frontiers | Research Topics

Education in obstetrics and gynecology

Edited by
Fedde Scheele and Florian Recker

Published in
Frontiers in Medicine

ISSN 1664-8714
ISBN 978-2-8325-5697-9
DOI 10.3389/978-2-8325-5697-9

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Education in obstetrics and gynecology

Topic editors

Fedde Scheele — VU Amsterdam, Netherlands

Florian Recker — University of Bonn, Germany

Citation

Scheele, F., Recker, F., eds. (2024). *Education in obstetrics and gynecology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5697-9

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 05 November 2024
DOI 10.3389/fmed.2024.1490673

OPEN ACCESS

EDITED AND REVIEWED BY
Michael J. Wolyniak,
Hampden–Sydney College, United States

*CORRESPONDENCE
Florian Recker
✉ florianrecker@icloud.com

RECEIVED 03 September 2024
ACCEPTED 21 October 2024
PUBLISHED 05 November 2024

CITATION
Recker F and Scheele F (2024) Editorial: Education in obstetrics and gynecology. *Front. Med.* 11:1490673. doi: 10.3389/fmed.2024.1490673

Editorial: Education in obstetrics and gynecology

Florian Recker[1]* and Fedde Scheele[2]

[1]Department of Obstetrics and Prenatal Medicine, University Hospital Bonn, Bonn, Germany, [2]Athena Institute, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

KEYWORDS

medical education, obstetrics, gynecology, continuing medical education, training

Editorial on the Research Topic
Education in obstetrics and gynecology

Medical education in obstetrics and gynecology (OB-GYN) has undergone significant transformations to reach its current state, with the field adapting to advancements in technology, evolving patient needs, and a growing emphasis on multidisciplinary approaches to healthcare. This dynamic field, crucial to women's health, now requires a more comprehensive and integrative educational framework to prepare future clinicians for the complex challenges they will face in their practice. High-quality education and training of medical professionals are fundamental pillars in ensuring the health and wellbeing of global populations. In particular, OB-GYN requires continuous advancements in educational strategies to address the unique challenges presented by a globalized and interconnected world. This Research Topic seeks to explore the multifaceted aspects of medical education within OB-GYN, spanning from undergraduate studies to postgraduate education, continuing medical education (CME), and beyond.

The journey to becoming an OB-GYN specialist begins with a robust undergraduate medical education. Here, it is imperative to develop a curriculum that not only covers the essential medical knowledge but also fosters critical thinking, empathy, and self-reflective skills (1). Modern medical schools are increasingly adopting integrated curricula that blend basic sciences with clinical practice early on (2). This approach helps students contextualize their learning and understand the relevance of theoretical knowledge in real-world scenarios. Innovative teaching methods, such as problem-based learning (PBL) and team-based learning (TBL), have shown significant promise in enhancing student engagement and retention of knowledge (3). These methodologies encourage collaborative learning, critical analysis, and the application of knowledge to practical problems, preparing students for the dynamic and often unpredictable nature of clinical practice in OB-GYN (4).

In the context of obstetrics and gynecology education, competency-based ultrasound education has emerged as a pivotal component of training programs (Weimer et al.). This approach ensures that learners develop the essential skills and knowledge required for effective ultrasound use, regardless of the time it takes to achieve these competencies. Rather than progressing through a fixed curriculum, students advance based on their demonstrated proficiency in various aspects of ultrasound, from image acquisition and interpretation to the integration of findings into clinical decision-making (5). Simulation-based training plays a significant role, providing learners with opportunities to practice and refine their skills in a controlled environment until they reach a level of mastery. This method not only enhances technical abilities but also promotes critical thinking and clinical judgment, ensuring that graduates are fully prepared to utilize ultrasound as a diagnostic and

therapeutic tool in their practice. By focusing on competency rather than time spent in training, this educational model better equips future obstetricians and gynecologists to meet the demands of modern healthcare.

Postgraduate education, particularly residency training, is where the foundational knowledge acquired during undergraduate studies is honed into specialized clinical skills (Plöger et al.). Residency programs for OB-GYN must be rigorous, inclusive, and comprehensive, ensuring that trainees are well-equipped to handle the complexities of the field. Standardized accreditation and validation processes are crucial to maintaining the quality and consistency of these programs globally. Simulation-based training has become a cornerstone in residency programs, providing a safe and controlled environment for residents to practice and refine their skills. High-fidelity simulations, including the use of virtual reality (VR) and augmented reality (AR), offer immersive experiences that mimic real-life clinical scenarios. These technologies not only enhance technical skills but also improve decision-making, teamwork, and communication—essential components of effective clinical practice.

The rapid pace of medical advancements necessitates a commitment to lifelong learning for OB-GYN professionals. Continuing medical education (CME) ensures that practitioners remain up to date with the latest developments, techniques, and best practices in the field. CME programs must be flexible, accessible, and relevant, catering to the diverse needs of healthcare professionals across different stages of their careers. Digital platforms have revolutionized CME, offering online courses, webinars, and virtual conferences that enable professionals to learn at their own pace and convenience (6). Additionally, mobile applications and e-learning modules provide on-the-go access to educational resources, allowing clinicians to seamlessly integrate learning into their daily routines.

The digitalization of medical education has opened up new avenues for teaching and training. VR and AR are transforming the way medical students and professionals learn, providing immersive and interactive experiences that enhance understanding and retention (7). These technologies can simulate complex OB-GYN procedures, offering hands-on practice without the risks associated with real-life interventions. Simulation centers equipped with high-fidelity mannequins and task trainers allow for the practice of intricate surgical techniques, obstetric emergencies, and patient management scenarios. These controlled environments enable learners to make mistakes and learn from them, building confidence and competence.

Healthcare is inherently multidisciplinary, and the education of OB-GYN professionals must reflect this reality. Inter-professional training, where medical students, nursing students, and other healthcare trainees learn together, promotes a collaborative approach to patient care. This model fosters mutual respect, understanding, and communication among different healthcare professionals, ultimately improving patient outcomes. Transdisciplinary approaches, which incorporate non-medical education such as ethics, communication, and leadership training, further enrich the learning experience. These elements are crucial in developing well-rounded healthcare professionals who are not only skilled clinicians but also empathetic leaders and effective communicators.

Qualitative and quantitative research in medical education is essential for continuous improvement and innovation. Studies that evaluate the effectiveness of different teaching methods, curricular designs, and training tools provide valuable insights that inform educational practices. Research also helps identify gaps and areas for improvement, ensuring that medical education evolves to meet the changing needs of society.

The education of healthcare professionals in Obstetrics and Gynecology is a dynamic and evolving field that requires a multifaceted approach. From undergraduate education to postgraduate training and continuing medical education, each stage plays a critical role in developing competent, compassionate, and resilient practitioners. The integration of innovative teaching tools, inter-professional education, and rigorous research is essential to meet the challenges of a globalized world and ensure the wellbeing of global populations. As we look to the future, it is imperative that medical education remains inclusive, collaborative, and forward-thinking. By embracing these principles, we can train the next generation of OB-GYN professionals who are not only skilled clinicians but also advocates for health and wellbeing in their communities and beyond. This Research Topic aims to shed light on these important Research Topics, providing a platform for sharing knowledge, experiences, and best practices in the education of OB-GYN professionals.

The 17 contributions for this Research Topic concerning education in obstetrics and gynecology can be classified into different sections. For that purpose, the article by Frenk et al. in the Lancet, is still valuable (8). He categorized learning into three areas: specific skills needed to become a medical expert, general competencies required for professionalism, and transformative, future-focused learning to develop as a change agent. Eleven articles concern program evaluation, of which two describe checklists to assess either a skill or a generic competency, i.e., teamwork, and two articles describe program evaluation focused on generic competencies. The third class of articles are those concerning the future of education. They concern the changing demographics of our workforce and the need for culture changes, as artificial intelligence and digital tools are introduced in both our learning and care provision systems. Learning for the future in a fast-changing world also demands elements of transformative learning to foster change-agents in the health sector. The item of transformative learning is still in its infancy and needs far more academic attention. Hopefully, in a next Research Topic of this journal, the connection between health system transformation and education may attract more articles from this emerging field. Finally, an article from Ethiopia reminds us that sociological, economic, and cultural features of society are extremely relevant for addressing health issues. Learning about public health, with a broad view on society and its influence on health care, is essential for the quality and efficiency of future health services.

The landscape of medical education in obstetrics and gynecology is characterized by technological innovation, a shift toward competency-based learning, interdisciplinary collaboration, and a strong commitment to global health and equity. These

advancements ensure that future OB-GYN practitioners are well-prepared to meet the complex demands of their field, providing high-quality care to women across diverse settings. As the field continues to evolve, so too will the educational strategies, maintaining a focus on excellence and adaptability in the face of new challenges.

We hope that both obstetrician-gynecologists and teachers will enjoy this Research Topic and get engaged in optimal training for the benefit of global health.

Author contributions

FR: Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology. FS: Writing – review & editing, Writing – original draft, Methodology, Investigation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Reed VR, Emery J, Farrell RM, Jelovsek JE. Tracking-a flexible obstetrics and gynecology residency curriculum. *Obstetr Gynecol.* (2019) 134:29S—33S. doi: 10.1097/AOG.0000000000003464

2. Meyer B, Riedel F, Amann N, Graf A, Stuehrenberg A, Ritter V, et al. Exploring the current state of clinical and practical teaching in obstetrics and gynecology in the era of competency-based education: a nationwide survey among German teaching coordinators. *BMC Med Educ.* (2024) 24:165. doi: 10.1186/s12909-024-05138-2

3. Recker F, Haverkamp N, Mustea A, Gembruch U, Raupach T. Application of test-enhanced learning (TEL) in obstetrics and gynecology: a prospective study. *Arch Gynecol Obstet.* (2022) 306:1563–71. doi: 10.1007/s00404-022-06656-4

4. Winder FM, Breuer G, Favero M, Foessleitner P, Friemann M, Krischer B, et al. Postgraduate medical education in obstetrics and gynaecology: where are we now and what do we need for the future? A study on postgraduate training in obstetrics and gynaecology in Germany, Austria and Switzerland. *GMS J Med Educ.* (2022) 39.

5. Schlapp MS, Wittek A, Neubauer R, Geipel A, Gembruch U, Strizek B, et al. Enhancing ultrasound education in obstetrics and gynecology in Germany: insights and innovations from postgraduate training. *Arch Gynecol Obstet.* (2024) 24:2. doi: 10.1007/s00404-024-07610-2

6. Matschl J, Gembruch U, Strizek B, Recker F. Shaping the future of obstetric/gynecological ultrasound training. *Ultrasound Obstet Gyne.* (2024) 63:717–22. doi: 10.1002/uog.27554

7. Neubacher M, Siebers P, Wittek A, Recker F. How to play a game properly—enhancing obstetrics and gynecology education through gamification: a scoping review. *Geburtshilfe Frauenheilkd.* (2024) 2024:2379–8729.

8. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet.* (2010) 376:1923–58. doi: 10.1016/S0140-6736(10)61854-5

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 15 March 2023
DOI 10.3389/fmed.2023.1122472

OPEN ACCESS

EDITED BY
Ray Samuriwo,
University of Bradford, United Kingdom

REVIEWED BY
Holger Hauch,
Justus Liebig University Giessen, Germany
Sarah Turner,
Coventry University, United Kingdom

*CORRESPONDENCE
Xiaoyan Yu
✉ yuxy@zju.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Healthcare Professions Education, a section of the journal Frontiers in Medicine

RECEIVED 13 December 2022
ACCEPTED 01 March 2023
PUBLISHED 15 March 2023

CITATION
Qian J, Chen S, Jevitt C, Sun S, Wang M and Yu X (2023) Experiences of obstetric nurses and midwives receiving a perinatal bereavement care training programme: A qualitative study.
Front. Med. 10:1122472.
doi: 10.3389/fmed.2023.1122472

Experiences of obstetric nurses and midwives receiving a perinatal bereavement care training programme: A qualitative study

Jialu Qian[1,2†], Shuyi Chen[3†], Cecilia Jevitt[2], Shiwen Sun[4], Man Wang[1] and Xiaoyan Yu[4*]

[1]Faculty of Nursing, Zhejiang University School of Medicine, Hangzhou, China, [2]Division of Midwifery, Department of Family Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada, [3]Department of Thyroid and Breast Surgery, Huashan Hospital Affiliated to Fudan University, Shanghai, China, [4]Department of Obstetrics, Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Aim: To explore obstetric nurses and midwifery professionals' experiences with the Perinatal Bereavement Care Training Programme (PBCTP) after implementation.

Design: A qualitative descriptive design was used.

Method: This qualitative study was conducted at a tertiary level maternity hospital in China. The PBCTP was implemented at Women's Hospital School of Medicine, Zhejiang University from March to May 2022. A total of 127 nurses and 44 midwives were invited to participate in the training. Obstetric nurses and midwives studied a 5-module training programme comprised of eight online theoretical courses and submitted a reflective journal after each session. Semi-structured interviews were conducted with 12 obstetric nurses and four midwives from May to July 2022 as a post-intervention evaluation. Thematic analysis was used in data analysis.

Findings: A total of 16 participants in this study ranged in age from 23 to 40 years [mean age (SD), 30 (4) years]. Six main themes within participants' experiences of PBCTP intervention were identified: participants' aims of undertaking the training; personal growth and practice changes after training; the most valuable training content; suggestions for training improvement; directions for practice improvement; influencing factors of practice optimization.

Conclusion: Nursing and midwifery professionals described the PBCTP as satisfying their learning and skills enhancement needs and supporting positive changes in their care providing for bereaved families. The optimized training programme should be widely applied in the future. More efforts from the hospitals, managers, obstetric nurses, and midwives are needed to jointly contribute to forming a uniform care pathway and promoting a supportive perinatal bereavement care practice.

KEYWORDS

bereavement care, qualitative, pregnancy loss, medical education, training

1. Introduction

Perinatal loss occurs worldwide in 25% of known pregnancies and includes miscarriage, stillbirth, therapeutic abortion, and neonatal death (1–3). Women who experience perinatal loss often receive unsatisfactory care, feeling that the care they received did not meet their expectations (4). An online survey done in 44 high-and middle-income countries reported that 25.4% of 3,769 bereaved parents reported disrespectful care and 23.5% reported disrespectful care of their baby (5).

Nurses and midwives play an important role in providing education, assistance, and treatment options to the women who experience perinatal loss (6). When the bereaved families experience congruence between the care they expect and the care they actually receive, the intensity of their grief is lower (7). However, nursing and midwifery professionals reported that they lacked sufficient confidence and capability to provide perinatal bereavement care (8–10), which may have a significant influence on the bereaved families' mental health (11–13). Nursing and midwifery professionals themselves bear a heavy emotional burden during the provision of perinatal bereavement care (14–16). It may further hinder the quality of care they offer (17, 18). Thus, the gap between women's needs for high-quality care and nursing and midwifery professionals' lack of perinatal bereavement care capability has become increasingly prominent (17, 19, 20).

A series of educational programmes in the UK, USA, and Ireland identified in a scoping review sought to prepare nurses and midwives to care for the bereaved parents (21). However, the programmes reviewed lacked a theoretical base and comprehensive training contents. Relevant guidelines have addressed the necessity of conducting continuous training for healthcare professionals in the field of perinatal bereavement care (22–24). Systematic and comprehensive perinatal bereavement care has not been developed in China. Therefore, there was urgency to establish and conduct the Perinatal Bereavement Care Training Programme (PBCTP) to prepare competent nursing and midwifery professionals to provide this specialized care in China. This qualitative study aimed to understand nursing and midwifery professionals' experiences with the PBCTP after its implementation at a tertiary level maternity hospital in China.

2. Materials and methods

2.1. Study design

A qualitative descriptive study design was used to record and analyze obstetric nurses' and midwives' experiences with the PBCTP. This approach is appropriate for research questions that aim to give a straightforward report of participant experience (25, 26). This study was approved by the Women's Hospital School of Medicine, Zhejiang University (IRB no. 20210091) Institutional Review Board. Results are reported according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) reporting guideline for qualitative studies (27).

2.2. The PBCTP intervention

The PBCTP intervention was specifically designed for nursing and midwifery professionals who provide perinatal bereavement care to parents who experience a perinatal loss. It aimed to meet nursing and midwifery professionals' knowledge and skills gaps and emotional support needs. There are five modules in this programme including programme introduction, general knowledge of perinatal bereavement care, practical skills of perinatal bereavement care, emotional support for nurses and midwives, and practices reflection and learning. Content was delivered *via* eight online theoretical courses, which were given every 2 days. Asynchronous online lectures, videos, workshops, mindfulness training, and other **Supplementary material**, were used in the teaching. Follow-up and questions were addressed through the WeChat communications app. More details of the PBCTP Intervention are described in this reference Qian et al. (28) and **Supplementary Table 1**. Participants received a WeChat prompt after each online course to submit a journal reflection. The aim of the reflective journal was to ensure participants' completion of the course. Mandarin was the language used in the PBCTP and reflective journal.

2.3. Setting and participants

This study was conducted in a delivery room and seven obstetric wards in a tertiary maternity hospital. There are approximately 140 hospital obstetric nurses and 65 midwives working in the study hospital. Nurses and midwives were enrolled in the PBCTP intervention between March and May 2022. The inclusion criteria were nurses or midwives (1) who have provided perinatal bereavement care and (2) who have over 1 year of nursing or midwifery experience. Nurses and midwives who could not complete the training or had received prior perinatal bereavement training were excluded.

2.4. Sampling and data collection

Figure 1 shows an overview of participants' inclusion. Purposive sampling was used to include both nurses and midwives who submitted eight reflective journal entries, and gave detailed, reflective comments in their journals instead of using a few simple words. After the PBCTP intervention, nurses and midwives received a phone call from the researchers (SS) inviting them to participate in the interview. Interviews were conducted *via* telephone using a semistructured interview guide (see **Supplementary Table 2**) and lasted 30 to 50 min. The interviews were conducted in Mandarin. At the beginning of the interview, the first author (JQ), provided a clear explanation of the objectives of the interview and informed participants about the sound recording. Written and verbal consents were obtained. Member checking was used to ensure the credibility of results (29).

2.5. Data analysis

Audio recordings were transcribed verbatim by the first author with personal identifiers removed. *A priori* codes were not used. Themes were identified through a six-stage thematic analysis (data familiarization, initial codes generation, searching for themes, carefully reviewing themes, defining and naming themes, and generating the report) (30). The data analysis was performed from August to October 2022. To fully understand the interviewees' experiences, the transcripts were read multiple times. Authors (JQ and MW) independently performed line-by-line coding to identify initial concepts adopting inductive and deductive theme approaches. In order to produce the initial codes, each code was highlighted with a different color to identify the most significant and meaningful elements. Then, codes were organized into meaningful categories. Higher-level clusters of categories were generated based on the associations between the codes and categories. Authors (JQ and MW) examined if the themes worked in relation to the coded extracts. After that, clear definitions for each theme were generated. Any disagreement was discussed by all the authors until consensus was reached. Data were managed using NVivo software, version 12 (QSR International).

3. Results

3.1. Demographic characteristics

We interviewed 12 (75%) hospital obstetric nurses and 4 (25%) midwives. One participant was male. The average age of participants was 30 years. All the participants held a baccalaureate degree. Five of the participants (31%) had ≤5 years clinical experience, 7 (44%) had 6–9 years clinical experience, and 4 (25%) had more than 10 years clinical experience.

Six dominant themes and subthemes were identified within our interviewees' experiences with the PBCTP intervention, which are shown in **Supplementary Table 3** with illustrative quotes.

3.2. Theme 1: Participants' aims of undertaking the training

3.2.1. Needs for knowledge and skill enhancement

Nursing and midwifery staff [6 participants (38%)] stated that in their clinical practice more and more women experience perinatal loss and their needs for knowledge and skill enhancement were proposed. Participants described that they were overwhelmed when encountering such situations. One participant claimed that she lacked communication skills with the bereaved women and their families. Nurses and midwives claimed that the lack of perinatal bereavement care ability made it difficult to obtain trust and cooperation of the women, which hinders their work efficiency.

> *"For such patients, we do not have a good way to cope with their emotional changes, which leads to some problems in mutual communication. It is difficult to get their cooperation, which will cause great problems for our work, so I think this course could help us to get a better sense of trust from the women" Nurse 8.*

3.2.2. Providing high-quality care

Due to sympathy for the bereaved women, nursing, and midwifery professionals [7 participants (44%)] instinctively wanted to give assistance to the women as much as possible. They wanted to alleviate women's negative emotions by providing high-quality care. They wanted to ameliorate women's negative feelings by providing high quality, emotionally sensitive care. However, the pervasive grief surrounding loss often caused care providers' usual communications to trigger more unpleasant feelings for women, resulting in dissatisfaction with the nursing and midwifery staff. Nurses and midwives realized that they needed to receive training to offer targeted care for the women and improve their satisfaction. They considered the PBCTP helpful to provide better humanistic care for the women.

