.7 to 0.9)	ΔISS: -0.2 (-0.3 to -0.0)	0.01	0.0 (-0.2 to 0.2) P[1] = 0.94
	OSS	1.1 (1.0 to 1.2)	0.9 (0.7 to 1.0)		0.8 (0.7 to 0.9)	0.9 (0.7 to 1.0.)	ΔISS: -0.2 (-0.3 to -0.1)	0.006	

CI: Confidence interval ; ISS: in situ simulation ; OSS: off site simulation ; STAI: Stress-Trait Anxiety Inventory (range 20-80) ; CA: cognitive appraisal (range 0,1-10).

[1)] CI and p-values based on generalised estimating equations to account for potential correlation within teams.
[2)] Adjusted for multiple testing.
[3)] Peak level is the maximum of the measurements obtained at 0, 5, and 10 minutes after the end of the simulation.

Intrinsic Motivation Inventory (IMI): No differences were found between the ISS and the OSS groups for the IMI score (table 6).

Table 6. Mean (95% CI) motivation after participation in either ISS or OSS. Exploratory analysis comprised a comparison of the mean IMI and the mean of the ISS and OSS groups

	Simulation intervention	IMI mean (1 week after)
Interest / Enjoyment	ISS	5.2 (4.9 to 5.5)
	OSS	5.3 (5.1 to 5.5)
		P[1] = 0.72
Perceived competence	ISS	5.1 (4.8 to 5.4)
	OSS	4.9 (4.7 to 5.1)
		P[1] = 0.24
Perceived choice	ISS	5.8 (5.6 to 6.1)
	OSS	5.5 (5.2 to 5.9)
		P[1] = 0.15
Pressure tension (reversed)	ISS	2.8 (2.5 to 3.1)
	OSS	2.9 (2.6 to 3.3)
		P[1] = 0.65

CI: confidence interval; ISS: in situ simulation; OSS: off site simulation; IMI: Intrinsic Motivation Inventory (range:1-7).

[1)] CI and p-values based on generalised estimating equations to account for potential correlation within teams.

Participant evaluations and perception: For almost all 20 questions in the evaluation questionnaire, the scores participants gave in the ISS and OSS groups did not differ significantly. However, the two questions addressing the authenticity fidelity of the simulations were scored significantly higher by the ISS participants compared to the OSS participants (table 7).

Table 7. Participant evaluations after participation in either ISS or OSS in medians with 25% and 75% quartiles. Exploratory analysis comprised a comparison of the evaluation medians of the ISS versus OSS group

	ISS	OSS	
Evaluation questions (shortened version, original version in Danish)	**Median (1st Q–3rd Q)**	**Median (1st Q–3rd Q)**	**P[1]**
1. Over all the training day was (1=very bad to 5=very good)	5 (4-5)	5 (4-5)	0.70
2. Multi-professional approach with all healthcare groups involved was (1=very bad to 5=very good)	5 (4-5)	5 (4-5)	0.70
3. I thought the level of education of the training was (1=very much over my level to 5=very much below my level)	3 (3-3)	3 (3-3)	0.70
4. Will recommend others to participate (1=never to 5=always)	5 (5-5)	5 (4-5)	0.70
5. Did simulations inspire you to change procedures or practical issues in the labour room or operating theatre (1=no ideas to 5=many ideas) (included open-ended questions)	3 (2-3)	3 (2-4)	0.70
6. Did simulations inspire you to change guidelines (1=no ideas to 5=many ideas) (included open-ended questions)	2 (1-2)	2 (1-2)	0.70
Simulation of an emergency CS			
7. Over all my learning was (1=very bad to 5=very good)	4 (3-4)	4 (3-4)	0.90
8. The authenticity of the CS simulation was (1=not at all authentic to 5=very authentic)	4 (3-4)	3 (3-4)	0.02
9. The authenticity of the CS simulation influenced my learning (1=not at all important to 5=very important)	4 (4-4.5)	4 (4-4)	0.65
10. Collaboration in the CS team was (1=very bad to 5=very good)	4 (4-4.5)	4 (3.8-4)	0.27
11. Communication in the CS team was (1=very bad to 5=very good)	4 (3-4)	4 (3-4)	0.23
12. The CS team leader was (1=very bad to 5=very good)	4 (3-4)	4 (3-4)	0.26
13. My learning at the debriefing after the CS was (1=very bad to 5=very good)	4 (4-5)	4 (4-4)	0.88
Simulation in PPH			
14. My learning overall was (1=very bad to 5=very good)	4 (4-4)	4 (4-4)	0.70
15. The authenticity of the PPH simulation was (1=not a tall authentic to 5=very authentic)	4 (3-4)	3 (3-4)	0.01
16. The authenticity of the simulation in PPH influenced my learning (1=not at all important to 5=very important)	4 (4-4.5)	4 (4-4)	0.23
17. Collaboration in the PPH team was (1=very bad to 5=very good)	4 (4-4.5)	4 (4-4)	0.64
18. Communication in the PPH team was (1=very bad to 5=very good)	4 (3.5-4)	4 (3-4)	0.64
19. The PPH team leader was (1=very bad to 5=very good)	4 (4-4)	4 (3-4)	0.23
20. My learning at the debriefing after the PPH was (1=very bad to 5=very good)	4 (4-4)	4 (4-4)	0.57

ISS: in situ simulation; OSS: off site simulation; 1st Q–3rd Q: 25% and 75% quartiles; CS: Emergency caesarean section; PPH: postpartum haemorrhage.

[1)] Adjusted for multiple testing.

Team Emergency Assessment Measure (TEAM): No significant differences were found in the team scoring of performance between the ISS and OSS groups (table 8).

Team Emergency Assessment Measure (TEAM) post hoc analysis: A significant increase was found in the team scoring of performance from the first simulation (emergency caesarean section) to the second (postpartum haemorrhage) (table 8).

Table 8. Mean (95% CI) of video assessment scores of performance with the TEAM scale. Four consultants recruited outside the research hospital did the video assessment scoring. Exploratory analysis comprised a comparison of the mean TEAM score of the ISS versus the OSS group

	ISS	OSS	
Video assessment scoring of performance	Mean	Mean	P
TEAM (means of item rating) Simulation in emergency CS[1]	2.6 (2.3 to 3.0)	2.4 (2.1 to 2.8)	
TEAM (means of item rating) Simulation in PPH[1]	2.9 (2.5 to 3.2)	2.8 (2.5 to 3.2)	
Estimated overall difference in mean between ISS and OSS[2]	0.1 (-0.2 to 0.5)		0.36
TEAM (global rating) Simulation in emergency CS[1]	6.1 (4.8 to 7.3)	5.3 (4.0 to 6.5)	
TEAM (global rating) Simulation in PPH[1]	6.8 (5.5 to 8.1)	6.3 (5.0 to 7.6)	
Estimated overall difference in mean between ISS and OSS[2]	0.7 (-0.4 to 1.7)		0.18
Differences in video assessment scores of performance between emergency CS (1st) and PPH (2nd) simulation scenarios			
Differences in mean of TEAM (means of item rating) of the simulation in emergency CS versus PPH[2]	0.3 (0.1 to 0.5)		0.0003
Differences in mean of TEAM (global rating) of the simulation in emergency CS versus PPH[2]	0.9 (0.3 to1.5)		0.005

CI: confidence interval to ISS: in situ simulation ; OSS: off site simulation ; Q: quartile ; TEAM: Team Emergency Assessment Measure (range for item rating:0-4 ; range for global rating:1-10) ; CS: Emergency caesarean section ; PPH: postpartum haemorrhage.

[1)] Means found from a linear mixed model including an interaction between simulation group (ISS and OSS) and simulation scenario (emergency CS and PPH).
[2)] Overall difference in means found from an additive linear mixed model based on simulation group and simulation scenario.

Organisational changes: A qualitative analysis showed that more ideas for organisational changes were suggested by ISS participants than OSS participants. For details see supplementary table S1. The quantitative analysis, however, showed that participants in the ISS and the OSS groups scored equally concerning whether the simulations inspired making changes in procedures or guidelines (table 7, question numbers 5 and 6).

DISCUSSION

In this randomised trial, we did not find that simulation-based medical education conducted as ISS compared with OSS led to different outcomes assessed on knowledge, patient safety attitude, stress, motivation, perceptions of the simulations, and team performance. Participant perception of the authenticity of the ISS and OSS differed significantly, but this had no influence on other individual or team outcomes. We observed that ISS participants provided more ideas for organisational changes than OSS participants did. This is in accordance with several non-randomised studies describing a positive impact of ISS on the organisation.[8,10,11,13,57-59]

In the evaluation questionnaire (table 7), participants were asked about their perceptions of the authenticity of the simulations, which can be interpreted as their perception of the simulation's fidelity. The participants scored fidelity to be significantly higher in ISS compared with OSS; however, there were no differences in any of the other outcomes between the ISS and OSS groups. The results from this randomised trial are not consistent with traditional situated learning theory, which states that increased fidelity leads to improved learning.[60,61] Conclusions from this trial, however, are in alignment with more recent empirical research and discussions on fidelity and learning.[6,62-64] Our study indicates that the change in simulation fidelity, as change in setting for simulation, does not necessarily translate into learning.[6] Another randomised trial, which compared OSS as in-house training at the hospital in rooms specifically allocated for training to OSS in a simulation centre, also showed that the simulation setting was of minor importance and that there was no additional benefit from training OSS in a simulation centre versus OSS in-house.[51,65]

The present trial involved simulation based training with six different healthcare professions. A relevant perspective is the discussion on expanding the traditional concept of fidelity to include sociological fidelity, which encompasses the relationship between the various healthcare professionals.[37;64] After completing the trial we decided to explore more closely the experiences between the healthcare professionals in a qualitative study *(Submitted BMJopen, March 2015, Sørensen JL et al. Clarifying the learning experiences of healthcare professionals with in situ versus off site simulation-based medical education: a qualitative study).*

Post hoc analyses showed similar educational effects in the ISS and the OSS groups with a knowledge gain of 13% in both groups. It can be argued that this knowledge gain was due to the test effect.[66,67] We believe, however, that the test effect was minimised as feedback was not given after the initial testing, which is viewed as crucial to learning from a test, and furthermore only one MCQ test was used.[67]

No differences were found in the mean Safety Attitudes Questionnaire (SAQ) score after simulation-based medical education in the ISS and OSS groups. Earlier studies have described that high SAQ values mean that SAQ cannot be influenced by an intervention.[68,69] The values for SAQ were generally high in this trial compared to various other studies from non-Scandinavian countries.[68-71] The post hoc analysis showed no effect of the simulation intervention on SAQ.

There were no differences in the stress level when measured as salivary cortisol levels, Stress-Trait Anxiety Inventory (STAI) and Cognitive Appraisal (CA) in the ISS and OSS groups. The post hoc analysis showed that simulation-based medical education triggered objective stress, measured by salivary cortisol, to the same extent in the ISS and OSS groups. As measured by STAI, subjective stress was only triggered in the first but not the second simulation for the ISS and OSS groups, respectively. This habituation to simulation is well known from the literature. CA seemed to be without discriminatory effect and a decrease was observed where an increase would have been expected, and the levels of CA were low compared to other studies. Previously used among students and medical trainees,[21,72,73] CA appeared to have a less discriminatory effect in these more senior groups of healthcare professionals.

Intrinsic Motivation Inventory[23,47] revealed no differences between ISS and OSS. Motivation has not previously been tested in educational simulation studies; hence there is no comparison for these results. It is argued a gap appears to exist in the simulation literature on motivational factors and further research has been encouraged. [5] Some argue that simulation in the clinical setting, as with ISS, should increase motivation,[14] but this was not confirmed by findings in the present trial.

The evaluation data showed no differences between ISS and OSS. Both the ISS and the OSS participants gave very high scores on the evaluation. This is in accordance with what is generally seen in inter-professional training.[74]

The team performance showed no differences between ISS and OSS. The post hoc analysis showed that teams performed statistically significantly better in the second compared to the first simulation, which indicates that the simulations were effective. Validated in previous studies, the Team Emergency Assessment Measure (TEAM) scale has been found reasonably intuitive to use,[48,49] which was also our impression in this study.

According to the participants own perceptions they found that ISS and OSS were equally inspirational with regard to suggesting organisational changes in the delivery room, operating theatre and for clinical guidelines. The qualitative analysis, however, revealed that particularly ISS participants provided more ideas for suggested changes, especially concerning technology and tools[56] in the delivery ward and the operating theatre. The organisational impact of ISS has also been found in previous non randomised studies, but it has never been confirmed in a randomised trial.[8,11,13,57]

Strength and limitations

This trial has several strengths. It was conducted with an adequate generation of allocation sequence; adequate allocation concealment; adequate reporting of all relevant outcomes; had very few drop-outs; and was conducted on a not-for-profit bias.[75,76] The trial was also blinded for data managers and statisticians. Generally, ISS programmes have been criticised for their lack of meaningful evaluations of the effectiveness of the programmes.[8] A strength of this trial was its use of a broad variety of outcome measures using previously validated scales to assess the effect on the individual, the team, and the organisational level.

A limitation of the study is the fact that the outcome was based only on immediate measurements of knowledge level and of team performance. Only perceptions of simulation were measured after one week (evaluation and motivation) and safety attitudes after one month. No clinical outcome was measured.

A strength of this trial is the involvement of authentic teams that mirrored teams in real life, which is important for the so-called sociological fidelity.[36,77] The teams in this trial were authentic in their design and hence resemble the real clinical setting in every possible way.[63,78] This kind of teams are called ‘add hoc’ on call teams and is very difficult to follow and observe in the real

clinical setting, and assessment of 'ad hoc' team's clinical performance for a long period is almost impossible.

An additional approach to the assessment could have been performance-based tests of clinical work, but this was considered unfeasible. Previous research on assessment suggests that knowledge-based written assessments can predict the results of performance-based tests, and hence knowledge-based assessment was used as a proxy for performance.[79-81]

There is a risk of type II error and the trial is most likely underpowered, as many educational trials are. On the other hand, it should be discussed whether performing a larger trial to detect a statistically significant effect of ISS is relevant or feasible, and appear to be without a clinically or educationally relevant effect.[82]

The improvements on knowledge and team performance may also be due to the Hawthorne effect, i.e., due to individuals changing behaviour as a result of their awareness of being observed.[83] From an educational perspective a major problem with the Hawthorne effect is an intervention group versus a control group, where the control group is given no intervention. [83] This issue was avoided in this trial as exactly the same intervention was used for both groups, the only difference being the physical setting, thus minimising the Hawthorne effect in our trial. [83]

Conclusion

This randomised trial compared ISS versus OSS, where OSS was provided as in-house training at the hospital in rooms specifically allocated for training. From the present trial we concluded that changes in settings from OSS to ISS do not provide key elements for improving simulation-based medical education. Although participant perception of the fidelity of ISS versus OSS differed significantly, there were no differences in knowledge, patient safety attitude, stress measurements, motivation and team performance between the groups, except that the ISS group generated more suggestions for organisational changes. The present trial indicated that the fidelity of the setting seemed to be of less importance for learning; however, more research is necessary to better understand which aspects of simulation that is the most important for learning.

