

# The Effect of Simulation-Based Neonatal Emergency Team Training on Clinical Performance and Patient Outcome: A Systematic Review

Morten S. Lindhard (✉ [mortjs@rm.dk](mailto:mortjs@rm.dk))

Regional Hospital of Randers <https://orcid.org/0000-0002-5685-8845>

**Signe Thim**

Aarhus University Hospital

**Henrik Sehested Laursen**

Medical Library, Regional Hospital Central Jutland

**Anders Wester Schram**

MidtSIM, Corporate HR, Central Region Denmark

**Charlotte Paltved**

MidtSIM, Corporate HR, Central Region Denmark

**Tine Brink Henriksen**

Aarhus University Hospital

---

## Research

**Keywords:** Simulation, team, training, neonatology, newborn, patient outcome, medical education

**Posted Date:** October 26th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-95526/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

## Background

A number of neonatal simulation-training programmes have been deployed during the last decade, and a growing number of studies have investigated effects of simulation-based team training. However, the body of evidence remains to be compiled. Therefore, we performed a systematic review on the effects of simulation-based team training on clinical performance and patient outcome.

## Methods

The review was conducted according to the preferred reporting items for systematic review and meta-analysis (PRISMA). We included studies on team training in emergency neonatal settings with reported outcome on clinical performance and patient outcome. Two reviewers independently selected articles and assessed risk-of-bias using the Cochrane risk-of-bias tool 2.0 and the Newcastle-Ottawa quality assessment scale. Kirkpatrick's model for evaluation of training programs provided the framework for a narrative synthesis.

## Results

We screened 1,434 titles and abstracts, evaluated 173 full-texts for eligibility, and included 24 studies. We identified only two studies with neonatal mortality outcome, and they had significant methodological limitations, and no conclusion could be reached regarding effects of simulation training in developed countries. Considering clinical performance, randomized studies showed improved team performance in simulated re-evaluations 3 and 6 months after the intervention.

## Conclusions

Simulation-based team training in neonatal resuscitation improves team performance and technical performance in simulation-based evaluations 3 to 6 months later. The current evidence was insufficient to conclude on neonatal mortality after simulation-based team training, since no studies were available from developed countries. Future research should include patient outcomes or clinical proxies of treatment quality whenever possible.

# Background

It is estimated that less than 1% of all newborns will need extensive neonatal resuscitation in the delivery room (1). The individual health care professional will therefore rarely experience this, and even more rarely a specific team of professionals will experience this together. In 2004, the Joint Commission for the Accreditation of Healthcare Organizations published a sentinel event alert indicating that ineffective communication within the neonatal resuscitation team played a role in almost three-quarters of perinatal deaths or permanent disabilities (2).

Before 2010 the Neonatal Resuscitation Program (NRP) have focussed on the acquisition of knowledge and technical skills pertinent to neonatal resuscitation (3,4). The sixth edition (2010) of NRP transitioned from instructor driven didactics and skills stations to interactive and simulation-based learning (4). The seventh edition (2016) introduced more focus on communication and team behaviours to the curriculum (4). Ten desired behavioural skills were adapted from crisis resource management principles (4). Optimal behaviour can be challenging in a high-stakes time sensitive critical situation. Thus, simulation training needs to expand beyond technical skill acquisition, and to use simulated environments to study human and system performance (5).

We were unable to identify a review on simulation-based team training in neonatal resuscitation and emergency situations to answer our questions: Does simulation-based team training improve the performance of the team? Does it improve patient outcome and safety? We therefore aimed to perform a comprehensive and high-quality systematic review in order to describe the current state of evidence, and to point out areas where more research is needed to pave way for future improvements of neonatal emergency team training and patient safety.

# Methods

We conducted and report this review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (6). We registered a study protocol at the International Prospective Register of Systematic Reviews (PROSPERO) repository (CRD42019128213); it was submitted June 24, 2019 and published September 6, 2019 (7).

## Study eligibility criteria

We included studies of all designs if they met the criteria stated below.

## *Population*

We included studies of active health care providers, e.g. nurses, doctors, midwives, and respiratory therapists, with clinical responsibilities in the delivery room, the neonatal intensive care unit (NICU), or in other hospital settings with emergency care for newborns. Studies focusing on pre-graduate simulation training were excluded.

### ***Intervention***

The intervention was simulation-based team training of neonatal clinical emergencies in situ or in a training facility. We defined team training as two or more health care providers in a critical situation requiring a coordinated effort for a successful outcome.

### ***Comparators***

We included studies comparing health care providers' performance before (comparator) and after simulation training (intervention) in pre-post designs with no control group. Studies of team skills retention with several evaluations over time (two or more) were also included. In randomized and non-randomized cohort studies the control group could be training or teaching as usual or no training at all. We also accepted comparators with different types or intensities of the simulation training.

### ***Outcomes***

We evaluated simulation training outcomes according to the framework of Kirkpatrick's four levels of training evaluation and transfer of learning to behaviour (8). Given our aim, we did not include reaction outcomes (level I), corresponding to providers' evaluation of the simulation training. Learning outcomes (level II) were included as self-reported changes in knowledge, attitude, confidence, preparedness, self-efficacy, and technical- and non-technical skills. Technical skills were defined as: "Adequacy of the actions taken from a medical and technical perspective"; non-technical skills were defined as: "Decision-making and team interaction processes used during the team's management of a scenario or a clinical situation" (9). Behaviour outcomes or clinical performance (level III) were included as observed technical and non-technical skills in the simulation setting or in the clinical setting. Patient outcomes (level IV) were included as a change in clinical parameters (e.g. time to critical task or action) or a change in patient outcomes (e.g. survival) of neonatal emergencies.

### **Search strategy**

We conducted a combined search aiming to identify all papers on simulation-based team training in neonatology and emergency paediatrics as simulation-training programs may include both neonatal cases and paediatric (non-neonatal) cases. An experienced medical librarian (HSL) and a subject specialist (MSL) devised a search strategy for Medline, Embase, CINAHL and Cochrane Library. Studies were limited to English language, however there was no limit on year of publication. Details of the search strategy are available in Additional file 1. The final search was run on March 6, 2019. Reference lists of the included studies were scrutinized to identify additional studies; identified studies were also subjected to the selection process as detailed below. After data extraction the papers were divided into a neonatal review and a paediatric (non-neonatal) review; two studies provided relevant data on both neonatal and non-neonatal simulations and were included in both reviews (Figure 1) (10,11).

### **Study selection**

Two authors (ST and MSL) independently screened titles and abstract of all identified studies. Studies meeting the inclusion criteria or potentially meeting the inclusion criteria, as well as studies with insufficient information, were included for full-text review. Any disagreements were resolved by discussion, and abstracts with non-consensus were also included for full-text review. Two authors (ST and MSL) independently performed full-text review of all included manuscripts. Any disagreements were resolved by discussion to consensus, or by consulting a third author in case of non-consensus (TBH). The study selection process was managed using the Covidence review platform (12), which facilitates co-reviewer blinded abstract- and full text screening, study inclusion, and resolution of conflicts.