> *"I found that a large part of the low satisfaction is from women who experienced perinatal loss. This kind of patient is particularly sensitive. If they do not receive proper care, they may have low opinion of the care" Nurse 8.*

3.3. Personal growth and practice changes after training

3.3.1. Caring from various perspectives

After training, nursing and midwifery staff [11 participants (69%)] improved their awareness of providing perinatal bereavement care and learned to care women from various perspectives. They began to pay attention to the various needs of the bereaved women, including emotional, spiritual, and other needs. Nurses and midwives more actively asked about women's needs, paid attention to the expression of language and showed sympathy to the women. They started to provide care based on the actual needs of women instead of their personal views. Nursing and midwifery staff seldom thought of providing care from the perspective of the grieving families' needs before the PBCTP. Following the PBCTP, they more easily integrated the families' psychological needs into their care. They thought that they should provide support for the entire family because of the important roles of family members in the women's psychological recovery.

> *"We can provide some help to family members, such as some emotional comfort. Because the support from family members is more important than medical staff. Some family members have a bad mood. It may bring women pressure" Nurse 7.*

3.3.2. Attaching importance to pain management

Some nursing and midwifery staff thought women who experience perinatal loss would have an easier birth than women who experience a full-term pregnancy with a full-grown baby. They underestimated the need for pain management among

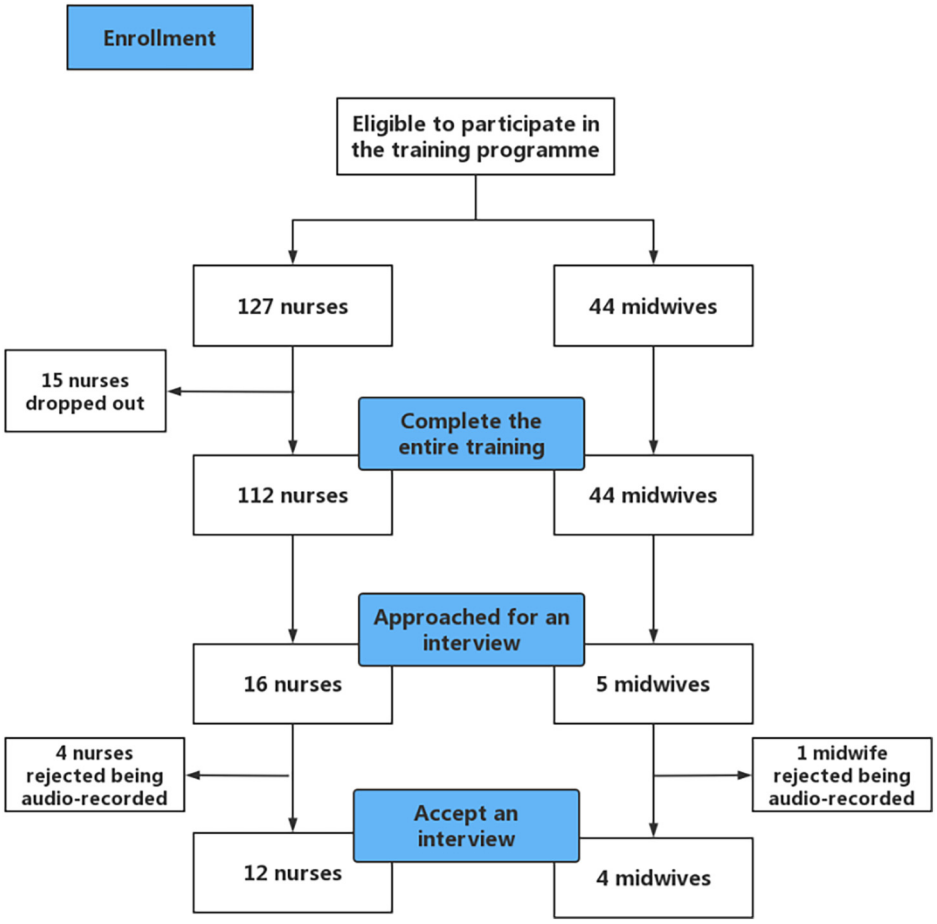


FIGURE 1
Flow diagram of participants' inclusion.

this population. After learning the course material, nursing and midwifery staff [6 participants (38%)] realized the importance of pain management and they suggested taking actions such as verbal comfort, physical pain management interventions, and analgesics.

> *"In the past, when women requested epidural pain relief, we suggested that they wait and told them that epidural use is not cost-effective and may affect progress of labor. Now we try to meet their analgesic needs as much as possible." Midwife 3 (Women were eligible for epidural pain relief if the gestational age was 28 weeks or greater. All women could receive pain relieving medications if they were unable to cope with the labor pain).*

3.3.3. More effective communication

Being afraid of communicating with the bereaved women was a notable problem in the care provided before the training. One participant described that she did not know how to start communication. After the training, nursing and midwifery staff [4 participants (25%)] gradually overcame their worries and started their communication by giving detailed health education. They would encourage women to recuperate and give instructions on future pregnancy plans.

> *"I didn't dare to communicate with them before, because I thought it would make them feel depressed. I left in a hurry after every ward inspection. Now when I see the bereaved women and their families, I will tell them the following precautions for labor induction. I have received training and learned many courses, so I have a certain knowledge reserve. I have the confidence and ability to communicate with them" Nurse 2.*

3.4. Theme 3: The most valuable training content

3.4.1. Respectful grief care

Nursing and midwifery staff [6 participants (38%)] found the content about grief care very practical. Nurses and midwives were more responsive to the women's needs for grief care. They learned to ask and respect women's grief care needs (e.g., provision of baby's footprint, dressing the baby, holding/seeing the baby, etc.). Nurses and midwives paid more attention to words and the expression of language. They said, "baby" rather than "a small corpse." When the women wanted to see the baby, they would also carefully prepare the baby's body for viewing.

> *"I usually don't pay attention to how to name the baby, maybe sometimes I just called the baby "a small corpse." The mother might be very upset to hear that. After we learned the course materials, we would say the baby's name. I'll also double-check if the mother wants to see the baby. We'll clean the baby when the mother needs to see her baby" Midwife 1.*

3.4.2. Increased medical knowledge

Nurses and midwives [5 participants (31%)] believed that learning more about the medical causes of pregnancy loss would help them better answer women's questions about pregnancy loss. Their basic education did not prepare them for understanding chromosomal abnormalities or fatal formation defects. The bereavement course provided nurses and midwives with more knowledge about the causes of pregnancy loss, and they felt more confident providing care after the course.

> *"I think we need to have more theoretical knowledge. For example, women often ask me questions about their babies who have chromosomal abnormalities. I lack theoretical knowledge in this aspect. I want to know what circumstances may lead to fetal abnormality. I want to learn something about doctors' theoretical knowledge" Midwife 2.*

3.4.3. Mindfulness breathing

After practicing mindfulness breathing in the course, the nursing and midwifery staff [5 participants (31%)] appreciated the effect of mindfulness breathing on relaxation. Mindfulness breathing, as a convenient skill, was recommended to patients by some nursing and midwifery staff. Nurses and midwives also described that they would use mindfulness breathing to release their negative emotions in life.

> *"In the provision of bereavement care, we also have negative emotions due to empathy. Providing empathetic bereavement care causes emotional distress in nurses and midwives. But we should manage emotions by ourselves. I think the mindfulness breathing in the course can be helpful in various events in life! Being immersed in sadness is not conducive to our work" Nurse 1.*

3.5. Theme 4: Suggestions for training improvement

3.5.1. Strengthening course interactions

Perinatal loss is a solemn topic. Nursing and midwifery staff reported that they do not communicate much about it. Nurses and midwives [7 participants (44%)] thought that online courses lacked communication and interaction between colleagues. They suggested that offline learning would be more intuitive and in-depth. Everyone could share experiences offline. One nurse suggested carrying out in-person courses so they could get immediate feedback on the questions.

> *"The courses are recorded, which is one-sided. If these courses can be carried out via several live broadcasts, the effect would be better. For example, if I don't get any feedback on a certain question, there may be a knowledge point that I don't understand. It seems that I have learned it, but I don't know how to use it" Nurse 8.*

3.5.2. Enriching clinical cases

Nursing and midwifery staff [5 participants (31%)] believed that more clinical cases should be added to the training. Theories and clinical cases should be combined because their understanding of perinatal bereavement concepts is theoretical in most cases, and they have doubts about how to apply specific theoretical knowledge to clinical practice.

> *"The course content could add more clinical cases. Some concepts are very theoretical and difficult to understand. Our practical experience is not extensive. It is not clear how to specifically apply it to clinical work" Nurse 8.*

3.5.3. Slowing the frequency of courses

Nursing and midwifery staff [3 participants (19%)] reported that the frequency of the course once every 2 days is a little too frequent. Because of their busy clinical work, participants sometimes forgot to study. Therefore, they suggested a slightly longer interval between the two courses.

> *"I think the frequency can be slightly slowed down. When we're learning, we might feel like we're starting a new lecture right after we've learned a lecture. Sometimes it's tiring to study this course after work. The interval can be slightly longer" Nurse 12.*

3.6. Theme 5: Directions for practice improvement

3.6.1. More appropriate content and form of perinatal bereavement care

The nursing and midwifery staff [4 participants (25%)] thought that bereavement care should be provided according to the needs of the bereaved women to be more easily accepted by them. They described that providing baby footprints as mementos is a good memento. In the process of understanding women's bereavement care needs, it was necessary to ask about their needs in a way that will not cause additional harm. Using a questionnaire that allowed the women to choose their preferred bereavement care was suggested.

> *"It would be nice to give the women a questionnaire about what kind of bereavement care they would like us to provide, such as the care of a small body or a farewell ceremony. Because she*

> *would feel (the nurse) presumptuous if you asked her directly" Nurse 3.*

3.6.2. Uniform service consensus

After receiving the training, nursing and midwifery staff [5 participants (31%)] believed that uniform and clear guidelines to support women experiencing a perinatal loss should be developed. Uniform guidelines would facilitate the provision of consistent and supportive care. More feedback is needed to guide the application of painless delivery during induction of labor.

> *"It would be better to have a uniform nursing process for this population. For example, what care should be provided for the women? Are there any special needs to be satisfied for them? We prefer something routine and then we could offer other assistance according to our clinical experience" Nurse 5.*

3.7. Theme 6: Influencing factors of practice optimization

3.7.1. Busy clinical work

Nursing and midwifery staff [4 participants (25%)] felt that the clinical workload was so heavy that they sometimes did not pay attention to their tone during communication and did not have enough time to provide satisfactory bereavement care. It's not that they did not respect the women. The nurses and midwives were intellectually and emotionally willing to provide high quality bereavement care, but their own exhaustion sometimes weakened their ability.

> *"After all, there are a lot of mothers who come to the department to give birth. Sometimes, when we are busy, we may not pay much attention to the bereaved women's emotions. Maybe at this time, they will feel even worse" Nurse 9.*

3.7.2. Environmental support from the hospital

Nurses and midwives [3 participants (19%)] felt that after training, it would be better if the hospital could provide a private, undisturbed place for perinatal bereavement care. Because postpartum rooms may be shared rooms, there is no private room for the farewell ceremony at the hospital. The cries of other babies can also affect the mood of the bereaved women. In rooms shared by several women, the women often do not want to share their experiences with professionals.

> *"If we give a woman who experiences the perinatal loss a separate room, she might be more willing to talk. Another important reason to a provide private room is that the bereaved women don't want to hear other babies' voices, because these voices are a stimulus to her" Nurse 3.*

4. Discussion

In this qualitative study, nurses and midwives described the PBCTP as supporting growth and changes in their behaviors of providing perinatal bereavement care. Nursing and midwifery staff reported the most valuable training content and suggestions for improving the training. Directions for practice improvement and relevant influencing factors were also proposed.

The reason why nursing and midwifery staff were willing to participate in the PBCTP was that they realized their need for skills enhancement in perinatal bereavement care. This finding is similar to previous research where nurses and midwives also felt inadequate in terms of genetic knowledge, psychological counseling, and communication skills (8, 31). They wished to enhance their ability to provide sufficient support for the bereaved women (9, 32, 33). One survey study found that training in perinatal bereavement care was an important influencing factor in nursing and midwifery staff's confidence of providing bereavement care (16). Therefore, our study and others highlight the significance of conducting this bereavement training among nurses and midwives.

The findings that nursing and midwifery professionals perceived personal growth and practice changes after training are consistent with other studies implementing this kind of training (34, 35). Positive feedback reflected the effectiveness of the PBCTP. The PBCTP intervention included comprehensive training content, based on a systematic scoping review (21) and the exploration of Chinese perinatal bereavement care (19), which enhanced the evidence-based rigor of the programme. Therefore, wider application and testing of the PBCTP intervention should be conducted with larger samples and in diverse communities to verify its effectiveness.

Respectful grief care increased medical knowledge and mindfulness breathing were considered as the most valuable contents of the training. In China, it is taboo to talk about death. Nurses and midwives were unfamiliar with providing respectful grief care. Nursing and midwifery staff in our study considered asking women's desire to see the baby after the delivery (36), calling a baby by name, and keeping baby's footprint (22) in this training as practical bereavement skills. In a previous study, nurses and midwives recommended training related to physical care issues and knowledge of genetics (37), which is similar to our study findings where nursing and midwifery staff acknowledged the value of learning more medical knowledge related to perinatal loss. Richer medical knowledge reserves ensure sufficient confidence to provide perinatal bereavement care.

Moreover, nursing and midwifery staff pointed out the value of practicing mindfulness breathing. A cross-sectional survey of 571 nurses and midwives showed approximately 50% reported average or severe secondary traumatic stress symptoms and emotional exhaustion after responding to providing perinatal bereavement care (16). Mindfulness training has been widely used for relieving healthcare professionals' traumatic psychological symptoms (38, 39). Mindfulness breathing in our training helped nursing and midwifery professionals to relieve negative emotions.

Due to the COVID-19 pandemic, the PBCTP intervention was implemented *via* online lectures. It reduced the course interactions

to some extent. A previous study found that blended learning using a mixture of online resources and materials combined with face-to-face meetings, and in-class interaction had active engagement (40). Organizing offline workshops and debriefing (41, 42) might enhance the learning effects in the future training. Case presentations and case-based discussions could help trainees to strengthen their understanding of abstract concepts (43). As suggested by nurses and midwives in our study, adding more clinical cases in the training could facilitate the combination of theory and clinical practice.

More appropriate content and form of perinatal bereavement care were suggested. Considering the sensitivity of providing perinatal bereavement care, it is important to evaluate women's bereavement needs in an appropriate way. The key to providing perinatal bereavement support is to respect women's preferences (44). Hospitals could develop tools such as a list of bereavement care services to help clinical nurses standardize the work process. It could help nursing and midwifery professionals better understand women's bereavement care needs in a way of avoiding awkwardness (45).

Busy clinical work and environmental support from the hospital were considered factors that limited practice optimization. In this study, nurses and midwives suggested slowing the frequency of courses. That suggestion reflected their busy clinical work. It is similar to previous findings where nurses and midwives also reported the heavy clinical workload was an influencing factor in the bereavement care they provided (19). The improvement of bereavement care involves the contribution of individuals, teams, institutions, and systems (46). Organizational support is necessary to create conditions that enable nursing and midwifery professionals to provide high-quality bereavement care (20). Effective and feasible knowledge should be integrated into the system and continuously monitored and evaluated to maintain the continuous application of knowledge (47). Hospital managers should optimize human resource allocation and create the necessary conditions to enable the implementation, monitoring, and evaluation of best practice perinatal bereavement care (22).

4.1. Strengths and limitations

The methodology of this study was rigorous, and a theory-driven approach was adopted for data analysis. This is the first implementation of PBCTP in China and the results were helpful to provide enlightenment about future training for healthcare providers in perinatal bereavement practice.

There are several limitations of this study. The number of participants represented was small. We included nurses and midwives who were expressive in their reflective journals. This purposive sampling may have generated favorably biased interview responses. Additionally, any improvement in perinatal bereavement care is recognized only from the perspectives of the nursing and midwifery professionals, as we did not obtain experiences from the bereaved families after the PBCTP training. Furthermore, the PBCTP was implemented only in a single hospital that included seven wards. Transferability of the findings to other health systems is unknown.

5. Conclusion

The PBCTP was described by nursing and midwifery professionals as satisfying their training needs and supporting their professional enhancement. It facilitated personal growth and positive changes in providing perinatal bereavement care. The PBCTP intervention should be optimized based on suggestions from nurses and midwives for wider dissemination and application to verify its effectiveness. More efforts from hospitals, managers, obstetric nurses, and midwives are needed to form a uniform care pathway and promote perinatal bereavement care practice.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Women's Hospital, School of Medicine, Zhejiang University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

JQ designed the content of the research, conducted all interviews, carried out the primary analysis, and wrote the manuscript. JQ and MW participated in the data analysis. SS was responsible for participant recruitment and data collection. CJ and XY contributed to the planning of the study and provided feedback on the study design, the results, and drafting of the manuscript. SC was responsible for the revision of the manuscript and consulted in data analysis. All authors were responsible for reviewing the results and contributed to the draft revision.

Funding

This study was funded by the Zhejiang Medical and Health Research Project (Foundation Number: 2022KY859) and the China Scholarship Council (Grant Number: 202106320274).

Acknowledgments

We thank all the nurses and midwives who gave their time to participate.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1122472/full#supplementary-material

References

1. Armstrong D. Impact of prior perinatal loss on subsequent pregnancies. *J Obstet Gynecol Neonatal Nurs.* (2004) 33:765–73. doi: 10.1177/0884217504270714

2. Christiansen D. Posttraumatic stress disorder in parents following infant death: a systematic review. *Clin Psychol Rev.* (2017) 51:60–74. doi: 10.1016/j.cpr.2016.10.007

3. Dimarco M, Renker P, Medas J, Bertosa H, Goranitis J. Effects of an educational bereavement program on health care professionals' perceptions of perinatal loss. *J Contin Educ Nurs.* (2002) 33:180. doi: 10.3928/0022-0124-20020701-10

4. Kuforiji O, Mills T, Lovell K. Women's experiences of care and support following perinatal death in high burden countries: a metasynthesis. *Women Birth.* (2022) 6:e195–202. doi: 10.1016/j.wombi.2022.07.170

5. Atkins B, Blencowe H, Boyle F, Sacks E, Horey D, Flenady V. Is care of stillborn babies and their parents respectful? Results from an international online survey. *BJOG.* (2022) 129:1731–9. doi: 10.1111/1471-0528.17138

6. Davoudian T, Gibbins K, Cirino N. Perinatal loss: the impact on maternal mental health. *Obstet Gynecol Surv.* (2021) 76:223. doi: 10.1097/OGX.0000000000000874

7. Hutti M. Perinatal bereavement care. *MCN Am J Matern Child Nurs.* (2019) 44:5. doi: 10.1097/NMC.0000000000000496

8. Armour S, Gilkison A, Hunter M. The lived experience of midwives caring for women facing termination of pregnancy in the late second and third trimester. *Women Birth.* (2018) 31:S14. doi: 10.1016/j.wombi.2018.08.048

9. Mauri P, Squillace F. The experience of Italian nurses and midwives in the termination of pregnancy: a qualitative study. *Eur J Contracept Reprod Health Care.* (2017) 22:227–32. doi: 10.1080/13625187.2017.1318846

10. Andersson I, Gemzell-Danielsson K, Christensson K. Caring for women undergoing second-trimester medical termination of pregnancy. *Contraception.* (2014) 89:460–5. doi: 10.1016/j.contraception.2014.01.012

11. Meredith P, Wilson T, Branjerdporn G, Strong J, Desha L. "Not just a normal mum": a qualitative investigation of a support service for women who are pregnant subsequent to perinatal loss. *BMC Pregnancy Childbirth.* (2017) 17:6. doi: 10.1186/s12884-016-1200-9

12. Bakhbakhi D, Burden C, Storey C, Siassakos D. Care following stillbirth in high-resource settings: latest evidence, guidelines, and best practice points. *Semin Fetal Neonatal Med.* (2017) 22:161–6. doi: 10.1016/j.siny.2017.02.008

13. Ravaldi C, Levi M, Angeli E, Romeo G, Biffino M, Bonaiuti R, et al. Stillbirth and perinatal care: are professionals trained to address parents' needs? *Midwifery.* (2018) 64:53–9. doi: 10.1016/j.midw.2018.05.008

14. Beck C, LoGiudice J, Gable R. A mixed-methods study of secondary traumatic stress in certified Nurse-Midwives: shaken belief in the birth process. *J Midwifery Womens Health.* (2015) 60:16–23. doi: 10.1111/jmwh.12221

15. Gardiner P, Kent A, Rodriguez V. Evaluation of an international educational programme for health care professionals on best practice in the management of a perinatal death: improving perinatal mortality review and outcomes via education (IMPROVE). *BMC Pregnancy Childbirth.* (2016) 16:376. doi: 10.1186/s12884-016-1173-8

16. Qian J, Cai W, Sun S, Wang M, Yu X. Influencing factors of perinatal bereavement care confidence in nurses and midwives: a cross-sectional study. *Nurse Educ Today.* (2022) 117:105479. doi: 10.1016/j.nedt.2022.105479

17. Hu H, Qin C, Tang S, Deng Y, Li Y, Gong N. Analysis of potential factors of healthcare violence from the perspective of patients with fetal malformations inducing labor based on qualitative research. *J Shaoyang Univ.* (2020) 6:97–104.