Acknowledgements

The authors would like to thank the doctors, midwives, and nurses who took part in the working committee and in planning the intervention, but especially midwife Pernille Langhoff-Roos (Department of Obstetrics, Rigshospitalet, University of Copenhagen) for her contributions to the detailed planning, recruitment of participants, and completion of the simulations. The authors would like to thank Jørn Wetterslev (Copenhagen Trial Unit) for advice on the potential impact of the clustering effect; Karl Bang Christensen (Section of Biostatistics, Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen) for advice on the statistical plan; Solvejg Kristensen (Danish National Clinical Registries) for her advice and for providing a Danish edition of Safety Attitudes Questionnaire; Per Bech (Psychiatric Research Unit, Mental Health Centre North Zealand) for advice and for providing a Danish edition of the State-Trait Anxiety Inventory; Claire Welsh (Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen) and Niels Leth Jensen (certified translator) for forward the back translation of Intrinsic Motivation Inventory; Jesper Sonne (Department of Clinical Pharmacology, Bispebjerg Hospital) for advice on medicine that can influence cortisol measurements; Anne Lippert and Anne-Mette Helsø (Danish Institute for Medical Simulation, Herlev Hospital, University of Copenhagen) for training the facilitators in conducting debriefing after simulation; medical student Veronica Markova for undertaking all practical issues on cortisol measurements; Bente Bennike (Laboratory of Neuropsychiatry, Rigshospitalet, University of Copenhagen) for managing the cortisol analyses; medical students Tobias Todsen and Ninna Ebdrup for video recordings of simulations; and Tobias Todsen for editing and preparing videos for assessment. Furthermore, the authors would like to thank Helle Thy Østergaard and Lone Fuhrmann (Department of Anesthesiology, Herlev Hospital, University of Copenhagen), Lone Krebs (Department of Obstetrics and Gynecology, Holbæk Hospital) and Morten Beck Sørensen (Department of Obstetrics and Gynecology, Odense University Hospital) for reviewing and assessing all videos. Finally, thank you to Nancy Aaen, (copy editor and translator) for revising the manuscript.

Contributors

JLS conceived of the idea for this trial. BO and CV supervised the trial. All authors made contributions to the design of the trial. JLS, assisted by BO, was responsible for acquiring funding. JLS, JL, and CG contributed to the sample size estimation and detailed designing of and execution

of the randomisation process. JLS, MJ, KE, DOE, and VL made substantial contributions to the practical and logistical aspects of the trial, while PW contributed to the discussion and practical and logistical issues concerning testing salivary cortisol. Jointly with JLS, SR and LS performed the statistical analysis. JLS wrote the draft manuscript. All authors provided critical review of this paper and approved the final manuscript.

Funding

The trial was funded by non-profit funds, including the Danish Regions Development and Research Foundation, the Laerdal Foundation for Acute Medicine and Aase and Ejnar Danielsen Foundation. None of the fundgivers had any role in study design; data collection, access, analysis, or interpretation; writing of the report; or the decision to publish

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf The principal investigator and lead author (JLS) reports non-profit funding mentioned above. Doris Østergaard reports board membership of Laerdal Foundation for Acute Medicine. Other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the this work in the previous three years; no other relationships or activities that could appear to have influenced the our work.

Ethical approval

Participants were healthcare professionals, and neither patients nor patient data were used in the trial. Approval from the Regional Ethics Committee (protocol number H-2-2012-155) and the Danish Data Protection Agency (Number 2007-58-0015) were obtained. Participants were assured that their personal data, data on questionnaires, salivary cortisol samples, and video recordings would remain anonymous during analyses and reporting. The participants were asked to respect the confidentiality of their observations about colleagues' performance in the simulated setting.

Data sharing

No consent for data sharing with other parties was obtained but the corresponding author may be contacted to forward requests for data sharing.

Transparency

The principal investigator and lead author (JLS) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported. No important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Reference List

[1] Issenberg SB, McGaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;27:10-28.

[2] McGaghie WC, Issenberg SB, Petrusa ER, et al. A critical review of simulation-based medical education research: 2003-2009. *Med Educ* 2010;44:50-63.

[3] Motola I, Devine LA, Chung HS, et al. Simulation in healthcare education: a best evidence practical guide. AMEE Guide No. 82. *Med Teach* 2013;35:e1511-e1530.

[4] Merien AE, Van der Ven J, Mol BW, et al. Multidisciplinary team training in a simulation setting for acute obstetric emergencies: a systematic review. *Obstet Gynecol* 2010;115:1021-31.

[5] McGaghie WC, Issenberg SB, Barsuk JH, et al. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ* 2014;48:375-85.

[6] Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;46:636-47.

[7] Beaubien JM, Baker DP. The use of simulation for training teamwork skills in health care: how low can you go? *Qual Saf Health Care* 2004;13 Suppl 1:i51-i56.

[8] Rosen MA, Hunt EA, Pronovost PJ, et al. In situ simulation in continuing education for the health care professions: a systematic review. *J Contin Educ Health Prof* 2012;32:243-54.

[9] Riley W, Davis S, Miller KM, et al. Detecting breaches in defensive barriers using in situ simulation for obstetric emergencies. *Qual Saf Health Care* 2010;19 Suppl 3:i53-i56.

[10] Guise JM, Lowe NK, Deering S, et al. Mobile in situ obstetric emergency simulation and teamwork training to improve maternal-fetal safety in hospitals. *Jt Comm J Qual Patient Saf* 2010;36:443-53.

[11] Walker ST, Sevdalis N, McKay A, et al. Unannounced in situ simulations: integrating training and clinical practice. *BMJ Qual Saf* 2013;22:453-8.

[12] Patterson MD, Geis GL, Falcone RA, et al. In situ simulation: detection of safety threats and teamwork training in a high risk emergency department. *BMJ Qual Saf* 2013;22:468-77.

[13] Patterson MD, Geis GL, Lemaster T, et al. Impact of multidisciplinary simulation-based training on patient safety in a paediatric emergency department. *BMJ Qual Saf* 2013;22:383-93.

[14] Stocker M, Burmester M, Allen M. Optimisation of simulated team training through the application of learning theories: a debate for a conceptual framework. *BMC Med Educ* 2014;14:69.

[15] Sullivan NJ, Duval-Arnould J, Twilley M, et al. Simulation Exercise to Improve Retention of Cardiopulmonary Resuscitation Priorities for In-Hospital Cardiac Arrests: A Randomized Controlled Trial. *Resuscitation* 2014;86C:6-13.

[16] Rubio-Gurung S, Putet G, Touzet S, et al. In situ simulation training for neonatal resuscitation: an RCT. *Pediatrics* 2014;134:e790-e797.

[17] LeBlanc VR. The effects of acute stress on performance: implications for health professions education. *Acad Med* 2009;84:S25-S33.

[18] LeBlanc VR, Manser T, Weinger MB, et al. The study of factors affecting human and systems performance in healthcare using simulation. *Simul Healthc* 2011;6 Suppl:S24-S29.

[19] Arora S, Sevdalis N, Aggarwal R, et al. Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surg Endosc* 2010;24:2588-93.

[20] Harvey A, Nathens AB, Bandiera G, et al. Threat and challenge: cognitive appraisal and stress responses in simulated trauma resuscitations. *Med Educ* 2010;44:587-94.

[21] Harvey A, Bandiera G, Nathens AB, et al. Impact of stress on resident performance in simulated trauma scenarios. *J Trauma Acute Care Surg* 2012;72:497-503.

[22] Finan E, Bismilla Z, Whyte HE, et al. High-fidelity simulator technology may not be superior to traditional low-fidelity equipment for neonatal resuscitation training. *J Perinatol* 2012;32:287-92.

[23] Kusurkar RA, Ten Cate TJ, van AM, et al. Motivation as an independent and a dependent variable in medical education: a review of the literature. *Med Teach* 2011;33:e242-e262.

[24] Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Med Teach* 2011;33:961-73.

[25] Ryan RM, Deci EL. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp Educ Psychol* 2000;25:54-67.

[26] Drife J. Reducing risk in obstetrics. *Qual Health Care* 1995;4:108-14.

[27] Gold KJ, Kuznia AL, Hayward RA. How physicians cope with stillbirth or neonatal death: a national survey of obstetricians. *Obstet Gynecol* 2008;112:29-34.

[28] Veltman LL. Getting to havarti: moving toward patient safety in obstetrics. *Obstet Gynecol* 2007;110:1146-50.

[29] Pronovost PJ, Holzmueller CG, Ennen CS, et al. Overview of progress in patient safety. *Am J Obstet Gynecol* 2011;204:5-10.

[30] Hove LD, Bock J, Christoffersen JK, et al. Analysis of 127 peripartum hypoxic brain injuries from closed claims registered by the Danish Patient Insurance Association. *Acta Obstet Gynecol Scand* 2008;87:72-5.

[31] Berglund S, Grunewald C, Pettersson H, et al. Severe asphyxia due to delivery-related malpractice in Sweden 1990-2005. *BJOG* 2008;115:316-23.

[32] Cantwell R, Clutton-Brock T, Cooper G, et al. Saving Mothers' Lives: Reviewing maternal deaths to make motherhood safer: 2006-2008. The Eighth Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. *BJOG* 2011;118 Suppl 1:1-203.

[33] Bodker B, Hvidman L, Weber T, et al. Maternal deaths in Denmark 2002-2006. *Acta Obstet Gynecol Scand* 2009;88:556-62.

[34] Johannsson H, Ayida G, Sadler C. Faking it? Simulation in the training of obstetricians and gynaecologists. *Curr Opin Obstet Gynecol* 2005;17:557-61.

[35] Sørensen JL, Van der Vleuten C, Lindschou J, et al. 'In situ simulation' versus 'off site simulation' in obstetric emergencies and their effect on knowledge, safety attitudes, team performance, stress, and motivation: study protocol for a randomized controlled trial. *Trials* 2013;14:220.

[36] Boet S, Bould MD, Layat BC, et al. Twelve tips for a successful interprofessional team-based high-fidelity simulation education session. *Med Teach* 2014;36:853-7.

[37] SimMom Mannequin. http://www.laerdal.com/dk/SimMom . 2014.

[38] Royal College of Obstetricans and Gynaecologists course: PROMPT, Train the trainers.https://www.rcog.org.uk/globalassets/documents/courses-exams-and-events/training-courses/prompt-2---train-the-trainers-mar2015.pdf. 2015.

[39] Rudolph JW, Simon R, Dufresne RL, et al. There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. *Simul Healthc* 2006;1:49-55.

[40] Thistlethwaite JE, Davies D, Ekeocha S, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Med Teach* 2012;34:e421-e444.

[41] Sørensen JL, Thellensen L, Strandbygaard J, et al. Development of a knowledge test for multi-disciplinary obstetric-anesthesia emergency training. *Acta Anaesthesiol Scand* 2015;59:123-33.

[42] Kristensen S, Sabroe S, Bartels P, et al. Adaption and validation of the Safety Attitudes Questionnaire for the Danish hospital setting. *Clin Epidemiol* 2015;7:149-60.

[43] Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;6:44.

[44] Deilkas ET, Hofoss D. Psychometric properties of the Norwegian version of the Safety Attitudes Questionnaire (SAQ), Generic version (Short Form 2006). *BMC Health Serv Res* 2008;8:191.

[45] Spielberger CD, Gorsuch RL, Lushene RE. Manual for the State-Trait Anxiety Inventory. Palo Alto, Calif.: Consulting Psychologists Press; 1970.

[46] Bech P. [Klinisk psykometri]. 1.udgave, 2 oplag ed. Munksgaard Danmark; 2011.

[47] Self-Determination Theory (SDT) http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/. 2015.

[48] Cooper S, Cant R, Porter J, et al. Rating medical emergency teamwork performance: development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;81:446-52.

[49] McKay A, Walker ST, Brett SJ, et al. Team performance in resuscitation teams: comparison and critique of two recently developed scoring tools. *Resuscitation* 2012;83:1478-83.

[50] Sorensen JL, Lokkegaard E, Johansen M, et al. The implementation and evaluation of a mandatory multi-professional obstetric skills training program. *Acta Obstet Gynecol Scand* 2009;88:1107-17.

[51] Crofts JF, Ellis D, Draycott TJ, et al. Change in knowledge of midwives and obstetricians following obstetric emergency training: a randomised controlled trial of local hospital, simulation centre and teamwork training. *BJOG* 2007;114:1534-41.

[52] Ukoumunne OC, Gulliford MC, Chinn S, et al. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;3:iii-92.

[53] Campbell MK, Piaggio G, Elbourne DR, et al. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;345:e5661.

[54] Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13-22.

[55] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal StatisticaI Society Series B (Methodological)* 1995;57:289-300.

[56] Carayon P, Schoofs HA, Karsh BT, et al. Work system design for patient safety: the SEIPS model. *Qual Saf Health Care* 2006;15 Suppl 1:i50-i58.

[57] Sorensen JL, Lottrup P, Vleuten van der C, et al. Unannounced in situ simulation of obstetric emergencies: staff perceptions and organisational impact. *Postgrad Med J* 2014;90:622-9.

[58] Kobayashi L, Parchuri R, Gardiner FG, et al. Use of in situ simulation and human factors engineering to assess and improve emergency department clinical systems for timely telemetry-based detection of life-threatening arrhythmias. *BMJ Qual Saf* 2013;22:72-83.

[59] Wheeler DS, Geis G, Mack EH, et al. High-reliability emergency response teams in the hospital: improving quality and safety using in situ simulation training. *BMJ Qual Saf* 2013;22:507-14.

[60] Durning SJ, Artino AR, Jr., Pangaro LN, et al. Perspective: redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med* 2010;85:894-901.

[61] Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. *Med Teach* 2011;33:188-99.

[62] Koens F, Mann KV, Custers EJ, et al. Analysing the concept of context in medical education. *Med Educ* 2005;39:1243-9.

[63] Teteris E, Fraser K, Wright B, et al. Does training learners on simulators benefit real patients? *Adv Health Sci Educ Theory Pract* 2012;17:137-44.

[64] Grierson LE. Information processing, specificity of practice, and the transfer of learning: considerations for reconsidering fidelity. *Adv Health Sci Educ Theory Pract* 2014;19:281-9.

[65] Ellis D, Crofts JF, Hunt LP, et al. Hospital, simulation center, and teamwork training for eclampsia management: a randomized controlled trial. *Obstet Gynecol* 2008;111:723-31.

[66] Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 2006;17:249-55.

[67] Larsen DP, Butler AC, Roediger HL, III. Test-enhanced learning in medical education. *Med Educ* 2008;42:959-66.

[68] Schwendimann R, Zimmermann N, Kung K, et al. Variation in safety culture dimensions within and between US and Swiss Hospital Units: an exploratory study. *BMJ Qual Saf* 2013;22:32-41.

[69] Sexton JB, Berenholtz SM, Goeschel CA, et al. Assessing and improving safety climate in a large cohort of intensive care units. *Crit Care Med* 2011;39:934-9.

[70] Shoushtarian M, Barnett M, McMahon F, et al. Impact of introducing Practical Obstetric Multi-Professional Training (PROMPT) into maternity units in Victoria, Australia. *BJOG* 2014;121:1710-8.

[71] Siassakos D, Fox R, Hunt L, et al. Attitudes toward safety and teamwork in a maternity unit with embedded team training. *Am J Med Qual* 2011;26:132-7.

[72] Pottier P, Hardouin JB, Dejoie T, et al. Stress responses in medical students in ambulatory and in-hospital patient consultations. *Med Educ* 2011;45:678-87.

[73] Piquette D, Tarshis J, Sinuff T, et al. Impact of acute stress on resident performance during simulated resuscitation episodes: a prospective randomized cross-over study. *Teach Learn Med* 2014;26:9-16.

[74] Reeves S, Goldman J. Medical education in an interprofessional context. In: Dornan T, Mann K, Scherpbier A., Spencer J, editors. Medical Education: Theory and Practice.Edinburgh, London, New York, Oxford, Philadelphia, St Louis, Sydney, Toronto: Churchill Livingstone Elsevier; 2011. p. 51-64.

[75] Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.

[76] Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429-38.

[77] Sharma S, Boet S, Kitto S, et al. Interprofessional simulated learning: the need for 'sociological fidelity'. *J Interprof Care* 2011;25:81-3.

[78] Lingard L. Rethinking Competence in the Context of teamwork. In: Hodges B, Lingard L, editors. The Question of Competence. Reconsidering Medical Education in the Twenty-First Century. First ed. New York: ILR Press an imprent of Cornell University Press; 2012. p. 42-69.