### **Data extraction**

Two authors (AWS and MSL) extracted data using a predefined template developed for the purpose. We did not contact authors of studies with missing or inadequate information. The following information was extracted from the included studies; author, year, country, simulation setting, study design, intervention, comparator, number of participants, professions of participants, participant/instructor ratio, simulation and debriefing duration, simulation frequency (if applicable), level of fidelity, in situ or simulation centre, re-test timing (if applicable), self-reported outcomes, observed simulation outcomes, observed clinical behaviour outcomes, change in clinical parameter outcomes, change in patient outcomes.

### **Risk of bias in individual studies**

Two authors (AWS and MSL) evaluated randomized studies according to the Cochrane risk-of-bias tool for randomized trials (ROB 2) (13). Separate score tools were applied for individually randomized parallel group trials and for cluster randomized trials. Each study received an overall risk-of-bias judgement of "low", "some concern", or "high". We evaluated non-randomized studies according to the Newcastle Ottawa quality assessment scale (NOS) (14). We used the NOS with the adaptations and operational definitions for evaluating medical education research as suggested by Cook and Reed (15). An overall score of 0 (low quality) to 6 (high quality) was assigned each study. The risk of bias evaluation was presented for all studies and used for the synthesis and discussion of the results. Any disagreements were resolved by discussion to consensus, or by consulting a third author in case of non-consensus (TBH). The bias evaluation had no impact on study inclusion or exclusion.

## **Synthesis of results**

We conducted a narrative synthesis of the included studies due to heterogenous interventions and outcomes. No single summary measure was applicable across studies. We used the guidance by Popay et al. (16), and focused the main synthesis based on the hierarchical Kirkpatrick levels; results (effects on patient outcomes); behaviour (effects on clinical performance); and learning (effects on knowledge). We presented all available studies for each level to provide full transparency. We then emphasized high quality studies in a narrative synthesis, but also integrated the remaining studies at hand.

We used a funnel plot to assess potential publication bias. We used the estimate from the main outcome of each included study. Ratio estimates (ORs, RRs, IRRs) were used directly, whereas a ratio was calculated for studies using continuous outcomes, e.g. intervention mean divided by comparator mean. A ratio above 1.0 favoured the intervention group or the post-intervention estimate.

We assessed selective reporting by inspecting pre-registered study or analysis protocols when available; for randomized studies this was part of the ROB 2 score. Further, we compared the specified analysis from the methods section with the reported outcomes in each included study.

## **Results**

### **Study selection**

A total of 2,315 studies were identified across databases. After duplicate removal, we screened titles and abstracts of 1,434 studies. A total of 173 full-text articles were assessed for eligibility. Of these, 88 met the inclusion criteria common to neonatal and paediatric populations, and 24 provided relevant data for this neonatal review (Figure 1). A list of excluded full-text manuscripts is provided in Additional file 2.

### **Study characteristics**

Included studies were published between 2008 and 2018. Most were conducted in developed countries; however, one study from Kenya, one from Guatemala, one from Lebanon and two from Mexico were included (Table 1). Nine studies had a control group, six of these used random allocation. The remaining 15 studies had a single group pre-post intervention design. Two studies provided patient outcome (Kirkpatrick level IV), 14 provided clinical performance outcome (level III), and 15 provided learning outcome (level II). Number of participating health care providers ranged from 16 to 305 (Table 1).

### **Risk of bias within studies**

The risk-of-bias judgements of the 6 randomized studies are presented in Table 2. Five studies received an overall judgement of "some concern" (17–21). One study received "high-risk" due to three domains with some concern, including the randomization domain (22). All but one study had low risk of bias in the missing outcome domain, reflecting limited dropout from the short-term educational interventions. Two studies had published their trial protocols (20,22), which was necessary to obtain "low risk" in the selected results domain.

The risk-of-bias scores for 18 non-randomized studies are presented in Table 3. As in the randomized studies the dropout was low in all studies. Two of three cohort studies received an overall risk-of-bias score of 4 (out of 6) (23,24). The third cohort study scored 2 due to suboptimal description of the intervention and control groups, and to non-blinded outcome assessment (25).

### **Results and synthesis of patient outcome (Kirkpatrick level IV)**

Two studies included patient outcomes (Table 1). Walker et al. conducted a cluster randomized study of 12 intervention hospitals matched (number of births, caesarean rate, mortality, complications, and number of operating rooms) with 12 control hospitals in Mexico (22). Hospitals with higher than average maternal mortality were selected from a list of government run facilities with 500-3,000 annual births. Intervention hospitals received 2+1 full days of training including interactive team and communication exercises, skills sessions, and in situ simulation of obstetric and neonatal emergencies. Control hospitals received no training during the study period. In total 305 (9%) of 3228 eligible health care professionals (nurses and doctors) received both training modules during 2010-2012. Statistical analysis adjusted for matching and presence

of a NICU due to imbalance at baseline; 83% of control hospitals and 42% of intervention hospitals. Incidence of hospital-based neonatal mortality tended to be lower at the intervention hospitals; IRR (4 months) 0.73 (95% CI 0.45-1.17), IRR (8 months) 0.59 (0.37-0.94), IRR (12 months) 0.83 (0.50-1.37). However, we were concerned about the validity because the study received an overall high risk-of-bias judgement (Table 2). The intended primary outcome of perinatal mortality was changed to in-hospital neonatal mortality due to poor reporting of stillbirths (not further specified). The intervention covered both obstetric and neonatal emergencies, thus any change in mortality really reflects the combined perinatal emergency care, not neonatal resuscitation alone.

Charafeddine et al. conducted a single group pre-post intervention study at 22 hospitals in Lebanon (26). The Hospitals were part of a larger National Collaborative Perinatal-Neonatal Network (NCPNN) covering 32 hospitals. Intervention was an 8-hour session including 40 minutes teaching including the NRP algorithm, hands-on simulation on low-fidelity manikins, and finally "megacode" simulations including all steps of neonatal resuscitation. Some 256 professionals (doctors, nurses, midwives) were trained during 2009-2011; the selection process and participation rate was not described. Patient outcomes were retrieved from surveillance data on mortality at hospital discharge and neonatal morbidity from the NCPNN network. The first intervention year (2009) was chosen as reference; mortality odds ratio (OR) decreased steadily from 1.53 (95% CI 1.18-1.98) in 2006 to 0.72 (0.54-0.96) in 2013. Furthermore, the years 2011-2013 had fewer infants requiring oxygen at birth, bag and mask ventilation, intubation, and chest compressions, compared with 2009. The study obtained a low NOS score of 1 out of maximum 3 for a study with no control group and non-blinded outcome ascertainment, indicating risk of bias. There was no presentation of neonatal mortality rates for the 10 non-participating network hospitals likely also contributing surveillance data. Thus, we are concerned that other factors than the simulation training may explain the observed change in mortality. We acknowledge that the authors also do not emphasize this finding, but rather the participants' change in knowledge (included below).

