

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

Jette Led Sørensen,<sup>1</sup> Cees van der Vleuten,<sup>2</sup> Susanne Rosthøj,<sup>3</sup> Doris Østergaard,<sup>4</sup> Vicki LeBlanc,<sup>5</sup> Marianne Johansen,<sup>6</sup> Kim Ekelund,<sup>7</sup> Liis Starkopf,<sup>3</sup> Jane Lindschou,<sup>8</sup> Christian Gluud,<sup>8</sup> Pia Weikop,<sup>9</sup> Bent Ottesen<sup>10</sup>

**To cite:** 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

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-008344>).

Received 28 March 2015

Revised 19 July 2015

Accepted 27 July 2015



► <http://dx.doi.org/10.1136/bmjopen-2015-008344>



CrossMark

For numbered affiliations see end of article.

#### Correspondence to

Jette Led Sørensen;  
jette.led.soerensen@regionh.dk

#### 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:** 100 participants in teams of 10, comprising midwives, specialised midwives, auxiliary nurses, nurse anaesthetists, operating theatre nurses, and consultant doctors and trainees in obstetrics and anaesthesiology.

**Interventions:** Two multiprofessional simulations (clinical management of an emergency caesarean section and a postpartum haemorrhage scenario) were conducted in teams of 10 in the ISS versus the OSS setting.

**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 did the participants in the OSS group. Expert video assessment of team performance showed no differences between the ISS versus the OSS group. The ISS group provided more ideas and suggestions for changes at the organisational level.

#### Strengths and limitations of this study

- To the best of 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. However, in this randomised trial, no significant differences were found regarding knowledge, patient safety attitude, stress measurements, motivation or team performance 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 did the participants in the off-site simulation group. However, this perception did not influence the individual and team outcomes. On the outcome on the 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 seems to be 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 1 week (evaluation and motivation) and safety attitudes after 1 month. No clinical outcome was measured.

**Conclusions:** In this randomised trial, no significant differences were found regarding knowledge, patient safety attitude, motivation 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 number:** NCT01792674.

## INTRODUCTION

Frequently recommended as a learning modality,<sup>1–5</sup> 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.”<sup>5</sup> However, its key elements 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 or physical fidelity, that is, does the simulation look realistic? (2) psychological fidelity, that is, does the simulator contain the critical elements to accurately simulate the behaviours required to complete a task?<sup>6 7</sup>

Simulation-based medical education has traditionally been conducted as an 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.”<sup>8–12</sup> An unanswered question is whether ISS is superior to OSS. It has been argued that ISS has more fidelity and can lead to better teaching and greater organisational impact compared with OSS.<sup>8–14</sup>

We hypothesised that the physical setting could influence fidelity, and hence ISS could be more effective for educational purposes. To the best of our knowledge, no randomised educational trials have been conducted comparing the ISS versus the OSS setting. Two articles that do use randomisation focused on frequency of training and not setting, and did not include a relevant control group.<sup>15 16</sup> 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.<sup>8</sup> A recent retrospective video-based study showed that the performance was similar in all the tested simulation settings, but the participants favoured ISS and the authors argued that prospective studies are needed.<sup>17</sup>

Human factors such as stress and motivation impact learning,<sup>18–26</sup> 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 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 workplaces, where patient safety is high on the agenda and unexpected emergencies occur.<sup>27–34</sup> Simulation-based medical education is thus argued to be an essential learning strategy for labour wards.<sup>4 35</sup> 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 multiprofessional obstetric anaesthesia teams.