18. Garel M, Etienne E, Blondel B, Dommergues M. French midwives' practice of termination of pregnancy for fetal abnormality. At what psychological and ethical cost? *Prenatal Diag.* (2010) 27:622–8. doi: 10.1002/pd.1755

19. Qian J, Wang W, Sun S, Wu M, Liu L, Sun Y, et al. Exploring interactions between women who have experienced pregnancy loss and obstetric nursing staff: a descriptive qualitative study in China. *BMC Pregnancy Childbirth.* (2022) 22:450. doi: 10.1186/s12884-022-04787-9

20. Ellis A, Chebsey C, Storey C, Bradley S, Jackson S, Flenady V, et al. Systematic review to understand and improve care after stillbirth: a review of parents' and healthcare professionals' experiences. *BMC Pregnancy Childbirth.* (2016) 16:16. doi: 10.1186/s12884-016-0806-2

21. Qian J, Sun S, Wu M, Liu L, Yaping S, Yu X. Preparing nurses and midwives to provide perinatal bereavement care: a systematic scoping review. *Nurs Educ Today.* (2021) 103:104962. doi: 10.1016/j.nedt.2021.104962

22. Boyle F, Horey D, Middleton P, Flenady V. Clinical practice guidelines for perinatal bereavement care — an overview. *Women Birth.* (2020) 33:107–10. doi: 10.1016/j.wombi.2019.01.008

23. Bereavement Care Standards Group. *National standards for bereavement care following pregnancy loss and perinatal death.* Bootle: HSE (2016).

24. Sands Australia. *Sands Australian principles of bereavement care.* Box Hill, VC: Sands Australia (2018).

25. Bradshaw C, Atkinson S, Doody O. Employing a qualitative description approach in health care research. *Glob Qual Nurs Res.* (2017) 4:2333393617742282. doi: 10.1177/2333393617742282

26. Sandelowski M. What's in a name? Qualitative description revisited. *Res Nurs Health.* (2010) 33:77–84. doi: 10.1002/nur.20362

27. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care.* (2007) 19:349–57. doi: 10.1093/intqhc/mzm042

28. Qian J, Sun S, Wang M, Liu L, Yu X. Effectiveness of the implementation of a perinatal bereavement care training programme on nurses and midwives: protocol for a mixed-method study. *BMJ Open.* (2022) 12:e59660. doi: 10.1136/bmjopen-2021-059660

29. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? *Qual Health Res.* (2016) 26:1802–11. doi: 10.1177/1049732316654870

30. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* (2006) 3:77–101. doi: 10.1191/1478088706qp063oa

31. Qian J, Pan P, Wu M, Zheng Q, Sun S, Liu L, et al. The experiences of nurses and midwives who provide surgical abortion care: a qualitative systematic review. *J Adv Nurs.* (2021) 77:3644–56. doi: 10.1111/jan.14853

32. Strefling I, Lunardi Filho W, Kerber N, Soares M, Ribeiro J. Nursing perceptions about abortion management and care: a qualitative study. *Texto Context Enferm.* (2015) 24:784–91. doi: 10.1590/0104-07072015000940014

33. Parker A, Swanson H, Frunchak V. Needs of labor and delivery nurses caring for women undergoing pregnancy termination. *J Obstet Gynecol Neonatal Nurs.* (2014) 43:478. doi: 10.1111/1552-6909.12475

34. Slade P, Sheen K, Collinge S, Butters J, Spiby H. Acceptability of a programme for the prevention of post-traumatic stress disorder in midwifery: a qualitative investigation with midwives and midwifery managers following feasibility testing. *MIDIRS Midwifery Digest.* (2020) 30:27–32.

35. Ratislavova K, Stipkova M. The perinatal loss care educational programme and its evaluation. *Zdrav Varst.* (2020) 59:1–7. doi: 10.2478/sjph-2020-0001

36. Kingdon C, Givens J, O'Donnell E, Turner M. Seeing and holding baby: systematic review of clinical management and parental outcomes after stillbirth. *Birth.* (2015) 42:206–18. doi: 10.1111/birt.12176

37. Chiappetta-Swanson C. *The process of caring: Nurses' perspectives on caring for women who end pregnancies for fetal anomaly*. Hamilton, ON: McMaster University (2001). 249 p.

38. Fiol-DeRoque M, Serrano-Ripoll M, Jiménez R, Zamanillo-Campos R, Yáñez-Juan A, Bennasar-Veny M, et al. A mobile phone-based intervention to reduce mental health problems in health care workers during the COVID-19 pandemic (PsyCovidApp): randomized controlled trial. *JMIR Mhealth Uhealth.* (2021) 9:e27039. doi: 10.2196/27039

39. Gerhart J, O' Mahony S, Abrams I, Grosse J, Greene M, Levy M. A pilot test of a mindfulness-based communication training to enhance resilience in palliative care professionals. *J Context Behav Sci.* (2016) 5:89–96.

40. Protsiv M, Rosales-Klintz S, Bwanga F, Zwarenstein M, Atkins S. Blended learning across universities in a South-North-South collaboration: a case study. *Health Res Policy Syst.* (2016) 14:67. doi: 10.1186/s12961-016-0136-x

41. Sorce G, Chamberlain J. Evaluation of an education session using standardized patients and role play during perinatal bereavement. *J Neonatal Nurs.* (2019) 25:145–51. doi: 10.1016/j.jnn.2018.11.007

42. Doherty J, Cullen S, Casey B, Lloyd B, Sheehy L, Brosnan M, et al. Bereavement care education and training in clinical practice: supporting the development of confidence in student midwives. *Midwifery.* (2018) 66:1–9. doi: 10.1016/j.midw.2018.06.026

43. Karim M. Using clinical cases to restore basic science immunology knowledge in physicians and senior medical students. *Front Immunol.* (2020) 11:1756. doi: 10.3389/fimmu.2020.01756

44. NBCP. *Termination of pregnancy for fetal anomaly(ToPFA) bereavement care pathway*. London: NBCP (2022).

45. Lu L, Qiong Z, Jialu Q, Mengwei W, Yaping S, Xiaoyan Y. A qualitative study on the needs of grief counseling for women with fetal abnormalities induced labor. *Chin J Nurs.* (2021) 56:1509–15.

46. Graham I, Logan J. Innovations in knowledge transfer and continuity of care. *Can J Nurs Res.* (2004) 36:89–103.

47. Graham I, Logan J, Harrison M, Straus S, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? *J Contin Educ Health Prof.* (2006) 26:13–24. doi: 10.1002/chp.47

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 07 December 2023
DOI 10.3389/fmed.2023.1207388

OPEN ACCESS

EDITED BY
Florian Recker,
University of Bonn, Germany

REVIEWED BY
Melissa Neubacher,
Heinrich-Heine-University Hospital, Germany
Claus-Juergen Bauer,
University Hospital Bonn, Germany
Ricarda Neubauer,
University Hospital Bonn, Germany, in collaboration with reviewer C-JB

*CORRESPONDENCE
Isabel Graul
✉ isabel.graul@icloud.com

RECEIVED 18 April 2023
ACCEPTED 18 July 2023
PUBLISHED 07 December 2023

CITATION
Hertling S, Kaiser M, Schleußner E, Loos FM, Eckhardt N and Graul I (2023) Gender gap–Gender-specific development in the field of obstetrics and gynecology in Germany in the last 20 years.
Front. Med. 10:1207388.
doi: 10.3389/fmed.2023.1207388

Gender gap–Gender-specific development in the field of obstetrics and gynecology in Germany in the last 20 years

Stefan Hertling[1,2], Mario Kaiser[3], Ekkehard Schleußner[1], Franziska Maria Loos[4], Niklas Eckhardt[5] and Isabel Graul[2,6]*

[1]Department of Obstetrics and Gynecology, University Hospital Jena, Jena, Germany, [2]Department of Orthopedic, Campus Eisenberg, University Hospital Jena, Eisenberg, Germany, [3]Jenoptik GmbH, Jena, Germany, [4]Practice for Orthopedics and Shoulder Surgery Leipzig, Leipzig, Germany, [5]Institute for Diagnostic and Interventional Radiology, University Hospital Jena, Jena, Germany, [6]Department of Trauma-, Hand and Reconstructive Surgery, University of Jena, Jena, Germany

Background: Gender Gap refers to differences between men and women in terms of access to medical education, career development, and leadership positions in medical practice and research. Although women now make up most medical school graduates in many countries, they are often underrepresented in higher positions.

Objective: The aim of this study is therefore to analyze the gender-specific development in the field of Obstetrics and Gynecology in Germany over the past 20 years and to survey the current *status quo*.

Materials and methods: An narrative review was carried out on the development of female graduates of human medicine, the proportion of women in contract medical care and clinical care, as well as the gender-specific evaluation of obtaining a gynecological/obstetric additional qualification. habilitation figures in the field of Obstetrics and Gynecology were evaluated about gender distribution. All data were received from federal institutes.

Results: A total of 46.7% (n = 95,234) of all inpatient doctors were female. A total of 46.7% (n = 95,234) of the physicians in hospitals were female. A total of 46% (1,832/3,958) were the portion of females as assistant physicians, 39.8% (n = 45.551) as specialists, 35.3% (n = 18789) as senior physicians, 25.1% (n = 2394) as first senior physicians and 25% (n = 10) as chief physicians in hospital. A total of 64.6% (n = 3958) of the physicians in Obstetrics and Gynecology were female. A total of 46% (1,832/3,958) were the portion of females as assistant physicians, 64.6% (n = 3958) as specialists, 65.0% (n = 1919) as senior physicians, 26.4% (n = 207) first senior physicians and 25% (n = 10) as chief physicians in Obstetrics and Gynecology.

Discussion: The problem with the gender gap in medicine, does not seem to be access to teaching or starting a residency. But in the functions with increasing responsibility and management functions, e.g., as senior physicians, women are already rarely seen. In Obstetrics and Gynecology, too, there is a shortage of women in leading positions, despite the relatively high numbers, for example as senior physicians. Factors like maternity and establishing a family are points mentioned therefore, but also stereotypes seem to be considerable facts.

Conclusion: However, it is important to recognize the need for more women in higher positions in medicine and actively work to encourage more women to choose a career in medicine.

KEYWORDS

gender gap, gynecology, obstetrics, Germany, women, leadership

Introduction

The term "Gender Gap" refers to the gender-specific inequality that exists in many areas of society. Essentially, Gender Gap means that there are differences between genders that can affect various aspects of life, such as access to education, employment opportunities, wages, political representation, or healthcare (1). The Gender Gap can have both positive and negative impacts on society, depending on whether the differences between genders result in one group being disadvantaged or preferred. Eliminating the Gender Gap is an important goal of gender equality efforts to ensure that all people have equal rights and opportunities regardless of their gender (2). There is also a Gender Gap among physicians. This Gender Gap refers to differences between men and women in terms of access to medical education, career development, and leadership positions in medical practice and research. Although women now make up most medical school graduates in many countries, they are often underrepresented in higher positions (3). After medical school, physicians in Germany must complete a residency, which lasts 6 years for obstetrics and gynecology. During this time, the physicians are employed as assistant physicians. After completion of the residency, the physician is a specialist. In addition to the acquisition of the qualification, there are also hierarchical structures in the clinic, so a doctor can also become a senior physician in recognition of his performance and assumption of responsibilities. The qualification of specialist is not necessary for this, but often already exists. At the top of the hierarchy is the chief physician.

For example, there are fewer female chief physicians, professors, or senior female scientists in medical research. The reasons for this can be diverse, including gender biases, discrimination, and structural barriers such as inadequate support for work-life balance. Eliminating the Gender Gap among physicians is important to ensure that women are equally represented in medical practice and research, and to ensure that all patients, regardless of their gender, receive equal medical care. The consequences of this inequality are far-reaching and affect both healthcare and society (4, 5). It is therefore crucial that measures are taken to improve women's access to medical education and career development and to promote gender equality in medicine. Although the Gender Gap in medicine varies in different specialties and countries, there are some areas that are particularly affected. One example is surgery, where women are often underrepresented, especially in surgical subspecialties such as cardiothoracic surgery or neurosurgery (6, 7). In other areas such as oncology, cardiology, and critical care medicine, women are also frequently underrepresented in higher positions and leadership roles (8–10). These differences can have negative effects on healthcare as women may not be able to benefit equally from certain specialties and treatment options. Therefore, it is important for the medical community to recognize the challenges of the Gender Gap in various specialties and take appropriate measures to ensure that women have equal career opportunities and access to medical care. Although gynecology is a specialty that is traditionally dominated by women, there is still a Gender Gap in this field especially in the leading positions. One challenge is that there are often few career opportunities in gynecology beyond clinical practice, which means that women are often underrepresented in higher positions such as chief physicians or professorships in gynecology (11). Furthermore, there are also challenges regarding the balance of family and career that can affect the career development of women in gynecology. For example, pregnancies and childcare can make it difficult for women to maintain or advance in higher positions. It is therefore important that the medical community is aware of the gender gap in gynecology and takes measures to ensure that women have equal career opportunities and that their health needs are adequately considered (12). The aim of this study is therefore to analyze the gender-specific development in the field of Obstetrics and Gynecology in Germany over the past 20 years and to survey the current *status quo*.

Materials and methods

For the purpose of the study, an analysis was carried out on the development of female graduates of human medicine, the proportion of women in contract medical care and clinical care, as well as the gender-specific evaluation of obtaining a gynecological/obstetric additional qualification. Habilitation figures in the field of Obstetrics and Gynecology were evaluated about gender distribution. The hypothesis of the present study was that the number of female doctors in the field of Obstetrics and Gynecology is increasing, but there remains a significant imbalance in gender distribution.

Ethics commission

The responsible Ethics Committee of the University of Jena was informed and did not raise any objections to the study (Reg. -Nr.:2019-1456-Bef).

Data collection

Between January and April 2023, this study analyzed publicly available data on gender distribution in Obstetrics and Gynecology

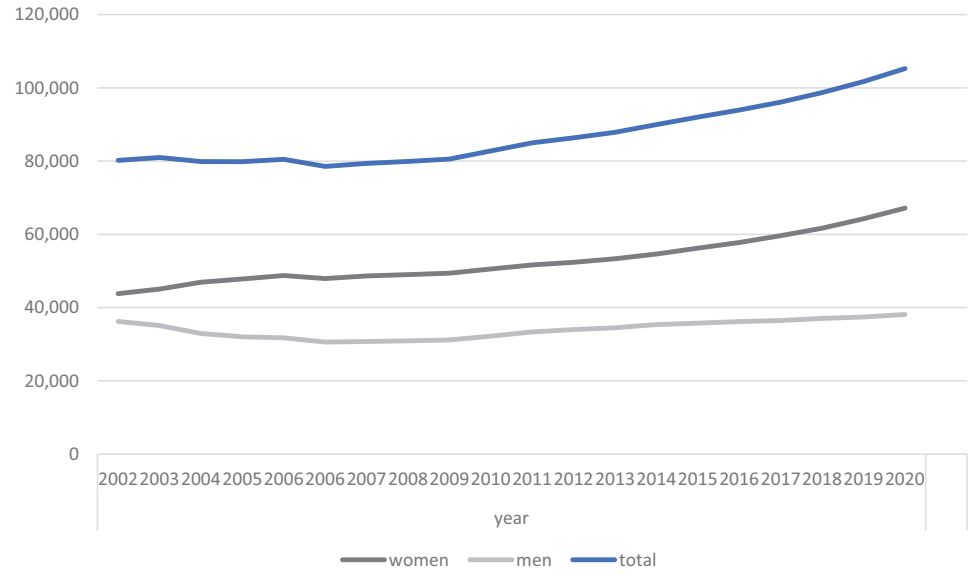


FIGURE 1
Medical students in Germany subdivided by gender over the years 2002–2020 in total amounts.

in Germany over the past 20 years from various national and international sources. These publicly available sources were:

- Statistical data on graduates of the human medicine program in Germany [Health Reporting of the Federal Government, (11)].
- Development in German hospitals [Information System of Health Reporting of the Federal Government, supported by the Robert Koch Institute (RKI) and the Federal Statistical Office, (12)].
- Development in outpatient care [Federal Physician Register of the National Association of Statutory Health Insurance Physicians, (13)].
- Data on the recognition of additional training in Obstetrics and Gynecology from 2016 to 2018 [German Medical Association, (13, 14)].
- Gender analysis of the leadership of a gynecological/obstetric university hospital (website of the 37 German university hospitals where the study of human medicine is state-approved).
- Data on successfully completed habilitations in the field of Obstetrics and Gynecology from 2002 to 2020 [University Statistics of the Federal Statistical Office, (15, 16)].

Data analysis

Statistical analysis was performed using IBM SPSS Statistics 29.0. Nominally scaled data were analyzed using the chi-square test. Measured metric values were examined for significant differences with the Mann-Whitney-*U*-test. The significance level was set at the 5% level ($p < 0.05$). Charts were produced using Microsoft Excel (Excel 2019, Microsoft Corporation, Redmond, Washington, DC, USA).

Results

Medical education

Currently, around 98,700 people are studying human medicine in Germany–of which approximately 63 percent are women (62,586/98,733). The number of first-year students has slightly increased over the past 10 years and decreased in 2021 compared to 2020. In 2021, around 1,750 more people were studying in the first semester than in 2012. The number of applicants for a medical degree program continues to far exceed the number of available spots. In 2021, 12,433 female and 6,337 male human medicine students completed their studies. Over the past 20 years, the number of female human medicine students has steadily increased from 43,807 in 2002 to 67,149 in 2020. This represents an increase of almost 53% (see **Figure 1**).

Specialist training/residency

Currently, there are a total of 409,121 physicians working in Germany. Of these, 48.2% (n = 197,036) are female doctors, with 13,422 working as specialists in gynecology and obstetrics. In general, the proportion of physicians under the age of 35 has decreased from almost 25 to 19% in the last 25 years. The number of physicians who have been employed in an outpatient setting has increased by 660% in the last 25 years. In the field of gynecology and obstetrics, the highest proportion of practicing gynecologists and obstetricians are aged between 50 and 59 years (n = 6,609). The proportion of practicing physicians in the field of Obstetrics and Gynecology under the age of 34 is the lowest (n = 684). Epidemiologically, almost 60% of all specialists in Obstetrics and Gynecology are over 50 years old.

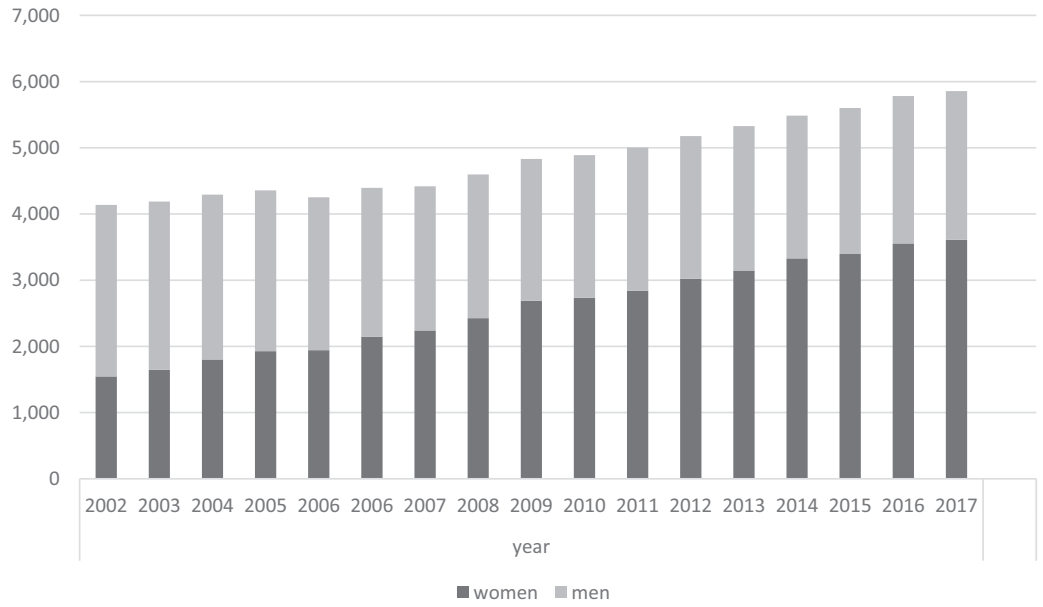


FIGURE 2
The development of the number of physicians of obstetrics and gynecology employed in the hospital from 2002 to 2017 in Germany is shown in figure.