In summary of patient outcomes, we identified one randomized study and one single group pre-post study that reported a measure of neonatal mortality (22,26). Both studies were from developing countries and indicated lower hospital-based neonatal mortality after simulation-based training. Both studies had a high risk of bias, and the randomized study by Walker et al. also included obstetric emergency training, thus hampering interpretation of effects of neonatal team training.

### **Results and synthesis of clinical performance (Kirkpatrick level III)**

We included 14 studies with clinical performance outcomes (Table 4). Eight studies had a control group, five with random allocation. Six studies were single group pre-post design. One study by LeFlore et al. simulated neonatal transport cases (24), the rest simulated neonatal (delivery room) resuscitation.

Two randomized studies evaluated the effects of simulation-based team training after approximately 3 months (18,19). Rubio-Gurung et al. conducted a large cluster randomized study in 12 hospitals in France (19). They compared 4-hour high fidelity in situ multidisciplinary team trainings of 6 professionals with no simulation training. They trained 80% of the delivery room staff within 1 month, amounting to 202 professionals in 6 intervention hospitals. Simulation-based evaluations were conducted for a random sample of professionals before (n= 116) and after (n= 114) intervention. No differences in baseline evaluations were observed. Significant improvements were demonstrated 3 months later for technical skills, team performance and global performance (Table 4). The study received a low risk-of-bias judgement in 5 of 6 domains (Table 2). Overall, we consider the study by Rubio-Gurung et al. important and well conducted. Lee et al. conducted a randomized study of 27 emergency medicine residents (18). They were randomized to a 4-hour high fidelity simulation-based session (45 min. didactics) on neonatal resuscitation, or the standard emergency medicine curriculum, including monthly paediatric (occasional neonatal) simulations. Baseline data were similar in both groups. Simulation-based evaluation after 16 weeks demonstrated no change in the neonatal resuscitation score for the control group, but a 12-percentage points improvement in the intervention group (Table 4). The intervention group also significantly reduced the time to warm, dry, stimulate, and hat on the infant compared to controls (p= 0.017). The study received an overall risk-of-bias judgement of some concern (Table 2). The randomized studies by Rubio-Gurung et al. and Lee et al. supports improved team performance and technical skills 3 months after simulation-based team training (18,19).

Two randomized studies extended re-testing to 6 months (20,21). Sawyer et al. conducted a study of 30 residents randomized to either standard oral debriefing or video-assisted debriefing at 3 high-fidelity simulation sessions approximately 2 months apart (21). Baseline data were similar in both groups. No significant differences in neonatal resuscitation performance score and time to perform critical actions were observed at 6 months comparison of standard oral and video-assisted debriefing groups (Table 4). The study received a low risk-of-bias judgement in 4 of 5 domains (Table 2). Thomas et al. randomized 98 interns to either standard NRP training (comparator) or NRP plus 2 hour session on communication and teamwork and low- (intervention group 1) or high (intervention group 2) fidelity simulation (20). At 6 months follow-up the intervention groups (analysed together to increase power) exhibited more teamwork behaviours per minute (11.8) than controls (10.0) (p=0.03). However, no differences were observed for NRP performance score, duration of resuscitation, vigilance, or workload management. The study received low risk-of-bias in 3 of 5 domains (Table 2).

Bender et al. investigated whether a NRP booster at 9 months improved performance at 15 months evaluation (17); 50 residents were randomized to either a half-day NRP booster with high-fidelity simulations (intervention) or to routine clinical duties (comparator). At 15-months evaluation the intervention group scored higher on both technical score and team performance score (Table 4). The study was of some risk-of-bias concern (Table 2).

Two cohort studies explored minor interventions related to simulation-based team training. Rovamo et al. studied 99 doctors, nurses and midwives on a one day high-fidelity neonatal resuscitation course (23). Both intervention and control groups had the same simulation training, but in addition the intervention group received a 1-hour interactive lecture on crisis resource management (CRM) and anaesthesia non-technical skills principles. There was no difference in team performance score in the two groups after the lecture (Table 4). The study received a NOS bias score of 4 of 6, indicating low to moderate risk-of-bias (Table 3). LeFlore et al. studied a neonatal transport team over 2 years (24); the first year they trained with high fidelity simulation and self-paced modular learning (control), the second year they used high fidelity simulation and expert modelled learning (intervention). Some, but not all team members participated both years. There was no significant change in team performance score (Table 4). The study received a NOS bias score of 4 of 6, indicating low to moderate risk-of-bias (Table 3).

Barry et al. studied a group of 28 first year residents (intervention) and compared them to a group of 24 senior residents (control) (25). The intervention was half-a-day equipment workshop and in situ simulation-based team training. The control group was senior residents with NRP course and routine clinical duties. Re-testing was done after 1 month, and after 1-2 years. The intervention group's global performance score increased from a lower level before training to the same level as the senior residents after training (Table 4). The study received a NOS score of 2 of 6 indicating moderate to high concern of bias (Table 3).

Dadiz et al studied 228 perinatal health care professionals over a 3-year period (27); 90-minute multidisciplinary high-fidelity trainings were conducted in a simulated delivery room. Over the years, increasing communication checklist scores was observed (Table 4). The study received a NOS risk-of-bias score of 2 (Table 3). Five single group pre-post design studies of simulation-based team training by Walker et al., Sawyer et al., and Cordero et al. observed improved team performance scores 0-6 months later (Table 4) (28–32). NOS risk-of-bias scores ranged from 1-3 of 6 (Table 3).

In summary of clinical performance, randomized studies showed effects of team training in simulated re-evaluations after 3 and 6 months. Booster simulation sessions 9 months after NRP improved performance at 15 months evaluation. One randomized study showed no differences in team performance comparing standard oral debriefing and video-assisted debriefing. One single-group study showed steadily improving communication performance during a 3-year intervention with yearly simulation training. Six smaller single group studies showed that simulation-based team training improved team performance scores 0-6 months later.

## **Results and synthesis of learning (Kirkpatrick level II)**

Two small studies with random allocation to the intervention and control group presented self-reported outcome on knowledge and confidence. Bender et al. observed no significant difference in knowledge 15 months after a simulation-based NRP boost at nine months (17). Lee et al. observed significant improvements in confidence in neonatal resuscitation after 16 weeks in both the intervention and control group, but no statistically significant difference between groups (18). Both studies had limited power to detect a difference.

A total of 13 studies with single group design presented self-reported learning outcome, we briefly summarize the findings of 7 studies with more than 50 participants (Table 1). They all evaluated outcome immediately after the simulated neonatal resuscitation team training intervention or within a 2-3 months period. Self-assessed improvements were reported for neonatal resuscitation knowledge (26,28,33–35), self-efficacy (28,34), communication (33,36), and leadership, confidence and technical skills (36). Dadiz et al. specifically trained and studied delivery room communication, and interestingly the health care professionals reported significant improvements in team communication in real clinical situations over a 3-year study period (27).