## METHODS

### Design

An investigator-initiated, single-centre randomised superiority educational trial was previously described in a design article.<sup>36</sup>

### 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 6300 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.<sup>36</sup>

### 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.<sup>36</sup>

### Interventions

The experimental intervention was a preannounced ISS,<sup>8 9</sup> that is, 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 ISS and OSS.<sup>36</sup> 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 interprofessional skills and communication, the scenarios gave each healthcare profession a significant role to play.<sup>37</sup>

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 the ISS and OSS settings. In the real and the fictitious operating theatre, a full-body birthing simulator, a SimMom, was used for parts of the simulation scenario.<sup>38</sup> Recruited from the working committee, the instructors conducting the simulations were trained in facilitating simulations and doing debriefings. The working committee was trained in local organised courses and attended a British National train the trainers course: PROMPT (PRactical Obstetric Multi-Professional Training).<sup>39</sup> 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 min and comprised three phases: description, analysis and application.<sup>40</sup> In addition to the simulation-based medical education, the training day also included video-based, case-based<sup>41</sup> 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.<sup>42</sup> 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.<sup>43</sup> Furthermore, previously used knowledge tests could be used for inspiration and for sample size calculation.<sup>44 45</sup> 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.<sup>46</sup> It included 33 items covering five dimensions: (1) team work climate; (2) safety climate; (3) job satisfaction; (4) stress recognition; and (5) work conditions.<sup>47 48</sup> The participants did the SAQ 1 month prior to and 1 month after participating in the training day.

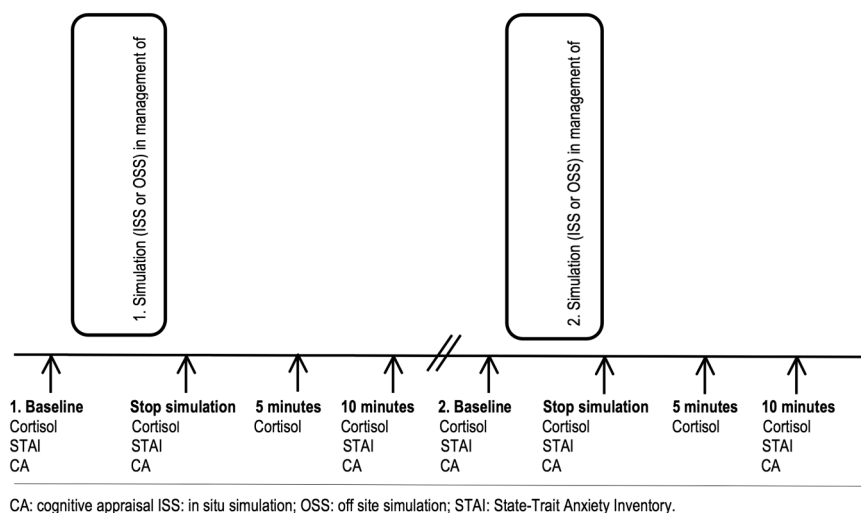
**Stress:** Salivary cortisol levels were used as an objective measure of physiological stress.<sup>36</sup> The salivary cortisol samples were obtained as a baseline before the first and the second simulation and at three additional times after the two simulations (figure 1). The subjective stress level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

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).<sup>51</sup>

**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.<sup>36</sup>

**Team performance** was video recorded and assessed by experts using a Team Emergency Assessment Measure

**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.



(TEAM).<sup>36 52 53</sup> 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 times 3 h 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 management of an emergency caesarean section and a postpartum haemorrhage, respectively.

*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.<sup>44 45</sup> We assumed the distribution of the primary outcome (the percentage of correct MCQ answers) to be normally distributed with an SD 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%. Since the interventions were delivered in teams (clusters), observations from the same team were likely to be correlated.<sup>54 55</sup> 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.<sup>55</sup> 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 10. 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,

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

	ISS group	OSS group
Number of participants	48*	49†
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‡		
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§	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/estradiol acetate	0	1
Oral contraceptives	1	3
Beta blockers (metoprolol)	0	1
Antidepressants (nortriptyline, fluoxetine)	2	0

\*Not included due to illness: A consultant obstetrician and an operating room nurse (n=2).

†Not included due to illness: An auxiliary nurse (n=1).

‡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.