[79] van der Vleuten C, Van Luyk SJ, Beckers HJ. A written test as an alternative to performance testing. *Med Educ* 1989;23:97-107.

[80] Kramer AW, Jansen JJ, Zuithoff P, et al. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. *Med Educ* 2002;36:812-9.

[81] Ram P, van der Vleuten C, Rethans JJ, et al. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. *Med Educ* 1999;33:197-203.

[82] Jakobsen JC, Gluud C, Winkel P, et al. The thresholds for statistical and clinical significance - a five-step procedure for evaluation of intervention effects in randomised clinical trials. *BMC Med Res Methodol* 2014;14:34.

[83] Fraenkal JR, Wallen NE, Hyde A. Internal Validity. In: McGraw-Hill, editor. How to Design and Evaluate Research in Education.New York: 2012. p. 166-83.

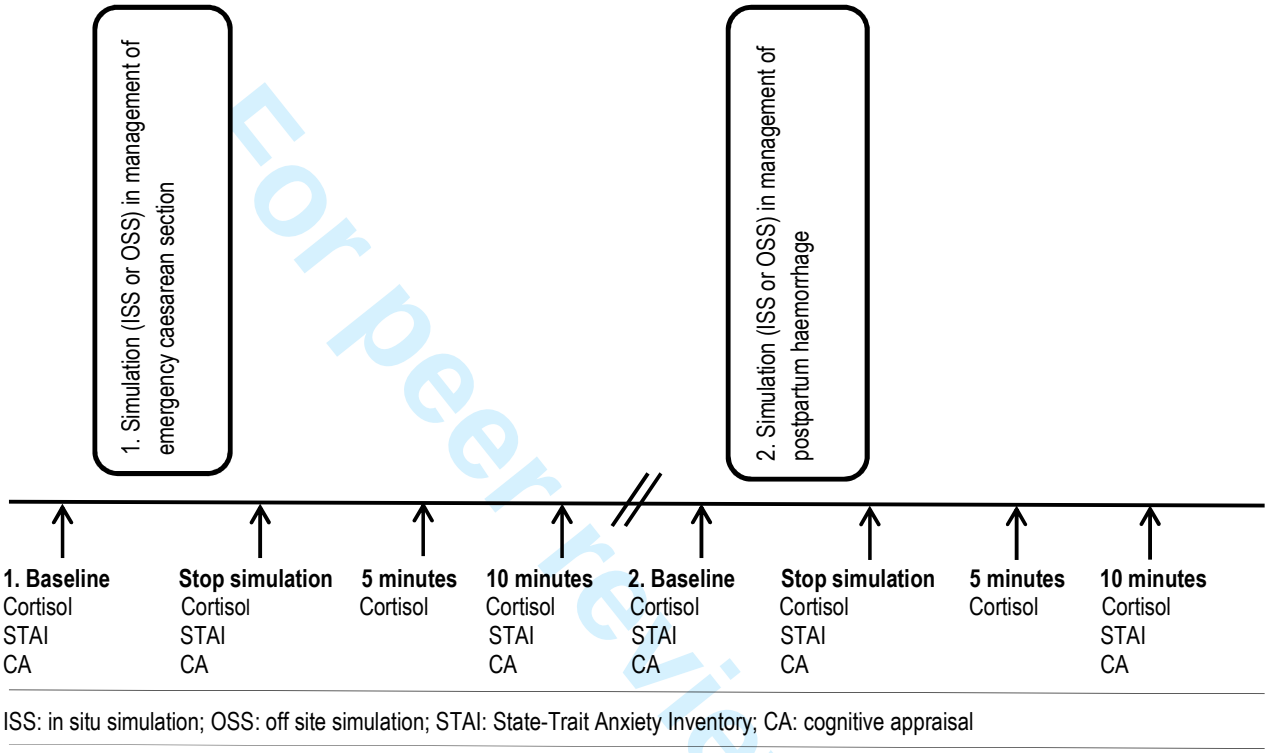
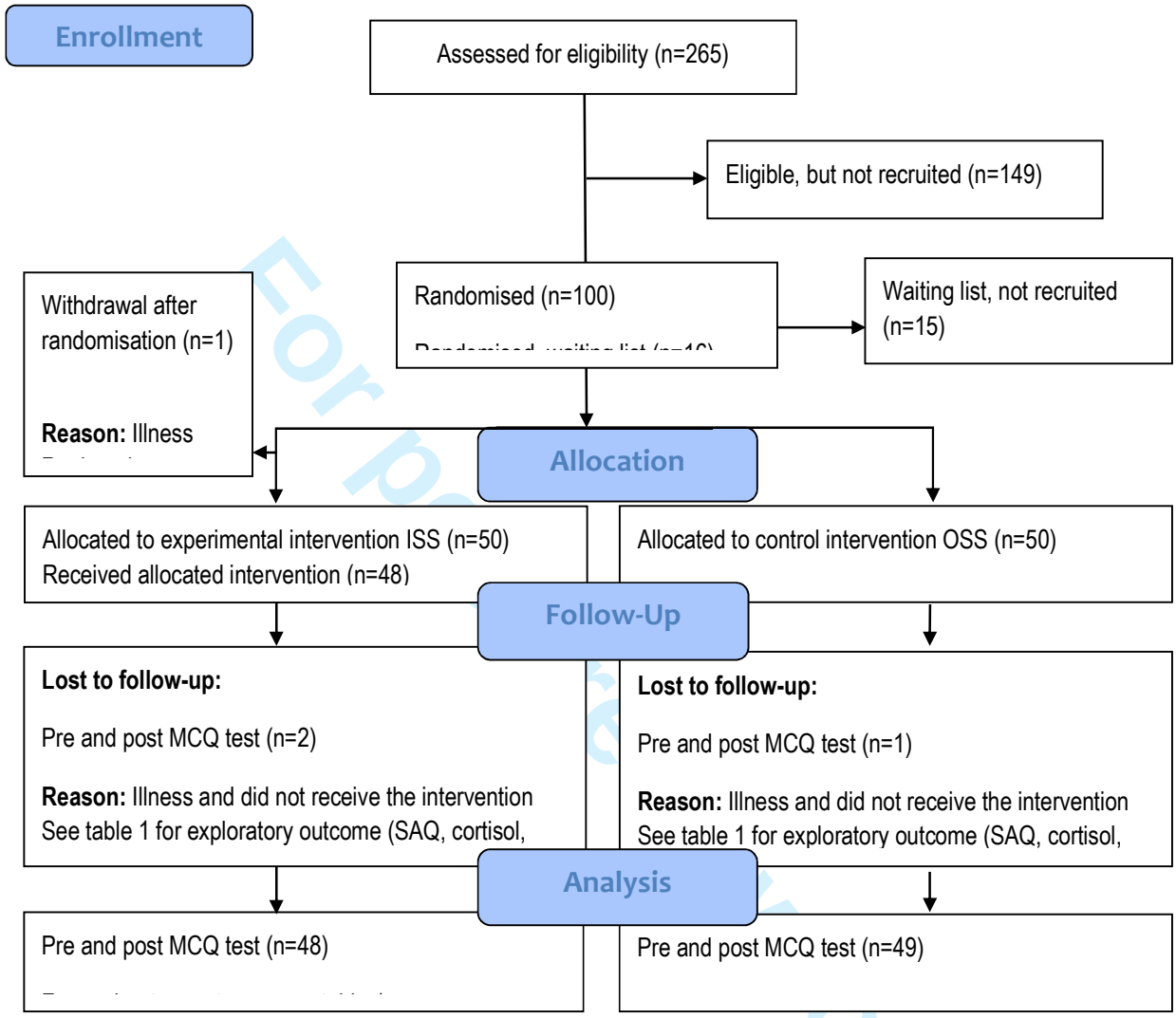


Figure 1. Timing of the simulations and measurement of stress: Objective stress was measured by salivary cortisol and subjective stress was measured by State-Trait Anxiety Inventory and cognitive appraisal



ISS: in situ simulation; OSS: off site simulation; MCQ: multiple choice question; SAQ: Safety Attitudes Questionnaire; STAI: Stress-Trait Anxiety Inventory; CA: cognitive appraisal; IMI: Intrinsic Motivation Inventory; TEAM: Team Emergency Assessment Measure.

Figure 2. Flow diagram for participants in a trial determining the effects of ISS versus OSS on 1) primary outcome: knowledge (MCQ test); and 2) exploratory outcomes: patient safety attitudes (SAQ), stress (salivary cortisol, STAI, CA), motivation (IMI), perceptions of simulation (evaluation questionnaire), video-assessed team performance (TEAM), and organisational impact

Table S1. Suggestions for practical and organisational changes identified during ISS and OSS classified according to the model of work system or structure from the Systems Engineering Initiative for Patient Safety. [55]

	Identified during		SEIPS component
	OSS	ISS	
Introduction of new employees			
Better introduction of new employees, including guided tour of trauma centre, acute admission centre, and blood bank	X	X	Person
Identification of staff			
Name badges visible on uniforms at all times and improved name badges for easier identification of the various healthcare professions	X	X	Person
Inclusion of people's names and also healthcare profession during staff presentation rounds in emergency situations	X	X	Organisation
Call systems, telephones and telephone numbers			
Pros and cons involved in changing the emergency call system from calling people individually to calling them as a group	X	X	Organisation
Request for more well-defined telephone chain for emergency calls; doctors preferred briefing to be from doctor to doctor	X	X	Organisation
More well-defined criteria for using emergency alarm button in delivery room	X	X	Organisation
Identical numbers to call night and day for anaesthesia assistance in obstetric emergencies	X	X	Organisation
A clearly visible list of relevant numbers in operating theatre for emergency situations		X	Tools & technology
Stickers with brief list of emergency numbers on back of name badges		X	Tools & technology
Clinical handover in emergency situations			
Repetition upon arrival in delivery room and operating theatre of clinical details and indication for procedures provided in telephone handover	X	X	Organisation
Consistent use of terminology from local guidelines and when grading emergency caesarean sections	X	X	Organisation
Patient identification and "time out" in operating theatre in emergency situations			
More clearly defined designation of who is responsible for identifying the patient and confirming the indication for procedure	X	X	Organisation
Improvement of computer system that is too difficult and slow for emergency situations		X	Tools & technology
Presence of partners during emergency caesarean sections			
Various opinions on whether partners should be allowed in operating theatre; more well-defined criteria for designating who communicates with partners	X	X	Organisation
Medication – postpartum haemorrhage			
Placement of tranexamic acid in the haemorrhage medication box; clinical guidelines on its administration should be made easily accessible	X	X	Tools & technology
Pre-prepared drips with oxytocin	X	X	Tools & technology
Midwives generally found administering medicine in operating theatre difficult; requested more clarity for designating who is responsible for the haemorrhage medication box there	X	X	Person, Task
Clarification of who is to document administration of medicine in operating theatre, especially when administered directly in the uterus and/or per rectum	X	X	Tools & technology

	Identified during		SEIPS components
	ISS	OSS	
Medication – emergency caesarean sections			
Pros and cons involved in administering terbutaline for tocolysis during foetal distress; informing anaesthesia staff important due to subsequent risk of maternal tachycardia	X	X	Tools & technology
Placement of terbutaline for tocolysis (intrauterine resuscitation) in delivery room and operating theatre to allow quick administration	X	X	Tools & technology
Address the knowledge gap among auxiliary nurses and midwives on administration of sodium citrate to prevent aspiration during general anaesthesia	X	X	Person, Task
Amend action card and clinical guidelines on emergency caesarean sections to specifically address oral administration of sodium citrate to prevent aspiration during general anaesthesia	X	X	Organisation
Make sodium citrate more easily accessible in the delivery room	X	X	Tools & technology
Staff members in operating theatre during postpartum haemorrhage			
Two operating theatre nurses ideally present in severe cases of postpartum haemorrhage	X	X	Person, Task
Obstetric team members (midwife, specialised midwife, and auxiliary nurse) should ideally prioritise remaining in operation theatre to assist	X	X	Person, Task
Fluid resuscitation and blood transfusion in operating theatre			
Clarification of whether lactated Ringer's solution is superior to sodium chloride	X	X	Tools + technology
Easy access to a blood heater and pressure bags to improve IV infusion in delivery room	X	X	Tools + technology
Clarification of which healthcare professional should ideally collect blood at the blood bank in an emergency situation	X	X	Person, Task
Clearly posted telephone number in operating theatre for the blood bank and its location	X	X	Tools + technology
Training/retraining of midwives in management of blood transfusions to allow them to assist the anaesthesia team correctly	X	X	Person, Task
Transfer of patient from delivery room to operating theatre and type of operating table			
Clarification of who is responsible for birthing bed (preparations for transport)		X	Person Tools & technology
Mechanisms to ensure clear passage (e.g. no beds, transport cages) along corridors for emergency transport of patients on delivery ward		X	Person, Tools & technology
Clarification of when and how to ideally transfer patients from delivery room to operating theatre; clarification of who determines timing of patient transport in emergency situations	X	X	Person, Organisation
Improved standards for where to keep the remote control for the operating table and for recharging its batteries; have easy-to-use instructions available		X	Tools & technology
Pros and cons of continuous use of birthing beds in emergency situations when patient transferred to operating theatre; address the knowledge gap on functionality of birthing beds among operating theatre staff; establish standards for their use		X	Tools & technology
Practical issues in operating theatre			
Better labelling of equipment in operating theatre and standardised placement of equipment in the two operating theatres for obstetric emergencies to streamline management	X	X	Tools & technology
Improve use of remote control to the operation table, which is difficult due to a lack of clarity about which direction the table moves		X	Person, Tools & technology
A more suitable walking distance between the cabinet with surgical caps and the operating theatre		X	Tools and technology

	Identified during		SEIPS components
	ISS	OSS	
Operating theatre nurses			
Clarification on role of responsibility for clinical decision making for urinary bladder catheter and/or shaving the pubic area, communication and actual management hereof; general agreement that obstetricians make the clinical decision and then inform the operating theatre nurse, who then places the catheter and/or shaves the area	X	X	Person, Organisation
Anaesthesia team in operating theatre			
More assistance from midwives when transferring patients from birthing bed to operating table and with positioning of patient; midwives requested more guidance from the anaesthesia team on the ideal way to do transfers	X	X	Person, Organisation
Improved procedures for checking equipment to ensure that it works (e.g. problem with no light in a laryngoscope)		X	Tools & technology, Organisation
More detailed introduction of new employees, including presentation of equipment for management of the difficult airway and equipment for blood heating and rapid infusion		X	Person
Observation charts and boards			
Improved observation charts for emergency situations, especially for postpartum haemorrhage	X	X	Tools & technology, Organisation
Greater use of white boards in delivery rooms in emergency situations for temporary observational charting	X	X	Tools & technology, Organisation
Use of white boards in operating theatre in emergency situations	X	X	Tools & technology, Organisation
Mode of anaesthesia in emergency situations			
Determine who makes final clinical decision about mode of anaesthesia; generally agreed to be the anaesthetist's responsibility	X	X	Person
Preoxygenation necessary prior to induction of spinal anaesthesia in case general anaesthesia is required	X	X	Task, Organisation
Clear communication on mode of anaesthesia to all staff in the room crucial so operating theatre nurses can prepare for e.g. sterile drapes, leg holders	X	X	Task, Organisation
Use, when feasible, obstetric manoeuvres like bimanual compression with severe postpartum haemorrhage and replacement of foetal head during cord prolapse with the parturient woman in side position (for attempt of spinal anaesthesia); communicate this during training/retraining of staff and address in clinical guidelines	X	X	Task, Organisation
Guidelines			
Greater clarity in postpartum haemorrhage guidelines on indications and general clinical management principles for blood product transfusion and risk of hypothermia	X	X	Task, Organisation
Addition of pointers in local clinical guidelines on how to choose the best team leader and this individual's role in emergency situations	X	X	Task, Organisation