In summary of learning outcomes, the single-group design studies all reported significant improvements in self-reported outcomes, but 2 small randomized studies found no difference in improvements between the intervention and control groups.

## **Risk of bias across studies**

Within each group of studies (according to Table 1) the funnel plots were quite symmetric, and no major concern about publication bias was raised (Figure 2). We found no indication of selective reporting bias, as methods section and reported result were coherent in all studies. Pre-published study protocols were available for only 2 of 6 randomized studies, which was reflected in the risk-of-bias score.

# **Discussion**

## **Summary of evidence**

We encountered an evolving research field with the earliest included study published in 2008 and the majority after 2012. One main finding of this review was the lack of evidence to support effects of team training on patient outcome (Kirkpatrick level IV). We identified only one randomized study and one single group pre-post study that reported a measure of neonatal mortality (22,26). Both studies indicated lower hospital-based neonatal mortality after simulation-based training. However, both studies had a high risk of bias. The randomized study by Walker et al. also included obstetric emergency training, which complicates interpretation of isolated effects of neonatal resuscitation team training. Thus, the evidence from the two studies was inconclusive, and they were conducted in developing countries.

Another main finding based on five randomized studies was improved team performance (Kirkpatrick level III) in simulated re-evaluations 3 to 6 months after the intervention simulation training (18–20). Booster simulation sessions 9 months after NRP improved performance at 15 months evaluation (17). The study by Rubio-Gurung et al. stands out by being sufficiently powered (n= 114), well designed, and by evaluating technical and team performance in a simulation-based setting before and after intervention (19). They were thus able to compare changes in neonatal resuscitation team performance between intervention and control hospitals using randomly selected providers available on the day of evaluation. Three of the randomized studies were small (n ≤ 50) and had limited power (17,18,21).

## **Strengths**

This systematic review applied a comprehensive search strategy in four medical databases by an experienced medical librarian. We followed a pre-specified study protocol registered in the PROSPERO repository. Two reviewers independently screened and selected studies performed data extraction and risk-of-bias scoring. We presented the data according to the PRISMA guidelines. The funnel plot was not indicative of publication bias.

## **Limitations**

Meta-analysis was impossible due to heterogenous interventions and outcomes. Instead, we performed a narrative synthesis structured by outcome Kirkpatrick level and with emphasis on studies of high quality. For transparency, every included study was presented and cited (Table 1). Although this process is less standardized than a meta-analysis, we do consider it reproducible and open to scrutiny.

The majority of included studies had significant methodological limitations. Fifteen of 24 studies had no control group, which is concerning when using a self-reported endpoint or a simulation-based endpoint; familiarization with the simulation setting may improve performance during subsequent simulations. Most non-randomized studies had inadequate adjustment for potential confounding factors such as years of clinical experience, profession, team composition, age and gender. Small sample size (n< 50 in 12 studies) may likely have resulted in inadequate power and inability to perform adjusted statistical analyses.

## **Recommendations for future research**

Measuring effects of simulation-based training on clinical outcomes of neonates should be preferred whenever possible. Newborn morbidity and mortality is obviously of primary interest, but also clinical information that may serve as proxy measures of treatment quality, for example time to critical tasks during neonatal resuscitation. We acknowledge that such measures are difficult and time consuming to obtain given the paucity and volatile nature of real-life critical events. We advocate for the use of control groups when designing new studies, and the use of random allocation to intervention and control whenever possible. The study protocol should be published, to avoid issues of selective reporting. When applying simulation-based evaluation of the impact of training, use video recordings and blinded scoring by accepted and validated protocols. Compared to planned simulation-based re-testing, the use of unannounced simulated mock codes may mimic real clinical encounters more closely, because clinical professionals often prepare for planned testing, which may bias results. Describe the theoretical framework and process of debriefing, as this is an important part of simulation-based training that impacts the learning outcome.

## **Conclusion**

This systematic review compiles the first decade of research on simulation-based team training in neonatal medicine emergencies. We were unable to reveal effects of team training on neonatal morbidity and mortality, as we identified only two studies both conducted in developing countries and both with significant methodological limitations. However, five randomized studies showed improved team performance in simulation-based re-evaluations 3 to 6 months after the intervention simulation training. Future research should include patient outcomes or clinical proxy measures of treatment quality whenever possible.

## **Abbreviations**

In order of appearance:

NRP: Neonatal Resuscitation Program

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

PROSPERO: Prospective Register of Systematic Reviews

NICU: Neonatal intensive care unit

ROB 2: Cochrane risk-of-bias tool for randomized trials

NOS: Newcastle Ottawa quality assessment scale

OR: Odds ratio

RR: Risk ratio

IRR: Incidence rate ratio

NCPNN: National Collaborative Perinatal-Neonatal Network

CRM: Crisis resource management

## Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: Not applicable.

Competing interests: All authors declare that they have no competing interests.

Funding: The study was supported by Corporate HR, MidtSim, Central Region Denmark. The funding body had no influence on the design of the study, data collection, analysis, interpretation of data, manuscript drafting or conclusions.

Authors' contributions: CP initiated the study and provided intellectual support for all parts of it. MSL, ST and TBH designed and planned the study. HSL and MSL specified the literature search and HSL conducted and documented the search. MSL and ST conducted the abstract and full-text screening for manuscript inclusion. MSL and AWS conducted risk-of-bias scoring and data extraction process. MSL drafted the first manuscript version. All authors provided intellectual contribution and read and approved the final manuscript.

Acknowledgements: We acknowledge and appreciate the scientific contribution of the authors of all manuscripts included in this review.

## References

1. Wyckoff Myra H., Aziz Khalid, Escobedo Marilyn B., Kapadia Vishal S., Kattwinkel John, Perlman Jeffrey M., et al. Part 13: Neonatal Resuscitation. *Circulation*. 2015;132(18\_suppl\_2):S543–60.
2. Sentinel Event Alert, Issue 30: Preventing infant death and injury during delivery. 2004. [https://www.jointcommission.org/sentinel\\_event\\_alert\\_issue\\_30\\_preventing\\_infant\\_death\\_and\\_injury\\_during\\_delivery/](https://www.jointcommission.org/sentinel_event_alert_issue_30_preventing_infant_death_and_injury_during_delivery/). Accessed 20 Sep 2019.
3. Halamek LP. The simulated delivery-room environment as the future modality for acquiring and maintaining skills in fetal and neonatal resuscitation. *Semin Fetal Neonatal Med*. 2008;13(6):448–53.
4. Ades A, Lee HC. Update on simulation for the Neonatal Resuscitation Program. *Semin Perinatol*. 2016;40(7):447–54.
5. Halamek LP. Simulation and debriefing in neonatology 2016: Mission incomplete. *Semin Perinatol*. 2016;40(7):489–93.
6. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):6.
7. Thim S, Henriksen TB, Laursen H, Paltved C, Lindhard MS. Simulation-based team training in emergency pediatrics and neonatology – a systematic review. PROSPERO. 2019. [https://www.crd.york.ac.uk/prospero/display\\_record.php?ID=CRD42019128213](https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019128213). Accessed 17 Sep 2019.
8. Kirkpatrick, D.L. K JD. Evaluating training programs: The four levels. 3rd ed. San Francisco, CA: Berrett-Koehler Publishers inc.; 2006.
9. Gaba DM MD, Howard SK MD, Flanagan B FANZCA, Smith BE MD, Fish KJ MD, Botney R MD. Assessment of Clinical Performance during Simulated Crises Using Both Technical and Behavioral Ratings. *Anesthesiol J Am Soc Anesthesiol*. 1998;89(1):8–18.