§Intrauterine contraceptive devices, angiotensin II receptor antagonists, angiotensin-converting-enzyme inhibitors, simvastatin, alendronate, pantoprazole, antihistamine and tinzaparine.

ISS, in situ simulation; OSS, off-site simulation.



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

Owing 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 the inventories SAQ, IMI and STAI were excluded from the overall calculation of the summary scores.

Calculation of 95% CI obtained after the simulation intervention (post-MCQ, post-SAQ, stress measurements, IMI) was based on generalised estimating equations (GEE)<sup>56</sup> 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.<sup>57</sup>

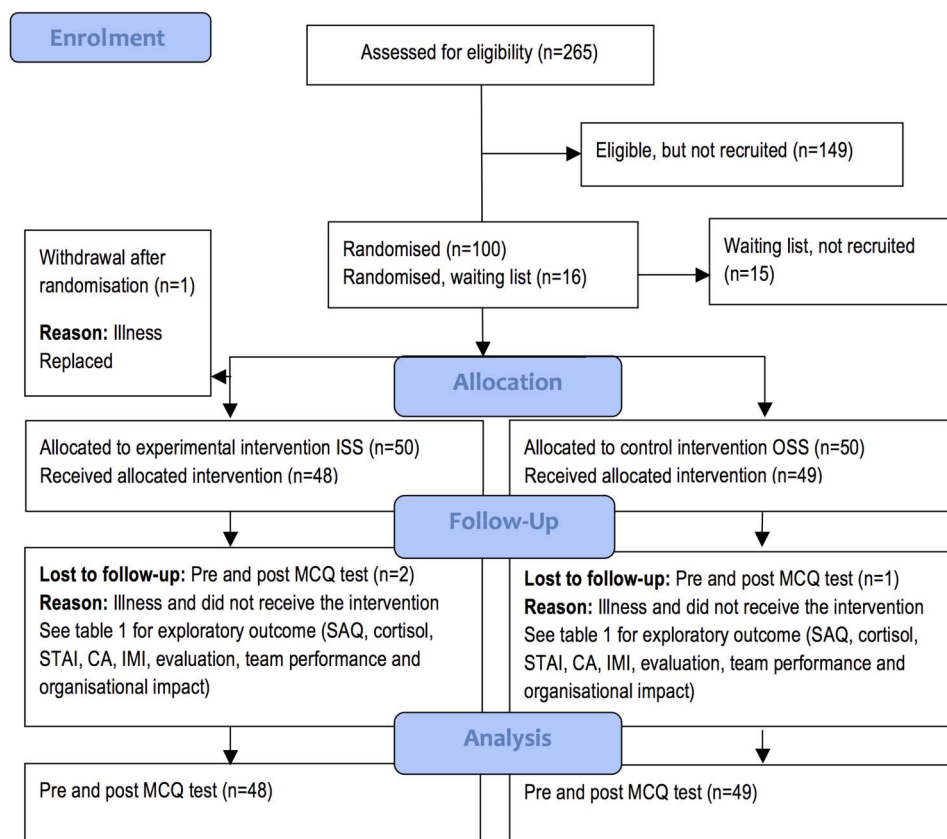
The mean outcomes obtained after the simulation intervention (postmeasurements) in the two intervention groups were compared by a linear model including

intervention and baseline (premeasurements) 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 postmeasurements 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 postmeasurements on the premeasurements were performed.

The team data, that is, 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 the 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

**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.



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

framework from the Systems Engineering Initiative for Patient Safety model.<sup>58</sup>

SAS V.9.2, R V.3.0.2 and IBM SPSS Statistics V.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).

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

### Intervention delivery

The trial was conducted from April to June 2013. Out of 100 participants included, 97 participated (tables 1 and 2 and figure 2). The 10 simulations were conducted as planned, although one ISS had to be postponed for 15 min 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 min, respectively (*p*=0.70), while the mean for the postpartum haemorrhage simulation was 26 and 24 min, respectively, (*p*=0.40).