Working place hospital

In 2021, a total of 203,286 male and female doctors were employed in the inpatient sector in Germany 46.7% (n = 95,234) of them were female.

A total of 43.7% (n = 88,803) were in specialist training, 56.0% (n = 49692) of them were female. Of these, 56.3% (n = 114,483) were male and female doctors with completed specialist training, 39.8% (n = 45.551) of them were female.

In 2021, 26.2% (n = 53,283) of the hospital-employed doctors were working as senior physicians and 35.3% (n = 18789) of them were female. Of these, nearly 1 percent did not have completed specialist training. A total of 8.0% (n = 53.283) were employed as first senior doctors with leading functions. Of these, 99 percent had completed specialist training and in total 14.7% (n = 2394) of the first senior physicians were female. Around 5% are working as female chief physicians. A total of 133,731 male and female assistant doctors were employed in German hospitals in 2021. The development of the number of physicians employed in the hospital over time is shown in Figure 2.

A total of 3.0% (n = 6,127) of the specialists were employed as specialist doctors in Obstetrics and Gynecology in hospitals and 64.6% (n = 3958) were female. Also 65.0% (n = 1919) of senior physicians in Obstetrics and Gynecology (n = 2954) were female. But only 26.4% (n = 207) of first senior physicians in Obstetrics and Gynecology (n = 784) were female.

Compared to the total number of women or senior physicians working in hospitals, statistically more women are employed as senior physicians in the field of Obstetrics and Gynecology (p > 0.001). Out of the 95,234 female doctors working in hospitals, 74,060 (77.8%) are working as assistant physicians. Compared to the total number, roughly 46% (1,832/3,958) of women within gynecology are working as assistant physicians,

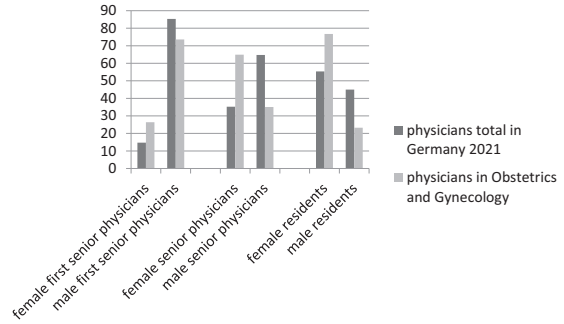


FIGURE 3
Proportion of genders in the positions: residents, senior physicians and first senior physicians in medicine and in obstetrics and gynecology in 2021 in Germany.

while nearly 78% of female doctors working in hospitals are working as assistant physicians (Figure 3). Therefore, statistically more assistant physicians are working in other fields than in Obstetrics and Gynecology (p > 0.001).

Specialists in obstetrics and gynecology

The number of specialist doctors in Obstetrics and Gynecology who work in a hospital setting has increased for both genders in the last 20 years, with the female proportion rising by almost 60% from 2002 to 2017 (from 1,546 to 3,612). In comparison to male specialists in gynecology, the proportion of female specialists in gynecology has shifted from being in the minority to being in the majority: in 2002, the number of female gynecologists was 1,546, and in 2017 it was 3,612. In 2021, 6,127 male and female specialist

doctors in Obstetrics and Gynecology were employed in a hospital setting, of which 3,958 were female specialist doctors in Obstetrics and Gynecology working full-time.

Leadership in gynecology/obstetrics at German universities

At the 39 state-run university hospitals in Germany, out of the 40 chief physician positions, 25% (n = 10) are held by women and 75% (n = 30) by men. Out of the 80 leading senior physician positions to be filled, 37.5% (n = 30) are held by women and 62.5% (n = 50) by men.

Specialization

A total of three specialized designations can be achieved in the field of Obstetrics and Gynecology: Endocrinology and Reproductive Medicine, Gynecologic Oncology, and Special Obstetrics and Perinatal Medicine. In 2020, a total of 1,157 doctors in the field of Obstetrics and Gynecology had a specialized designation: 18.9% (n = 219) in Endocrinology and Reproductive Medicine, 42.3% (n = 489) in Gynecologic Oncology, and 38.8% (n = 449) in Special Obstetrics and Perinatal Medicine. Of these 1,157 doctors, 52% (n = 599) were female and 48% (n = 558) were male with a specialized designation. Of the 219 doctors with the specialized designation in Endocrinology and Reproductive Medicine, 78% (n = 171) were female doctors of the 489 doctors with the specialized designation in Gynecologic Oncology, 33% (n = 161) were female doctors and of the 449 doctors with the specialized designation in Special Obstetrics and Perinatal Medicine, 59% (n = 267) were female doctors (Figure 4).

In 2020, a total of 700 doctors achieved the specialist title in gynecology and obstetrics. Of these, 16% (n = 116) were men and 84% (n = 584) were women (Figure 5).

Habilitations

From 2002 to 2020, a total of 872 doctors in Germany habilitated in the field of gynecology and obstetrics. Of these, 25% (n = 217) were women and 75% (n = 655) were men. The number of habilitations has decreased in recent years for both genders (Figure 6).

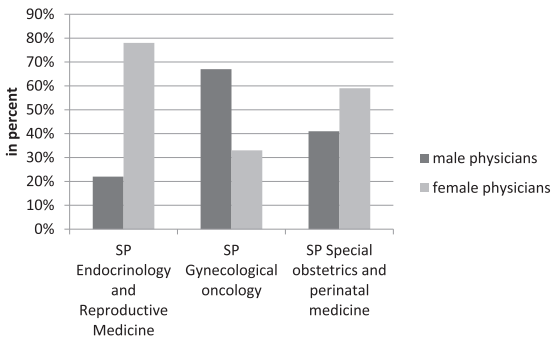


FIGURE 4
Additional designation of specialization in obstetrics and gynecology divided by gender.

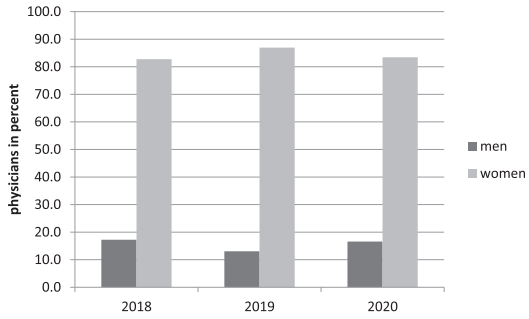


FIGURE 5
Physicians with specialization in obstetrics and gynecology 2018–2020 divided by gender.

Discussion

The gender gap in medicine has implications for many areas of healthcare (17). Here are some examples:

Diagnosis and treatment

Studies have shown that men and women often have different symptoms for the same conditions. If doctors do not take these differences into account, they may make incorrect diagnoses or prescribe ineffective treatments (18).

Research

Because women are often underrepresented in medical research, study results may not be applicable to both genders. This can lead to suboptimal treatments for women and impede understanding of certain diseases (19).

Career opportunities

Women are often underrepresented in the medical industry and have fewer opportunities for leadership positions. This can result in women having less influence over decisions that impact patient health (20).

Salaries

Women in medicine often earn less than their male counterparts. This can discourage women from pursuing a career in medicine or staying in the field (21).

These are just a few examples of how gender differences can impact healthcare. It is important to raise awareness of these

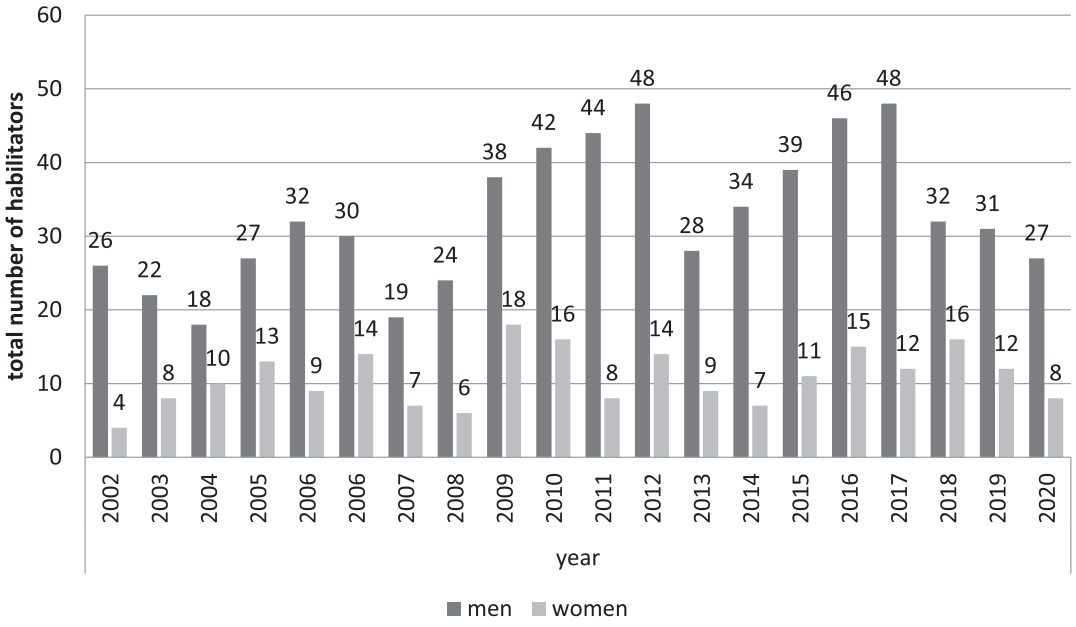


FIGURE 6
Habilitators in obstetrics and gynecology from 2002 to 2020 in Germany divided by gender.

issues and take action to ensure that women and men have equal opportunities and access to medical care (22). There is a growing number of female doctors in medicine.

In many countries, there are now more women than men who study medicine and work as doctors. This is a positive development as it increases diversity in the medical field and can contribute to better patient care. Women may approach certain aspects of healthcare differently than men, and their presence in the medical field can help better consider certain needs and perspectives (22). However, there are still some challenges that female doctors face, particularly about balancing family and career. In some medical specialties, women are still underrepresented, and there are still obstacles that can make it more difficult to pursue a successful career in medicine (23).

The problem with the gender gap in medicine, does not seem to be access to teaching, the majority of students are female. Starting a residency is not the problem either, again the majority is female. But by the time residency is completed, it no longer shows a majority of initially starting female physicians. In the functions in the hospital with increasing responsibility and management functions, e.g., as senior physicians, women are already rarely to be seen. Among first senior physicians and chiefs, women can be found in single digits. So it can be concluded that on the way to senior positions, women are not climbing the career ladder. But why is that? Morrison et al. in his recently published study for orthopedic surgeons, showed that maternity was a fact of being more difficult to advance in surgical specialties (24). This is an often cited problem that women have to take care of family and job is a point here (25).

In the study by Smith et al. it is also shown that the pandemic has had a further detrimental effect on women in leadership positions. thus, additional praying tasks have fallen on women. Women need up to 15 h per week to care for the household and children (26).

However, this does not seem to be the only factor. In particular, the stereotypes of the woman by herself, plays a major role. Thus, women more often go part-time and take jobs in medicine outside the hospital and often away from patient care (27).

Karaharju-Suvanto et al. showed gender differences in the career choices of dentists in his study. Women saw themselves as comforters, while men saw themselves more as technicians. Cultural ideals regarding appropriate occupations for men and women may have influenced their career choices. It appeared that male graduates seemed to concentrate in the private sector and chose more financially rewarding specialties that required technical skills, while women tended to prefer the public sector and focused on specialties that focused on social care and health promotion (28).

In the field of gynecology, there is a higher proportion of female doctors compared to other medical specialties. This is likely because gynecology deals with the health and wellbeing of women, and only few women prefer to be treated by a female doctor (29). According to a survey by the German Medical Association in 2020, about 69% of specialist doctors in Obstetrics and Gynecology are female, compared to an average female representation of about 46% among all specialist doctors. In other specialties such as surgery, orthopedics, or urology, the proportion of women is significantly lower (14).

In Obstetrics and Gynecology, too, there is a shortage of women in senior positions, despite the relatively high numbers, for example as senior physicians.

A greater number of women in this field could also help bring a more diverse perspective to medical practice and improve overall healthcare (30).

It is important to note that hiring doctors, whether male or female, depends primarily on their skills and qualifications, not their gender. The best person for the job should always be selected, regardless of gender, race, or other personal characteristics. Nevertheless, it is important to encourage women to choose a

career in medicine and actively address the obstacles and barriers they may face. This can help create a fairer and more diverse medical profession that better considers the needs and perspectives of all patients and doctors (31). The gender of a physician has no impact on their abilities as a gynecologist. A gynecologist, whether male or female, should have the necessary expertise, experience, and empathy to care for women in all aspects of reproductive health. Both male and female gynecologists can provide excellent care, and it is important for women to choose the physician with whom they feel most comfortable and can trust. Trust and open communication between physician and patient are crucial for good medical care. There are an increasing proportion of women in healthcare professions such as nursing and medical care (23). This change has various impacts on healthcare:

Improved patient care

Research has shown that female healthcare professionals often have a stronger ability to empathize and communicate with patients, leading to higher patient satisfaction and better outcomes.

Gender pay gap

Although women make up the majority of workers in healthcare, they are often concentrated in lower-paid roles such as nursing, while men are more likely to be found in higher-paid positions such as medicine and surgery. This leads to a gender pay gap (32).

Workload

As women often work in patient care and nursing, they often have a higher workload and must deal with stressful situations that can lead to burnout.

Potential changes in healthcare policy

With women playing a greater role in healthcare, their interests and needs can be better represented in healthcare policy.

Overall, the higher amount of women in healthcare has both positive and negative impacts. While women can often provide better patient care, they must also deal with challenges such as the gender pay gap and burnout (31, 32). The disadvantages of the gender gap are certainly individual for women in terms of career development, financial consideration and personal health. But the facts are also negative from a societal perspective, as less income is available for the economic system and the number of illnesses and thus healthcare costs increase due to psychological challenges such as mobbing and burnout (33).

There is a shortage of women in the areas of research work and the end point of habilitation. Only a quarter of habilitators in Obstetrics and Gynecology are women, even though they make up a majority of physicians. Often this is an additional work and can usually no longer be done by women in the medical professions in their quitting time.

In Germany, the problem for women in obstetrics and gynecology does not seem to be access to education or even the further path to becoming a senior physician, but here, too, there are no women in management positions.

Conclusion

The hiring of doctors should always be based on their skills, experience, and qualifications, not on their gender. However, it is important to recognize the need of women in medicine and actively work to encourage more women to choose a higher career in medicine.

This goal is difficult to achieve, it is essential to show the possibility of reconciling work and family, for example with kindergartens with appropriate opening times, breastfeeding options and flexible working hours. But that's not enough, especially stereotype thinking in Germany in relation to the additional challenges that women have to meet must be reduced. Women should be given the opportunity to continue their career path without being pushed into other jobs or part-time, which often block managerial positions. Furthermore, the option of an academic career must also be supported, if necessary with mentoring programs, provided working hours, but above all networking. Only a rethinking of the women concerned and the men in decision-making positions can reduce the gender gap in managerial positions in obstetrics and gynecology.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

SH, IG, ES, and FL contributed to the design and implementation of the research, analysis of the results, and the writing of the manuscript. NE analyzed the data, supervised the study, and edited the manuscript. MK performed the statistical analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the Else-Kröner-Fresenius Stiftung to IG.

Acknowledgments

We acknowledge support by the German Research Foundation Projekt-Nr. 512648189 and the Open Access Publication Fund of the Thüringer Universitäts- und Landesbibliothek Jena.

Conflict of interest

MK was employed by company Jenoptik GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Lönnqvist J. The gender gap in political psychology. *Front Psychol.* (2022) 13:1072494. doi: 10.3389/fpsyg.2022.1072494

2. Misra V, Safi F, Brewerton K, Wu W, Mason R, Chan A, et al. Gender disparity between authors in leading medical journals during the COVID-19 pandemic: a cross-sectional review. *BMJ Open.* (2021) 11:e051224. doi: 10.1136/bmjopen-2021-051224

3. Brown M, Erdman M, Munger A, Miller A. Despite growing number of women surgeons, authorship gender disparity in orthopaedic literature persists over 30 years. *Clin Orthop Relat Res.* (2020) 478:1542–52. doi: 10.1097/CORR.0000000000000849

4. Madden C, O'Malley R, O'Connor P, O'Dowd E, Byrne D, Lydon S. Gender in authorship and editorship in medical education journals: a bibliometric review. *Med Educ.* (2021) 55:678–88. doi: 10.1111/medu.14427

5. Bram J, Magee L, Parambath A, Bauer A, Lawler E, Miller P, et al. Glass ceiling in hand surgery: publication trends by gender. *Iowa Orthop J.* (2022) 42:3–9.

6. Buda A, Pendleton A, El-Gabri D, Miranda E, Bowder A, Dua A. Analysis of authorship trends in vascular surgery demonstrates a sticky surgical floor for women. *J Vasc Surg.* (2022) 75:20–8. doi: 10.1016/j.jvs.2021.07.228

7. Shah A. Women in Neurosurgery - the road less traveled. *Neurol India.* (2022) 70:1340–3. doi: 10.4103/0028-3886.355170

8. Mitchell C, Roussel M, Walsh L, Weeraratna A. Women in cancer research. *Nat Rev Cancer.* (2019) 19:547–52. doi: 10.1038/s41568-019-0176-y

9. Casadei B. Women in cardiology. *Hellenic J Cardiol.* (2021) 62:366–7. doi: 10.1016/j.hjc.2020.12.002

10. Joseph M, Ahasic A, Clark J, Templeton K. State of women in medicine: history, challenges, and the benefits of a diverse workforce. *Pediatrics.* (2021) 148:e2021051440C. doi: 10.1542/peds.2021-051440C

11. Destatis. *Gender pay gap*. Wiesbaden: Destatis (2023).

12. Deutscher-Ärztinnenbund. *Medical women on top*. (2016). Available online at: https://www.aerztinnenbund.de/downloads/4/WoT.pdf. Z. (accessed on Apr 18, 2023).

13. Destatis. *Studierende insgesamt und Studierende Deutsche im Studienfach Medizin nach Geschlecht*. Wiesbaden: Destatis (2023).

14. Kassenärztliche-Bundesvereinigung. *Statistische Information aus dem bundesarztregister*. (2023). Available online at: https://gesundheitsdaten.kbv.de/cms/html/16396.php. 2020 (accessed on Apr 18, 2023).

15. Bundesärztekammer. *Anerkennung von Zusatz-Weiterbildungen*. (2023). Available online at: https://www.bundesaerztekammer.de/ueber-uns/aerztestatistik/. (accessed on Apr 18, 2023)

16. Destatis. *Hochschulstatistik*. Wiesbaden: Destatis (2020).

17. Kiely K, Brady B, Byles J. Gender, mental health and ageing. *Maturitas.* (2019) 129:76–84. doi: 10.1016/j.maturitas.2019.09.004

18. Johnson P, Echanique K, Chitguppi C, Lee J, Toskala E. Gender gap among rhinology and skull base surgeons. *Int Forum Allergy Rhinol.* (2022) 12:1303–6. doi: 10.1002/alr.22986

19. Raguindin P, Muka T, Glisic M. Sex and gender gap in spinal cord injury research: focus on cardiometabolic diseases. A mini review. *Maturitas.* (2021) 147:14–8. doi: 10.1016/j.maturitas.2021.03.004

20. Gadek L, Dammann C, Savich R, Mmuo-Oji C, Barrera L, Gallagher P, et al. Gender analysis of Journal of Perinatology authorship during COVID-19. *J Perinatol.* (2023) 43:518–22. doi: 10.1038/s41372-022-01551-x

21. Ramakrishnan A, Sambuco D, Jagsi R. Women's participation in the medical profession: insights from experiences in Japan, Scandinavia, Russia, and Eastern Europe. *J Womens Health.* (2014) 23:927–34. doi: 10.1089/jwh.2014.4736

22. Warner A, Lehmann L. Gender wage disparities in medicine: time to close the gap. *J Gen Intern Med.* (2019) 34:1334–6. doi: 10.1007/s11606-019-04940-9

23. Chatterton C. Women in mental health nursing: angels or custodians? *Int Hist Nurs J.* (2000) 5:11–9.

24. Morrison L, Abbott A, Mack Z, Schneider P, Hiemstra L. What are the challenges related to family planning, pregnancy, and parenthood faced by women in orthopaedic surgery? A systematic review. *Clin Orthop Relat Res.* (2023) 481:1307–18. doi: 10.1097/CORR.0000000000002564

25. Alomar A. Fellowship and future career plans for orthopedic trainees: gender-based differences in influencing factors. *Heliyon.* (2022) 8:e10597. doi: 10.1016/j.heliyon.2022.e10597

26. Smith S, Sinkford J. Gender equality in the 21st century: overcoming barriers to women's leadership in global health. *J Dent Educ.* (2022) 86:1144–73. doi: 10.1002/jdd.13059

27. Rogus-Pulia N, Humbert I, Kolehmainen C, Carnes M. How gender stereotypes may limit female faculty advancement in communication sciences and disorders. *Am J Speech Lang Pathol.* (2018) 27:1598–611. doi: 10.1044/2018_AJSLP-17-0140

28. Karaharju-Suvanto T, Choroszewicz M, Näpänkangas R, Suominen A, Tolvanen M, Koivumäki J. The reproduction of gender differences in early career choices and professional identity of young dentist in Finland. *Eur J Dent Educ.* (2021) 25:451–9. doi: 10.1111/eje.12620

29. Johnson A, Schnatz P, Kelsey A, Ohannessian C. Do women prefer care from female or male obstetrician-gynecologists? A study of patient gender preference. *J Am Osteopath Assoc.* (2005) 105:369–79.