10. Ross J, Rebella G, Westergaard M, Damewood S, Hess J. Simulation Training to Maintain Neonatal Resuscitation and Pediatric Sedation Skills for Emergency Medicine Faculty. *WMJ Off Publ State Med Soc Wis.* 2016;115(4):180–4.
11. Bragard I, Seghaye M-C, Farhat N, Solowianiuk M, Saliba M, Etienne A-M, et al. Implementation of a 2-Day Simulation-Based Course to Prepare Medical Graduates on Their First Year of Residency. *Pediatr Emerg Care.* 2018;34(12):857–61.
12. Covidence. Melbourne, Australia. <https://www.covidence.org>. Accessed 18 Aug 2019.
13. Higgins J, Savović J, Page M, Elbers R, Blencowe N, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ.* 2019;366:l4898
14. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. Ottawa Hospital Research Institute. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. [http://www.ohri.ca/programs/clinical\\_epidemiology/oxford.asp](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). Accessed 20 Aug 2019.
15. Cook DA, Reed DA. Appraising the Quality of Medical Education Research Methods: The Medical Education Research Study Quality Instrument and the Newcastle–Ottawa Scale-Education. *Acad Med.* 2015;90(8):1067–76.
16. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. Version 1. A Product from the ESRC Methods Programme; 2006.
17. Bender J, Kennally K, Shields R, Overly F. Does simulation booster impact retention of resuscitation procedural skills and teamwork?. *J Perinatol Off J Calif Perinat Assoc.* 2014;34(9):664–8.
18. Lee MO, Brown LL, Bender J, Machan JT, Overly FL. A medical simulation-based educational intervention for emergency medicine residents in neonatal resuscitation. *Acad Emerg Med Off J Soc Acad Emerg Med.* 2012;19(5):577–85.
19. Rubio-Gurung S, Putet G, Touzet S, Gauthier-Moulinier H, Jordan I, Beissel A, et al. In situ simulation training for neonatal resuscitation: an RCT. *Pediatrics.* 2014;134(3):e790-7.
20. Thomas EJ, Williams AL, Reichman EF, Lasky RE, Crandell S, Taggart WR. Team Training in the Neonatal Resuscitation Program for Interns: Teamwork and Quality of Resuscitations. *Pediatrics.* 2010;125(3):539–46.
21. Sawyer T, Sierocka-Castaneda A, Chan D, Berg B, Lustik M, Thompson M. The effectiveness of video-assisted debriefing versus oral debriefing alone at improving neonatal resuscitation performance: a randomized trial. *Simul Healthc J Soc Simul Healthc.* 2012;7(4):213–21.
22. Walker DM, Cohen SR, Fritz J, Olvera-Garcia M, Zelek ST, Fahey JO, et al. Impact Evaluation of PRONTO Mexico: A Simulation-Based Program in Obstetric and Neonatal Emergencies and Team Training. *Simul Healthc J Soc Simul Healthc.* 2016;11(1):1–9.
23. Rovamo L, Nurmi E, Mattila M-M, Suominen P, Silvennoinen M. Effect of a simulation-based workshop on multidisciplinary teamwork of newborn emergencies: an intervention study. *BMC Res Notes.* 2015;8(101462768):671.
24. LeFlore JL, Anderson M. Effectiveness of 2 methods to teach and evaluate new content to neonatal transport personnel using high-fidelity simulation. *J Perinat Neonatal Nurs.* 2008;22(4):319–28.
25. Barry JS, Gibbs MD, Rosenberg AA. A delivery room-focused education and deliberate practice can improve pediatric resident resuscitation training. *J Perinatol.* 2012;32(12):920–6.
26. Charafeddine L, Badran M, Nakad P, Ammar W, Yunis K. Strategic assessment of implementation of neonatal resuscitation training at a national level. *Pediatr Int Off J Jpn Pediatr Soc.* 2016;58(7):595–600.
27. Dadiz R, Weinschreider J, Schriefer J, Arnold C, Greves CD, Crosby EC, et al. Interdisciplinary simulation-based training to improve delivery room communication. *Simul Healthc J Soc Simul Healthc.* 2013;8(5):279–91.
28. Walker D, Cohen S, Fritz J, Olvera M, Lamadrid-Figueroa H, Cowan JG, et al. Team training in obstetric and neonatal emergencies using highly realistic simulation in Mexico: impact on process indicators. *BMC Pregnancy Childbirth.* 2014;14(100967799):367.
29. Sawyer T, Laubach VA, Hudak J, Yamamura K, Pocrnich A. Improvements in teamwork during neonatal resuscitation after interprofessional TeamSTEPPS training. *Neonatal Netw NN.* 2013;32(1):26–33.
30. Cordero L, Hart BJ, Hardin R, Mahan JD, Nankervis CA. Deliberate Practice Improves Pediatric Residents' Skills and Team Behaviors During Simulated Neonatal Resuscitation. *Clin Pediatr (Phila).* 2013;52(8):747–52.
31. Cordero L, Hart BJ, Hardin R, Mahan JD, Giannone PJ, Nankervis CA. Pediatrics residents' preparedness for neonatal resuscitation assessed using high-fidelity simulation. *J Grad Med Educ.* 2013;5(3):399–404.
32. Sawyer T, Sierocka-Castaneda A, Chan D, Berg B, Lustik M, Thompson M. Deliberate practice using simulation improves neonatal resuscitation performance. *Simul Healthc J Soc Simul Healthc.* 2011;6(6):327–36.
33. Dettinger J.C., Kamau S., Calkins K., Cohen S.R., Cranmer J., Kibore M., et al. Measuring movement towards improved emergency obstetric care in rural Kenya with implementation of the PRONTO simulation and team training program. *Matern Child Nutr.* 2018;14(Supplement 1):e12465.