### Primary outcome

*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 7–11 of the 40 MCQ items were disputable.<sup>42</sup>

*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).

### Exploratory outcomes

*SAQ:* No differences were found in the ISS versus the OSS group for any of the post-SAQ dimensions (table 4).

*Salivary cortisol, STAI and 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 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).

*IMI:* No differences were found in the ISS versus the OSS group for the IMI score (table 6).

*Participant evaluations and perception:* For almost all 20 questions in the evaluation questionnaire, the ISS and

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

	ISS group	OSS group
Pre-MCQ test	2 (2%)*	1 (1%)*
Post-MCQ test	2 (2%)*	1 (1%)*
Salivary cortisol level at emergency caesarean section simulation	2 (2%)*	3 (3%)*†
Salivary cortisol level at postpartum haemorrhage simulation	2 (2%)*	2 (2%)*‡
STAI at emergency caesarean section simulation	2 (2%)*	1 (1%)*
STAI at postpartum haemorrhage simulation	2 (2%)*	2 (2%)*‡
CA at caesarean section simulation	2 (2%)*	1 (1%)*
CA at postpartum haemorrhage simulation	2 (2%)*	2 (2%)*‡
Evaluation questionnaire	3 (3%)*§	1 (1%)*
IMI	4 (4%)*¶	1 (1%)*
Pre-SAQ	1 (1%)**	4 (4%)*††
Post-SAQ	5 (5%)*‡‡	4 (4%)*‡‡

\*Participants ill and did not participate (n=3).

†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).

‡Since 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).

§Questionnaires not returned (n=1).

¶Questionnaires not returned (n=2).

\*\*Of the individuals who did not participate due to illness (n=3), one filled out the pre SAQ anyhow.

††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).

‡‡Questionnaires not returned (n=6).

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

**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

MCQ test Per cent correct	Simulation intervention	Descriptive statistics		
		Pre-MCQ mean* (start of training day)	Post-MCQ mean* (end of training day)	Mean difference*†
	ISS	69.4 (65.4 to 73.4)	82.6 (79.3 to 85.8)	−0.02 (−2.13 to 2.09) p=0.98
	OSS	70.6 (66.0 to 75.2)	83.3 (80.4 to 86.1)	

\*Based on generalised estimating equations to account for potential correlation within teams.

†Adjusted for pre MCQ (ANCOVA).

ANCOVA, analysis of covariance; ISS, in situ simulation; MCQ, multiple choice question (range: 0–100%); OSS, off-site simulation.

OSS groups did not differ significantly. However, the two questions addressing the authenticity or fidelity of the simulations were scored significantly higher by the ISS participants compared with the OSS participants (table 7).

**TEAM:** No significant differences were found in the team scoring of performance between the ISS versus the OSS group (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).

**Organisational changes:** A qualitative analysis showed that more ideas for organisational changes were suggested by ISS participants than OSS participants. For details, see online supplementary table S1. The quantitative analysis, however, showed that participants in the ISS and 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 did OSS participants. This is in accordance with several non-randomised studies describing a positive impact of ISS on the organisation.<sup>8 10 11 13 59–61</sup>

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.<sup>62 63</sup> The conclusions from this trial, however, are in alignment with more recent empirical research and discussions on fidelity and learning.<sup>6 64–66</sup> Our study indicates that the change in simulation fidelity, that is change in setting for simulation, does not necessarily translate into learning. Another randomised trial, which

**Table 4** Means (95% CI) of SAQ (converted to percentages) for five dimensions 1 month before (pre-SAQ) and 1 month after (post-SAQ) the simulation training day with ISS and OSS