30. Pearlman J. Gender differences in the impact of job mobility on earnings: the role of occupational segregation. *Soc Sci Res.* (2018) 74:30–44. doi: 10.1016/j.ssresearch.2018.05.010

31. Magnusson C, Nermo M. Gender, parenthood and wage differences: the importance of time-consuming job characteristics. *Soc Indic Res.* (2017) 131:797–816. doi: 10.1007/s11205-016-1271-z

32. Harvey E, Ball C. Gender pay gap in medicine: not always an hours worked issue. *Can J Surg.* (2022) 65:E843–4. doi: 10.1503/cjs.016322

33. Newman C, Templeton K, Chin E. Inequity and women physicians: time to change millennia of societal beliefs. *Perm J.* (2020) 24:1–6. doi: 10.7812/TPP/20.024

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 13 December 2023
DOI 10.3389/fmed.2023.1296615

OPEN ACCESS

EDITED BY
Omneya Attallah,
Arab Academy for Science, Technology and Maritime Transport (AASTMT), Egypt

REVIEWED BY
Anna Siri,
University of Genoa, Italy
Hosna Salmani,
Iran University of Medical Sciences, Iran

*CORRESPONDENCE
Maximilian Riedel
✉ maximilian.riedel@mri.tum.de

RECEIVED 18 September 2023
ACCEPTED 27 November 2023
PUBLISHED 13 December 2023

CITATION
Riedel M, Kaefinger K, Stuehrenberg A, Ritter V, Amann N, Graf A, Recker F, Klein E, Kiechle M, Riedel F and Meyer B (2023) ChatGPT's performance in German OB/GYN exams – paving the way for AI-enhanced medical education and clinical practice. *Front. Med.* 10:1296615. doi: 10.3389/fmed.2023.1296615

ChatGPT's performance in German OB/GYN exams – paving the way for AI-enhanced medical education and clinical practice

Maximilian Riedel[1]*, Katharina Kaefinger[1], Antonia Stuehrenberg[1], Viktoria Ritter[1], Niklas Amann[2], Anna Graf[1], Florian Recker[3], Evelyn Klein[1], Marion Kiechle[1], Fabian Riedel[4] and Bastian Meyer[1]

[1]Department of Gynecology and Obstetrics, Klinikum Rechts der Isar, Technical University Munich (TU), Munich, Germany, [2]Department of Gynecology and Obstetrics, Friedrich–Alexander-University Erlangen–Nuremberg (FAU), Erlangen, Germany, [3]Department of Gynecology and Obstetrics, Bonn University Hospital, Bonn, Germany, [4]Department of Gynecology and Obstetrics, Heidelberg University Hospital, Heidelberg, Germany

Background: Chat Generative Pre-Trained Transformer (ChatGPT) is an artificial learning and large language model tool developed by OpenAI in 2022. It utilizes deep learning algorithms to process natural language and generate responses, which renders it suitable for conversational interfaces. ChatGPT's potential to transform medical education and clinical practice is currently being explored, but its capabilities and limitations in this domain remain incompletely investigated. The present study aimed to assess ChatGPT's performance in medical knowledge competency for problem assessment in obstetrics and gynecology (OB/GYN).

Methods: Two datasets were established for analysis: questions (1) from OB/GYN course exams at a German university hospital and (2) from the German medical state licensing exams. In order to assess ChatGPT's performance, questions were entered into the chat interface, and responses were documented. A quantitative analysis compared ChatGPT's accuracy with that of medical students for different levels of difficulty and types of questions. Additionally, a qualitative analysis assessed the quality of ChatGPT's responses regarding ease of understanding, conciseness, accuracy, completeness, and relevance. Non-obvious insights generated by ChatGPT were evaluated, and a density index of insights was established in order to quantify the tool's ability to provide students with relevant and concise medical knowledge.

Results: ChatGPT demonstrated consistent and comparable performance across both datasets. It provided correct responses at a rate comparable with that of medical students, thereby indicating its ability to handle a diverse spectrum of questions ranging from general knowledge to complex clinical case presentations. The tool's accuracy was partly affected by question difficulty in the medical state exam dataset. Our qualitative assessment revealed that ChatGPT provided mostly accurate, complete, and relevant answers. ChatGPT additionally provided many non-obvious insights, especially in correctly answered questions, which indicates its potential for enhancing autonomous medical learning.

Conclusion: ChatGPT has promise as a supplementary tool in medical education and clinical practice. Its ability to provide accurate and insightful responses showcases its adaptability to complex clinical scenarios. As AI technologies continue to evolve, ChatGPT and similar tools may contribute to more efficient and personalized learning experiences and assistance for health care providers.

KEYWORDS

artificial intelligence, ChatGPT, medical education, machine learning, obstetrics and gynecology, students

Introduction

Chat Generative Pre-Trained Transformer (ChatGPT) is an artificial learning and large language model tool that was first released by OpenAI on 30 November 2022. Its web-browser-based conversational interface was built atop the large language model Generative Pretrained Transformer 3 (GPT-3), which had first been released in 2020 (1). Reinforcement learning techniques were further utilized to train the model on a dataset of 570 gigabytes of text, which allowed ChatGPT to establish probabilistic relationships between words and to perform natural language processing and generation tasks (2).

The theoretical foundation for the development of ChatGPT was grounded on the idea that language can be learned using patterns and rules found in large text corpora (3). ChatGPT generates its responses using deep learning algorithms that are trained on vast sets of text data, which enables the model to understand the structure, syntax, and semantics of natural language (4, 5). ChatGPT analyzes the input text and generates a response based on the patterns and rules it has learned during the training process (3). The model selects the most probable answer from a large set of potential responses and ranks these responses according to their likelihood of being coherent with and appropriate for the input text so as to mimic human language and provide relevant information or assistance. Prior to ChatGPT, large language models had predominantly been confined to the artificial intelligence (AI) research community. However, such models were not widely adopted by the general public due to their technical complexity and lack of accessibility. However, ChatGPT was different because it introduced a conversational interface that enabled users to interact with the AI in a more human-like manner (6).

The inclusion of ChatGPT or other AI applications in medical education and clinical practice may have the potential to transform the way in which students – as well doctors and patients – acquire knowledge in biomedical sciences (7, 8). In theory, by leveraging their vast knowledge base and real-time information processing capabilities, AI applications can offer personalized learning experiences to medical students. They can adapt to individual learning styles, providing tailored resources and interactive simulations to enhance understanding. Additionally, AI applications could provide instant feedback on students' performance, identifying areas of weakness and suggesting targeted improvement strategies. This continuous, adaptive learning approach promises to significantly improve the quality and effectiveness of medical education. This development is still at an early stage, and potential applications and benefits remain rather hypothetical due both to the short time for which they have been available and to the lack of impactful studies on the topic. At this early stage of implementation of AI in medical education and clinical practice, our objective was to gain insights into both the capabilities and limitations of ChatGPT in this regard. One significant challenge is the lack of nuanced understanding and contextual judgment that AI applications currently have, which is crucial in medical training. AI systems might struggle to replicate the complex decision-making processes and ethical considerations inherent in clinical practice. Furthermore, issues around data privacy and security are paramount in medical education, where sensitive health information is involved.

We sought to assess the tool's ability to demonstrate medical knowledge and to evaluate its performance in the context of problem assessment in obstetrics and gynecology (OB/GYN). In order to achieve this goal, we investigated ChatGPT's performance by analyzing two datasets: first, the OB/GYN course examinations at the Technical University (TU) Munich, and second, the German medical state licensing exam, with a focus on OB/GYN-related questions. We conducted a quantitative and qualitative analysis and compared the results obtained by ChatGPT with those obtained by the students. We hypothesized that at this early stage of its development, ChatGPT should be capable both of processing medical questions and problems well and of providing a high-quality informative output for medical education. Therefore, we aimed to address and discuss three core questions in our study: (i) How well does ChatGPT perform on standardized written OB/GYN exam questions? (ii) Can ChatGPT potentially be applied as a teaching and learning tool in medical education? (iii) What are the potential applications and limitations of ChatGPT for future use in medical education and clinical practice?

Materials and methods

Data acquisition and processing

We established two datasets for analyzing and assessing ChatGPT's ability to understand medical topics and issues in OB/GYN. The first dataset was obtained from the exams in the OB/GYN course during the clinical stage of medical studies at the University Hospital of the Technical University Munich. The corresponding exams were held on 6 February 2023 (with 154 participants), 22 July 2022 (with 125 participants), 7 February 2022 (with 185 participants), 14 July 2021 (with 149 participants), and 1 February 2021 (with 173 participants). The exams are mandatory for passing the OB/GYN course and are usually held in the eighth semester. They test theoretical knowledge across the entire spectrum of OB/GYN that is taught in lectures and small-group seminars during the semester. Topics include general gynecology, prenatal and perinatal medicine, gynecologic oncology, endocrinology, and reproductive medicine. Multiple-choice questions are provided with five answer options each, only one of which is correct. The exams test both the knowledge and understanding of clinical concepts through clinical case presentations. The answer sheets were evaluated anonymously for our study, and the mean rate of correct answers was calculated for each question individually.

The data of the OB/GYN-related state exam questions were obtained from the online teaching and learning software AMBOSS®. This commercial learning platform is particularly well-regarded for its integrated learning system, which includes detailed medical

articles, an extensive question bank for exam preparation, and interactive case studies to enhance clinical understanding. One of the standout features of AMBOSS® is its cross-linked articles, which allow users to easily navigate between related topics, making it an efficient tool for both studying and quick reference in clinical practice (9). The state exam ("Zweiter Abschnitt der Ärztlichen Prüfung") typically consists of just over 300 multiple-choice questions with five answer options each, only one of which is correct. The individual specialties and topic areas are represented differently based on their general clinical relevance. OB/GYN-related questions account for approximately 5% of all questions on the state exam. The questions consist partly of general-knowledge questions as well as of questions in the form of clinical case presentations. The questions are compiled anew biannually for the state exam by the Institute for Medical and Pharmaceutical Exam Questions (IMPP). The exams are uniform across Germany and are conducted over three consecutive days. Students can answer the multiple-choice questions from the state exams of recent years through the online platform AMBOSS®, where they can receive annotated explanations and links to teaching content for each answer option. AMBOSS® also offers the option to indicate the relative difficulty of a question on a scale of 1–5 for personal self-assessment and provides the mean rate of correct answers for each question on the exam. For our dataset, we used the latest 104 questions available on AMBOSS® at the time of data acquisition, and we limited the questions to those related to the field of OB/GYN from the state exams from autumn 2022, spring 2022, autumn 2021, and spring 2021.

We standardized our input formats for both datasets in ChatGPT in line with Gilson et al. (10). This standardization was crucial as it is widely recognized that the phrasing of a prompt can significantly influence the AI's response or output. Consequently, we excluded questions that contained images because ChatGPT only accepted textual inputs. Additionally, we removed questions with answers presented in table format. The questions were formatted by presenting the question first, followed by the possible multiple-choice answers, each on a new line with a clear numeration of the answer options. A new chat in ChatGPT was opened for each question.

After establishing the two datasets as described above, we investigated ChatGPT's ability to correctly answer the exam questions. To do so, we entered the corresponding questions with their five answer options into the chat interface of ChatGPT between 10 and 28 February 2023 during several input sessions (not counted). The response time by ChatGPT is immediate as it requires only seconds to process the question and create a cohesive response without the need for corrections. Questions and answers were obtained in the German language. Access was granted through free registration and login via the website https://openai.com/blog/chatgpt. The most recent version of ChatGPT-3 was used at the time of data analysis, and the tool's response was documented immediately for each individual question. Version 3 of ChatGPT was the first publicly available version capable of comprehensively answering users' questions formulated in a free and open-ended manner. At the time of data collection for the present study, the most recent text corpora on which ChatGPT had been trained stemmed from September 2021. Therefore, more recent or innovative medical findings after that date may not have been incorporated into the AI's knowledge base that could have a negative effect on the quality of output. After data acquisition, data analysis followed as described below.

Data analysis

In order to compare the accuracy of ChatGPT with that of the students from both datasets, we evaluated each answer manually and only considered a response to be correct if it clearly matched one of the five options. During this stage of data analysis, we disregarded any explanations or justifications of the correct answer as well as comments on the incorrect choices. Furthermore, we classified the level of difficulty for each question based on whether it had been answered correctly by more or fewer than the mean number of students, which was 83% of students for the OB/GYN course and 73% of students for the state exam questions. Thus, we established a simple, yet objective dichotomous characterization of a level of difficulty for each question for further statistical analysis. We then examined whether ChatGPT exhibited any significant differences in its rate of correct responses for the group of easy or difficult questions. Furthermore, we examined whether there was a correlation between the numerical difficulty score (from 1 = "very easy" to 5 = "very difficult") that had been assigned by AMBOSS® for each question on the one hand and ChatGPT's performance on the other hand.

In order to further investigate ChatGPT's ability to process information and provide useful medical information, the multiple-choice questions from both the OB/GYN course and the state exam were assigned to the following two groups:

1 General knowledge questions; for example: *"Which statement about cervical cancer is correct?"*
2 Clinical case presentation; for example: *"A 40-year-old gravida I / para 0 presents to the labor ward on the weekend at 34 + 4 weeks of gestation with complaints of having felt unwell for two days. She reports increasing lower limb edema, nausea, and vomiting since yesterday evening, intermittent visual disturbances and tinnitus since a few hours ago, and right upper quadrant abdominal discomfort. Which diagnostic measure is NOT helpful in this situation?"*

We evaluated ChatGPT's responses to each question based on five variables that are indicative of data quality. For our analysis, we orientated ourselves around the conceptual framework created by Richard Wang and Diane Strong (11). Their work, often cited in the field of information systems, identifies several key dimensions of data quality, which are critical for ensuring that the information being used or analyzed is reliable, accurate, and useful. By using these categories derived from Wang and Strong's framework, our evaluation of ChatGPT's responses not only adheres to established principles of data quality but also provides a structured and thorough method for assessing the effectiveness and reliability of AI-generated information.

We chose five categories to characterize the answers given by ChatGPT:

1 *Ease of understanding*: Was the answer clearly and precisely formulated in a way that was easy to understand?
2 *Concise representation*: Was the answer clearly structured and divided into sections that facilitated readability?
3 *Accuracy*: Did the facts mentioned in the answer correspond to the current scientific literature? Were the statements logical and understandable?

4 *Completeness*: Was the answer complete, and were all aspects of the question adequately addressed? Was important information omitted, or were there unnecessary details?
5 *Relevance*: Was the answer directly related to the question asked, or was there any ambiguity in the answer?

Three medical experts in the field of OB/GYN with long clinical experience individually assessed each answer independently with regard to the five items above using a five-point Likert scale (ranging from 1 = "completely disagree" to 5 = "completely agree"). Mean values from their responses were used for further statistical analysis to ensure consistency of the data.

Subsequently, for every question that had received an incorrect response, we categorized the reason for the error into one of the three options listed below. The authors responsible for conducting the qualitative analysis of the responses (i.e., MR, BM, and FR) collaborated on the analysis and resolved any ambiguous labels.

1 *Incorrect external information.*
2 For the question, ChatGPT used incorrect external information that could not be directly derived from the content of the question.
3 *Failure to consider information within the question.*
4 ChatGPT did not consider information mentioned in the question when generating the answer.
5 *Incorrect linkage of external knowledge with information within the question.*
6 ChatGPT considered the information in the question but combined it incorrectly with correct external knowledge.

ChatGPT's responses to every question in both datasets were further analyzed for the presence and quantity of "non-obvious insights" as described and applied by Kung et al. (12). In the authors' recent publication, an insight was defined by a section of the answer that was characterized by the following four items:

1 *Non-definitional*: Did not simply define a term in the input question.
2 *Unique*: A single insight may have been used to eliminate several answer choices.
3 *Non-obvious*: Required deduction or knowledge external to the question.
4 *Valid*: Clinically or numerically accurate; preserved directionality.

Using Kung et al. as an example, we then established an index ("density of insights") by normalizing the number of insights to the word count for each response generated by ChatGPT (number of insights/word count *100). The significance of the "density of insights" index in our case lies in its ability to quantify the richness and depth of information provided by ChatGPT. In medical contexts, where every word can carry significant weight and the efficiency of communication is crucial, this index helps in assessing whether ChatGPT is providing dense and meaningful content that goes beyond superficial explanations. By focusing on non-obvious insights—those elements of a response that are not immediately apparent, require deeper knowledge or deductive reasoning, and are clinically or numerically accurate—the index ensures that the evaluated content is not only informative but also relevant and applicable in a real-world setting.

Statistical analysis

The data were evaluated descriptively using Excel (Microsoft®) or Prism (Version 9, GraphPad®). Unpaired chi-square tests were used to determine whether the question difficulty or the type of question (i.e., general knowledge vs. clinical case) had significantly affected ChatGPT's performance in either dataset. These tests are ideal for comparing categorical variables and are appropriate for assessing whether there are significant differences in the distribution of correct and incorrect answers across these categories. The McNemar statistical test was used to assess whether the availability of multiple-choice answer options had an impact on the accuracy of ChatGPT in answering the questions. The test is used for paired nominal data. Comparing the performance of the same entity (ChatGPT) under two scenarios (with and without multiple-choice options), this test is suitable for evaluating whether these conditions lead to a statistically significant difference in performance. Unpaired *t*-tests were used to compare the word count between correct and incorrect answers generated by ChatGPT, the expert assessment of ChatGPT's responses via Likert scale ratings and to analyze the density of insight. These tests are used to compare the means of two independent groups. All p-values < 0.05 were defined as statistically significant. Tables and figures were generated in Word (Microsoft®) and Prism (Version 9, GraphPad®).

Results

ChatGPT delivered comparable and consistent results

The dataset of questions from the OB/GYN course included a total of 160 questions, each of which contained five multiple-choice answer options. On average, the medical students in our survey answered 83.1% (95% CI = 80.0–86.2%) of these questions correctly. Average student results for questions ($n = 35$) from spring semester 2022 were unavailable. In the same dataset, ChatGPT provided correct answers 85.6% ($n = 137$) of the time. In order to test the consistency of ChatGPT's answers, we conducted a second individual validation round. In this second round, ChatGPT achieved similarly good results for the dataset, with 88.7% ($n = 142$) of answers being correct. Overall, ChatGPT provided consistent results for 91.6% ($n = 145$) of the questions in the validation round compared with in the first round of testing.

The dataset of the medical sate exam included a total of 104 questions. Students answered 73.4% (95% CI = 69.0–77.8%) of these questions correctly. Four questions had to be removed because they included images that ChatGPT could not process. Moreover, two questions were removed because ChatGPT could not commit to one single answer option. Of the remaining 98 questions, ChatGPT answered a total of 70.4% ($n = 69$) correctly.

ChatGPT maintained its performance without the need for multiple-choice answers

In order to further explore how the answer choices in the exam questions influence the way ChatGPT forms answers and solutions to clinical case presentations, ChatGPT had to solve the questions again, but this time without the five provided multiple-choice answers for each question. Of a total of 46 clinical case presentation questions from the OB/GYN course dataset, five negative questions ("which answer is incorrect?") had to be excluded because the answers given by ChatGPT could not be assessed without the multiple-choice answer options. ChatGPT's performance tended to be slightly worse (80.5%; $n = 33$ out of 41) without the multiple-choice answer options in comparison with its performance when the answer choices were provided for the whole group of clinical case presentations (84.8%; $n = 39$ out of 46). However, this difference was not statistically significant (p-value = 0.13).