ISS: in situ simulation; OSS: off site simulation; SEIPS: Systems Engineering Initiative for Patient Safety

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	1
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	7-8
	2b	Specific objectives or hypotheses	8
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	8
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	None
Participants	4a	Eligibility criteria for participants	8
	4b	Settings and locations where the data were collected	8
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	8-9
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	9-10 Fig.1
	6b	Any changes to trial outcomes after the trial commenced, with reasons	None
Sample size	7a	How sample size was determined	10-11
	7b	When applicable, explanation of any interim analyses and stopping guidelines	None
Randomisation:			11
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	11
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	11
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	11
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	None

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	8-9
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	11-12
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	11-12
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	12 Fig.2 Table 3
	13b	For each group, losses and exclusions after randomisation, together with reasons	12 Fig.2 Table 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	12
	14b	Why the trial ended or was stopped	Not done
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 2
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Fig 2 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Page 13-14 Table 3,4,5,6,7,8
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	None
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	Page 13-14 Table 4,5,6,7,8 Online table
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	None
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	17-18
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	14-15, 17
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	14-18
Other information			
Registration	23	Registration number and name of trial registry	NCT01792674
Protocol	24	Where the full trial protocol can be accessed, if available http://www.rigshospitalet.dk/NR/rdonlyres/1F6BDE2B-CA37-4017-8A42-CF94F7FB31AE/0/SIM5_Engelsk_protokol_Insitu_versus_offsite_CTU_obanopsimulation_2013_02_153.pdf	Ref 36 protocol article

Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	20

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

Simulation based multi-professional obstetric anaesthesia training conducted in situ versus off site leads to similar individual and team outcomes: a randomised educational trial

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2015-008344.R1
Article Type:	Research
Date Submitted by the Author:	19-Jul-2015
Complete List of Authors:	Sorensen, Jette Led; Rigshospitalet, University of Copenhagen, Juliane Marie Centre for Children, Women and Reproduction Van der Vleuten, Cees; Maastricht University, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences Rosthøj, Susanne; University of Copenhagen, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences Østergaard, Doris; Capital Region of Denmark and University of Copenhagen, Danish Institute for Medical Simulation, Herlev Hospital Leblanc, Vicki; University of Ottawa, University of Ottawa Skills and Simulation Centre, The Ottawa Hospital Johansen, Marianne; Rigshospitalet, University of Copenhagen, Department of Obstetrics, Juliane Marie Centre for Children, Women and Reproduction Ekelund, Kim; Rigshospitalet, University of Copenhagen, Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction Starkopf, Liis; University of Copenhagen, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences Lindschou, Jane; Rigshospitalet, University of Copenhagen, Copenhagen Trial Unit, Centre for Clinical Intervention Research Gluud, Christian; Rigshospitalet, Copenhagen University Hospital, The Copenhagen Trial Unit, Centre for Clinical Intervention Research Weikop, Pia; Rigshospitalet, University of Copenhagen, Laboratory of Neuropsychiatry Ottesen, Bent; Rigshospitalet, University of Copenhagen, Juliane Marie Centre for Children, Women and Reproduction
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Obstetrics and gynaecology, Anaesthesia, Emergency medicine
Keywords:	OBSTETRICS, MEDICAL EDUCATION & TRAINING, Anaesthesia in obstetrics < ANAESTHETICS, patient simulation, in situ simulation, interprofessional

Title page

Title

Simulation based multi-professional obstetric anaesthesia training conducted in situ versus off site leads to similar individual and team outcomes: a randomised educational trial

Corresponding author

Jette Led Sørensen, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Blegdamsvej 9, 2100 Copenhagen, Denmark.

jette.led.soerensen@regionh.dk

Telephone +45 35458901 Mobile phone +45 26711484

Authors name

Jette Led Sørensen, consultant obstetrician, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Cees Van der Vleuten, professor, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, 6200 Maastricht, Netherlands.

Susanne Rosthøj, PhD, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, 1014 Copenhagen, Denmark.

Doris Østergaard, professor, Danish Institute for Medical Simulation, Herlev Hospital, Capital Region of Denmark and University of Copenhagen, 2730 Herlev, Denmark.

Vicki LeBlanc, director, University of Ottawa Skills and Simulation Centre, The Ottawa Hospital, University of Ottawa, K1Y 4E9 Ottawa, Canada.

Marianne Johansen, consultant obstetrician, Department of Obstetrics, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Kim Ekelund, consultant anaesthetist, Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Liis Starkopf, computing assistant, Section of Biostatistics, Institute of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, 1014 Copenhagen, Denmark.

Jane Lindschou, clinical research consultant, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.

Christian Gluud, head of department, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.
Pia Weikop, senior scientist Laboratory of Neuropsychiatry, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark.
Bent Ottesen, professor, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, 2100 Copenhagen, Denmark

Keywords Patient simulation, in situ simulation, emergencies, interprofessional, stress

Word count abstract: 297

Word count main text (excluding title page, abstract, references, figures and tables): 4,219
Figures: 2
Tables: 8
Supplementary material (online): 1 table

ABSTRACT

Objective

To investigate the effect of in situ simulation (ISS) versus off site simulation (OSS) on knowledge, patient safety attitude, stress, motivation, perceptions of simulation, team performance and organisational impact.

Design

Investigator-initiated single-centre randomised superiority educational trial.

Setting

Obstetrics and anaesthesiology departments, Rigshospitalet, University of Copenhagen, Denmark.

Participants

One-hundred participants in teams of ten, comprising midwives, specialised midwives, auxiliary nurses, nurse anaesthetists, operating theatre nurses, and consultant doctors and trainees in obstetrics and anaesthesiology.

Interventions

Two multi-professional simulations (clinical management of an emergency caesarean section and a postpartum haemorrhage scenario) were conducted in teams of ten in the ISS versus the OSS settings.

Primary outcome

Knowledge assessed by a multiple choice question test.

Exploratory outcomes

Individual outcomes: scores on the Safety Attitudes Questionnaire, stress measurements (State-Trait Anxiety Inventory, cognitive appraisal and salivary cortisol), Intrinsic Motivation Inventory and perceptions of simulations. Team outcome: video assessment of team performance. Organisational impact: suggestions for organisational changes.

Results

The trial was conducted from April to June 2013. No differences between the two groups were found for the multiple choice question test, patient safety attitude, stress measurements, motivation, or the evaluation of the simulations. The participants in the ISS group scored the authenticity of the simulation significantly higher than the participants in the OSS group did. Expert video assessment of team performance showed no differences between the ISS versus the OSS groups. The ISS group provided more ideas and suggestions for changes at the organisational level.

Conclusion

In this randomised trial no significant differences were found regarding knowledge, patient safety attitude, or stress measurements when comparing ISS versus OSS. Although participant perception of the authenticity of ISS versus OSS differed significantly, there were no differences in other outcomes between the groups except that the ISS group generated more suggestions for organisational changes.

Trial registration

ClincialTrials.gov NCT01792674

Article summary

Strengths and limitation

- To our knowledge this is the first randomised trial conducted to assess the effects of two different simulation settings, in situ simulation versus off site simulation, on a broad variety of outcomes.
- Previous non-randomised studies have recommended in situ simulation. But in this randomised trial no significant differences regarding knowledge, patient safety attitude, stress measurements or team performance were found when comparing in situ simulation versus off site simulation. The participants in the in situ group scored the authenticity of the simulation significantly higher than the participants in the off site simulation group did. However, this perception did not influence the individual and team outcomes. On the outcome on organisational level the in situ group generated more suggestions for organisational changes.
- A strength of this trial is the involvement of authentic teams that mirrored teams in real life that resembles the real clinical setting in every possible way. This was important for the so-called sociological fidelity.
- A limitation of the trial is the fact that the outcomes were based only on immediate measurements of knowledge level and of team performance. Only perceptions of simulation were measured after one week (evaluation and motivation) and safety attitudes after one month. No clinical outcome was measured.

INTRODUCTION

Frequently recommended as a learning modality,[1-5] simulation-based medical education is described as "devices, trained persons, lifelike virtual environments and contrived social situations that mimic problems, events, or conditions that arise in professional encounters".[5] Its key elements, however, remain to be studied in depth in order to improve simulation-based medical education. One potential aspect that may influence the effect of this kind of education is the level of fidelity, or authenticity in more layman's terms. Fidelity is traditionally described to be assessed on two levels: 1) engineering fidelity, i.e. does the simulation look realistic, 2) psychological fidelity, i.e. does the simulator contain the critical elements to accurately simulate the behaviours required to complete a task.[6,7]

Simulation-based medical education has traditionally been conducted as off site simulation (OSS), either at a simulation centre or in facilities in the hospital set up for the purpose of simulation. Recently, in situ simulation (ISS) has been introduced and described as "a team based simulation strategy that occurs on the actual patient care units involving actual healthcare team members within their own working environment".[8-12] An unanswered question is whether ISS is superior to OSS. It has been argued that ISS has more fidelity compared to OSS and that ISS can lead to better teaching and a greater organisational impact.[8-14]

We hypothesised that the physical setting could influence fidelity, and hence that ISS could be more effective for educational purposes. To our knowledge, no randomised educational trials have been conducted comparing the ISS versus the OSS setting. Two articles that do use randomisation and compare ISS with OSS focus on frequency of training and not setting,[15] nor did they include a relevant control group.[16] Previous studies have been criticised for having small sample sizes, weak study designs and a lack of meaningful evaluations of the effectiveness of the programmes.[8] A recent retrospective video-based study showed that the performance were similar in all the tested simulation settings, but the participants favoured the ISS and the authors argued that prospective studies are needed.[17]

Human factors such as stress and motivation impact learning,[18-26] which is why we set out to investigate how stress and motivation were affected by ISS versus OSS. We anticipated that the participants would experience ISS as more demanding and as creating higher levels of stress and

motivation, which might enhance their learning. Furthermore, we hypothesised that ISS might provide the investigators with more information on changes needed in the organisation to improve quality of care.

In this trial, we wanted to apply simulation-based medical education in the field of obstetrics, as delivery wards are challenging work places, where patient safety is high on the agenda and unexpected emergencies occur.[27,27-34] Simulation-based medical education is thus argued to be an essential learning strategy for labour wards.[4,35] The objective of this randomised educational trial was to investigate the effect of ISS versus OSS on knowledge, patient safety attitude, stress, motivation, perception of the simulation, team performance and organisational impact among multi-professional obstetric anaesthesia teams.

METHODS

Design

An investigator-initiated, single-centre randomised superiority educational trial previously described in a design article.[36]

Setting and participants

The setting was the Department of Obstetrics and the Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, which has approximately 6,300 deliveries per year. Participants were healthcare professionals who worked in shifts on the labour ward: consultant and trainee doctors in obstetrics and anaesthesiology, midwives, specialised midwives, auxiliary nurses, nurse anaesthetists and operating theatre nurses. Participants gave written informed consent. Exclusion criteria were lack of informed consent, employees with managerial and staff responsibilities, staff members involved in the design of the trial and employees who did not work in shifts.[36]

Recruitment of participants

Eligible participants were provided with information via meetings, a website and personal letters but additional verbal and written information could also be obtained from the principal investigator (JLS). Informed written consent was obtained if people decided to participate in the trial.[36]

Interventions

The experimental intervention was a pre-announced ISS,[8,9] i.e. simulation-based medical education in the delivery room and operating theatre. The control intervention was an OSS, which took place in hospital rooms set up for the occasion but away from the patient care unit.

An appointed working committee consisting of representatives from all the healthcare professionals participating in the trial developed its aims and objectives, and they designed simulated scenarios for the ISS and the OSS.[36] The two simulation scenarios were: 1) management of an emergency caesarean section after a cord prolapse; and 2) a postpartum haemorrhage including surgical procedures to evacuate the uterus. Focusing mainly on inter-professional skills and communication, the scenarios gave each healthcare profession a significant role to play.[37]

All participants recruited for a training day were told to arrive at a specific time dressed in work clothes, but had not been told what kind of simulation they were randomised to. The OSS room that was to function as the delivery room was in the doctors' on-call room, which was small compared to the usual delivery room. A roller table prepared with the usual labour ward equipment had been placed in the room. The OSS room that was to function as the operating theatre was set up in the corner of a lecture hall. An anaesthetic trolley with the usual equipment was placed in the room and equipment for the operating theatre nurses was placed on a roller table. An introductory presentation was given to all participants on how the simulation was organised and then the participants recruited for OSS were shown the fictitious delivery room and fictitious operating theatre.

In the first part of the simulation in the delivery room, someone who has been instructed in role playing acted as the patient in both the ISS and the OSS setting. In both the real and the fictitious operating theatre a full-body birthing simulator, a SimMom, was used for parts of the simulation scenario.[38] Recruited from the working committee, the instructors conducting the simulations were trained in facilitating simulations and doing debriefings. The working committee were trained in local organised courses and attended a British National train the trainers course: PROMPT (PRactical Obstetric Multi-Professional Training).[39] They worked in groups of two comprising either a consultant obstetrician with a nurse anaesthetist or a consultant anaesthetist with a midwife. The debriefings lasted 50 to 60 minutes and comprised three phases: description, analysis and

application.[40] In addition to the simulation-based medical education, the training day also included video-based, case-based,[41] and lecture-based teaching sessions.

Primary outcome

The primary outcome was the results from a knowledge test based on a 40-item multiple choice question (MCQ) test developed specifically for this trial.[42] The choice of a knowledge test as the primary outcome was mainly a pragmatic choice. MCQ testing is feasible for testing many participants in a relatively short time and at a low cost.[43] Furthermore previously used knowledge tests could be used for inspiration and for sample size calculation [44,45]. The participants completed the MCQ test at the beginning and at the end of the training day. They were asked not to discuss the MCQ test with other participants or instructors during the training day.

Exploratory outcomes

The Safety Attitudes Questionnaire (SAQ) is validated in a Danish context.[46] It included 33 items covering five dimensions: 1) team work climate; 2) safety climate; 3) job satisfaction; 4) stress recognition; and 5) work conditions.[47,48] The participants did the SAQ one month prior to and one month after participating in the training day.

Stress: Salivary cortisol levels were used as an objective measure of physiological stress. [36] The salivary cortisol samples were obtained as a baseline before the first and the second simulation and at three additional times during the two simulations (figure 1). The subjective stress level was measured using the Stress-Trait Anxiety Inventory (STAI) and cognitive appraisal (CA) (figure 1).[21,23,49,50]

Intrinsic Motivation Inventory (IMI) included 22 items with four dimensions: 1) interest/ enjoyment; 2) perceived competence; 3) perceived choice; and 4) pressure or tension (reversed scale).[51]

Evaluation questionnaire: Together with the IMI, each participant received an evaluation questionnaire at the end of the training day and they were asked to return it within a week.[36]

Team performance was video recorded and assessed by experts using a Team Emergency Assessment Measure (TEAM).[36,52,53] The TEAM scale was used in the original version in English and supplemented with a translated Danish version. The scoring of team performance was done by two consultant anaesthetists and two consultant obstetricians from outside the trial hospital. All four video assessors jointly attended two three-hour training sessions on video rating but assessment of the trial videos was conducted individually. Each video-assessor received an external hard disc with 20 simulated scenarios in random order of teams and scenarios of respectively management of an emergency caesarean section and a postpartum haemorrhage.

Organisational outcomes were registered using: 1) two open-ended questions included in the evaluation questionnaire on suggestions for organisational changes; and 2) debriefing and evaluation at the end of the training day, where participants reported ideas for organisational changes. The principal investigator (JLS) took notes during these sessions, which were then discussed in the previously mentioned working committee, which included authors MJ and KE.

