

Simulation-Based Ultrasound Training and Clinical Performance in Obstetric Residents: A Prospective Controlled Exploratory Pilot Study

Livia Schuler

livia.schuler@luks.ch

Lucerne Cantonal Hospital

Sara Ardabili

Lucerne Cantonal Hospital

Markus Hodel

Lucerne Cantonal Hospital

Benedict Krischer

Lucerne Cantonal Hospital

Research Article

Keywords: ultrasound training, simulation, residency teaching, ultrasound competency

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11144471/v1>

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

Additional Declarations: No competing interests reported.

Abstract

Background:

Obstetric ultrasound is an essential diagnostic tool, but performance depends substantially on examiner experience and technical skills. Simulation-based training may provide a standardized environment for deliberate practice independent of patient availability. This study evaluated the effect of structured ultrasound simulation training on clinical ultrasound performance in residents.

Methods

In this prospective controlled pilot study, ten residents (five per group) underwent standardized ultrasound assessments at baseline (T1) and after 6 weeks (T2). Five less experienced residents received weekly 15-minute simulation-based ultrasound training sessions, whereas five more experienced residents served as controls and continued routine clinical training. The primary outcome was the change in ultrasound performance score. Secondary outcomes included OSAUS score, supervisor rating, self-assessment, and domain-specific measures.

Results

From T1 to T2, ultrasound performance improved more in the intervention group (Δ 28.4 vs. 8.1 points; $g = 0.88$), without reaching statistical significance ($p = 0.164$). Improvements were also greater for OSAUS (Δ 3.8 vs. 1.8; $g = 0.48$) and supervisor ratings (Δ 6.6 vs. 0.6; $g = 0.76$). In domain-specific analyses, magnification improved significantly ($p = 0.034$; $g = 1.52$), while biometry showed a large effect ($g = 1.02$) without statistical significance. Gain and plane acquisition did not differ. OSAUS correlated strongly with supervisor ratings ($p = 0.86$ – 0.87 , $p = 0.001$).

Conclusion

Six weeks of structured simulation-based training was associated with consistent improvements in ultrasound performance, particularly in technical skills, with moderate to large effect sizes. The findings support the potential effectiveness of structured simulation training, although larger studies are required to confirm these results.

Background

Ultrasound imaging is an essential diagnostic modality in obstetrics. The quality of an ultrasound examination depends on multiple factors, including the examiner's technical skills and level of experience, which may directly influence diagnostic accuracy and subsequent clinical management. Training in obstetric ultrasound diagnostics has traditionally been conducted predominantly through hands-on practice on patients. However, due to variability between patients, this approach is not always objective, standardized, or reproducible.

The European Board and College of Obstetrics and Gynaecology (EBCOG) has developed recommendations for basic education in obstetrics and gynecology, in where trainees should acquire comprehensive knowledge of normal and pathological fetal anatomy, biometry, and fetal development¹. Similarly, the ISUOG (the International Society of Ultrasound and Obstetrics and Gynecology) has proposed minimum standards to ensure structured and consistent ultrasound training for residents². Despite these recommendations, substantial variation in ultrasound training programs persists across Europe, partly because of differences in national training structures and requirements³.

In recent years, the use of ultrasound simulation has been established as an effective training tool. Simulation allows for the practice of standardized clinical scenarios independent of patient availability and different ultrasound conditions, thereby providing a reproducible and objective learning environment. Initial studies have demonstrated that simulators can generate realistic imaging conditions and enable valid assessment of both technical and cognitive skills⁴. A meta-analysis including 15 studies showed that most evidence favors simulation-based ultrasound training over clinical training alone or theoretical teaching⁵.

Particularly in the field of fetal ultrasound, important skills such as anatomical orientation and measurement accuracy can be practiced and assessed in simulated environments⁶. Simulator-based obstetric ultrasound training has been found to perform comparably to real patient training and to be especially beneficial for inexperienced trainees⁷. Even short-term simulation-based training can improve skills in both inexperienced and experienced trainees⁸.