The difficulty of the questions interfered with ChatGPT's performance in the OB/GYN course, but not in the state exam questions

We aimed to investigate whether the accuracy of ChatGPT's answers depended on the difficulty level of the clinical questions. Initially, we categorized questions as *easy* or *difficult* based on their rate of correct answers, which had to be either above or below the mean of correct answers achieved by the medical students across all questions in the datasets (83% for the OB/GYN course dataset and 73% for the state exam dataset). We found that ChatGPT's performance in the OB/GYN course dataset did not significantly depend on the level of difficulty of the question (p-value = 0.1). However, in the medical state exam dataset, ChatGPT's performance on questions that had been defined as *easy* was significantly better (p-value <0.01) (Table 1). Furthermore, we assigned a numerical difficulty score (ranging from 1 = "very easy" to 5 = "very difficult") to each question in the medical state exam dataset and correlated it with ChatGPT's performance. We observed a decline in the rate of correct answers from $n = 27$ (27.6%) for Level 1 difficulty to $n = 1$ (1.0%) for Level 5 difficulty. Correspondingly, there was a rise in the rate of incorrect answers from $n = 2$ (2.0%) for Level 1 difficulty to $n = 8$ (8.2%) for Level 5 difficulty (p-value <0.001) (Table 2).

ChatGPT answered simple knowledge and patient case questions equally well

For both datasets, we separated questions into one of the following groups: general knowledge questions or clinical case studies. We tested whether the type of question (i.e., knowledge vs. clinical case) altered ChatGPT's performance. For both the OB/Gyn course (p-value = 0.84) and the state exam questions (p-value = 0.42), we did not find significant differences between the two types of questions (Table 3).

ChatGPT delivered high-quality answers, especially for questions that were answered correctly

In our qualitative analysis of the reasons for ChatGPT's incorrect answers in the dataset of the OB/GYN course, the main reason was found to be "incorrect internal knowledge" ($n = 20$; 76.9%), followed by "incorrect connection of internal knowledge with external information in the question" ($n = 4$; 15.4%) and "no consideration of the external information provided in the question" ($n = 2$; 7.7%).

For further qualitative analysis, three medical experts from the field of OB/GYN assessed ChatGPT's answers using a five-point Likert scale (from 1 = "completely disagree" to 5 = "completely agree"). The experts found a positive assessment for both correct and incorrect answers with regard to the qualities of "ease of understanding" (mean Likert scale = 4.8 for correct answers and 4.6 for incorrect answers) and "concise representation" (mean Likert scale = 4.2 for correct answers and 3.9 for incorrect answers). By contrast, the qualities of "accuracy," "completeness," and "relevance" were found to have been significantly better assessed (p-value <0.0001) for the correct answers compared with for the incorrect answers (Table 4).

ChatGPT provided a high density of non-obvious insights

After assessing the quality and consistency of ChatGPT, we continued to evaluate the tool's potential capacity to enhance autonomous medical learning in the field of OB/GYN. ChatGPT provided answers with a significantly larger mean word count (p-value <0.0001) of 93 words for the state exam questions compared with 63 words for the OB/GYN course questions (Figure 1A). We further analyzed the quantity of non-obvious insights provided by ChatGPT. More than 85% of both the correctly and incorrectly

TABLE 1 ChatGPT's performance on questions from the OB/GYN course and the state exam.

		Performance	Question difficulty		
		Overall, *n* (%)	"easy," *n* (%)	"difficult," *n* (%)	*p*-value
OB/GYN course	Correct	104 (83.2)	69 (55.2)	35 (28.0)	0.1
	Incorrect	21 (16.8)	10 (8.0)	11 (8.8)	
State exam	Correct	69 (70.4)	51 (52.0)	20 (20.5)	< 0.001
	Incorrect	29 (29.6)	9 (9.1)	18 (18.4)	

The level of difficulty for each question was based on whether it had been answered correctly by more or less than the mean rate of students, which was 83% for the OB/GYN course and 73% for the state exam questions.

TABLE 2 ChatGPT's performance on the state exam questions.

		Question difficulty, *n* (%)					
		1	2	3	4	5	*p*-value
State exam	Correct	27 (27.6)	22 (22.4)	13 (13.3)	6 (6.1)	1 (1.0)	< 0.001
	Incorrect	2 (2.0)	8 (8.2)	6 (6.1)	5 (5.1)	8 (8.2)	

The level of difficulty for each question is depicted from 1 = "very low" to 5 = "very high," as assigned by AMBOSS®. For statistical analysis, the numerical difficulty scores 1–3 were depicted as "easy" and 4–5 as "difficult." Unpaired chi-square test was used to calculate *p*-values.

TABLE 3 Correct and incorrect answers by ChatGPT for each type of question (general knowledge or clinical case presentation).

		Type of question		
		Knowledge, *n* = (%)	Clinical case presentation, *n* = (%)	*p*-value
OB/GYN course	Correct	98 (61.3)	39 (24.3)	0.84
	Incorrect	16 (10.0)	7 (4.4)	
State exam	Correct	30 (30.6)	41 (41.8)	0.42
	Incorrect	9 (9.2)	18 (18.4)	

Unpaired chi-square test was used to calculate *p*-values.

answered OB/GYN course questions provided at least one non-obvious insight. With regard to the state exam question, however, the incorrect answers, in particular, demonstrated a decrease in the number of questions with at least one significant insight (Figure 1B). In order to better assess non-obvious insights generated by ChatGPT, we established a density index by normalizing the number of insights to the word count for each response generated by ChatGPT. Thus, we aimed to quantify ChatGPT's ability to provide knowledge in OB/GYN in a correct, concise, and relevant manner for medical education. In so doing, we noticed that the density of the insight index was significantly higher for questions that had been answered correctly compared with for those that had been answered incorrectly for both the state exam (p-value <0.0001) and the OB/GYN course question (p-value <0.0045) (Figure 1C).

Discussion

ChatGPT has gained significant attention from the public and media alike since its release in late 2022 (13–15). The tool surpassed 100 million users in January 2023, making it the fastest-growing consumer application to date (16). ChatGPT has been recognized both for its potential to revolutionize the way in which we interact with machines and for its ability to pave the way for the transformation of entire industries, such as media, marketing, and computer science (17). Some have even argued that ChatGPT may represent the beginning of a new industrial age in which AI and its applications will have a similar impact on economies and societies as the Internet had at the beginning of the 1990s (18). The fact that this new technology could also expand to other fields – such as medical education and clinical practice – is therefore not only possible, but also rather likely (19, 20).

Our study yielded novel findings that highlight the proficiency of ChatGPT to perform intricate tasks related to handling complex medical and clinical information in the field of OB/GYN. ChatGPT provided consistent answers and explanations to medial problems and did not require help in finding correct solutions to medical problems in multiple-choice answer options. Remarkably, ChatGPT demonstrated comparable results with those achieved by medical students in the setting of a real exam after extensive studying and exam preparation. Indeed, referring to the usual exam passing score of 60%, ChatGPT was capable of passing both our OB/GYN course examination (83.1%) and the OB/GYN share of the German medical licensing examination (73.4%). It is important to note that we refrained from providing any prompts or training to the AI during the experiment, and we systematically cleared the AI session before inputting each question variant in order to avoid a chain-of-thought bias (12).

Our results from the OB/GYN-related exams are in line with or slightly better than ChatGPT's performance in recent publications with regard to questions from the United States Medical Licensing Exam (USMLE) and the National Board of Medical Examiners (NBME) (10, 12). Interestingly, ChatGPT demonstrated particularly good results on questions from the OB/GYN course in our study. We hypothesize that the characteristics of the exam questions were crucial for obtaining such a result. In comparison with the state exam question, the OB/GYN course questions were significantly shorter (mean word count = 23 vs. 102 words), and the quality of the questions was different. The OB/GYN course questions were rather direct and required generally factual knowledge (e.g., "What is NOT a risk factor for the development of ovarian cancer?"), whereas the state exam questions addressed more complex medical questions (e.g., "A 64-year-old woman presents at a gynecological practice with vaginal bleeding that has been ongoing for 14 days. The patient weighs 90 kg and is 150 cm tall. She has been suffering from type 2 diabetes mellitus for 5 years, which is being treated with antidiabetic medication. The patient reports having consistently high blood pressure but is not following an antihypertensive therapy prescription. The suspected diagnosis of an early-stage malignancy (FIGO stage IA) has been confirmed. Which of the therapeutic options listed below is indicated next?"). The questions from the state exam seem to have been more challenging for both the medical students and ChatGPT. This higher level of difficulty of the state exam questions could be explained by the fact that they require both more complex processing and the weighing of multiple pieces of clinical information as compared with the OB/GYN course questions. This finding corresponds to the main reason that we identified for the incorrect answers given by ChatGPT: The primary cause was attributed to the processing of "incorrect internal knowledge," which accounted for 76.9% of the errors, while a significantly smaller percentage (7.7%) of the errors were attributed to "not considering the external information provided in the question." ChatGPT's insufficiencies could have been caused by an inadequately trained model, for example, due to the underrepresentation of clinically relevant and distinctive medical knowledge. Additionally, the human factor may have contributed to this finding because insufficient or invalid human judgment during the initial reinforcement stages could have affected the model's performance (12).

TABLE 4 Mean Likert scale ("1" = completely disagree; "5" = completely agree), with standard deviation (SD) for correctly and incorrectly answered questions by ChatGPT and the mean assessment by three medical experts depicted for each question.

		Ease of understanding	Concise representation	Accuracy	Completeness	Relevance
Incorrect	Mean	4.6	3.9	2.0	3.5	3.6
	SD	0.4	0.8	0.8	1.3	1.3
Correct	Mean	4.8	4.2	4.8	4.6	4.8
	SD	0.3	0.7	0.3	0.4	0.3
p-value		0.0268	0.0390	< 0.0001	< 0.0001	< 0.0001

Unpaired *t*-test was used to calculate *p*-values.

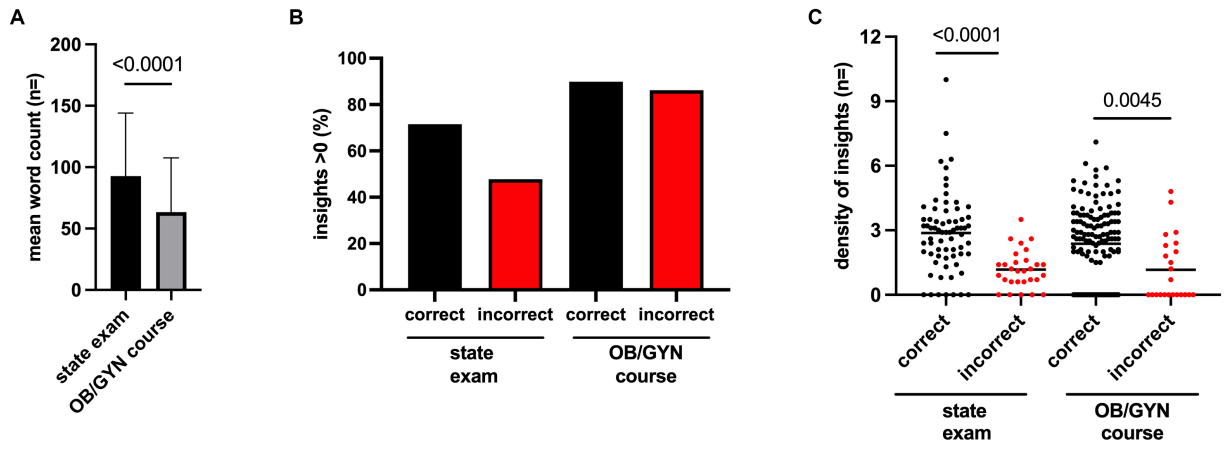


FIGURE 1
(A) Bar diagram depicting the mean word count, with standard deviation for ChatGPT's answers to the state exam and OB/GYN course questions. **(B)** Bar diagram depicting the percentage with at least a non-obvious insight for questions from the state exam and the OB/GYN course, differentiated by questions answered incorrectly or correctly by ChatGPT. **(C)** Scatter plot depicting the "density of insights" (number of insights/word count *100) of the responses by ChatGPT for the correctly or incorrectly answered questions from the state exam or the OB/GYN course. Each dot represents one answer by ChatGPT; the horizontal line represents mean. Unpaired *t*-test was used to calculate *p*-values. All *p*-values are shown as full digits.

Our analysis underlines the positive aspects of ChatGPT and its potential "real-world" application in medical education and clinical practice. With its use of machine-learning techniques, ChatGPT demonstrated high adaptability and accuracy in resolving medical problems. By implementing this approach, ChatGPT is able to thoroughly assess the context of a query and generate tailored responses that are customized to individual users (21). Moreover, the introduction of a conversational interface in ChatGPT enhances its usability, and the ability to ask follow-up questions enables users to gain a more comprehensive understanding of the concepts addressed in their queries, thereby enabling the tool to do more than merely output answers. However, it is important to acknowledge that the answers provided by ChatGPT are limited to the data on which the tool was trained and rely on patterns learned from vast amounts of text data. As a consequence, ChatGPT's responses may not always be up-to-date or entirely accurate for all scenarios (21). In a recent publication, Weng et al. stated that ChatGPT had failed Taiwan's Family Medicine Board Exam (22). The authors hypothesized that possible factors that had contributed to these results included the challenging nature of the specialist exam in family medicine and the limited availability of comprehensive traditional Chinese-language resources for processing medical problems. Despite these challenges, the authors argued that ChatGPT had demonstrated satisfactory performance in handling negatively phrased questions, mutually exclusive questions, and case-scenario questions, thereby suggesting its potential as a valuable learning and exam-preparation tool (22).

The results of our study indicate that ChatGPT is at the forefront of advancements in the area of machine learning tools and that it displays significant improvements and capabilities in answering medical issues because open-domain-question answering models have faced considerable challenges in solving medical problems until recently. For instance, Jin et al. achieved an accuracy rate of 36.7% on a dataset of 12,723 questions derived from Chinese medical licensing exams in 2020 (23). Similarly, Ha et al. reported only 29% accuracy on 454 USMLE Step 1 and Step 2 questions in 2019 (24). Improved results were seen in the biomedical question answering (QA) dataset collected from PubMed abstracts from 2019 (PubMedQA). This model attained an accuracy rate of 68.1% and was designed to answer only yes-or-no questions by using information sourced from the corpus of PubMed-available abstracts (25). Interestingly, ChatGPT outperformed PubMedGPT, a language model with a similar neural structure but that had been trained solely on biomedical literature and that had achieved an accuracy rate of only 50.3% (26). As argued by Kung et al., this difference in performance may be attributed to PubMedGPT's domain-specific training, from which it might have absorbed

ambiguous information from ongoing academic discourse, thereby leading to less conclusive responses (12).

The landscape of medical problem-solving tools is diverse, and comparing ChatGPT's capabilities to other commercially available solutions is vital for a comprehensive understanding of its position. Traditional Medical Decision Support Systems (MDSS) like UpToDate (27) and ClinicalKey (28), which rely on curated, evidence-based content, offer structured and highly processed medical information. While these systems are renowned for their accuracy and reliability, they lack the conversational interface and adaptability of AI-driven tools like ChatGPT. On the other hand, emerging AI-driven solutions, such as IBM Watson for Health (29), offer more dynamic interactions and the ability to process natural language queries, but they may still face challenges in areas like context understanding and the latest data incorporation. Unlike these systems, ChatGPT brings a unique blend of conversational ease and a vast database of information, though it may currently not match the specialized, up-to-date medical knowledge of dedicated MDSS.

Despite its great potential, the integration of AI tools like ChatGPT in medicine necessitates careful consideration of several ethical concerns (30). Firstly, in medical decision-making, the risk of over-reliance on AI could compromise the essential human element in healthcare, particularly in nuanced fields like OB/GYN. While AI can augment the decision-making process, it cannot replace the critical judgment and empathetic understanding inherent to medical professionals. Secondly, the accuracy and currency of data in AI systems are pivotal, especially in fast-evolving fields like medicine. AI's reliance on historical data may lead to outdated or incomplete medical advice, underscoring the need for continual updates and human oversight. Thirdly, data privacy and security are paramount in handling sensitive medical information. AI systems must ensure robust protection against data breaches and adhere to stringent privacy regulations.

The COVID-19 pandemic has shed a light on the lack of digitalization in medical education and training both in Germany and worldwide. New initiatives have focused on moving away from the conventional lecture-based teaching model and instead prioritize self-directed learning methods that cater to the individual needs that students have (31). Additionally, these endeavors incorporate the use of innovative (online) technologies for enhancing overall educational success (32, 33). The experiences during the COVID-19 pandemic have given rise to various experimental teaching concepts. For instance, in a sub-Saharan African setting, instant messaging platforms such as WhatsApp® have been utilized for distance teaching (34). Furthermore, one German university hospital introduced realistic e-learning cases within a symptom-based curriculum for internal medicine (35). Another approach involved the implementation of virtual "serious gaming" as an alternative to intensive small-group teaching (36). In spite of their innovative ideas, however, none of these concepts is likely to be implemented on a broader scale in medical education. This situation highlights an advantage that ChatGPT and similar AI-based applications have because they are freely available over the Internet, which makes them easily accessible for practical use in students' everyday learning.

As an interactive learning aid, AI can provide an accessible platform for reviewing and dissecting complex clinical cases and patient scenarios that medical students might not yet encounter in their training (37). Simulated conversations and diagnostic exercises are another area where AI can play a pivotal role. Through these simulations, students can practice their diagnostic skills, receive immediate feedback, and learn to navigate patient interactions effectively. These simulations can also be tailored to mimic a wide range of clinical situations, from common ailments to rare diseases, providing a safe and controlled environment for students to work on their clinical reasoning and decision-making skills. Moreover, AI can be integrated into various educational formats, such as virtual classrooms, online courses, and mobile applications, offering flexibility and convenience for students. It also provides an opportunity for continuous learning outside of traditional classroom settings, making medical education more accessible and adaptable to individual learning preferences (31).

ChatGPT shows special potential as a surrogate for small-group learning (10), which has been proven to be a highly effective teaching approach (36, 38). Small-group learning is characterized by three distinct elements: namely active participation, clear and specific tasks, and facilitated reflection by the participants (39, 40). ChatGPT can demonstrate all three of these essential characteristics of small-group learning, thereby making it a viable alternative for small-group education. One of the significant advantages of using ChatGPT in this role is its accessibility and availability. Small-group learning can be challenging to organize due to scheduling conflicts or limited resources, but ChatGPT offers an on-demand and self-paced learning experience. Learners can interact with the AI model at their convenience, thereby enabling flexibility in their learning journey. In our own study from 2022 that investigated the learning experiences of medical students at our faculty during the COVID-19 pandemic, the aspects of time and spatial flexibility were particularly praised and valued by our students (41). Furthermore, ChatGPT can cater to individual learning needs. In small-group settings, the pace and content of discussions might be influenced by the dynamics of the group, thereby leaving some students with unaddressed questions or uncertainties (42). ChatGPT, on the other hand, can provide personalized responses to individual queries, thus ensuring that each student's specific knowledge gaps are filled. Another advantage of ChatGPT lies in its potential for a more inclusive learning environment. In some small-group settings, students might feel hesitant to participate actively due to various factors, such as shyness or language barriers. As an AI interface, ChatGPT eliminates such barriers and provides a non-judgmental and non-intimidating platform with which learners can engage and ask questions freely. However, it is also important to acknowledge the limitations of ChatGPT as a surrogate for small-group learning. Indeed, ChatGPT lacks the interactivity and dynamic discussions of small-group settings. Learning in small groups allows for collaborative problem-solving, peer-to-peer feedback, and the exchange of diverse perspectives. These interactive elements foster critical thinking and deeper understanding, which ChatGPT cannot fully replicate. There is also a potential risk of undermining critical thinking and problem-solving skills among students. The ease of accessing information from AI could lead to a dependency that detracts from deeper engagement and intellectual development. Balancing these aspects is crucial for the responsible and effective use of AI in these critical sectors.

The fact that the application of AI will extend beyond academic support to direct patient care and clinical decision-making is obvious. As an advanced AI-driven tool, it can assist healthcare providers by offering quick access to medical information, suggesting potential

diagnoses, and providing drug information, thereby acting as a supportive tool for decision-making (30). In patient interactions, ChatGPT could be employed for patient education, explaining medical conditions and treatments in easily understandable language, and for gathering preliminary patient histories, thus streamlining the pre-consultation process. In the near future, we can envision AI systems aiding in diagnostic processes by analyzing patient data, symptoms, and medical histories, thereby providing clinicians with potential diagnoses or highlighting overlooked aspects of a patient's condition. This has been thoroughly discussed in transformative fields like radiology and pathology, where AI's image recognition capabilities can augment human expertise (43, 44). Moreover, AI can play a crucial role in personalized medicine. By analyzing large datasets, including genetic information, AI can help tailor treatments to individual patients, enhancing the efficacy and reducing the side effects of therapies (45). In patient management, AI tools can assist in monitoring chronic conditions, alerting healthcare providers to changes in a patient's status, and suggesting adjustments to treatment plans (46). Furthermore, ChatGPT could be integrated into electronic health records (EHRs) systems, assisting with documentation tasks, reducing administrative burdens, and allowing clinicians more time for patient care (47). However, it is crucial that these applications are monitored and guided by healthcare professionals to ensure accuracy and ethical use, particularly in dealing with sensitive patient data and making clinical decisions. The integration of ChatGPT into medicine and clinical care promises not only to enhance efficiency but also to improve the quality of patient education and engagement, ultimately contributing to better healthcare outcomes.