Sample size calculation

We chose data from knowledge tests from previous studies to conduct our sample size estimation.[44,45] We assumed the distribution of the primary outcome (the percentage of correct MCQs answers) to be normally distributed with a standard deviation of 24%. If a difference in the percentage of correct MCQ answers between the two groups (ISS and OSS) was 17%, then 64 participants had to be included to be able to reject the null hypothesis with a power of 80%. As the interventions were delivered in teams (clusters), observations from the same team were likely to be correlated.[54,55] The reduction in effective sample size depends on the cluster correlation coefficient, which is why the crude sample size had to be multiplied by a design effect. With a design effect of 0.05 the minimum sample size was increased to 92.8 participants.[55] We therefore decided to include a total of 100 participants.

Randomisation and blinding

Randomisation was performed by the Copenhagen Trial Unit using a computer-generated allocation sequence concealed to the investigators. The randomisation was conducted in two steps. First, the participants were individually randomised 1:1 to the ISS versus the OSS group. The allocation sequence consisted of nine strata, one for each healthcare professional group. Each stratum was composed of one or two permuted blocks with the size of ten. Second, the participants in each

group were then randomised into one of five teams for the ISS and OSS settings using simple randomisation that took into account the days they were available for training.

Questionnaire data were transferred from the paper versions and coded by independent data managers. The intervention was not blinded for the participants, instructors providing the educational intervention, the video assessors or the investigators drawing the conclusions. The data managers and statisticians were blinded to the allocated intervention groups.

Data analysis and statistical methods

Due to the low number of missing values, no missing data techniques were applied. Single missing items in the Multiple Choice Question (MCQ) test or more than one answer to an MCQ item were treated as incorrect answers. Single missing items in inventories as Safety Attitudes Questionnaire (SAQ), Intrinsic Motivation Inventory (IMI) Stress-Trait Anxiety Inventory (STAI) were excluded from the calculation of the summary scores.

Calculation of 95% confidence intervals (CI) obtained after the simulation intervention (post MCQ, post-SAQ, stress measurements, IMI) was based on generalised estimating equations (GEE)[56] since observations from individuals on the same team were potentially correlated.

The evaluation data measured on a Likert scale were analysed as comparisons of location of the ordinal responses from items in the evaluation questionnaire performed by the Kruskal-Wallis rank sum test, and the P-values were adjusted for multiple testing using the Benjamini-Hochberg method.[57]

The mean outcomes obtained after the simulation intervention (post measurements) in the two intervention groups were compared by a linear model including intervention and baseline (pre measurements) as explanatory variables (Analysis of covariance (ANCOVA)), and inferences were based on GEE to account for the potential correlation within teams. To assess whether there was a difference in mean between pre and post measurements in each of the intervention groups, overall tests of whether the intercept equals 0 and the slope equals 1 from a linear model of the post measurements on the pre measurements were performed.

The team data, i.e. the ratings from the four assessors, were analysed using linear mixed models to take into account the repeated measurements on the teams by the same assessors. Random effects for each team nested in the randomisation group and in each assessor were included. A model including the interaction between randomisation group and simulation was used to estimate means, whereas an additive model was used to determine the overall difference in mean between the ISS versus the OSS intervention and the first (emergency caesarean section) and the second (postpartum haemorrhage) simulation (no interaction between randomisation and simulations was found).

Ideas for organisational changes were registered by participants and the reported suggestions were categorised as qualitative data and analysed using part of the framework from the Systems Engineering Initiative for Patient Safety model.[58]

SAS version 9.2, R version 3.0.2 and IBM SPSS Statistics 20 were used for statistical analysis. Two-sided P-values <0.05 were considered significant.

RESULTS

Recruitment, basic characteristics and follow-up of participants

Informed written consent for participation in the trial was provided by 116 healthcare professionals. The two randomised intervention groups were comparable (table 1).

Table 1. Baseline characteristics of participants in the ISS and OSS groups (n=100)

	ISS group	OSS group
Number of participants	48[1]	49[2]
Number of females/males	42/6	43/6
Median age (range)	44.5 (26-63)	42 (27-65)
Median years of obstetric work experiences (range)	7 (0.6-38)	7 (0.6-39)
Previous simulation experiences[3]:		
No experience	8	10
Simple simulation	25	24
Full-scale simulation	15	15
Pregnant participants	2	2
Participants on any kind of medication	19	20
Participants on medication with no expected influence on cortisol measurement[4]	12	9
Participants on medication with potential influence on cortisol measurement	7	11
Intranasal and inhaled corticosteroids (mometasone furoate, budesonide/formoterol, budesonide, fluticasone/salmeterol)	2	3
Levothyroxine	1	2
Metformin	1	1
Norethisterone/estradiolacetate	0	1
Oral contraceptives	1	3
Beta blockers (metoprolol)	0	1
Antidepressants (nortriptyline, fluoxetine)	2	0

ISS: in situ simulation; OSS: off site simulation

[1] Not included due to illness: A consultant obstetrician and an operating room nurse (n=2).
[2] Not included due to illness: An auxiliary nurse (n=1).
[3] A simple simulation experience is, for example skills training using a low-tech delivery mannequin and no video recording of the simulation scenario. Full-scale simulation is for example done in teams with fully interactive mannequins and video recorded scenarios.
[4] Intrauterine contraceptive devices, angiotensin II receptor antagonists, angiotensin-converting-enzyme inhibitors, simvastatin, alendronate, pantoprazole, antihistamine and tinzaparine.

The flow of participants is described in figure 2 and in table 2.

Table 2. Reasons for lost to follow-up (n/100 randomised participants (%))

	ISS group	OSS group
Pre MCQ test	2 (2%)[1]	1 (1%)[1]
Post MCQ test	2 (2%)[1]	1 (1%)[1]
Salivary cortisol level at emergency caesarean section simulation	2 (2%)[1]	3 (3%)[1,2]
Salivary cortisol level at postpartum haemorrhage simulation	2 (2%)[1]	2 (2%)[1,3]
STAI at emergency caesarean section simulation	2 (2%)[1]	1 (1%)[1]
STAI at postpartum haemorrhage simulation	2 (2%)[1]	2 (2%)[1,3]
CA at caesarean section simulation	2 (2%)[1]	1 (1%)[1]
CA at postpartum haemorrhage simulation	2 (2%)[1]	2 (2%)[1,3]
Evaluation questionnaire	3 (3%)[1,4]	1 (1%)[1]
IMI	4 (4%)[1,5]	1 (1%)[1]
Pre SAQ	1 (1%)[7]	4 (4%)[6,1]
Post SAQ	5 (5%)[1,8]	4 (4%)[1,8]

CA: cognitive appraisal; IMI: Intrinsic Motivation Inventory; MCQ: multiple choice question; SAQ: Safety Attitudes Questionnaire; STAI: Stress-Trait Anxiety Inventory

[1)] Participants ill and did not participate (n=3).
[2)] Two measurements were clear outliers. A re-evaluation of the data collection indicated that the two samples had most likely been swapped between two participants, which is why these measurements were excluded from all analyses (n=2).
[3)] Because one participant was temporarily called away for clinical work, the cortisol measurement after the simulation in postpartum haemorrhage is lacking and he was unable to answer parts of the questionnaires (n=1).
[4)] Questionnaires not returned (n=1).
[5)] Questionnaires not returned (n=2).
[6)] For three participants pre SAQ data were excluded because these participants were employed in other departments prior to participating in the training days, hence their responses did not refer to the department in question (n=3).
[7)] Of the individuals who did not participate due to illness (n=3), one filled out the pre SAQ anyhow.
[8)] Questionnaires not returned (n=6).

Intervention delivery

The trial was conducted from April to June 2013. Out of 100 participants included, 97 participated (table 1 and 2, figure 2). The ten simulations were conducted as planned, although one ISS had to be postponed for 15 minutes due to an ongoing, real emergency caesarean section. The mean number of minutes spent on the caesarean section simulation in ISS and OSS was 18 and 15 minutes, respectively (P=0.70), while the mean for the postpartum haemorrhage simulation was 26 and 24 minutes, respectively (P=0.40).

Primary outcome

Multiple Choice Question (MCQ) test: There was no difference in mean post MCQ scores between the ISS versus the OSS group adjusted for the pre MCQ scores (table 3). Additional analyses based on the MCQ test, including 33 or 29 of the 40 items, gave similar results (data not shown). These additional analyses were performed because validation of the MCQ test revealed that seven to eleven of 40 MCQ items were disputable.[41]

Post hoc analysis: The average increase in percentage of correct answers in the MCQ test following training was 13.1% (95% CI, 11.0% to 15.3%) in the ISS group and 12.7% (95% CI, 10.3% to 15.2%) in the OSS group (overall tests of no difference between pre and post MCQ: both $p<0.0001$).

Table 3. Means (95% CI) of percentages of correct answers in the MCQ test before (pre MCQ) and after (post MCQ) in the ISS and OSS groups

Descriptive statistics				
MCQ test % correct	Simulation intervention	Pre MCQ mean[1] (start of training day)	Post MCQ mean[1] (end of training day)	Mean difference [1,2]
	ISS	69.4 (65.4 to 73.4)	82.6 (79.3 to 85.8)	-0.02 (-2.13 to 2.09)
	OSS	70.6 (66.0 to 75.2)	83.3 (80.4 to 86.1)	
				P = 0.98

CI: confidence interval; ISS: in situ simulation; MCQ: multiple choice question (range: 0-100%); OSS: off site simulation

[1)] Based on generalised estimating equations to account for potential correlation within teams.

[2)] Adjusted for pre MCQ (ANCOVA).

Exploratory outcomes

Safety Attitudes Questionnaire (SAQ): No differences were found in the ISS versus OSS groups for any of the post-SAQ dimensions (table 4).

Table 4. Means (95% CI) of SAQ (converted to percentages) for five dimensions one month before (pre SAQ) and one month after (post SAQ) the simulation training day with ISS and OSS

		Descriptive statistics		
	Simulation intervention	Pre SAQ mean (1 month before)	Post SAQ mean[1] (1 month after)	Mean difference[1,2]
SAQ Team work Climate	ISS	80.5 (76.7 to 84.3)	81.1 (76.7 to 85.5)	-1.38 (-5.8 to 3.05)
	OSS	78.4 (74.1 to 82.2)	81.2 (77.5 to 85.0)	
				P = 0.54
SAQ Safety Climate	ISS	66.7 (61.8 to 71.6)	70.6 (65.9 to 75.2)	1.57 (-2.0 to 5.1)
	OSS	69.2 (65.4 to 73.0)	70.8 (66.8 to 74.8)	
				P = 0.39
SAQ Job Satisfaction	ISS	86.4 (82.9 to 89.8)	87.5 (83.3 to 91.7)	0.6 (-2.9 to 4.1)
	OSS	85.6 (81.6 to 89.6)	85.7 (81.9 to 89.5)	
				P = 0.74
SAQ Stress recognition	ISS	69.7 (63.5 to 76.0)	68.8 (62.4 to 75.1)	-2.6 (-9.2 to 4.0)
	OSS	67.3 (61.2 to 73.3)	69.2 (64.0 to 74.4)	
				P = 0.44
SAQ Work condition	ISS	66.4 (60.8 to 72.1)	64.9 (59.0 to 70.8)	-0.32 (-5.7 to 5.1)
	OSS	65.9 (59.9 to 71.8)	64.0 (58.1 to 69.8)	
				P = 0.91

CI: confidence intervals; ISS: in situ simulation; OSS: off site simulation; SAQ: Safety Attitudes Questionnaire (range: 0-100%).

[1)] Based on generalised estimating equations to account for potential correlation within teams.
[2)] Adjusted for pre SAQ (ANCOVA).

Salivary cortisol, Stress-Trait Anxiety Inventory (STAI) and cognitive appraisal (CA): The mean change in baseline to peak was similar for ISS versus OSS for both the first (caesarean section) and the second (postpartum haemorrhage) simulation (table 5).

Post hoc analysis: The salivary cortisol and STAI levels increased significantly from baseline to peak in the ISS and OSS groups following both the first (caesarean section) and the second

(postpartum haemorrhage) simulation (overall tests for no difference between pre and post: all p<0.0001). CA decreased significantly from baseline to peak in the ISS and OSS settings in both the caesarean section and in the postpartum haemorrhage simulations (p<0.0001).

Table 5. Mean (95% CI) of salivary cortisol (nmol/L), STAI and CA during simulation in management of an emergency caesarean section and postpartum haemorrhage conducted as ISS and OSS

		Baseline	Post-simulation 0 min Mean[1]	Post-simulation 5 min Mean[1]	Post-simulation 10 min Mean[1]	Peak-level mean[1,2]	Mean difference of baseline to peak of ΔOSS vs. ΔISS[1,3]
1st simulation: Emergency caesarean section							
Corti-sol	ISS	7.0 (6.3 to 7.8)	8.9 (7.2 to 10.6)	8.1 (6.6 to 9.6)	8.1 (6.6 to 9.5)	9.3 (7.6 to 11.0)	-0.5 (-1.6 to 2.5) P = 0.64
	OSS	7.3 (5.3 to 9.2)	8.2 (6.3 to 10.2)	7.8 (6.1 to 9.6)	8.0 (6.2 to 9.8)	9.0 (6.9 to 11.1)	
STAI	ISS	32.2 (30.4 to 34.0)	34.8 (32.7 to 37.0)		31.3 (29.5 to 33.1)	36.5 (34.3 to 38.7)	-0.22 (-2.1 to 2.5) P = 0.85
	OSS	33.1 (31.1 to 35.0)	34.8 (32.2 to 37.3)		30.7 (29.0 to 32.4)	37.0 (34.7 to 39.3)	
CA	ISS	1.0 (0.9 to 1.1)	0.8 (0.7 to 1.0)		0.8 (0.7 to 0.9)	0.8 (0.7 to 1.0.)	0.0 (-0.2 to 0.2) P = 0.93
	OSS	1.0 (1.0 to 1.1)	0.8 (0.7 to 0.9)		0.8 (0.6 to 0.9)	0.9 (0.7 to 0.9)	
2nd simulation: Postpartum haemorrhage							
Corti-sol	ISS	7.4 (6.5 to 8.3)	9.2 (7.7 to 10.7)	7.7 (6.6 to 8.8)	7.4 (6.3 to 8.5)	9.4 (7.9 to 10.9)	-1.2 (-0.1to 0.2.5) P = 0.07
	OSS	6.9 (5.9 to 7.9)	7.5 (6.6 to 8.4)	6.7 (5.8 to 7.7)	6.8 (6.0 to 7.6)	7.7 (6.7 to 8.7)	
STAI	ISS	31.8 (30.0 to 33.6)	31.8 (30.1 to 33.6)		28.5 (27.3 to 29.7)	32.2 (30.5 to 33.9)	-0.5 (-2.2 to1.3) P = 0.61
	OSS	32.1 (29.9 to 34.2)	32.4 (30.5 to 34.3)		30.1 (28.5 to 31.8)	32.8 (31.0 to 34.7)	
CA	ISS	1.0 (0.9 to 1.1)	0.8 (0.7 to 0.9)		0.8 (0.7 to 0.9)	0.8 (0.7 to 0.9)	0.1 (-0.2 to 0.1) P = 0.56
	OSS	1.1 (1.0 to 1.2)	0.9 (0.7 to 1.0)		0.8 (0.7 to 0.9)	0.9 (0.7 to 1.0)	

CA: cognitive appraisal (range 0.1-10); CI: Confidence interval; ISS: in situ simulation; OSS: off site simulation; STAI: Stress-Trait Anxiety Inventory (range 20-80).

1) Based on generalised estimating equations to account for potential correlation within teams.
2) Peak level is the maximum of the measurements obtained at 0, 5 and 10 minutes after the end of the simulation.
3) Adjusted for pre cortisol, pre STAI and pre CA (ANCOVA).

Intrinsic Motivation Inventory (IMI): No differences were found in the ISS versus the OSS groups for the IMI score (table 6).