One limitation of simulation-based training is the lack of patient interaction. However, this aspect may be addressed through complementary methods such as role-playing⁹. Although simulators provide standardized conditions and a safe learning environment, real-life ultrasound examinations are influenced by patient-specific variability, suboptimal acoustic conditions, and situational pressure¹⁰.

Recently, Zhou et al. reported a randomized study of 22 obstetric ultrasound residents comparing simulator-based training with real-life ultrasound exams. Following two hours of theoretical instruction and five hours of hands-on practice, the simulator group achieved greater improvements in both simulator-based and patient-based assessments. These findings suggest that skills acquired during an intensive simulation curriculum can be demonstrated in clinical scanning¹¹.

The scope of our exploratory pilot study was to evaluate whether a structured six-week ultrasound simulation program was associated with improvements in clinical ultrasound performance among less experienced obstetric and gynecologic residents.

We additionally investigated changes in OSAUS scores, supervisor assessments, self-assessments, and specific technical domains of ultrasound performance

Methods

This prospective, controlled, exploratory pilot study included two groups of five resident physicians. The study was conducted at the Department of Obstetrics and Gynecology, Lucerne Cantonal Hospital, Switzerland. The intervention group consisted of resident physicians with 2–3 years of postgraduate training. These participants underwent a structured ultrasound simulation training program consisting of weekly 15-minute sessions (± 5 minutes) over a period of six weeks. Training was conducted in a one-to-one setting under the supervision of an experienced tutor and included the following components: (1) systematic examination workflow, (2) acquisition of standard planes, (3) image optimization, and (4) identification of anatomical structures.

An Ultrasound Mentor (Surgical Science, Sweden) was used for the simulation training. It consisted of a pregnant mannequin, a computer with a keyboard, mouse, a monitor as well as a dummy ultrasound transducer whose movements translate to movements on the screen, and it provides a realistic reconstruction of the fetal anatomy. The system allows to adjust the gain, magnification, and measure fetal biometrics. In our simulation training, we used a healthy 20-week fetus (with fetal movements during the examination). The participants were asked to obtain the standardized ISUOG planes and were trained to improve the image quality.

The control group included resident physicians with at least 3 years of postgraduate training and a minimum of four months of clinical rotation in obstetrics with supervision in obstetric ultrasound, but without exposure to simulation training. Participants in the control group documented the number of ultrasound examinations performed during the study period.

Assessments were conducted at two time points (T1 and T2) in both groups. Each assessment comprised (1) a real patient ultrasound examination and (2) a simulation-based ultrasound examination, both evaluated using predefined criteria and standardized scoring systems. Additional measures included subjective self-assessment (Likert scale 1–7), subjective external assessment, and the Objective Structured Assessment of Ultrasound Skills (OSAUS), a validated tool for ultrasound competency evaluation^{12,13}.

The primary outcome was a Clinical Ultrasound Quality Score, a composite score based on 24 standardized ISUOG planes, with each criterion rated on a scale from 0 to 2. The score comprised the following subscores: biometrics (including caliper placement; maximum 24 points), correct anatomical plane acquisition (48 points), magnification (48 points), and gain/image optimization (46 points), resulting in a total score of 158 points.

Secondary outcomes included the OSAUS score (range 0–25 points, scoring card in Appendix), consisting of five domains rated from 0 to 5 (applied knowledge of equipment, image optimization, systematic examination, interpretation, and documentation). Additional secondary endpoints comprised subjective self-assessment (0–56 points; 8 items rated 0–7), assessing participants' perceived competence in fetal ultrasound, including anatomy, screening, image quality, device handling, interpretation, and communication, and supervisor assessment (0–56 points; 8 items rated 0–7, see Appendix), reflecting external evaluation of the same competencies. Correlations between subjective and objective performance measures were also analyzed.