34. Walker DM, Holme F, Zelek ST, Olvera-Garcia M, Montoya-Rodriguez A, Fritz J, et al. A process evaluation of PRONTO simulation training for obstetric and neonatal emergency response teams in Guatemala. *BMC Med Educ.* 2015;15(101088679):117.
35. Letcher DC, Roth SJ, Varenhorst LJ. Simulation-Based Learning: Improving Knowledge and Clinical Judgment Within the NICU. *Clin Simul Nurs.* 2017;13(6):284–90.
36. Malmstrom B, Nohlert E, Ewald U, Widarsson M. Simulation-based team training improved the self-assessed ability of physicians, nurses and midwives to perform neonatal resuscitation. *Acta Paediatr Oslo Nor 1992.* 2017;106(8):1273–9.
37. Raffaelli G, Ghirardello S, Vanzati M, Baracetti C, Canesi F, Conigliaro F, et al. Start a Neonatal Extracorporeal Membrane Oxygenation Program: A Multistep Team Training. *Front Pediatr.* 2018;6(101615492):151.
38. Hossino D, Hensley C, Lewis K, Frazier M, Domanico R, Burley M, et al. Evaluating the use of high-fidelity simulators during mock neonatal resuscitation scenarios in trying to improve confidence in residents. *SAGE Open Med.* 2018;6(101624744):2050312118781954.

## Tables

**Table 1. The 24 included studies arranged by study design, outcome Kirkpatrick level and number of participants.**

Author	Year	Country	Design	No.	Outcome Kirkpatrick level <sup>a</sup>			Ref. <sup>b</sup>
					II	III	IV	
Studies with control group (randomized / non-randomized)								
Walker, DM	2016	Mexico	Cluster randomized	305			X	(22)
Rubio-Gurung, S	2014	France	Cluster randomized	114		X		(19)
Thomas, EJ	2010	USA	Randomized, 3 arms	98		X		(20)
Bender, J	2014	USA	Randomized, 2 arms	50	X	X		(17)
Sawyer, T	2012	USA	Randomized, 2 arms	30		X		(21)
Lee, MO	2012	USA	Randomized, 2 arms	27	X	X		(18)
Rovamo, L	2015	Finland	Cohort	99		X		(23)
LeFlore, JL	2008	USA	Cohort	72		X		(24)
Barry, JS	2012	USA	Cohort	52		X		(25)
Studies with no control group and outcome level III – IV <sup>a</sup>								
Charafeddine, L	2016	Lebanon	Pre-post	256	X		X	(26)
Walker, D	2014	Mexico	Pre-post	305	X	X		(28)
Dadiz, R	2013	USA	Pre-post	228	X	X		(27)
Sawyer, T	2013	USA	Pre-post	42	X	X		(29)
Cordero, L	2013(B)	USA	Pre-post	33		X		(30)
Sawyer, T	2011	USA	Pre-post	30		X		(32)
Cordero, L	2013(A)	USA	Pre-post	26	X	X		(31)
Studies with no control and outcome level II <sup>a</sup>								
Dettinger, J	2018	Kenya	Pre-post	182	X			(33)
Walker, D	2015	Guatemala	Pre-post	159	X			(34)
Letcher, D	2017	USA	Pre-post	130	X			(35)
Malmström, B	2017	Sweden	Pre-post	92	X			(36)
Raffaeli, G	2018	Italy	Pre-post	28	X			(37)
Hossino, D	2018	USA	Pre-post	26	X			(38)
Ross, J	2016	USA	Pre-post	17	X			(10)
Bragard, I	2018	Belgium	Pre-post	16	X			(11)

<sup>a</sup>Kirkpatrick level II (learning), level III (clinical performance), and level IV (patient outcome)

<sup>b</sup>Reference number

**Table 2. Risk of bias judgement for included randomized studies using the revised Cochrane risk-of-bias tool for randomized trials (ROB 2).**

Author	Year	Design	Subdomain judgement of risk-of-bias						Overall judgement of Risk-of-bias
			Domain 1	Domain 1b	Domain 2	Domain 3	Domain 4	Domain 5	
			Randomization	Recruitment	Intervention	Missing outcome	Measuring outcome	Selected results	
Walker, DM	2016	Cluster randomized	Some concern	Some concern	Low risk	Low risk	Low Risk	Some concern	High risk
Rubio-Gurung, S	2014	Cluster randomized	Low risk	Low risk	Low risk	Low risk	Low risk	Some concern	Some concern
Thomas, EJ	2010	Randomized, 3 arms	Low risk	N/A	Some concern	Some concern	Low risk	Low risk	Some concern
Bender, J	2014	Randomized, 2 arms	Low risk	N/A	Low risk	Low risk	Some concern	Some concern	Some concern
Sawyer, T	2012	Randomized, 2 arms	Low risk	N/A	Low risk	Low risk	Low risk	Some concern	Some concern
Lee, MO	2012	Randomized, 2 arms	Low risk	N/A	Some concern	Low risk	Low risk	Some concern	Some concern

**Table 3. Risk-of-bias for non-randomized studies using the Newcastle-Ottawa quality assessment Scale (NOS) adapted to educational research.**

Author	Year	Design	NOS subdomain risk-of-bias score					Study Retention	Outcome Assessment	Overall assessment score (0-6)
			Intervention group	Comparison group	Comparison group					
			Representative	Selection	Comparability					
Rovamo, L	2015	Cohort	0	1	1	1	1	4		
LeFlore, JL	2008	Cohort	0	1	1	1	1	4		
Barry, JS	2012	Cohort	0	0	1	1	0	2		
Charafeddine, L	2016	Pre/post	0	0	0	1	0	1		
Walker, D	2014	Pre/post	0	0	0	1	0	1		
Dadiz, R	2014	Pre/post	1	0	0	1	0	2		
Sawyer, T	2013	Pre/post	0	0	0	1	0	1		
Cordero, L	2013B	Pre/post	0	0	0	1	0	1		
Sawyer, T	2011	Pre/post	1	0	0	1	1	3		
Cordero, L	2013A	Pre/post	0	0	0	1	0	1		
Dettinger, J	2018	Pre/post	1	0	0	1	0	2		
Walker, D	2015	Pre/post	0	0	0	1	0	1		
Letcher, D	2017	Pre/post	0	0	0	1	0	1		
Malmström, B	2017	Pre/post	1	0	0	1	0	2		
Raffaeli, G	2018	Pre/post	1	0	0	1	0	2		
Hossino, D	2018	Pre/post	0	0	0	1	0	1		
Ross, J	2016	Pre/post	0	0	0	1	0	1		
Bragard, I	2018	Pre/post	1	0	0	1	0	2		

**Table 4. Summary of 14 studies with observed clinical performance outcome during neonatal emergency team simulations.**