	Simulation intervention	Descriptive statistics		
		Pre-SAQ mean (1 month before)	Post-SAQ mean* (1 month after)	Mean difference*†
SAQ teamwork climate	ISS	80.5 (76.7 to 84.3)	81.1 (76.7 to 85.5)	−1.4 (−5.8 to 3.1) p=0.54
	OSS	78.4 (74.1 to 82.2)	81.2 (77.5 to 85.0)	
SAQ safety climate	ISS	66.7 (61.8 to 71.6)	70.6 (65.9 to 75.2)	1.6 (−2.0 to 5.1) p=0.39
	OSS	69.2 (65.4 to 73.0)	70.8 (66.8 to 74.8)	
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) p=0.74
	OSS	85.6 (81.6 to 89.6)	85.7 (81.9 to 89.5)	
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) p=0.44
	OSS	67.3 (61.2 to 73.3)	69.2 (64.0 to 74.4)	
SAQ work condition	ISS	66.4 (60.8 to 72.1)	64.9 (59.0 to 70.8)	−0.3 (−5.7 to 5.1) p=0.91
	OSS	65.9 (59.9 to 71.8)	64.0 (58.1 to 69.8)	

\*Based on generalised estimating equations to account for potential correlation within teams.

†Adjusted for pre SAQ (ANCOVA).

ANCOVA, analysis of covariance; ISS, in situ simulation; OSS, off-site simulation; SAQ, Safety Attitudes Questionnaire (range: 0–100%).

**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	Postsimulation 0 min Mean*	Postsimulation 5 min Mean*	Postsimulation 10 min Mean*	Peak-level mean*†	Mean difference*‡ of baseline to peak of $\Delta$ OSS versus $\Delta$ ISS
First simulation: emergency caesarean section							
Cortisol	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)	-0.2 (-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)	
Second simulation: postpartum haemorrhage							
Cortisol	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 (-0.1 to 0.25) 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 to 1.3) 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)	

\*Based on generalised estimating equations to account for potential correlation within teams.

†Peak level is the maximum of the measurements obtained at 0, 5 and 10 min after the end of the simulation.

‡Adjusted for pre-cortisol, pre-STAI and pre-CA (ANCOVA).

ANCOVA, analysis of covariance; CA, cognitive appraisal (range 0.1–10); ISS, in situ simulation; OSS, off-site simulation; STAI, Stress-Trait Anxiety Inventory (range 20–80).



**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)*
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

\*Based on generalised estimating equations to account for potential correlation within teams.

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

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.<sup>44 67</sup>

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 the recently introduced term sociological fidelity, which encompasses the relationship between the various healthcare professionals.<sup>37 68</sup> After completing the trial, we decided to explore more closely the experiences between the healthcare professionals in a qualitative study.<sup>69</sup>

Post hoc analyses showed similar educational effects in the ISS and 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.<sup>70 71</sup> 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.<sup>71</sup>

No differences were found in the mean SAQ score after simulation-based medical education in the ISS versus the OSS group. Earlier studies have described that high SAQ values mean that SAQ cannot be influenced by an intervention.<sup>72 73</sup> The values for SAQ were generally high in this trial compared with various other studies from non-Scandinavian countries.<sup>72–75</sup>

There were no differences in the stress level when measured as salivary cortisol levels, STAI and CA in the ISS versus the OSS group. 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 with other studies. Previously used among students and medical trainees,<sup>22 76 77</sup> CA appeared to have a less discriminatory effect in these more senior groups of healthcare professionals.

IMI<sup>24 51</sup> revealed no differences between ISS versus OSS. Motivation has not previously been tested in educational simulation studies, and it is argued that a gap appears to exist in the simulation literature on motivational factors and further research has been encouraged.<sup>25</sup> Some argue that simulation in the clinical setting, as with ISS, should increase motivation,<sup>14</sup> but this was not confirmed by findings in this trial.