All clinical and simulation-based examinations were independently evaluated by two blinded reviewers. Ratings were performed separately. In cases where the difference between the two ratings was ≤ 15 points, the mean score was used. If the difference exceeded 15 points, a structured consensus discussion was conducted. If no agreement could be reached, a third reviewer served as a tie-breaker. The final score was defined as the mean of the two concordant ratings. Interrater reliability was evaluated using the intraclass correlation coefficient (ICC [2,k], absolute agreement). Additionally, Bland–Altman analyses were performed for clinical scores, complemented by Spearman correlation coefficients.

All statistical analyses were conducted using RStudio (Version 4.5.2, R Foundation for Statistical Computing, Vienna, Austria).

Change-score analyses were performed only for paired measurements (T1 and T2). The primary analysis assessed the between-group difference in change (difference-in-differences, $\Delta\Delta$). Effect estimates were reported with 95% confidence intervals derived via bootstrap resampling (2,000 resamples). P-values were obtained using permutation testing. Given the small sample size, no assumptions of normal distribution were made.

Secondary analyses were conducted analogously for OSAUS scores, simulation scores, subscores and subjective measures using change-score comparisons.

Prior to the study, the local ethics committee was consulted and confirmed that ethical approval was not required in accordance with the Declaration of Helsinki, as no health-related personal data were collected or analyzed in this study (Ethics Committee Northwest and Central Switzerland, Req-2025-01321). All participants gave written informed consent.

Results

All ten residents (five in each group) completed both T1 and T2 assessments.

At baseline, all outcome measures were substantially lower in the intervention group compared to controls, reflecting differences in prior clinical experience as per study design. At T1, residents' baseline US score was 106.1 ± 8.0 (mean, SD) for the control group and 86.4 ± 10.5 for the intervention group, OSAUS

Score was 18.8 ± 1.5 for the control group and 14.8 ± 2.9 for the intervention group, mean supervisor rating was 42.6 ± 2.6 vs 34.4 ± 6.1, respectively, and subjective self-rating was 39.7 ± 5.7 vs 34.0 ± 7.0, respectively (conf. Table 1).

Table 1 inter-group comparison at T1 and T2 for US score, OSAUS, reviewer score and self-rating										
group	US_T1	US_T2	US_delta	OSAUS_T1	OSAUS_T2	OSAUS_delta	Supervisor_T1	Supervisor_T2	Supervisor_delta	Self_
Control	106.1 ± 8.0	114.2 ± 19.1	8.1	18.8 ± 1.5	20.6 ± 3.4	1.8	42.6 ± 2.6	43.2 ± 8.7	0.6	39.7 ± 5.7
Intervention	86.4 ± 10.5	114.8 ± 9.3	28.4	14.8 ± 2.9	18.6 ± 2.2	3.8	34.4 ± 6.1	41.0 ± 5.3	6.6	34.0 ± 7.0

Inter-rater agreement for the ultrasound total score was good, with an ICC of 0.81 (95% CI 0.42–0.95) at T1.

From T1 to T2, the increase in ultrasound performance score was larger in the intervention group compared to controls (Δ 28.4 vs. 8.1), corresponding to a large effect size (Hedges’ g = 0.88, 95% CI – 2.6–46.0), although this did not reach statistical significance (Welch p = 0.164).

Improvements in OSAUS score were greater in the intervention group (Δ 3.8 vs. 1.8), corresponding to a small-to-moderate effect size (Hedges’ g = 0.48, 95% CI – 0.79–1.73). Similarly, supervisor ratings increased more in the intervention group (Δ 6.6 vs. 0.6), with a moderate effect size (Hedges’ g = 0.76, 95% CI – 0.55–2.04).

In the domain-specific analyses (conf. Table 2), magnification improved significantly in the intervention group (p = 0.034) and showed a large effect size (g = 1.52). Biometry showed a similarly large effect size (Hedge’s g 1.02, CI – 1.2–9.4) but did not reach statistical significance, suggesting a likely underpowered effect. No meaningful differences were observed for gain or plane acquisition.