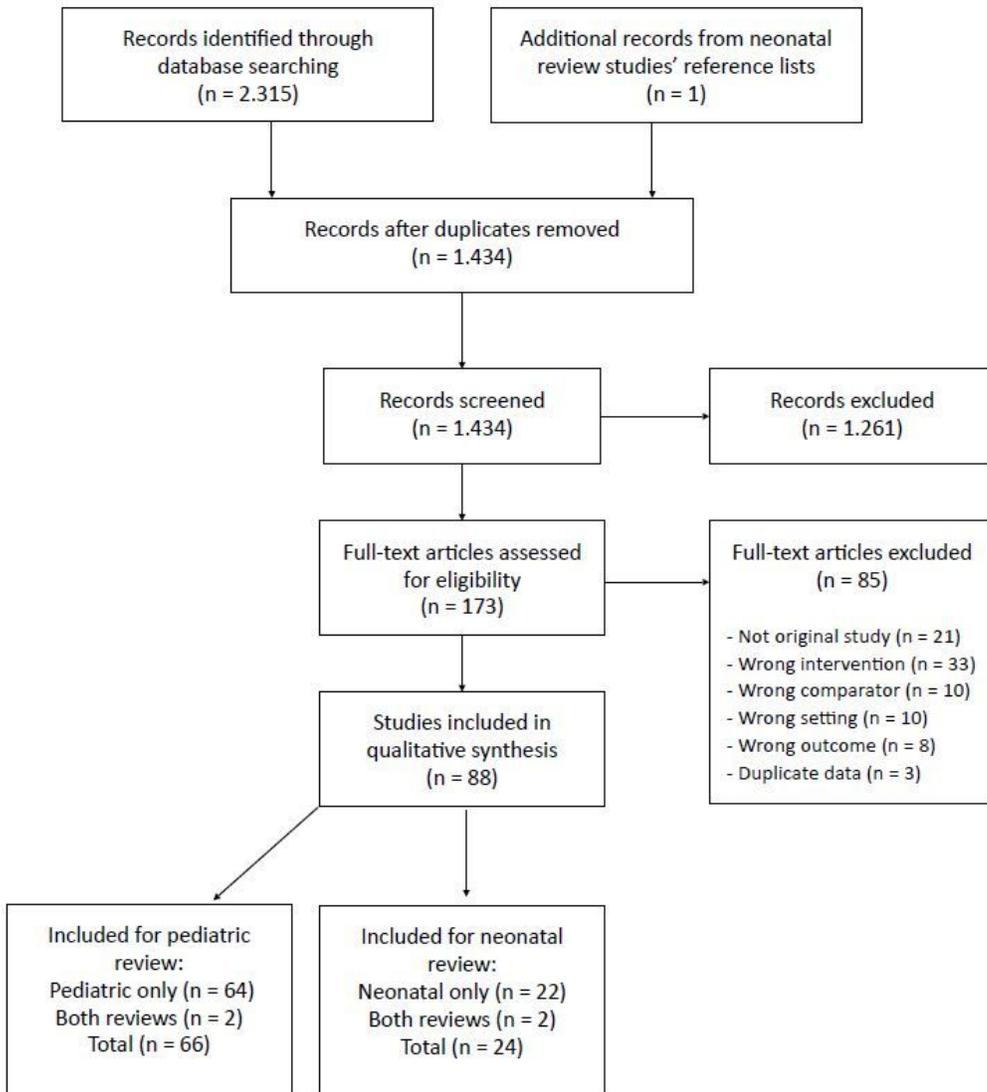
Author	Setting (annual births)	Design (No.)	Intervention	Comparator	Participants (team size)	Re-test timing	Cases (keywords)	Outcomes (observed performance)
Rubio-Gurung, S 2014	12 maternities (>1,000)	Cluster RCT (114)	4 hr simulation sessions High fidelity in situ 80% trained within 1 month	No simulation training	Doctors Nurses Midwives (6)	3 months	Resuscitation Asphyxia Meconium	TS 1: I 24.4 / C 17.4 (p= 0.01) TS 2: I 22.7 / C 17.5 (p= 0.004) TPS: I 31.1 / C 19.9 (p< 0.001) GPS: I 19.9 / C 6.7 (p= 0.001)
Thomas, EJ 2010	University of Texas Medical school	RCT, 3 arms (98)	2 hr session on communication and teamwork, and 1) low or 2) high fidelity skills session	Standard NRP with low fidelity skill sessions	Interns (3-4)	6 months	Resuscitation Haemorrhage Immaturity	Teamwork behaviours/min: I 11.8 / C 10.0 (p=0.03)  No difference in NRP PS, duration, vigilance, or workload management
Bender, J 2014	Level III NICU (9,000) and Level II Nursery (<600)	RCT, 2 arms (50)	Half-day NRP booster session High fidelity simulated OR and delivery room	Routine clinical duties	Residents	9 months booster, 15 months evaluation	Resuscitation Meconium Dystocia	TS: I 171.6 / C 64.4 (p= 0.02) TPS: I 18.8 / C 16.2 (p= 0.02)
Sawyer, T 2012	Army Medical Center (3,000)	RCT, 2 arms (30)	Video-assisted debriefing, 3 simulation sessions (30 min), high fidelity, simulation center	Oral debriefing, 3 simulation sessions (30 min), high fidelity, simulation center	Residents (2)	Two re-tests 2-4 months apart	Resuscitation	NRP PS improvement:  Video 12%, oral 8% (p=0.59) No difference in time to perform critical tasks
Lee, MO 2012	Acedemic medical center, level 1 trauma	RCT, 2 arms (27)	4 hr session, including 45 min didactic  Several high fidelity in situ simulations, procedural practice	Standard emergency medicine resident curriculum (including monthly paediatric simulation)	Residents (2-3)	16 weeks	Resuscitation	Neonatal resuscitation score change: I +11.8 / C -0.5, difference 12.3 (p= 0.056), the I group performed 2.31 more critical actions (p= 0.017)
Rovamo, L 2015	2 hospitals (6,000 and 3,800)	Cohort (99)	One day high fidelity in situ simulation + 1 hr interactive lecture on CRM and ANTS	One day high fidelity in situ simulation	Doctors NICU nurses Midwives (5-7)	Immediate	Resuscitation Respiratory distress Asphyxia	No difference in TEAM score between I and C groups