Limitations

The present study on ChatGPT's performance in clinical problems entailed several potential limitations that should be considered in the context of medical education and clinical practice. The limited dataset size is a first point of concern. Our reliance on a dataset of OB/GYN course questions and state exams may not have adequately captured the vast range of medical knowledge and question types that are encountered in real-world scenarios. Therefore, the use of a more extensive and diverse dataset would have provided a more comprehensive evaluation of ChatGPT's abilities. Moreover, the fact that the study focused solely on course questions and state exams in the field of OB/GYN may call into question the generalizability of our data. Each medical specialty has its unique complexities, terminologies, and practice nuances, which may not be uniformly comprehensible or addressable by a generalized AI tool like ChatGPT. For instance, the diagnostic reasoning in psychiatry, in surgical fields, or in emergency medicine present different challenges that may not be fully captured in the OB/GYN examination context. However, recent research in medical education and the potential application of ChatGPT in diverse educational and clinical settings has shown that ChatGPT shows general applicability and could serve as a valuable tool for medical educators, students, and clinicians alike (6, 48–51). In fact, a fundamental advantage of AI is its high flexibility, allowing it to be applied effectively in various settings without diminishing its impact.

The study's results regarding the impact of question difficulty on ChatGPT's performance may also raise concerns. While ChatGPT demonstrated consistent performance in the OB/GYN course dataset, its effectiveness was influenced by the complexity of the questions in the state exam. This finding raises doubts about ChatGPT's adaptability to varying levels of difficulty and emphasizes the need both for further fine-tuning and for incorporating data that could enhance the tool's performance across all contexts. Another potential limitation is the inability of ChatGPT to process images. In medical education, visual information – such as radiological images and anatomical diagrams – plays a vital role. The fact that ChatGPT struggles with image processing currently restricts its applicability in real-world medical scenarios in which such visual content is very commonly encountered. However, the tool's newest version – ChatGPT-4 – addresses this shortcoming by implementing image-analysis features that are able to identify objects in photos (52). This innovation may indeed only mark the beginning of further advancements in this area.

All of these limitations are, however, temporary. In the next decade, ChatGPT is expected to undergo transformative advancements, primarily driven by breakthroughs in artificial intelligence, machine learning, and natural language processing. We can anticipate a more nuanced understanding of language and context, including better handling of cultural nuances and idioms, achieved through sophisticated algorithms and diverse, expansive training datasets. Enhanced contextual awareness will make interactions more coherent over longer conversations, while personalization features will tailor responses to individual user preferences and histories. Multimodal capabilities will likely be integrated, allowing ChatGPT to process and respond to text, voice, and visual inputs. Real-time learning from interactions will continuously update its knowledge base, and efforts have to be made to ensure ethical use and bias mitigation.

Conclusion

Our data provide valuable insights into ChatGPT's role in medical education and clinical practice, particularly in the field of OB/GYN. The findings suggest that ChatGPT demonstrates promising potential as a supplementary tool for medical students and healthcare professionals alike. ChatGPT's ability to provide accurate and insightful responses to medical questions showcases its adaptability to complex clinical scenarios. The study's insights into ChatGPT's capabilities within OB/GYN are particularly significant, hinting at broader implications for its application across various medical settings. As AI technology continues to advance, it is necessary to continue evaluating and refining tools like ChatGPT to ensure they meet the evolving needs of medical education and practice, while also addressing ethical considerations and maintaining the highest standards of patient care.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The present study was conducted in accordance with the Declaration of Helsinki (2008 version). Since the data stem from anonymously

analyzed exam questions, no further consultation or approval by the ethics committee was required, as is outlined in the online guidelines by the Joint Ethics Committee of the Universities of Bavaria (https://www.gehba.de/fileadmin/daten/Gehba/GEHBa-FAQ_2.1.pdf).

Author contributions

MR: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft, Methodology, Project administration. KK: Data curation, Investigation, Writing – review & editing. AS: Data curation, Investigation, Writing – review & editing. VR: Data curation, Investigation, Writing – review & editing. NA: Data curation, Investigation, Writing – review & editing. AG: Data curation, Investigation, Writing – review & editing. FlR: Supervision, Writing – review & editing. EK: Resources, Supervision, Writing – review & editing. MK: Resources, Supervision, Writing – review & editing. FaR: Conceptualization, Writing – original draft. BM: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. OpenAI. (2023). OpenAI ChatGPT: Optimizing language models for dialogue. Available at: https://openai.com/blog/chatgpt/ (Accessed September 17, 2023).

2. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. (2020). Language models are few-shot learners. Available at: https://arxiv.org/abs/2005.14165 (Accessed September 17, 2023).

3. Tamkin A, Brundage M, Clark J, Ganguli D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. Available at: https://arxiv.org/abs/2102.02503 (Accessed September 17, 2023).

4. Dai Z, Yang Z, Yang Y, Carbonell J, Le QV, Salakhutdinov R (2019). Transformer-XL: Attentive language models beyond a fixed-length context. Available at: https://arxiv.org/abs/1901.02860 (Accessed September 17, 2023).

5. Keskar NS, Mc Cann B, Varshney LR, Xiong C, Socher R. (2019). CTRL: A conditional transformer language model for controllable generation. Available at: https://arxiv.org/abs/1909.05858 (Accessed September 17, 2023).

6. Ge J, Lai JC. Artificial intelligence-based text generators in hepatology: ChatGPT is just the beginning. *Hepatol Commun*. (2023) 7:e0097. doi: 10.1097/HC9.0000000000000097

7. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Med Educ*. (2023) 9:e46885. doi: 10.2196/46885

8. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)*. (2023) 11:887. doi: 10.3390/healthcare11060887

9. AMBOSS. (2020). Medical knowledge distilled. AMBOSS. Available at: https://www.amboss.com/us (Accessed September 17, 2023).

10. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. (2023) 9:e45312. doi: 10.2196/45312

11. Wang R, Strong D. *Beyond accuracy: What data quality means to data consumers*. Armonk, NY: M.E. Sharpe, Inc (1996).

12. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. (2023) 2:e0000198. doi: 10.1371/journal.pdig.0000198

13. The Economist. (2022). How good is ChatGPT? Available at: https://www.economist.com/business/2022/12/08/how-good-is-chatgpt (Accessed September 17, 2023).

14. Süddeutsche Zeitung. (2022). Die gefühlte Revolution. Available at: https://www.sueddeutsche.de/kultur/chatgpt-kuenstliche-intelligenz-1.5708877

15. Wang LKP, Paidisetty PS, Cano AM. The next paradigm shift? ChatGPT, artificial intelligence, and medical education. *Med Teach*. (2023) 1:1. doi: 10.1080/0142159X.2023.2256961

16. Reuters. (2023). ChatGPT sets record for fastest-growing user base – analyst note. Available at: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (Accessed September 17, 2023).

17. PECB Insights. (2023). The rise of ChatGPT: Paving the way for large language model adoption. Available from: https://insights.pecb.com/rise-chatgpt-paving-way-large-language-model-adoption/ (Accessed September 17, 2023).

18. Cognizant. (2023). ChatGPT and the generative AI revolution. Available at: https://www.cognizant.com/us/en/insights/perspectives/chatgpt-and-the-generative-ai-revolution-wf1532750 (Accessed September 17, 2023).

19. Biswas S. ChatGPT and the future of medical writing. *Radiology*. (2023) 307:e223312. doi: 10.1148/radiol.223312

20. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT – reshaping medical education and clinical management. *Pak J Med Sci*. (2023) 39:605–7. doi: 10.12669/pjms.39.2.7653

21. Temsah O, Khan SA, Chaiah Y, Senjab A, Alhasan K, Jamal A, et al. Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. *Cureus*. (2023) 15:e37281. doi: 10.7759/cureus.37281

22. Weng TL, Wang YM, Chang S, Chen TJ, Hwang SJ. ChatGPT failed Taiwan's family medicine board exam. *J Chin Med Assoc*. (2023) 86:865. doi: 10.1097/JCMA.0000000000000956

23. Jin D, Pan E, Oufattole N, Weng WH, Fang H, Szolovits P. (2020). What disease does this patient have? A large-scale open domain question answering dataset from medical exams. Available at: https://arxiv.org/abs/2009.13081 (Accessed September 17, 2023).

24. Le AH, Victoria Y. "Automatic Question Answering for Medical MCQs: Can It go Further than Information Retrieval?,"in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*. (Varna, Bulgaria: INCOMA Ltd) (2019) 418–22.

25. Jin Q, Dhingra B, Liu Z, Cohen WW, Lu X. (2019). PubMedQA: A dataset for biomedical research question answering. Available at: https://arxiv.org/abs/1909.06146 (Accessed September 17, 2023).

26. Stanford CRFM. (2022). Introduces PubMedGPT 2.7B. Available at: https://hai.stanford.edu/news/stanford-crfm-introduces-pubmedgpt-27b (Accessed September 17, 2023).

27. UpToDate. (2023) Surgeon viewing UpToDate on tablet UpToDate: Interaktive Unterstützung klinischer Entscheidungen. Available at: https://www.wolterskluwer.com/de-de/solutions/uptodate (Accessed September 17, 2023).

28. ClinicalKey. (2023) Diagnose and treat your patients with confidence. Available at: https://www.clinicalkey.com/#!/ (Accessed September 17, 2023).

29. IBM Watson. (2023) Conversational AI for fast and friendly customer care. Available at: https://www.ibm.com/products/watsonx-assistant (Accessed September 17, 2023).

30. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. (2023) 6:1169595. doi: 10.3389/frai.2023.1169595

31. Riedel M, Amann N, Recker F, Hennigs A, Heublein S, Meyer B, et al. The COVID-19 pandemic and its impact on medical teaching in obstetrics and gynecology-a nationwide expert survey among teaching coordinators at German university hospitals. *PLoS One*. (2022) 17:e0269562. doi: 10.1371/journal.pone.0269562

32. Skochelak SE, Stack SJ. Creating the medical schools of the future. *Acad Med*. (2017) 92:16–9. doi: 10.1097/ACM.0000000000001160

33. Irby DM, Cooke M, O'Brien BC. Calls for reform of medical education by the Carnegie Foundation for the Advancement of Teaching: 1910 and 2010. *Acad Med*. (2010) 85:220–7. doi: 10.1097/ACM.0b013e3181c88449

34. Enyama D, Balti EV, Simeni Njonnou SR, Ngongang Ouankou C, Kemta Lekpa F, Noukeu Njinkui D, et al. Use of WhatsApp®, for distance teaching during COVID-19 pandemic: experience and perception from a sub-Saharan African setting. *BMC Med Educ*. (2021) 21:517. doi: 10.1186/s12909-021-02953-9

35. Rahm AK, Töllner M, Hubert MO, Klein K, Wehling C, Sauer T, et al. Effects of realistic e-learning cases on students' learning motivation during COVID-19. *PLoS One*. (2021) 16:e0249425. doi: 10.1371/journal.pone.0249425

36. Middeke A, Anders S, Schuelper M, Raupach T, Schuelper N. Training of clinical reasoning with a serious game versus small-group problem-based learning: a prospective study. *PLoS One*. (2018) 13:e0203851. doi: 10.1371/journal.pone.0203851

37. Lee H. The rise of ChatGPT: exploring its potential in medical education. *Anat Sci Educ*. (2023). doi: 10.1002/ase.2270

38. Springer L, Stanne ME, Donovan SS. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. *Rev Educ Res*. (1999) 69:21–51. doi: 10.3102/00346543069001021

39. Robinson L. Small groups, big possibilities: radical pedagogical approaches to critical small-group learning in medical education. *Can Med Educ J*. (2023) 14:178–9. doi: 10.36834/cmej.76464

40. Jones RW. Learning and teaching in small groups: characteristics, benefits, problems and approaches. *Anaesth Intensive Care*. (2007) 35:587–92. doi: 10.1177/0310057X0703500420

41. Riedel M, Eisenkolb G, Amann N, Karge A, Meyer B, Tensil M, et al. Experiences with alternative online lectures in medical education in obstetrics and gynecology during the COVID-19 pandemic-possible efficient and student-orientated models for the future? *Arch Gynecol Obstet*. (2021) 305:1041–53. doi: 10.1007/s00404-021-06356-5

42. Burgess A, van Diggele C, Roberts C, Mellis C. Facilitating small group learning in the health professions. *BMC Med Educ*. (2020) 20:457. doi: 10.1186/s12909-020-02282-3

43. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging*. (2023) 104:269–74. doi: 10.1016/j.diii.2023.02.003

44. Schukow C, Smith SC, Landgrebe E, Parasuraman S, Folaranmi OO, Paner GP, et al. Application of ChatGPT in routine diagnostic pathology: promises, pitfalls, and potential future directions. *Adv Anat Pathol*. (2023). doi: 10.1097/PAP.0000000000000406

45. Lacalamita A, Serino G, Pantaleo E, Monaco A, Amoroso N, Bellantuono L, et al. Artificial intelligence and complex network approaches reveal potential gene biomarkers for hepatocellular carcinoma. *Int J Mol Sci*. (2023) 24:15286. doi: 10.3390/ijms242015286

46. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. *Am J Obstet Gynecol*. (2023) 228:696–705. doi: 10.1016/j.ajog.2023.03.009

47. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. A large language model for electronic health records. *NPJ Digit Med*. (2022) 5:194. doi: 10.1038/s41746-022-00742-2

48. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof*. (2023) 20:1. doi: 10.3352/jeehp.2023.20.1

49. Huang H, Zheng O, Wang D, Yin J, Wang Z, Ding S, et al. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model. *Int J Oral Sci*. (2023) 15:29. doi: 10.1038/s41368-023-00239-y

50. Beam K, Sharma P, Kumar B, Wang C, Brodsky D, Martin CR, et al. Performance of a large language model on practice questions for the neonatal board examination. *JAMA Pediatr*. (2023) 177:e232373. doi: 10.1001/jamapediatrics.2023.2373

51. Moshirfar M, Altaf AW, Stoakes IM, Tuttle JJ, Hoopes PC. Artificial intelligence in ophthalmology: a comparative analysis of GPT-3.5, GPT-4, and human expertise in answering StatPearls questions. *Cureus*. (2023) 15:e40822. doi: 10.7759/cureus.40822

52. Waisberg E, Ong J, Masalkhi M, Kamran SA, Zaman N, Sarker P, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci*. (2023). doi: 10.1007/s11845-023-03377-8

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 21 December 2023
DOI 10.3389/fmed.2023.1263862

OPEN ACCESS

EDITED BY
Jacqueline G. Bloomfield,
The University of Sydney, Australia

REVIEWED BY
Mohd Nasri Awang Besar,
National University of Malaysia, Malaysia
Majed Wadi,
Qassim University, Saudi Arabia

*CORRESPONDENCE
Florian Recker
florian.recker@ukbonn.de

RECEIVED 15 August 2023
ACCEPTED 13 November 2023
PUBLISHED 21 December 2023

CITATION
Plöger R, Abramian A, Egger EK, Mustea A, Sänger N, Plöger H, Weber E, Gembruch U, Walter A, Strizek B and Recker F (2023) Evaluation of an OSCE's implementation and a two-step approach for a theoretical and practical training program in Obstetrics and Gynecology.
Front. Med. 10:1263862.
doi: 10.3389/fmed.2023.1263862

Evaluation of an OSCE's implementation and a two-step approach for a theoretical and practical training program in Obstetrics and Gynecology

Ruben Plöger[1], Alina Abramian[2], Eva Katharina Egger[3], Alexander Mustea[3], Nicole Sänger[4], Hannah Plöger[5], Eva Weber[6], Ulrich Gembruch[7], Adeline Walter[7], Brigitte Strizek[7] and Florian Recker[7]*

[1]Department of Obstetrics and Gynecology, University Hospital Bonn, Bonn, Germany, [2]Department of Senology, University Hospital Bonn, Bonn, Germany, [3]Department of Gynecology and Gynecological Oncology, University Hospital Bonn, Bonn, Germany, [4]Department of Gynecological Endocrinology and Reproductive Medicine, University Hospital Bonn, Bonn, Germany, [5]Department of Neonatology and Pediatric Intensive Care, University Hospital Bonn, Bonn, Germany, [6]Division of Prenatal Medicine, Gynecological Ultrasound and Fetal Surgery, Department of Obstetrics and Gynecology, University of Cologne, Cologne, Germany, [7]Department of Obstetrics and Perinatal Medicine, University Hospital Bonn, Bonn, Germany

Objective structured clinical examination (OSCE) is a well-known assessment method to evaluate clinical skills and competence in healthcare. Following the recently reformed National Competence-Based Catalog of Learning Objectives in Medicine, the implementation of this assessment method in the training program for medical students is now obligatory in Germany. This major change requires a reorganization not only of the training programs but also of the students themselves and the way they learn. We performed a poll evaluating the students' opinions regarding these major changes and the implementation of the OSCE with a new training program. To implement this assessment method and to evaluate the OSCE, Kern's six-step approach comprising (1) problem identification and general needs assessment, (2) needs assessment of the targeted learners, (3) goals and objectives, (4) educational strategies, (5) implementation, and (6) evaluation and feedback was applied. To evaluate and gather feedback, a poll was used to analyze the student's opinions regarding OSCE in gynecology and obstetrics and OSCE in general, in addition to the regular analysis of the students' results. To reform the educational strategy, a two-step approach was developed: First, the students completed the regular training program and a written examination, and second, they participated in a 1-week clerkship, in small group teaching, and in the OSCE. The OSCE stations were developed primarily based on the National Competence-Based Catalog and the German Catalog of Learning Objectives in Medicine, as well as on the feedback of experts reflecting their expectations for physicians beginning their careers. The students performed well in the OSCE and gave positive feedback regarding this examination method. Furthermore, they welcomed the upcoming changes by considering OSCE a valuable assessment tool, and they showed appreciation for the two-step approach by supporting the combination of an OSCE and a written examination. Thus, this article presents the implementation of an OSCE and a strategy for the adaptation of the curriculum to fulfill the new OSCE requirements and—to our knowledge—reveals students' primary opinions regarding the changes in their medical training program for the first time.

KEYWORDS

OSCE, implementation, interprofessional training, undergraduate medical education, national curriculum reform, transition, gynecology and obstetrics

1 Introduction

The objective structured clinical examination (OSCE) is a method to evaluate the examinee's clinical skills and competency (1). An OSCE often has numerous stations each with a task, for example, obtaining a patient's history, which the student conducts by acting as a physician. An examiner evaluates the student by considering certain predefined factors such as the student's introduction or demeanor. As a result, the OSCE provides greater impartiality and equity than alternative approaches to practical assessments (2). The students are asked to apply knowledge in a situation likely to occur in a physician's daily work routine instead of simply demonstrating the memorization of facts, as is appraised by most traditional written examinations (3). Therefore, the OSCE allows for a clinical competence-based assessment at a high level of Miller's pyramid (4). In the last few decades, OSCEs have become a popular assessment method used in almost every medical discipline, such as dentistry (5), and in other healthcare professions such as nursing (6) or midwifery (7). Therefore, the objectives of OSCEs are adapted to each discipline. In obstetrics and gynecology, the reported OSCE includes topics covering a wide spectrum of clinical skills, such as a pelvic examination (8, 9), simulation of delivery (10), examination of a vaginal wet mount to diagnosing vaginitis and/or a sexually transmitted disease (8), and a breast examination (8), and assesses competencies such as the interpretation of cardiotocograph or laboratory results, for example, reflecting ovarian failure (11), management of risk factors and their complications in pregnancy (12), or the treatment and handling of a miscarriage (13). Virtual OSCEs have recently been described as a response to the COVID-19 pandemic (14). As a result, multiple OSCE stations depicting the breadth of the obstetric and gynecologic disciplines have been recorded.

In Germany, the National Competence-Based Catalog of Learning Objectives in Medicine is being reformed in line with the "Master Plan for Medical Studies 2020" (15, 16). In the revised catalog, OSCEs are required as an assessment instrument. As a result, several colleges have begun to implement OSCEs: for example, the Faculty of Medicine at the Goethe University in Frankfurt and the University of Dresden both provide OSCE preparatory courses. Other universities, such as the Friedrich-Alexander University in Erlangen's Faculty of Medicine and the Charitè in Berlin, have changed their didactical conceptions to a more skill-related model (17).