Table 6. Mean (95% CI) motivation after participation in either ISS or OSS. Analysis comprised a comparison of the mean IMI and the mean of the ISS and OSS groups

	Simulation intervention	IMI mean (1 week after) [1]
Interest/Enjoyment	ISS	5.2 (4.9 to 5.5)
	OSS	5.3 (5.1 to 5.5)
		P = 0.72
Perceived competence	ISS	5.1 (4.8 to 5.4)
	OSS	4.9 (4.7 to 5.1)
		P = 0.24
Perceived choice	ISS	5.8 (5.6 to 6.1)
	OSS	5.5 (5.2 to 5.9)
		P = 0.15
Pressure tension (reversed)	ISS	2.8 (2.5 to 3.1)
	OSS	2.9 (2.6 to 3.3)
		P = 0.65

CI: confidence interval; IMI: Intrinsic Motivation Inventory (range:1-7); ISS: in situ simulation; OSS: off site simulation.

[1)] Based on generalised estimating equations to account for potential correlation within teams.

Participant evaluations and perception: For almost all 20 questions in the evaluation questionnaire, the ISS and OSS groups did not differ significantly. However, the two questions addressing the authenticity fidelity of the simulations were scored significantly higher by the ISS participants compared with the OSS participants (table 7).

Table 7. Participant evaluations after participation in either ISS or OSS in medians with 25% and 75% quartiles. Analysis comprised a comparison of the evaluation medians of the ISS versus OSS group

	ISS	OSS	
Evaluation questions (shortened version, original version in Danish)	**Median (1st Q–3rd Q)**	**Median (1st Q–3rd Q)**	**P[1]**
1. Over all the training day was (1=very bad to 5=very good)	5 (4-5)	5 (4-5)	0.70
2. Multi-professional approach with all healthcare groups involved was (1=very bad to 5=very good)	5 (4-5)	5 (4-5)	0.70
3. I thought the level of education of the training was (1=very much over my level to 5=very much below my level)	3 (3-3)	3 (3-3)	0.70
4. Will recommend others to participate (1=never to 5=always)	5 (5-5)	5 (4-5)	0.70
5. Did simulations inspire you to change procedures or practical issues in the labour room or operating theatre (1=no ideas to 5=many ideas) (included open-ended questions)	3 (2-3)	3 (2-4)	0.70
6. Did simulations inspire you to change guidelines (1=no ideas to 5=many ideas) (included open-ended questions)	2 (1-2)	2 (1-2)	0.70
Simulation of an emergency CS			
7. Over all my learning was (1=very bad to 5=very good)	4 (3-4)	4 (3-4)	0.90
8. The authenticity of the CS simulation was (1=not at all authentic to 5=very authentic)	4 (3-4)	3 (3-4)	0.02
9. The authenticity of the CS simulation influenced my learning (1=not at all important to 5=very important)	4 (4-4.5)	4 (4-4)	0.65
10. Collaboration in the CS team was (1=very bad to 5=very good)	4 (4-4.5)	4 (3.8-4)	0.27
11. Communication in the CS team was (1=very bad to 5=very good)	4 (3-4)	4 (3-4)	0.23
12. The CS team leader was (1=very bad to 5=very good)	4 (3-4)	4 (3-4)	0.26
13. My learning at the debriefing after the CS was (1=very bad to 5=very good)	4 (4-5)	4 (4-4)	0.88
Simulation in PPH			
14. My learning overall was (1=very bad to 5=very good)	4 (4-4)	4 (4-4)	0.70
15. The authenticity of the PPH simulation was (1=not a tall authentic to 5=very authentic)	4 (3-4)	3 (3-4)	0.01
16. The authenticity of the simulation in PPH influenced my learning (1=not at all important to 5=very important)	4 (4-4.5)	4 (4-4)	0.23
17. Collaboration in the PPH team was (1=very bad to 5=very good)	4 (4-4.5)	4 (4-4)	0.64
18. Communication in the PPH team was (1=very bad to 5=very good)	4 (3.5-4)	4 (3-4)	0.64
19. The PPH team leader was (1=very bad to 5=very good)	4 (4-4)	4 (3-4)	0.23
20. My learning at the debriefing after the PPH was (1=very bad to 5=very good)	4 (4-4)	4 (4-4)	0.57

CS: caesarean section; ISS: in situ simulation; OSS: off site simulation; 1st Q–3rd Q: 25% and 75% quartiles; PPH: postpartum haemorrhage.

[1] Kruskal-Wallis rank sum test. P-values adjusted for multiple testing.

Team Emergency Assessment Measure (TEAM): No significant differences were found in the team scoring of performance between the ISS versus the OSS groups (table 8).

TEAM post hoc analysis: A significant increase was found in the team scoring of performance from the first simulation (emergency caesarean section) to the second (postpartum haemorrhage) (table 8).

Table 8. Mean (95% CI) of video assessment performance scores with the TEAM scale. Four consultants recruited outside the research hospital did the video assessment scoring. Analysis comprised a comparison of the mean TEAM score of the ISS versus the OSS group

	ISS	OSS	
Video assessment scoring of performance	Mean	Mean	P[1]
TEAM (means of item rating) Simulation in emergency CS[1]	2.6 (2.3 to 3.0)	2.4 (2.1 to 2.8)	
TEAM (means of item rating) Simulation in PPH[1]	2.9 (2.5 to 3.2)	2.8 (2.5 to 3.2)	
Estimated overall difference in mean between ISS and OSS[2]	0.1 (-0.2 to 0.5)		0.36
TEAM (global rating) Simulation in emergency CS[1]	6.1 (4.8 to 7.3)	5.3 (4.0 to 6.5)	
TEAM (global rating) Simulation in PPH[1]	6.8 (5.5 to 8.1)	6.3 (5.0 to 7.6)	
Estimated overall difference in mean between ISS and OSS[2]	0.7 (-0.4 to 1.7)		0.18
Differences in video assessment scores of performance between emergency CS (1st) and PPH (2nd) simulation scenarios			
Differences in mean of TEAM (means of item rating) of the simulation in emergency CS versus PPH[2]	0.3 (0.1 to 0.5)		0.0003
Differences in mean of TEAM (global rating) of the simulation in emergency CS versus PPH[2]	0.9 (0.3 to1.5)		0.005

CI: confidence interval; CS: caesarean section; ISS: in situ simulation; OSS: off site simulation; PPH: postpartum haemorrhage; Q: quartile; TEAM: Team Emergency Assessment Measure (range for item rating: 0-4; range for global rating:1-10);

[1)] Means found from a linear mixed model including an interaction between simulation group (ISS and OSS) and simulation scenario (emergency CS and PPH).
[2)] Overall difference in means found from an additive linear mixed model based on simulation group and simulation scenario.

Organisational changes: A qualitative analysis showed that more ideas for organisational changes were suggested by ISS participants than OSS participants. For details see supplementary table S1. The quantitative analysis, however, showed that participants in the ISS and the OSS groups scored equally concerning whether the simulations inspired making changes in procedures or guidelines (table 7, questions 5 and 6).

DISCUSSION

In this randomised trial, we did not find that simulation-based medical education conducted as ISS compared with OSS led to different outcomes assessed on knowledge, patient safety attitude, stress, motivation, perceptions of the simulations and team performance. Participant perception of the authenticity of the ISS and OSS differed significantly, but this had no influence on other individual or team outcomes. We observed that ISS participants provided more ideas for organisational changes than OSS participants did. This is in accordance with several non-randomised studies describing a positive impact of ISS on the organisation.[8,10,11,13,59-61]

In the evaluation questionnaire (table 7), participants were asked about their perceptions of the authenticity of the simulations, which can be interpreted as their perception of the simulation's fidelity. The participants scored the authenticity to be significantly higher in ISS compared with OSS; however, there were no differences in any of the other outcomes between the ISS and OSS groups. The results from this randomised trial are not consistent with traditional situated learning theory, which states that increased fidelity leads to improved learning.[62,63] Conclusions from this trial, however, are in alignment with more recent empirical research and discussions on fidelity and learning.[6,64-66] Our study indicates that the change in simulation fidelity, as change in setting for simulation, does not necessarily translate into learning. Another randomised trial, which compared OSS as in-house training at the hospital in rooms specifically allocated for training with OSS in a simulation centre, also showed that the simulation setting was of minor importance and that there was no additional benefit from training OSS in a simulation centre versus OSS in-house.[44,67]

The present trial involved simulation based training with six different healthcare professions. A relevant perspective is the discussion on expanding the traditional concept of fidelity to include sociological fidelity, which encompasses the relationship between the various healthcare professionals.[37,68] After completing the trial we decided to explore more closely the experiences between the healthcare professionals in a qualitative study.[69]

Post hoc analyses showed similar educational effects in the ISS and the OSS groups with a knowledge gain of approximately 13% in both groups. It can be argued that this knowledge gain was due to the test effect.[70,71] We believe, however, that the test effect was minimised as

feedback was not given after the initial testing, which is viewed as crucial to learning from a test, and furthermore only one MCQ test was used.[71]

No differences were found in the mean Safety Attitudes Questionnaire (SAQ) score after simulation-based medical education in the ISS versus OSS groups. Earlier studies have described that high SAQ values mean that SAQ cannot be influenced by an intervention.[72,73] The values for SAQ were generally high in this trial compared to various other studies from non-Scandinavian countries.[72-75] .

There were no differences in the stress level when measured as salivary cortisol levels, Stress-Trait Anxiety Inventory (STAI) and cognitive appraisal (CA) in the ISS versus OSS groups. The post hoc analysis showed that simulation-based medical education triggered objective stress, measured by salivary cortisol, to the same extent in the ISS and OSS groups. CA seemed to be without discriminatory effect and a decrease was observed where an increase would have been expected, and the levels of CA were low compared to other studies. Previously used among students and medical trainees,[22,76,77] CA appeared to have a less discriminatory effect in these more senior groups of healthcare professionals.

Intrinsic Motivation Inventory [24,51] revealed no differences between ISS versus OSS. Motivation has not previously been tested in educational simulation studies and it is argued a gap appears to exist in the simulation literature on motivational factors and further research has been encouraged. [2] Some argue that simulation in the clinical setting, as with ISS, should increase motivation,[14] but this was not confirmed by findings in the present trial.

The evaluation data showed no differences between ISS and OSS. Both the ISS and the OSS participants gave very high scores on the evaluation. This is in accordance with what is generally seen in inter-professional training.[78]

The team performance showed no differences between ISS versus OSS. The post hoc analysis showed that teams performed statistically significantly better in the second compared to the first simulation, which indicates that the simulations were effective. Validated in previous studies, the

Team Emergency Assessment Measure scale has been found reasonably intuitive to use,[52,53] which was also our impression in this study.

According to the participants own perceptions they found that ISS and OSS were equally inspirational with regard to suggesting organisational changes in the delivery room, operating theatre and for clinical guidelines. The qualitative analysis, however, revealed that ISS participants provided more ideas for suggested changes, especially concerning technology and tools in the delivery ward and the operating theatre.[58] Previous non randomised studies have suggested that ISS has an impact on organisations, but this has to our knowledge never been confirmed in a randomised trial.[8,11,13,59]

Strength and limitations

This trial has several strengths. It was conducted with an adequate generation of allocation sequence; adequate allocation concealment; adequate reporting of all relevant outcomes; had very few drop-outs; and was conducted on a not-for-profit bias.[79-81] The trial was also blinded for data managers and statisticians. Generally, ISS programmes have been criticised for their lack of meaningful evaluations of the effectiveness of the programmes.[8] A strength of this trial was its use of a broad variety of outcome measures using previously validated scales to assess the effect on the individual, the team and the organisational level.

A limitation of the study is the fact that the outcome was based only on immediate measurements of knowledge level and of team performance. Only perceptions of simulation were measured after one week (evaluation and motivation) and safety attitudes after one month. No clinical outcomes or patient safety data were measured.

A strength of this trial is the involvement of authentic teams that mirrored teams in real life, which is important for the so-called sociological fidelity.[37,68] The teams in this trial were authentic in their design and hence resemble the real clinical setting in every possible way.[65,82] These kinds of teams are called ‘add hoc’ on-call teams and are very difficult to follow and observe in the real clinical setting, and assessment of the clinical performance of ad hoc teams for a long period is almost impossible. The authentic teams may also be a limitation because two-thirds of the participants had some simulation experiences. The findings in this trial therefore need to be

confirmed among other kinds of healthcare professionals with less experience in simulation-based education.

Previous research on assessment suggests that knowledge-based written assessments can predict the results of performance-based tests, and hence knowledge-based assessment could be used as a proxy for performance.[83-85] However a better approach to the assessment could have been performance-based tests of clinical work, but this was considered unfeasible.

In this trial we did not measure long-term retention. Literature on retention of skills suggests that deterioration of the non-used skills appears to occur about three to 18 months after training. More research within the field of retention and on the effect of short booster courses is necessary.[45,86-88]

There is a risk of type II error and the trial is most likely underpowered, as many randomised trials are. On the other hand it should be discussed whether performing a larger trial to detect a statistically significant effect of ISS is relevant or feasible and appears to have a clinically or educationally relevant effect.[89]

The improvements on knowledge and team performance may also be due to the Hawthorne effect, i.e. due to individuals changing behaviour as a result of their awareness of being observed.[90] From an educational perspective a major problem with the Hawthorne effect is an intervention group versus a control group, where the control group is given no intervention. [90] This issue was avoided in this trial as exactly the same intervention was used for both groups, the only difference being the physical setting, thus likely minimising the Hawthorne effect in our trial. [90]

Conclusions

This randomised trial compared ISS versus OSS, where OSS was provided as in-house training at the hospital in rooms specifically allocated for training. From the present trial we concluded that changes in settings from OSS to ISS do not seem to provide key elements for improving simulation-based medical education. Although participant perception of the fidelity of ISS versus OSS differed significantly, there were no differences in knowledge, patient safety attitude, stress measurements, motivation and team performance between the groups, except that the ISS group generated more suggestions for organisational changes. The present trial indicated that the fidelity of the setting

seemed to be of less importance for learning; however, more research is necessary to better understand which aspects of simulation that is the most important for learning.

Acknowledgements

The authors would like to thank the doctors, midwives and nurses who took part in the working committee and in planning the intervention, especially midwife Pernille Langhoff-Roos (Department of Obstetrics, Rigshospitalet, University of Copenhagen) for her contributions to the detailed planning, recruitment of participants and completion of the simulations. The authors would like to thank Jørn Wetterslev (Copenhagen Trial Unit) for advice on the potential impact of the clustering effect; Karl Bang Christensen (Section of Biostatistics, Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen) for advice on the statistical plan; Solvejg Kristensen (Danish National Clinical Registries) for her advice and for providing a Danish edition of Safety Attitudes Questionnaire; Per Bech (Psychiatric Research Unit, Mental Health Centre North Zealand) for advice and for providing a Danish edition of the State-Trait Anxiety Inventory; Claire Welsh (Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen) and Niels Leth Jensen (certified translator) for forwarding the back translation of Intrinsic Motivation Inventory; Jesper Sonne (Department of Clinical Pharmacology, Bispebjerg Hospital) for advice on medicine that can influence cortisol measurements; Anne Lippert and Anne-Mette Helsø (Danish Institute for Medical Simulation, Herlev Hospital, University of Copenhagen) for training the facilitators in conducting debriefing after simulation; medical student Veronica Markova for undertaking all practical issues on cortisol measurements; Bente Bennike (Laboratory of Neuropsychiatry, Rigshospitalet, University of Copenhagen) for managing the cortisol analyses; medical students Tobias Todsen and Ninna Ebdrup for video recordings of simulations; and Tobias Todsen for editing and preparing videos for assessment. Furthermore, the authors would like to thank Helle Thy Østergaard and Lone Fuhrmann (Department of Anaesthesiology, Herlev Hospital, University of Copenhagen), Lone Krebs (Department of Obstetrics and Gynecology, Holbæk Hospital) and Morten Beck Sørensen (Department of Obstetrics and Gynecology, Odense University Hospital) for reviewing and assessing all videos. Finally, thank you to Nancy Aaen (copy editor and translator) for revising the manuscript.