The evaluation data showed no differences between ISS and OSS. Both the ISS and OSS participants gave very high scores on the evaluation. This is in accordance with what is generally seen in interprofessional training.<sup>78</sup>

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 scale has been found to be reasonably intuitive to use,<sup>52 53</sup> 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 and 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.<sup>58</sup> Previous non-randomised studies have suggested that ISS has an impact on organisations, but this has, to the best of our knowledge, never been confirmed in a randomised trial.<sup>8 11 13 17 59</sup>

### 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 dropouts; and was conducted on a not-for-profit basis.<sup>79–81</sup> 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.<sup>8</sup> 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

**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 the OSS group

	ISS Median (1st Q–3rd Q)	OSS Median (1st Q–3rd Q)	p Value*
<i>Evaluation questions (shortened version, original version in Danish)</i>			
1. Overall, 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
<i>Simulation of an emergency CS</i>			
7. Overall, 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
<i>Simulation in PPH</i>			
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 at all 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

\*Kruskal-Wallis rank sum test. p Values adjusted for multiple testing.

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

of simulation were measured after 1 week (evaluation and motivation) and safety attitudes after 1 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 seem to be of importance for the so-called sociological fidelity.<sup>37 68</sup> The teams in this trial were authentic in their design and hence resemble the real clinical setting in every possible way.<sup>65 82</sup> These kinds of teams are called ‘ad 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.<sup>83–85</sup> 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. The literature on retention of skills suggests that deterioration of the non-used skills appears to occur about 3–18 months after training. More research within the field of retention and on the effect of short booster courses is necessary.<sup>45 86–88</sup>

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

**Table 8** Mean (95% CI) of video assessment performance scores with the TEAM scale

	ISS Mean	OSS Mean	p Value*
<i>Video assessment scoring of performance</i>			
TEAM (means of item rating)	2.6 (2.3 to 3.0)	2.4 (2.1 to 2.8)	
Simulation in emergency CS*			
TEAM (means of item rating)	2.9 (2.5 to 3.2)	2.8 (2.5 to 3.2)	
Simulation in PPH*			
Estimated overall difference in mean between ISS and OSS†	0.1 (−0.2 to 0.5)		0.36
TEAM (global rating)	6.1 (4.8 to 7.3)	5.3 (4.0 to 6.5)	
Simulation in emergency CS*			
TEAM (global rating)	6.8 (5.5 to 8.1)	6.3 (5.0 to 7.6)	
Simulation in PPH*			
Estimated overall difference in mean between ISS and OSS†	0.7 (−0.4 to 1.7)		0.18
<i>Differences in video assessment scores of performance between emergency CS (1st) and PPH (2nd) simulation scenarios</i>			
Differences in mean of TEAM (means of item rating) of the simulation in emergency CS versus PPH†	0.3 (0.1 to 0.5)		0.0003
Differences in mean of TEAM (global rating) of the simulation in emergency CS versus PPH†	0.9 (0.3 to 1.5)		0.005

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.

\*Means found from a linear mixed model including an interaction between the simulation group (ISS and OSS) and simulation scenario (emergency CS and PPH).

†Overall difference in means found from an additive linear mixed model based on the simulation group and simulation scenario.

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).

of ISS is relevant or feasible and appears to have a clinically or educationally relevant effect.<sup>89</sup>

The improvements on knowledge and team performance may also be due to the Hawthorne effect, that is, due to individuals changing behaviour as a result of their awareness of being observed.<sup>90</sup> 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.<sup>90</sup> 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 most likely minimising the Hawthorne effect in our trial.<sup>90</sup>

## 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 this 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 authenticity or 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. This trial indicated that the physical fidelity of the setting seemed to be of less importance for learning; however, more research is necessary to better understand which aspects of simulation to be most important for learning.