							Hypovolemic shock	
LeFlore, JL 2008	Metropolitan children's hospital	Cohort (72)	Expert modelled learning + high fidelity simulation	Self-paced modular learning + high fidelity simulation	Nurses Respiratory therapists Paramedics (3)	Immediate	Sepsis Meconium PPHN	TPS: I 21.7 / C 24.7 (p= 0.14) TS: I group used more UVC (p= 0.001) and less paralytics (P= 0.04)
Barry, JS 2012	University hospital (>3,000)	Cohort (52)	Afternoon equipment workshop and in situ simulation with low fidelity mannequin	Senior residents with NRP course and clinical duties	Residents	1 month 1-2 years	Resuscitation Meconium Prematurity Hypovolemic shock	GPS I-pre: 76% GPS I-1MO: 85% GPS C: 81% GPS I1-2Y: 85% I-pre vs. I-1MO (p= 0.001)
Walker, D 2014	12 hospitals (750-4,500)	Pre/post (305)	2+1 day workshop, minimal didactics, high fidelity in situ simulation	No control group	Nurses Doctors	3 months	Resuscitation Dystocia Haemorrhage	TPS-pre: 3.90 (baseline) TPS-post: 6.68 (p< 0.001) TPS-3MO: 6.94 (p< 0.001)
Dadiz, R 2013	University hospital, level IV NICU	Pre/post (228)	90 min high fidelity in simulated delivery room	No control group	Perinatal health care professionals (9-14)	Yearly for 3 years	Resuscitation Dystocia Maternal and newborn codes	Communication checklist score: Median (IQR) Y1: 6 (4), Y2 8(4), Y3 11(6) (p< 0.001)
Sawyer, T 2013	Army medical center (3,000)	Pre/post (42)	6 hr session, 4 hr didactic teamSTEPPS course, 2 hr high fidelity simulation and testing	No control group	NICU staff (4)	Immediate	Resuscitation	Team structure 2.5 to 4.2 Leadership 2.6 to 4.4 Situation monitoring 2.5 to 4.3 Mutual support 2.9 to 4.3 Communication 3.0 to 4.4 (All comparisons p< 0.001)
Cordero, L 2013B	University medical center	Pre/post (33)	Two 90 min sessions in simulated delivery room High fidelity, 1.5-2.0 hr deliberate	No control group	Residents Interns (3)	2-3 weeks	Resuscitation Placental abruption	Acceptable performances pre /post: TS: 36% / 91% (p= 0.04)

			practice between simulation sessions					Timeliness: 45% / 45% (p= 1.0) TPS: 27% / 100% (p= 0.01)
Sawyer, T 2011	Army medical center (3,000)	Pre/post (30)	Three 30 min sessions in simulated delivery room, high fidelity	No control group	Residents (2)	6 months	Resuscitation Hypoxemia Placental abruption Septic shock	NRP PS pre/post: 82.5% / 92.5% (p= 0.024)
Cordero, L 2013A	University medical center	Pre/post (26)	Two 90 min sessions in simulated delivery room, high fidelity	No control group	Residents Interns (3)	2-3 weeks	Resuscitation Placental abruption	Acceptable performances pre /post: TS: 45% / 64% (p= 0.68) Timeliness: 36% / 27% (p> 0.99) TPS: 45% / 73% (p= 0.37)

Abbreviations alphabetical: ANTS: Anaesthesia non-technical skills, CRM: crisis resource management, GPS: Global performance score, I: Intervention, IQR: interquartile range, C: Control, NICU: Neonatal intensive care unit, NRP: Neonatal resuscitation program, OR: Operating room, Pre/post: Single group Pre/post intervention comparison, PS: Performance score, PPHN: Persistent pulmonary hypertension of the newborn, RCT: Randomized controlled trial, TPS: Team performance score, TS: Technical score, UVC: Umbilical venous catheter.

## Figures



**Figure 1**

PRISMA flowchart of the study selection process.

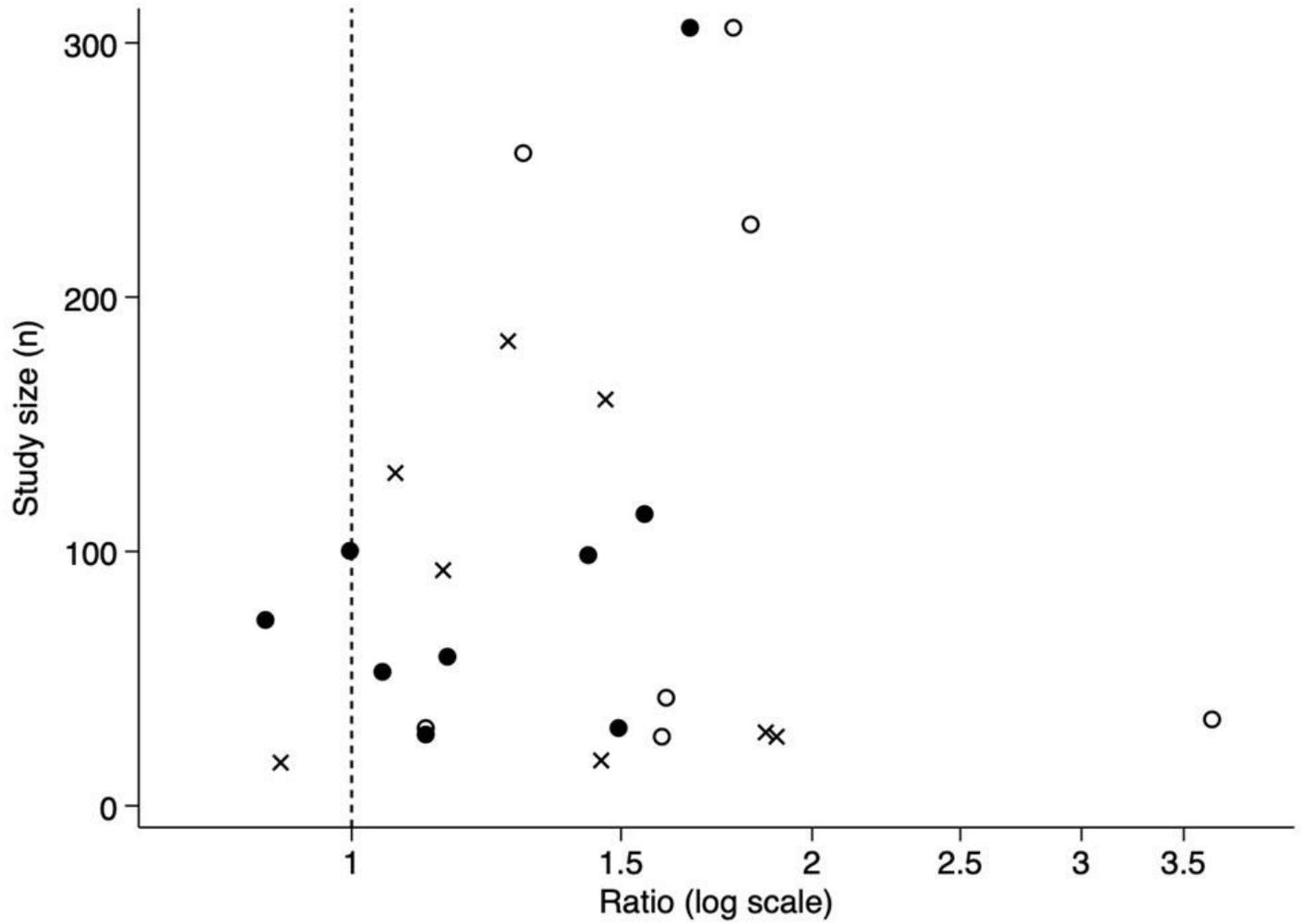


Figure 2

Funnel plot of the 24 included studies. (•) Studies with control group (+/- randomization). (○) Studies with no control group and clinical performance outcome or patient outcome (level III-IV). (x) Studies with no control group and learning outcome (level II).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.xlsx](#)