Contributors

JLS conceived of the idea for this trial. BO and CV supervised the trial. All authors made contributions to the design of the trial. JLS, assisted by BO, was responsible for acquiring funding. JLS, JL and CG contributed to the sample size estimation and detailed designing of and execution

of the randomisation process. JLS, MJ, KE, DOE and VL made substantial contributions to the practical and logistical aspects of the trial, while PW contributed to the discussion and practical and logistical issues concerning testing salivary cortisol. Jointly with JLS, SR and LS performed the statistical analysis. JLS wrote the draft manuscript. All authors provided critical review of this paper and approved the final manuscript.

Funding

The trial was funded by non-profit funds, including the Danish Regions Development and Research Foundation, the Laerdal Foundation for Acute Medicine and Aase and Ejnar Danielsen Foundation. None of the fund givers had any role in trial design; data collection, access, analysis, or interpretation; writing of the report; or the decision to publish.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf. The principal investigator and lead author (JLS) reports non-profit funding mentioned above. Doris Østergaard reports board membership of Laerdal Foundation for Acute Medicine. Other authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the this work in the previous three years; no other relationships or activities that could appear to have influenced the our work.

Ethical approval

Participants were healthcare professionals, and neither patients nor patient data were used in the trial. Approval from the Regional Ethics Committee (protocol number H-2-2012-155) and the Danish Data Protection Agency (Number 2007-58-0015) were obtained. Participants were assured that their personal data, data on questionnaires, salivary cortisol samples and video recordings would remain anonymous during analyses and reporting. The participants were asked to respect the confidentiality of their observations about their colleagues' performance in the simulated setting.

Data sharing

No consent for data sharing with other parties was obtained but the corresponding author may be contacted to forward requests for data sharing.

Transparency

The principal investigator and lead author (JLS) affirms that the manuscript is an honest, accurate and transparent account of the study being reported. No important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Figure 1 and 2 Legends

Figure 1. Timing of the simulations and measurement of stress: Objective stress was measured by salivary cortisol and subjective stress was measured by State-Trait Anxiety Inventory and cognitive appraisal.

Figure 2. Flow diagram for participants in a trial determining the effects of ISS versus OSS on 1) primary outcome: knowledge (MCQ test); and 2) exploratory outcomes: patient safety attitudes (SAQ), stress (salivary cortisol, STAI, CA), motivation (IMI), perceptions of simulation (evaluation questionnaire), video-assessed team performance (TEAM), and organisational impact

Reference List

[1] Issenberg SB, McGaghie WC, Petrusa ER, et al. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach* 2005;27:10-28.

[2] McGaghie WC, Issenberg SB, Petrusa ER, et al. A critical review of simulation-based medical education research: 2003-2009. *Med Educ* 2010;44:50-63.

[3] Motola I, Devine LA, Chung HS, et al. Simulation in healthcare education: a best evidence practical guide. AMEE Guide No. 82. *Med Teach* 2013;35:e1511-e1530.

[4] Merien AE, Van der Ven J, Mol BW, et al. Multidisciplinary team training in a simulation setting for acute obstetric emergencies: a systematic review. *Obstet Gynecol* 2010;115:1021-31.

[5] McGaghie WC, Issenberg SB, Barsuk JH, et al. A critical review of simulation-based mastery learning with translational outcomes. *Med Educ* 2014;48:375-85.

[6] Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;46:636-47.

[7] Beaubien JM, Baker DP. The use of simulation for training teamwork skills in health care: how low can you go? *Qual Saf Health Care* 2004;13 Suppl 1:i51-i56.

[8] Rosen MA, Hunt EA, Pronovost PJ, et al. In situ simulation in continuing education for the health care professions: a systematic review. *J Contin Educ Health Prof* 2012;32:243-54.

[9] Riley W, Davis S, Miller KM, et al. Detecting breaches in defensive barriers using in situ simulation for obstetric emergencies. *Qual Saf Health Care* 2010;19 Suppl 3:i53-i56.

[10] Guise JM, Lowe NK, Deering S, et al. Mobile in situ obstetric emergency simulation and teamwork training to improve maternal-fetal safety in hospitals. *Jt Comm J Qual Patient Saf* 2010;36:443-53.

[11] Walker ST, Sevdalis N, McKay A, et al. Unannounced in situ simulations: integrating training and clinical practice. *BMJ Qual Saf* 2013;22:453-8.

[12] Patterson MD, Geis GL, Falcone RA, et al. In situ simulation: detection of safety threats and teamwork training in a high risk emergency department. *BMJ Qual Saf* 2013;22:468-77.

[13] Patterson MD, Geis GL, Lemaster T, et al. Impact of multidisciplinary simulation-based training on patient safety in a paediatric emergency department. *BMJ Qual Saf* 2013;22:383-93.

[14] Stocker M, Burmester M, Allen M. Optimisation of simulated team training through the application of learning theories: a debate for a conceptual framework. *BMC Med Educ* 2014;14:69.

[15] Sullivan NJ, Duval-Arnould J, Twilley M, et al. Simulation Exercise to Improve Retention of Cardiopulmonary Resuscitation Priorities for In-Hospital Cardiac Arrests: A Randomized Controlled Trial. *Resuscitation* 2014;86C:6-13.

[16] Rubio-Gurung S, Putet G, Touzet S, et al. In situ simulation training for neonatal resuscitation: an RCT. *Pediatrics* 2014;134:e790-e797.

[17] Couto TB, Kerrey BT, Taylor RG, et al. Teamwork Skills in Actual, In Situ, and In-Center Pediatric Emergencies: Performance Levels Across Settings and Perceptions of Comparative Educational Impact. *Simul Healthc* 2015;10:76-84.

[18] LeBlanc VR. The effects of acute stress on performance: implications for health professions education. *Acad Med* 2009;84:S25-S33.

[19] LeBlanc VR, Manser T, Weinger MB, et al. The study of factors affecting human and systems performance in healthcare using simulation. *Simul Healthc* 2011;6 Suppl:S24-S29.

[20] Arora S, Sevdalis N, Aggarwal R, et al. Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surg Endosc* 2010;24:2588-93.

[21] Harvey A, Nathens AB, Bandiera G, et al. Threat and challenge: cognitive appraisal and stress responses in simulated trauma resuscitations. *Med Educ* 2010;44:587-94.

[22] Harvey A, Bandiera G, Nathens AB, et al. Impact of stress on resident performance in simulated trauma scenarios. *J Trauma Acute Care Surg* 2012;72:497-503.

[23] Finan E, Bismilla Z, Whyte HE, et al. High-fidelity simulator technology may not be superior to traditional low-fidelity equipment for neonatal resuscitation training. *J Perinatol* 2012;32:287-92.

[24] Kusurkar RA, Ten Cate TJ, van AM, et al. Motivation as an independent and a dependent variable in medical education: a review of the literature. *Med Teach* 2011;33:e242-e262.

[25] Ten Cate TJ, Kusurkar RA, Williams GC. How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Med Teach* 2011;33:961-73.

[26] Ryan RM, Deci EL. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemp Educ Psychol* 2000;25:54-67.

[27] Drife J. Reducing risk in obstetrics. *Qual Health Care* 1995;4:108-14.

[28] Gold KJ, Kuznia AL, Hayward RA. How physicians cope with stillbirth or neonatal death: a national survey of obstetricians. *Obstet Gynecol* 2008;112:29-34.

[29] Veltman LL. Getting to havarti: moving toward patient safety in obstetrics. *Obstet Gynecol* 2007;110:1146-50.

[30] Pronovost PJ, Holzmueller CG, Ennen CS, et al. Overview of progress in patient safety. *Am J Obstet Gynecol* 2011;204:5-10.

[31] Hove LD, Bock J, Christoffersen JK, et al. Analysis of 127 peripartum hypoxic brain injuries from closed claims registered by the Danish Patient Insurance Association. *Acta Obstet Gynecol Scand* 2008;87:72-5.

[32] Berglund S, Grunewald C, Pettersson H, et al. Severe asphyxia due to delivery-related malpractice in Sweden 1990-2005. *BJOG* 2008;115:316-23.

[33] Cantwell R, Clutton-Brock T, Cooper G, et al. Saving Mothers' Lives: Reviewing maternal deaths to make motherhood safer: 2006-2008. The Eighth Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. *BJOG* 2011;118 Suppl 1:1-203.

[34] Bodker B, Hvidman L, Weber T, et al. Maternal deaths in Denmark 2002-2006. *Acta Obstet Gynecol Scand* 2009;88:556-62.

[35] Johannsson H, Ayida G, Sadler C. Faking it? Simulation in the training of obstetricians and gynaecologists. *Curr Opin Obstet Gynecol* 2005;17:557-61.

[36] Sørensen JL, Van der Vleuten C, Lindschou J, et al. 'In situ simulation' versus 'off site simulation' in obstetric emergencies and their effect on knowledge, safety attitudes, team performance, stress, and motivation: study protocol for a randomized controlled trial. *Trials* 2013;14:220.

[37] Boet S, Bould MD, Layat BC, et al. Twelve tips for a successful interprofessional team-based high-fidelity simulation education session. *Med Teach* 2014;36:853-7.

[38] SimMom Mannequin. http://www.laerdal.com/dk/SimMom . 2014.

[39] Royal College of Obstetricans and Gynaecologists course: PROMPT, Train the trainers.https://www.rcog.org.uk/globalassets/documents/courses-exams-and-events/training-courses/prompt-2---train-the-trainers-mar2015.pdf. 2015.

[40] Rudolph JW, Simon R, Dufresne RL, et al. There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. *Simul Healthc* 2006;1:49-55.

[41] Thistlethwaite JE, Davies D, Ekeocha S, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Med Teach* 2012;34:e421-e444.

[42] Sørensen JL, Thellensen L, Strandbygaard J, et al. Development of a knowledge test for multi-disciplinary emergency training: a review and an example. *Acta Anaesthesiol Scand* 2015;59:123-33.

[43] Remmen R, Scherpbier A, Denekens J, et al. Correlation of a written test of skills and a performance based test: a study in two traditional medical schools. *Med Teach* 2001;23:29-32.

[44] Crofts JF, Ellis D, Draycott TJ, et al. Change in knowledge of midwives and obstetricians following obstetric emergency training: a randomised controlled trial of local hospital, simulation centre and teamwork training. *BJOG* 2007;114:1534-41.

[45] Sørensen JL, Løkkegaard E, Johansen M, et al. The implementation and evaluation of a mandatory multi-professional obstetric skills training program. *Acta Obstet Gynecol Scand* 2009;88:1107-17.

[46] Kristensen S, Sabroe S, Bartels P, et al. Adaption and validation of the Safety Attitudes Questionnaire for the Danish hospital setting. *Clin Epidemiol* 2015;7:149-60.

[47] Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;6:44.

[48] Deilkas ET, Hofoss D. Psychometric properties of the Norwegian version of the Safety Attitudes Questionnaire (SAQ), Generic version (Short Form 2006). *BMC Health Serv Res* 2008;8:191.

[49] Spielberger CD, Gorsuch RL, Lushene RE. Manual for the State-Trait Anxiety Inventory. Palo Alto, Calif.: Consulting Psychologists Press; 1970.

[50] Bech P. [Klinisk psykometri]. 1.udgave, 2 oplag ed. Munksgaard Danmark; 2011.

[51] Self-Determination Theory (SDT) http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/. 2015.

[52] Cooper S, Cant R, Porter J, et al. Rating medical emergency teamwork performance: development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;81:446-52.

[53] McKay A, Walker ST, Brett SJ, et al. Team performance in resuscitation teams: comparison and critique of two recently developed scoring tools. *Resuscitation* 2012;83:1478-83.

[54] Ukoumunne OC, Gulliford MC, Chinn S, et al. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;3:iii-92.

[55] Campbell MK, Piaggio G, Elbourne DR, et al. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;345:e5661.

[56] Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13-22.

[57] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal StatisticaI Society Series B (Methodological)* 1995;57:289-300.

[58] Carayon P, Schoofs HA, Karsh BT, et al. Work system design for patient safety: the SEIPS model. *Qual Saf Health Care* 2006;15 Suppl 1:i50-i58.

[59] Sorensen JL, Lottrup P, Vleuten van der C, et al. Unannounced in situ simulation of obstetric emergencies: staff perceptions and organisational impact. *Postgrad Med J* 2014;90:622-9.

[60] Kobayashi L, Parchuri R, Gardiner FG, et al. Use of in situ simulation and human factors engineering to assess and improve emergency department clinical systems for timely telemetry-based detection of life-threatening arrhythmias. *BMJ Qual Saf* 2013;22:72-83.

[61] Wheeler DS, Geis G, Mack EH, et al. High-reliability emergency response teams in the hospital: improving quality and safety using in situ simulation training. *BMJ Qual Saf* 2013;22:507-14.

[62] Durning SJ, Artino AR, Jr., Pangaro LN, et al. Perspective: redefining context in the clinical encounter: implications for research and training in medical education. *Acad Med* 2010;85:894-901.

[63] Durning SJ, Artino AR. Situativity theory: a perspective on how participants and the environment can interact: AMEE Guide no. 52. *Med Teach* 2011;33:188-99.

[64] Koens F, Mann KV, Custers EJ, et al. Analysing the concept of context in medical education. *Med Educ* 2005;39:1243-9.

[65] Teteris E, Fraser K, Wright B, et al. Does training learners on simulators benefit real patients? *Adv Health Sci Educ Theory Pract* 2012;17:137-44.

[66] Grierson LE. Information processing, specificity of practice, and the transfer of learning: considerations for reconsidering fidelity. *Adv Health Sci Educ Theory Pract* 2014;19:281-9.

[67] Ellis D, Crofts JF, Hunt LP, et al. Hospital, simulation center, and teamwork training for eclampsia management: a randomized controlled trial. *Obstet Gynecol* 2008;111:723-31.

[68] Sharma S, Boet S, Kitto S, et al. Interprofessional simulated learning: the need for 'sociological fidelity'. *J Interprof Care* 2011;25:81-3.

[69] Sørensen JL, Navne LE, Martin HM, et al. Clarifying the learning experiences of healthcare professionals with in situ versus off site simulation-based medical education: a qualitative study. *Accepted for publication in BMJ Open* 2015.

[70] Roediger HL, Karpicke JD. Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci* 2006;17:249-55.

[71] Larsen DP, Butler AC, Roediger HL, III. Test-enhanced learning in medical education. *Med Educ* 2008;42:959-66.

[72] Schwendimann R, Zimmermann N, Kung K, et al. Variation in safety culture dimensions within and between US and Swiss Hospital Units: an exploratory study. *BMJ Qual Saf* 2013;22:32-41.

[73] Sexton JB, Berenholtz SM, Goeschel CA, et al. Assessing and improving safety climate in a large cohort of intensive care units. *Crit Care Med* 2011;39:934-9.

[74] Shoushtarian M, Barnett M, McMahon F, et al. Impact of introducing Practical Obstetric Multi-Professional Training (PROMPT) into maternity units in Victoria, Australia. *BJOG* 2014;121:1710-8.

[75] Siassakos D, Fox R, Hunt L, et al. Attitudes toward safety and teamwork in a maternity unit with embedded team training. *Am J Med Qual* 2011;26:132-7.

[76] Pottier P, Hardouin JB, Dejoie T, et al. Stress responses in medical students in ambulatory and in-hospital patient consultations. *Med Educ* 2011;45:678-87.