Table 2 sub-domain analysis							
Domain	Control (Δ mean ± SD)	Intervention (Δ mean ± SD)	Mean difference (I–C)	t-value	p-value	Hedges' g	95% CI (I–C)
Magnification	-0.3 ± 6.1	13.6 ± 10.0	13.9	-2.66	0.034	1.52	1.4 to 26.4
Gain	6.4 ± 10.2	8.1 ± 7.5	1.7	-0.30	0.772	0.17	-11.5 to 14.9
Biometry	-0.4 ± 3.7	3.7 ± 3.5	4.1	-1.78	0.113	1.02	-1.2 to 9.4
Plane	2.2 ± 9.4	3.9 ± 5.2	1.7	-0.35	0.735	0.20	-9.9 to 13.3

The OSAUS score showed a strong agreement with the supervisor’s assessment, both in change over time (Spearman ρ = 0.86, p = 0.001) and at follow-up (ρ = 0.87, p = 0.001), supporting convergent validity of these two measures of clinical competence. (conf. Figure 1)

Body mass index (BMI) of the ultrasound scan patients, gestational age at scan and scan duration did not differ significantly at T1 and T2. The mean time from T1 to T2 and the number of performed ultrasound exams in daily clinical work did not differ significantly between groups (control vs intervention group: 58.8 ± 28.5 vs. 56.2 ± 35.0 ultrasound exams in 64.6 ± 18.5 vs. 71.6 ± 12.5 days).

At baseline, large between-group differences were observed for both OSAUS (Hedges’ g = 1.55, 95% CI 0.17–2.86) and supervisor ratings (g = 1.58, 95% CI 0.19–2.90), reflecting the higher experience level of the control group.

At follow-up, these differences were markedly reduced (OSAUS: g = 0.63, 95% CI – 0.55–1.77; supervisor: g = 0.28, 95% CI – 0.86–1.39), indicating a substantial narrowing of the baseline performance gap.

Change score effect sizes favored the intervention group (OSAUS: g = – 0.44; supervisor: g = – 0.69), suggesting greater improvement compared to controls, although confidence intervals were wide. (conf. Figure 2 and Table 3)

Table 3 Convergence of ratings between T1 and T2												
Outcome	Delta_Control	Delta_Intervention	Gap_T1	Gap_T2	Welch_p	Welch_CI	g_T1	g_T1_CI	g_T2	g_T2_CI	g_Delta	g_De
OSAUS	1.8 ± 3.6	3.8 ± 4.6	4	2	0.4688	[-8.11, 4.11]	1.55	[0.17, 2.86]	0.63	[-0.55, 1.77]	-0.44	[-1.5, 0.72]
Supervisor	0.6 ± 6.5	6.6 ± 9.0	8.2	2.2	0.2642	[-17.63, 5.63]	1.58	[0.19, 2.90]	0.28	[-0.86, 1.39]	-0.69	[-1.8, 0.50]

Discussion

This prospective controlled exploratory pilot study found consistent improvements in clinical ultrasound performance following a structured six-week simulation-based training program. The largest effects were observed in technical domains, particularly magnification and fetal biometry. Global

performance measures, including the Clinical Ultrasound Quality Score, OSAUS, and supervisor assessment, also favored the intervention group, although the small sample size resulted in wide confidence intervals and no statistically significant difference in the primary outcome.

A convergence of performance between the less experienced and the more experienced trainees was observed from T1 to T2, with between-group effect sizes for OSAUS and supervisor ratings reducing substantially ($g = 1.55$ to 0.63 and $g = 1.58$ to 0.28 , respectively). While not statistically significant, this consistent pattern across independent outcome measures suggests a catch-up effect of the less experienced trainees. This is plausible given that learners at an earlier stage of skill acquisition tend to benefit disproportionately from structured, deliberate practice – as described in established frameworks of skill development^{14,15}.