## Author affiliations

<sup>1</sup>Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup>Faculty of Health, Medicine and Life Sciences, Department of Educational Development and Research, Maastricht University, Maastricht, The Netherlands

<sup>3</sup>Faculty of Health and Medical Sciences, Section of Biostatistics, Institute of Public Health, University of Copenhagen, Copenhagen, Denmark

<sup>4</sup>Danish Institute for Medical Simulation, Herlev Hospital, Capital Region of Denmark and University of Copenhagen, Herlev, Denmark

<sup>5</sup>University of Ottawa Skills and Simulation Centre, The Ottawa Hospital, University of Ottawa, Ottawa, Canada

<sup>6</sup>Department of Obstetrics, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>7</sup>Department of Anaesthesiology, Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>8</sup>Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>9</sup>Laboratory of Neuropsychiatry, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

<sup>10</sup>Juliane Marie Centre for Children, Women and Reproduction, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

**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. They would also 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 Todsén and Ninna Ebdrup for video recordings of simulations; and Tobias Todsén 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 CVdV 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 a 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 foundations had a role in the design or conduct of the study.

**Competing interests** All authors have completed the ICMJE uniform disclosure form at [http://www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf). The principal investigator and lead author (JLS) reports non-profit funding mentioned above. DØ reports board membership of Laerdal Foundation for Acute Medicine. Other authors declare no support from any organisation for the submitted work.

**Ethics approval** Participants were healthcare professionals, and neither patients nor patient data were used in the trial. Approvals 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.

**Provenance and peer review** Not commissioned; externally peer reviewed.

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

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

- 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.
- 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.
- 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:e1511–30.
- 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.
- 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.
- Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012;46:636–47.
- 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):i51–6.
- 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.
- 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):i53–6.
- 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.
- Walker ST, Sevdalis N, McKay A, *et al.* Unannounced in situ simulations: integrating training and clinical practice. *BMJ Qual Saf* 2013;22:453–8.
- 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.
- 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.
- 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.
- 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.
- Rubio-Gurung S, Putet G, Touzet S, *et al.* In situ simulation training for neonatal resuscitation: an RCT. *Pediatrics* 2014;134:e790–7.
- 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.
- LeBlanc VR. The effects of acute stress on performance: implications for health professions education. *Acad Med* 2009;84:S25–33.
- 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):S24–9.
- Arora S, Sevdalis N, Aggarwal R, *et al.* Stress impairs psychomotor performance in novice laparoscopic surgeons. *Surg Endosc* 2010;24:2588–93.
- 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.
- 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.
- 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.
- 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:e242–62.
- 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.
- Ryan RM, Deci EL. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemp Educ Psychol* 2000;25:54–67.
- Drife J. Reducing risk in obstetrics. *Qual Health Care* 1995;4:108–14.
- 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.
- 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. 2014. <http://www.laerdal.com/dk/SimMom>.
39. Royal College of Obstetricians and Gynaecologists course: PROMPT, Train the trainers. 2015. <https://www.rcog.org.uk/globalassets/documents/courses-exams-and-events/training-courses/prompt-2—train-the-trainers-mar2015.pdf>.
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:e421–44.
42. Sørensen JL, Thellens 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, CA: Consulting Psychologists Press, 1970.
50. Bech P. [Klinisk psykometri]. 1.udgave, 2 oplag ed. Munksgaard Danmark, København, Danmark, 2011.
51. Self-Determination Theory (SDT). 2015. <http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/>.
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. *J Roy Statist Soc Ser 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):i50–8.
59. Sørensen 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. *BMJ Open* 2015;5:e008345.
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;121:1710–18.
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: Doman T, Mann K, Scherpbier A, *et al.*, eds. *Medical education: theory and practice*. Edinburgh, London, New York, Oxford, Philadelphia, St Louis, Sydney, Toronto: Churchill Livingstone Elsevier, 2011:51–64.
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, eds. *The question of competence. Reconsidering medical education in the twenty-first century*. 1st edn. New York: ILR Press an imprint of Cornell University Press, 2012:42–69.





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, Zuihthoff P, *et al.* Predictive validity of a written knowledge test of skills for an OSCE in postgraduate training for general practice. *Med Educ* 2002;36:812–19.
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. *Hum Perform* 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: Fraenkal JR, Wallen NE, Hyde A, eds. *How to Design and Evaluate Research in Education*. New York: McGraw-Hill, 2012:166–83.