[77] Piquette D, Tarshis J, Sinuff T, et al. Impact of acute stress on resident performance during simulated resuscitation episodes: a prospective randomized cross-over study. *Teach Learn Med* 2014;26:9-16.

[78] Reeves S, Goldman J. Medical education in an interprofessional context. In: Dornan T, Mann K, Scherpbier A., Spencer J, editors. Medical Education: Theory and Practice.Edinburgh, London, New York, Oxford, Philadelphia, St Louis, Sydney, Toronto: Churchill Livingstone Elsevier; 2011. p. 51-64.

[79] Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601-5.

[80] Savovic J, Jones HE, Altman DG, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;157:429-38.

[81] Lundh A, Sismondo S, Lexchin J, et al. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2012;12:MR000033.

[82] Lingard L. Rethinking Competence in the Context of teamwork. In: Hodges B, Lingard L, editors. The Question of Competence. Reconsidering Medical Education in the Twenty-First Century. First ed. New York: ILR Press an imprent of Cornell University Press; 2012. p. 42-69.

[83] van der Vleuten C, Van Luyk SJ, Beckers HJ. A written test as an alternative to performance testing. *Med Educ* 1989;23:97-107.

[84] Kramer AW, Jansen JJ, Zuithoff P, et al. Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. *Med Educ* 2002;36:812-9.

[85] Ram P, van der Vleuten C, Rethans JJ, et al. Assessment in general practice: the predictive value of written-knowledge tests and a multiple-station examination for actual medical performance in daily practice. *Med Educ* 1999;33:197-203.

[86] Arthur W, Bennett W, Stanush PL, et al. Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance* 1998;11:57-101.

[87] Hamilton R. Nurses' knowledge and skill retention following cardiopulmonary resuscitation training: a review of the literature. *J Adv Nurs* 2005;51:288-97.

[88] Trevisanuto D, Ferrarese P, Cavicchioli P, et al. Knowledge gained by pediatric residents after neonatal resuscitation program courses. *Paediatr Anaesth* 2005;15:944-7.

[89] Jakobsen JC, Gluud C, Winkel P, et al. The thresholds for statistical and clinical significance - a five-step procedure for evaluation of intervention effects in randomised clinical trials. *BMC Med Res Methodol* 2014;14:34.

[90] Fraenkal JR, Wallen NE, Hyde A. Internal Validity. In: McGraw-Hill, editor. How to Design and Evaluate Research in Education.New York: 2012. p. 166-83.

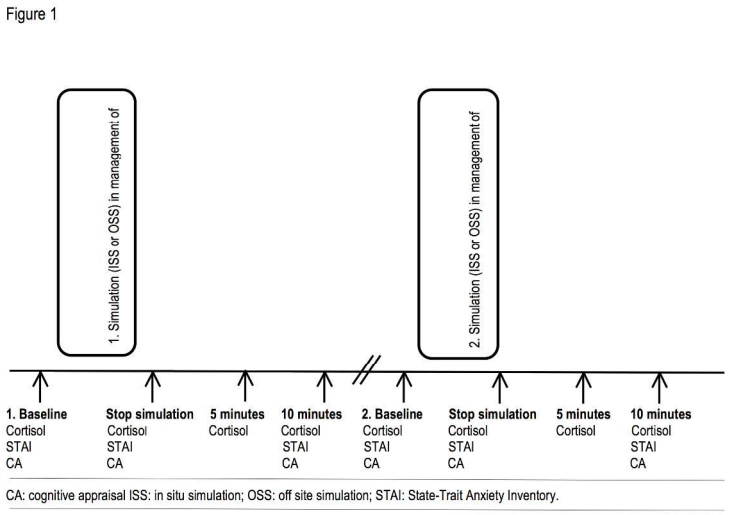


Figure 1. Timing of the simulations and measurement of stress: Objective stress was measured by salivary cortisol and subjective stress was measured by State-Trait Anxiety Inventory and cognitive appraisal.
209x296mm (300 x 300 DPI)

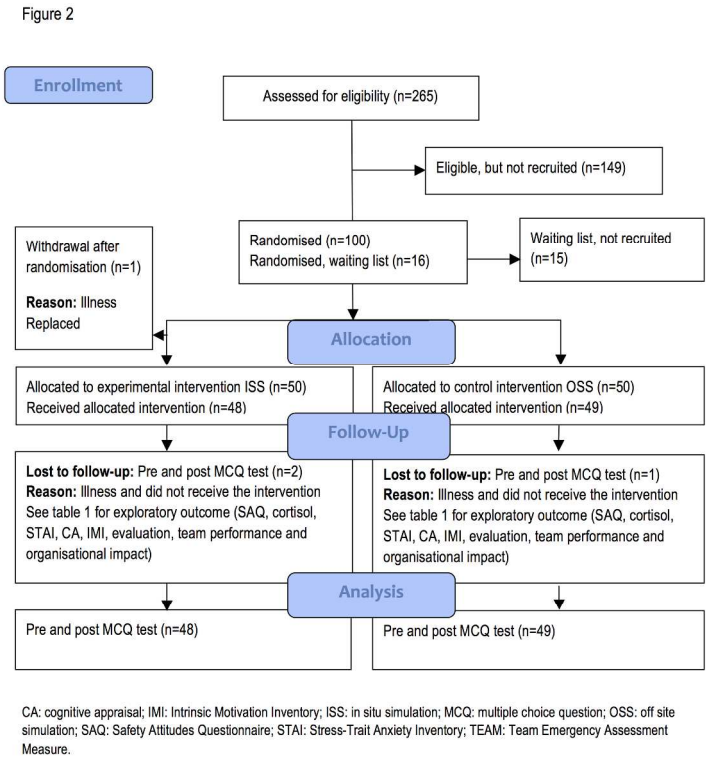


Figure 2. Flow diagram for participants in a trial determining the effects of ISS versus OSS on 1) primary outcome: knowledge (MCQ test); and 2) exploratory outcomes: patient safety attitudes (SAQ), stress (salivary cortisol, STAI, CA), motivation (IMI), perceptions of simulation (evaluation questionnaire), video-assessed team performance (TEAM), and organisational impact
209x296mm (300 x 300 DPI)

Sørensen JL et al, BMJ Open 2015: Simulation based multi-professional obstetric anaesthesia training conducted in situ versus off site leads to similar individual and team outcomes: a randomised educational trial.

Table S1. Suggestions for practical and organisational changes identified during ISS and OSS classified according to the model of work system or structure from the Systems Engineering Initiative for Patient Safety. [55]

	Identified during		SEIPS component
	OSS	ISS	
Introduction of new employees			
Better introduction of new employees, including guided tour of trauma centre, acute admission centre, and blood bank	X	X	Person
Identification of staff			
Name badges visible on uniforms at all times and improved name badges for easier identification of the various healthcare professions	X	X	Person
Inclusion of people's names and also healthcare profession during staff presentation rounds in emergency situations	X	X	Organisation
Call systems, telephones and telephone numbers			
Pros and cons involved in changing the emergency call system from calling people individually to calling them as a group	X	X	Organisation
Request for more well-defined telephone chain for emergency calls; doctors preferred briefing to be from doctor to doctor	X	X	Organisation
More well-defined criteria for using emergency alarm button in delivery room	X	X	Organisation
Identical numbers to call night and day for anaesthesia assistance in obstetric emergencies	X	X	Organisation
A clearly visible list of relevant numbers in operating theatre for emergency situations		X	Tools & technology
Stickers with brief list of emergency numbers on back of name badges		X	Tools & technology
Clinical handover in emergency situations			
Repetition upon arrival in delivery room and operating theatre of clinical details and indication for procedures provided in telephone handover	X	X	Organisation
Consistent use of terminology from local guidelines and when grading emergency caesarean sections	X	X	Organisation
Patient identification and "time out" in operating theatre in emergency situations			
More clearly defined designation of who is responsible for identifying the patient and confirming the indication for procedure	X	X	Organisation
Improvement of computer system that is too difficult and slow for emergency situations		X	Tools & technology
Presence of partners during emergency caesarean sections			
Various opinions on whether partners should be allowed in operating theatre; more well-defined criteria for designating who communicates with partners	X	X	Organisation
Medication – postpartum haemorrhage			
Placement of tranexamic acid in the haemorrhage medication box; clinical guidelines on its administration should be made easily accessible	X	X	Tools & technology
Pre-prepared drips with oxytocin	X	X	Tools & technology
Midwives generally found administering medicine in operating theatre difficult; requested more clarity for designating who is responsible for the haemorrhage medication box there	X	X	Person, Task
Clarification of who is to document administration of medicine in operating theatre, especially when administered directly in the uterus and/or per rectum	X	X	Tools & technology

	Identified during		SEIPS components
	ISS	OSS	
Medication – emergency caesarean sections			
Pros and cons involved in administering terbutaline for tocolysis during foetal distress; informing anaesthesia staff important due to subsequent risk of maternal tachycardia	X	X	Tools & technology
Placement of terbutaline for tocolysis (intrauterine resuscitation) in delivery room and operating theatre to allow quick administration	X	X	Tools & technology
Address the knowledge gap among auxiliary nurses and midwives on administration of sodium citrate to prevent aspiration during general anaesthesia	X	X	Person, Task
Amend action card and clinical guidelines on emergency caesarean sections to specifically address oral administration of sodium citrate to prevent aspiration during general anaesthesia	X	X	Organisation
Make sodium citrate more easily accessible in the delivery room	X	X	Tools & technology
Staff members in operating theatre during postpartum haemorrhage			
Two operating theatre nurses ideally present in severe cases of postpartum haemorrhage	X	X	Person, Task
Obstetric team members (midwife, specialised midwife, and auxiliary nurse) should ideally prioritise remaining in operation theatre to assist	X	X	Person, Task
Fluid resuscitation and blood transfusion in operating theatre			
Clarification of whether lactated Ringer's solution is superior to sodium chloride	X	X	Tools + technology
Easy access to a blood heater and pressure bags to improve IV infusion in delivery room	X	X	Tools + technology
Clarification of which healthcare professional should ideally collect blood at the blood bank in an emergency situation	X	X	Person, Task
Clearly posted telephone number in operating theatre for the blood bank and its location	X	X	Tools + technology
Training/retraining of midwives in management of blood transfusions to allow them to assist the anaesthesia team correctly	X	X	Person, Task
Transfer of patient from delivery room to operating theatre and type of operating table			
Clarification of who is responsible for birthing bed (preparations for transport)		X	Person Tools & technology
Mechanisms to ensure clear passage (e.g. no beds, transport cages) along corridors for emergency transport of patients on delivery ward		X	Person, Tools & technology
Clarification of when and how to ideally transfer patients from delivery room to operating theatre; clarification of who determines timing of patient transport in emergency situations	X	X	Person, Organisation
Improved standards for where to keep the remote control for the operating table and for recharging its batteries; have easy-to-use instructions available		X	Tools & technology
Pros and cons of continuous use of birthing beds in emergency situations when patient transferred to operating theatre; address the knowledge gap on functionality of birthing beds among operating theatre staff; establish standards for their use		X	Tools & technology
Practical issues in operating theatre			
Better labelling of equipment in operating theatre and standardised placement of equipment in the two operating theatres for obstetric emergencies to streamline management	X	X	Tools & technology
Improve use of remote control to the operation table, which is difficult due to a lack of clarity about which direction the table moves		X	Person, Tools & technology
A more suitable walking distance between the cabinet with surgical caps and the operating theatre		X	Tools and technology

	Identified during		SEIPS components
	ISS	OSS	
Operating theatre nurses			
Clarification on role of responsibility for clinical decision making for urinary bladder catheter and/or shaving the pubic area, communication and actual management hereof; general agreement that obstetricians make the clinical decision and then inform the operating theatre nurse, who then places the catheter and/or shaves the area	X	X	Person, Organisation
Anaesthesia team in operating theatre			
More assistance from midwives when transferring patients from birthing bed to operating table and with positioning of patient; midwives indicated more guidance from the anaesthesia team on the ideal way to do transfers	X	X	Person, Organisation
Improved procedures for checking equipment to ensure that it works (e.g. problem with no light in a laryngoscope)		X	Tools & technology, Organisation
More detailed introduction of new employees, including presentation of equipment for management of the difficult airway and equipment for blood heating and rapid infusion		X	Person
Observation charts and boards			
Improved observation charts for emergency situations, especially for postpartum haemorrhage	X	X	Tools & technology, Organisation
Greater use of white boards in delivery rooms in emergency situations for temporary observational charting	X	X	Tools & technology, Organisation
Use of white boards in operating theatre in emergency situations	X	X	Tools & technology, Organisation
Mode of anaesthesia in emergency situations			
Determine who makes final clinical decision about mode of anaesthesia; generally agreed to be the anaesthetist's responsibility	X	X	Person
Preoxygenation necessary prior to induction of spinal anaesthesia in case general anaesthesia is required	X	X	Task, Organisation
Clear communication on mode of anaesthesia to all staff in the room crucial so operating theatre nurses can prepare for e.g. sterile drapes, leg holders	X	X	Task, Organisation
Use, when feasible, obstetric manoeuvres like bimanual compression with severe postpartum haemorrhage and replacement of foetal head during cord prolapse with the parturient woman in side position (for attempt of spinal anaesthesia); communicate this during training/retraining of staff and address in clinical guidelines	X	X	Task, Organisation
Guidelines			
Greater clarity in postpartum haemorrhage guidelines on indications and general clinical management principles for blood product transfusion and risk of hypothermia	X	X	Task, Organisation
Addition of pointers in local clinical guidelines on how to choose the best team leader and this individual's role in emergency situations	X	X	Task, Organisation

ISS: in situ simulation; OSS: off site simulation; SEIPS: Systems Engineering Initiative for Patient Safety

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	1
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	7-8
	2b	Specific objectives or hypotheses	8
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	8
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	None
Participants	4a	Eligibility criteria for participants	8
	4b	Settings and locations where the data were collected	8
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	8-9
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	9-10 Fig.1
	6b	Any changes to trial outcomes after the trial commenced, with reasons	None
Sample size	7a	How sample size was determined	10-11
	7b	When applicable, explanation of any interim analyses and stopping guidelines	None
Randomisation:			11
Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	11
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	11
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	11
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	None

		assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	8-9
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	11-12
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	11-12
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	12 Fig.2 Table 3
	13b	For each group, losses and exclusions after randomisation, together with reasons	12 Fig.2 Table 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	12
	14b	Why the trial ended or was stopped	Not done
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 2
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Fig 2 Table 2
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Page 13-14 Table 3,4,5,6,7,8
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	None
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	Page 13-14 Table 4,5,6,7,8 Online table
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	None
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	17-18
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	14-15, 17
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	14-18
Other information			
Registration	23	Registration number and name of trial registry	NCT01792674
Protocol	24	Where the full trial protocol can be accessed, if available http://www.rigshospitalet.dk/NR/rdonlyres/1F6BDE2B-CA37-4017-8A42-CF94F7FB31AE/0/SIM5_Engelsk_protokol_Insitu_versus_offsite_CTU_obanopsimulation_2013_02_153.pdf	Ref 36 protocol article

Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	20

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

Correction: Simulation-based multiprofessional obstetric anaesthesia training conducted in situ versus off-site leads to similar individual and team outcomes: a randomised educational trial

Sørensen JL, van der Vleuten C, Rosthøj S, *et al*. Simulation-based multiprofessional obstetric anaesthesia training conducted in situ versus off-site leads to similar individual and team outcomes: a randomised educational trial. *BMJ Open* 2015;**5**:e008344. doi: 10.1136/bmjopen-2015-008344.

The table headings are incorrect on pages 2 and 3 of supplementary table S1. The sub-headings should read 'OSS' and 'ISS' - these headings are correct on page 1 of the Table but the wrong way round in pages 2 and 3.

BMJ Open 2017;**7**:e008344corr1. doi:10.1136/bmjopen-2015-008344corr1