Domain-specific findings suggest that simulation may be particularly effective for technical skills that can be practiced under standardized conditions. The particularly strong effect on magnification ($g = 1.52$, $p = 0.034$) may partly reflect a characteristic of the simulator: magnification produces directly visible image changes, whereas suboptimal acoustic conditions – which drive the clinical need for gain adjustment – are not fully replicated in the simulation environment. Plane acquisition showed minimal improvement in both groups, consistent with the notion that visuospatial and psychomotor skills underlying systematic plane navigation require longer training horizons than a six-week program can provide⁵.

The strong correlation between OSAUS and supervisor ratings supports the convergent validity of these instruments.

While the Clinical Ultrasound Quality Score focuses on technical image acquisition and quality, supervisor assessment reflects broader clinical competence; the measures therefore appear complementary. Our findings are consistent with previous studies demonstrating improved ultrasound skills following simulation training [7,8], while extending previous work by evaluating transfer to clinical performance in postgraduate obstetric and gynecologic trainees.

Strengths of this study include the prospective controlled design, blinded independent rating (which limits ascertainment bias in a domain with subjective assessment components), the comparison of two distinct experience levels, and the use of multiple validated outcome tools with a focus on clinical rather than simulation-only performance.

The main limitations are the small sample size, non-randomized design, baseline imbalance, and short follow-up. The small sample size substantially reduces statistical power and yields wide confidence intervals. Despite moderate to large effect sizes, no between-group comparison reached significance. The non-randomized design was a pragmatic necessity in a clinical training context but resulted in a baseline performance imbalance. Potential ceiling effects in the more experienced control group may have attenuated measurable gains in that group, making change-score analysis the more interpretable primary approach.

The six-week follow-up precludes conclusions about long-term retention. The study cannot establish the optimal frequency or duration of simulation training. Future studies should investigate dose–response relationships and assess skill retention after longer follow-up periods.

While the study by Zhou et al. provides an interesting juxtaposition, our two studies, albeit similarly in set-up, cannot be compared directly. While Zhou et al randomized two groups and started with a teaching-block, our concept were different groups at baseline and training spread over the course of six weeks. While both studies show improvement with simulation training, an interesting future approach would be the comparison of both concepts.

A recent study by Rameh et al. demonstrated the feasibility of simulation-based assessment of obstetric and gynecological ultrasound skills by evaluating three different simulators in a cohort of inexperienced trainees. However, a comparison to our study is difficult by differences in study design, outcome measures, simulator modalities, and the duration and structure of training. Nevertheless, our findings further support the potential role of simulation-based training in facilitating the acquisition of ultrasound skills, particularly in beginners¹⁶. Overall, our findings support the feasibility and potential educational value of short, structured simulation-based ultrasound training for obstetric and gynecologic residents. The apparent reduction in the performance gap between less experienced and more experienced trainees is encouraging, but should be considered hypothesis-generating rather than definitive evidence of a catch-up effect.

Future research should use larger, adequately powered randomized designs, standardized outcome measures, and longer follow-up. Studies should also examine the optimal training dose, long-term retention, and the extent to which individual ultrasound skills transfer from simulation to clinical practice.

Conclusion

Six weeks of structured simulation-based ultrasound training was associated with improvement in clinical ultrasound performance among less experienced obstetric and gynecologic residents, with the strongest effects observed in selected technical domains, particularly magnification and fetal biometry. The findings also suggest a potential reduction in the performance gap between less experienced and more experienced trainees.

Declarations

Ethics approval and consent to participate

The local ethics committee was consulted and confirmed that ethical approval was not required in accordance with the Declaration of Helsinki, as no health-related personal data were collected or analyzed in this study (Ethics Committee Northwest and Central Switzerland, Req-2025-01321). All participants gave written informed consent.

Consent for publication

Not applicable

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

Not applicable

Funding

There are no funding sources.

Author Contribution

Livia Schuler: Conceptualization, methodology, investigation, data curation, supervision, project administration, writing (original draft).

Sara Ardabili: Writing – review & editing.

Markus Hodel: Writing – review & editing.

Benedict Krischer: Conceptualization, methodology, investigation, data curation, analysis, supervision, project administration, writing (original draft).

Data Availability

All data supporting the findings of this study are available within the paper and its Supplementary Information.

References

1. Mcauliffe F, Scheele F, van der Aa J, Goverde A. Project for Achieving Consensus in Training (PACT) by the European Board and College of Obstetrics and Gynaecology [Internet]. 2026 [cited 2026 May 16]. Available from: <https://ebcog.eu/wp-content/uploads/2026/05/PACT-2026.pdf>
2. ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound. *Ultrasound Obstet Gynecol*. 2014;43(1):113–6. 10.1002/uog.13208.
3. Salvesen KÅ, Lees C, Tutschek B. Basic European ultrasound training in obstetrics and gynecology: where are we and where do we go from here? *Ultrasound Obstet Gynecol*. 2010;36(5):525–9. 10.1002/uog.8851.
4. Tolsgaard MG, Todsén T, Sørensen JL, Ringsted C, Lorentzen T, Ottesen B, et al. International Multispecialty Consensus on How to Evaluate Ultrasound Competence: A Delphi Consensus Survey. *PLoS ONE*. 2013;8(2):e57687. 10.1371/journal.pone.0057687.
5. Taksøe-Vester C, Dyre L, Schroll J, Tabor A, Tolsgaard M. Simulation-Based Ultrasound Training in Obstetrics and Gynecology: A Systematic Review and Meta-Analysis. *Ultraschall Med*. 2021;42(6):e42–54. 10.1055/a-1300-1680 . PubMed PMID: 33348415.
6. Patel H, Chandrasekaran D, Myriokefalitaki E, Gebeh A, Jones K, Jeve YB. The Role of Ultrasound Simulation in Obstetrics and Gynecology Training: A UK Trainees' Perspective. *Simul Healthcare: J Soc Simul Healthc*. 2016;11(5):340–4. 10.1097/SIH.0000000000000176.
7. Rosen H, Windrim R, Lee YM, Gotha L, Perelman V, Ronzoni S. Simulator Based Obstetric Ultrasound Training: A Prospective, Randomized Single-Blinded Study. *J Obstet Gynecol Can*. 2017;39(3):166–73. 10.1016/j.jogc.2016.10.009.
8. Zhao Y, Zhou P, Tang X, Zhu W, Liu W, Zhou J, et al. Short-term simulation training course improves basic obstetric ultrasound operational skills of residents: A multi-center study. *PLoS ONE*. 2025;20(12):e0328122. 10.1371/journal.pone.0328122.
9. Chalouhi GE, Bernardi V, Gueneuc A, Houssin I, Stirnemann JJ, Ville Y. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges and opportunities. *Am J Obstet Gynecol*. 2016;214(4):525. e1-525.e8.
10. Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sørensen JL, et al. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound Obstet Gynecol*. 2014;43(4):437–43. 10.1002/uog.13198.
11. Zhou J, Zhao Y, Zhou P, Tang X, Liu W, Guo S, Wang J, Li L. Evaluating the effectiveness of high-fidelity simulation-based training on clinical skill transfer in obstetric ultrasound residents: a prospective randomized controlled trial. *BMC Med Educ*. 2026;26(1):473. 10.1186/s12909-026-08835-2. PMID: 41699591.
12. Tolsgaard MG, Todsén T, Sørensen JL, Ringsted C, Lorentzen T, Ottesen B, et al. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PLoS ONE*. 2013;8(2):e57687. 10.1371/journal.pone.0057687 . PubMed PMID: 23469051.
13. Höhne E, Recker F, Dietrich CF, Schäfer VS. Assessment Methods in Medical Ultrasound Education. *Front Med (Lausanne)*. 2022;9. 10.3389/fmed.2022.871957.
14. Ericsson KA, Harwell KW. Deliberate Practice and Proposed Limits on the Effects of Practice on the Acquisition of Expert Performance: Why the Original Definition Matters and Recommendations for Future Research. *Front Psychol*. 2019;10:2396. 10.3389/fpsyg.2019.02396 . PubMed PMID: 31708836.

15. Tolsgaard MG. Assessment and learning of ultrasound skills in Obstetrics & Gynecology. Dan Med J. 2018;65(2). PubMed PMID: 29393042.

16. Rameh G, Gallo DM, Corroenne R, Helou N, Aboujaoude I, Seshadri S, Tolosa JE, Chalouhi G. Simulator-based assessment of obstetric and gynecological ultrasound skills: validity of three different simulators and feasibility of remote evaluation. Ultrasound Obstet Gynecol. 2026;67(4):542–7. <https://doi.org/10.1002/uog.70175>.

Figures

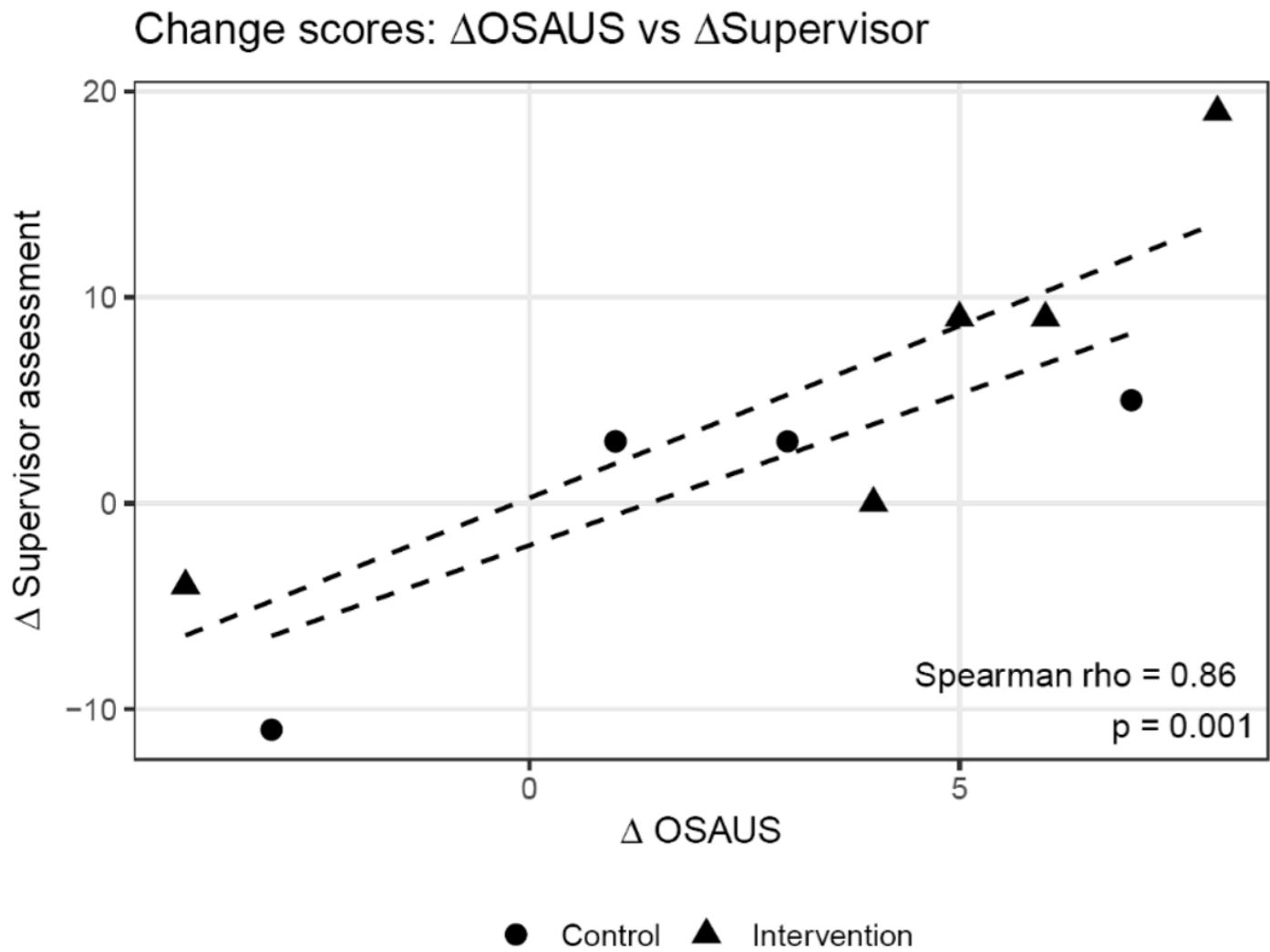


Figure 1

Comparison of OSAUS vs. Supervisor rating

Convergence of performance over time

Mean scores with 95% confidence intervals

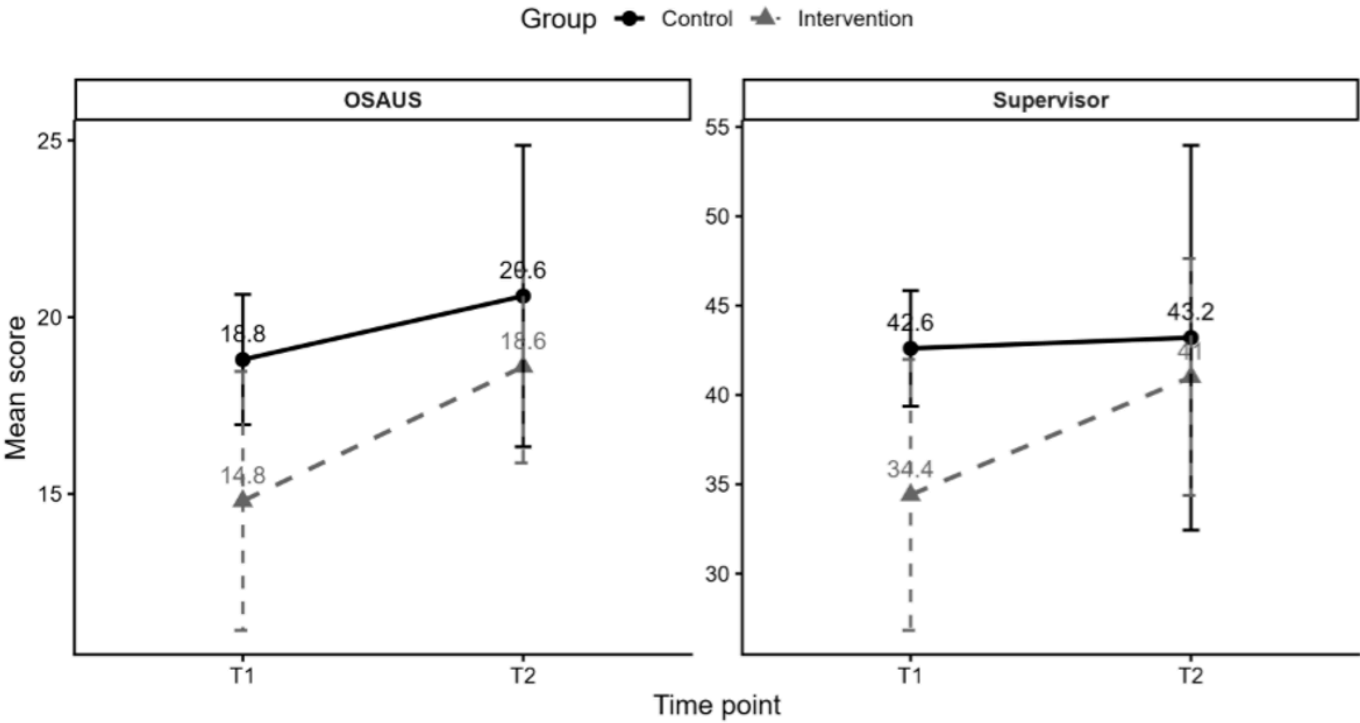


Figure 2
Convergence of performance over time

Supplementary Files

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

- [SIMULATEtables.docx](#)