This study focuses on (1) the development and implementation of an OSCE in gynecology and obstetrics in line with this new reform as a model for other training programs and (2) the reaction of the students regarding the OSCE implementation in their curriculum. Thus, an OSCE in gynecology and obstetrics was developed and implemented using Kern's six-step approach (18), and a poll was taken to assess the students' attitude toward the established changes. A two-step approach was developed as an educational strategy to prepare the students for the OSCE (19). The students' achieved adequate results in the OSCE and their feedback was positive. Regarding their reaction to the overall OSCE implementation in their curriculum, the students welcomed OSCE as a good assessment tool in general and specifically in obstetrics and gynecology. The two-step approach seems to be a suitable educational strategy, supported by the students' feedback that the combination of a written examination and an OSCE is useful to appropriately assess their knowledge. Overall, the students supported the OSCE's implementation, adapted quickly to the new teaching methods and assessment design, and appeared to benefit from the OSCE in general.

2 Methods

To implement an OSCE in the module "Obstetrics and Gynecology" we used Kern's six-step approach (18). Furthermore, the Delphi technique (20) was used to allow structured communication between highly qualified clinicians in senology, oncology, and perinatal medicine. The six-step approach was applied as follows (Figure 1):

2.1 Problem identification

The first step, named "problem identification," was used to describe the current situation of medical education and to identify the problem. In the past, studying human medicine has been focused on simply learning facts, rather than applying the knowledge in a practical way. Therefore, the recently formed "Master Plan for Medical Studies" now suggests a reform of medical education and practical training in all fields, including obstetrics and gynecology (15). The described changes in the "Master Plan for Medical Studies" were analyzed and compared with the current situation to identify the problem at hand.

2.2 Needs assessment of the targeted learners

We assessed the needs of the targeted learners. In this case, the needs of those affected by the upcoming reform in obstetrics and gynecology were discussed and identified.

2.3 Goals and objectives

The goals and objectives should be measurable and specify the cognitive, affective, and psychomotor skills necessary to fulfill the defined requirements for the targeted learners and to solve the identified problem. The updated German National Competence-Based Catalog (16) and German Catalog of Learning Objectives in Medicine (21) defined diseases and skills within obstetrics and gynecology to be known and mastered by every student after successful graduation. Thus, the National Competence-Based Catalog

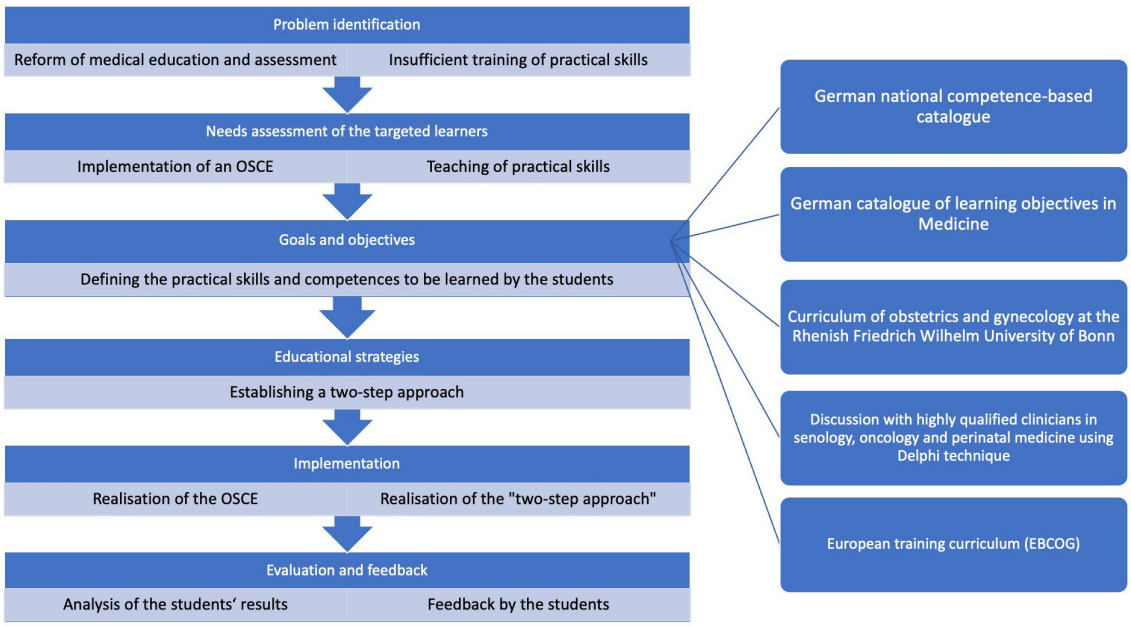


FIGURE 1
Overview of Kern's six-step approach applied for OSCE implementation.

and the German Catalog of Learning Objectives in Medicine were analyzed (16, 21). In the next step, the curriculum of obstetrics and gynecology at the Rhenish Friedrich Wilhelm University of Bonn was studied and compared with the requirements of both catalogs to identify discrepancies. Furthermore, the European Board and College of Obstetrics and Gynecology (EBCOG) examinations were evaluated regarding their content. Finally, clinical members with high qualifications in senology, oncology, and perinatal medicine were involved in defining key practical skills and competencies in obstetrics and gynecology using the Delphi technique (20).

2.4 Educational strategies

In the next step "educational strategies," the curriculum was revised to fulfill the educational objectives. Therefore, educational methods were discussed and conceptualized to adapt to the content and structure of the reformed curriculum.

2.5 Implementation

In this step, the implementation process of the OSCE after the realization of the curriculum changes was described.

2.6 Evaluation and feedback

Finally, the OSCE results and the feedback regarding the implementation of the OSCE were considered and supplemented by questions reflecting the student's overall opinion of OSCE as an assessment method. The students' OSCE results provided information regarding the quality of their practical and theoretical preparation for passing the OSCE. The poll with the questions was realized directly after the students had completed the OSCE. The organization committee handed the questionnaires out to the students who voluntarily answered the questions in a separate room. The questionnaires were anonymously collected in a box inside the room. The Ethics Commission of the University Hospital Bonn approved the approach (138/23).

3 Results

3.1 Problem identification

Prior to the reform, the students' training program for gynecology and obstetrics at the University Hospital Bonn consisted of lectures and seminars, followed by a written multiple-choice examination, and an internship on the wards, which students had to pass with an oral examination. In this oral examination, the students were asked to present a patient's history, which is clearly a vital component of a physician's daily practice. However, emphasis on practical skills and clinical competence was therefore inadequate in this training programm.

The "Master Plan for Medical Studies 2020" led to changes in the German National Competence-Based Catalog (16) and the German Catalog of Learning Objectives in Medicine (21) with more focus on the practical skills and competence that should be mastered by students. Therefore, the reform required an adaption of the current curriculum. Based on this new focus on practical skills and competencies, modified assessment methods, such as the OSCE, were needed, which allow an evaluation of the examinee's clinical skills and competency (1).

3.2 Needs assessment and targeted learners

In this case, the targeted learners were current medical students at the university hospital Bonn, who were enrolled in the module "gynecology and obstetrics" and thus mainly in the 10th semester. Two needs of the targeted learners were assessed: (1) Based on the reform, the students will have to face OSCE as a new assessment method in final examinations (15, 16). Therefore, the students should be confident prior to facing the new method and thus should be trained appropriately, so that the students could focus on the content of the OSCE station, instead of concerning themselves with the new examination setup. (2) As future physicians, the students needed to train and master their practical skills and their competencies. Therefore, students needed modules and clerkships, during which they could train their skills and competencies on models or on patients under supervision. In order to do justice to the students' needs, the training program had to evolve.

3.3 Goals and objectives

Several functions of a physician were mentioned in the German Catalog of Learning Objectives in Medicine (21), with the third description referring to the physician as a communicator (s. A 1.5). This high communication rating had not found a home in educational approaches thus far. Furthermore, the catalog emphasized the importance of medical students being able to gather patients' histories and execute examination methods (s. A 2.1.1, 2.4.1, 2.5.1, or 2.6.1). The National Competence-Based Catalog of Learning Objectives (16) also requires similar skills (III.8). Thus, successful students should be able to perform these skills during an assessment such as OSCE. In an OSCE, skills would be tested in regard to a certain subject and/or pathology, so that subjects had to be defined. Subjects for the OSCE were picked by overlapping topics from both catalogs (16, 21). For example, the station "speculum insertion" was created based on the listing of the disease: "cervix carcinoma" in the National Competence-Based Catalog of Learning Objectives [VI.06-01.4.12, (16)] and in the German Catalog of Learning Objectives in Medicine [Teil C:53 Krankheitsbilder (21)]. Furthermore, this station covers the item "prevention and screening" from the German Catalog of Learning Objectives in Medicine [Teil A, 2.6 (21)]. In the next step, the curriculum of obstetrics and gynecology at the Rhenish Friedrich Wilhelm University of Bonn and the content of the European Board and College of Obstetrics and Gynecology (EBCOG) examinations were studied and compared with the subjects of the catalogs. Following this, clinical members with high qualifications in senology, gynecological oncology, gynecological endocrinology, reproductive medicine, and perinatal medicine were involved in the process using the Delphi technique (20) to define key practical skills and competencies in obstetrics and gynecology and thus choose the subjects of the OSCE's stations (s. Table 1). Finally, using the gathered data and professional input, practical skills and competencies in obstetrics and gynecology were chosen, including breast and vaginal examination, delivery supervision, and medical care during pregnancy (s. Table 1). In the station "delivery," where the Apgar score should be determined, an interdisciplinary approach was established with colleagues from the Department of Neonatology and Pediatric Intensive Care (s. Table 1) in the respect to the subjects of neonatology in the catalogs [VIII.7. (16)].

3.4 Educational strategies

As an educational strategy, a two-step approach was implemented (19): In the first step, the module "obstetrics and gynecology" had to be passed by the students during their third clinical year. In the module, the students could attend lectures and seminars and had to pass a written examination. As the second step, the students had to attend a 1-week clerkship with an OSCE as an assessment tool. During the clerkship, the students rotated within the department, with each student assigned to a different physician each day. Thus, each day, the students were exposed to a specific region such as the prenatal diagnostic department, to a functional area such as the operating room or labor ward, or to patient wards such as the postnatal ward. During the clerkship, consultants from our clinic's various departments taught the students in small groups (up to 10 students). Vaginal examination, breast examination, birth mechanics, and contraception were among the topics covered. The consultants used models for training as much as possible, so that the students could perform vaginal and breast examinations as well as learn the hand grips important during labor. During the 1-week clerkship, these practical abilities might be honed further with training from the assigned physicians and separately with models in the teaching facility.

TABLE 1 Overviews of the stations, their affiliations to each department and the assessed skills and competence.

Stations	Breast cancer	Cervical Neoplasia	Delivery	Gestational diabetes	Preeclampsia
Departments	Senology	Gynecological Oncology	Obstetrics and pediatrics	Obstetrics	Obstetrics
Skills and competencies	Taking patient's history	Taking patient's history		Taking patient's history	Taking patient's history
	Breast examination	Vaginal examination Performing cervical smear Analysis of the smear	Vaginal examination Guiding a delivery Performing a first care of the newborn	Analysis of the pregnancy record Analysis of a cardiotocogram	Analysis of the pregnancy record Analysis of a cardiotocogram
	Report of the patient's history and result of the examination	Defining Pap-type	Taking the Apgar score	Diagnose gestational diabetes	Diagnose preeclampsia

Furthermore, the students could use the e-learning platform Amboss® or watch video podcasts reviewing topics such as "breast examination" and "concept of the pregnancy record" to accompany the training program.

3.5 Implementation

The OSCE was performed during the last term of medical school. The students (113 students) completed the module "obstetrics and gynecology" and a 1-week clerkship at the Department of Obstetrics and Gynecology prior to the OSCE, as described with two-step approach (s. above). For the OSCE, students were randomly assigned to one of 13 total exam groups. Every group had to travel through three of the five stations, being assigned again at random. Each station lasted 5 min. Students had time between stations to read the case vignette and take a breather before moving on to the next station. Examiners were required to grade students using a digital standardized questionnaire. The examiners received training to familiarize themselves with the digital grading tool prior to the OSCE. Students could reach a maximum of 25 points by fulfilling a checklist of items (22, 23). The tests were carried out on models.

In the station "senology," students were asked to envision themselves as physicians in a senologic outpatient clinic, where they should take a patient's medical history, perform an examination, describe a clinical report, and mention one key differential diagnosis. The station "speculum insertion" asked that the examinee obtain the patient's history, perform a speculum insertion, perform a cervical smear and Papanicolaou's test, and discuss the clinical results. In the station "delivery," students were asked to identify the phases of labor of a patient during delivery, to guide the delivery itself using a model, and to give first aid to the infant, which included calculating the Apgar score. In the station "gestational diabetes," students were confronted with a patient complaining of typical symptoms of this disease. The students were asked to identify their suspected diagnosis, outline the risk factors in the pregnancy record, and describe the cardiotocograph. In the station "preeclampsia," students had to identify the most likely diagnosis based on the pregnancy record and the patient's symptoms, indicating a hypertensive disorder during pregnancy. Finally, the students were asked to analyze the patient's cardiotocograph (s. Table 1).

3.6 Feedback

In total, 113 students had scores ranging from 39 to 75, with a maximum score of 75, resulting in an average score of 63.92. (IQR: 60–69). The average, maximum, and minimum scores for the 58 students who had completed the station "senology" were 21.97 (IQR: 21–23), 25, and 15, respectively. A total of 98 students achieved average, maximum, and minimum scores of 22.41 (IQR: 21–24), 25, and 13, respectively, at the station "speculum insertion." A total of 80 students completed the "delivery" station and received scores ranging from 7 to 25, with an average of 20.91 (IQR: 20–23). The station "gestational diabetes" was completed by 58 students with an average score of 20.75 (IQR: 19–24), a minimal score of 12, and a maximal score of 25. A total of 45 students passed through the station "preeclampsia" and achieved an average score of 19.47 (IQR: 17–22), a minimal score of 13, and a maximal score of 25 (s. Figure 2).

A total of 87 students who completed the OSCE (76.99%) answered the questionnaire (s. Figures 3, 4). The average age was 25 years (IQR: 24–26, youngest student: 23, oldest student: 35). There were 60 female students, 26 male students, and 1 gender-diverse student. In total, 20 students had work experience in healthcare prior to studying medicine, while 67 students had none. There was no statistically significant difference in the replies to these yes/no questions between the groups with and without experience. Figures 3, 4 depict the questions and responses of the students. Open inquiries on the positive and negative elements of the stations revealed the following aspect. Regarding the use of models, the students appreciated the utility of the breast model with the ability to inspect breast lumps in the station "breast examination," while the students complained that the model in the station "speculum insertion" was insufficient to demonstrate the investigation of the vagina.

4 Discussion

The OSCE was implemented in the current curriculum by applying Kern's six-step approach. Through the students' experiences of completing the OSCE, they developed a positive attitude toward this specific OSCE and toward OSCE in general, reflected by the poll data.

Kern's six-step approach is common for curriculum development worldwide (24–26) and has been recently used to implement new topics from the German National Competence-Based Catalog of Learning Objectives in Medicine (16), such as: "conflicts of interest and communicating risk" (27). Further tools to design health professional education curricula are emerging: One approach to modifying the curriculum is the "system thinking perspectives," by which the educator using the 3P-6Cs toolkit is able to understand the students' perspectives of the integration of elements within the medical education program and the impact of these elements for the lifelong practice of the student (28). Another approach applies the "twelve tips": (1) identify the purpose and scope of change, (2) create a vision, (3) develop a strategy for change involving key stakeholders, (4) importance of quick visible wins and communication, (5) analyze the internal environment and culture, (6) consider the external environment, cultural contexts, and political influences, (7) choose the right combination of approaches to change, (8) use project management techniques for operational planning and implementation, (9) acknowledge the psychological impact of changes, (10) plan for transition and loss of competence, (11) do not underestimate the complexity, and (12) celebrate success and shift from project to "New reality" (29). These approaches are based on different priorities such as the establishment of support in a department for the upcoming changes or focus on the student's perspectives. Kern's six-step approach was chosen based on its known establishment worldwide, its production of valid and reliable results, and its straightforward structure.

The development and implementation process of an OSCE in the revised curriculum using Kern's six-step approach was successful based on (1) the satisfactory results of the 113 students passing the OSCE (18, 19), (2) the results of the poll, and (3) the

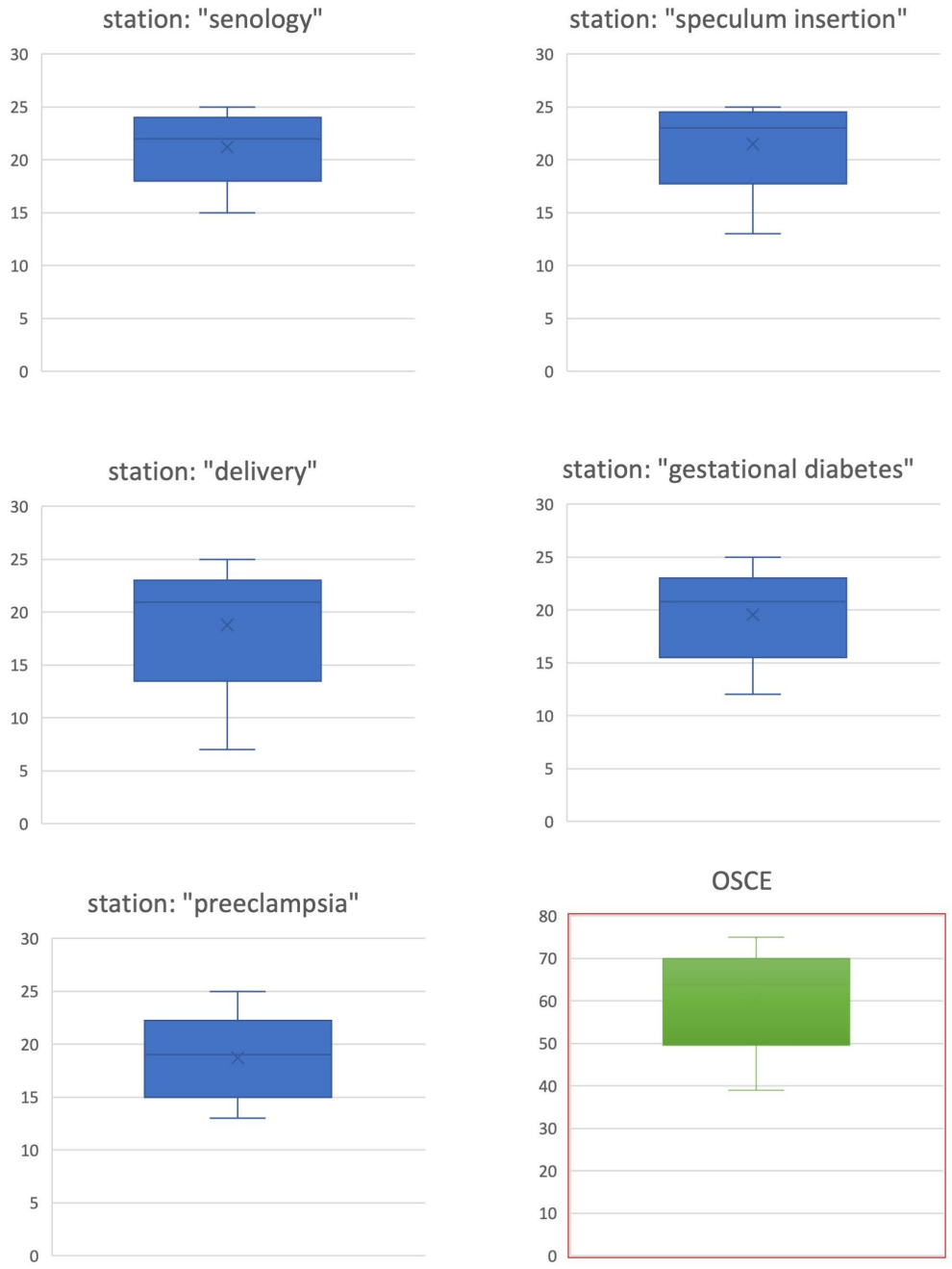


FIGURE 2
Overview of the OSCE results (box plots).

answers to the open inquiries: (1) The comparability with other OSCEs is demonstrated by a similar average result of 85% (10, 11, 19). This may be caused by the similarity of the OSCE stations and the tested skills and competencies required to treat a certain patient profile (8, 9, 11). Only one station included a stark difference to the previously published OSCEs in obstetrics and gynecology: This station devoted to a normal delivery also included the evaluation of the newborn's health by asking the students to determine an Apgar score. Nonetheless, the different stations seemed to have a similar level of difficulty based on the comparable average results (s. Figure 2).

(2) The response rate of 76.99% in this poll is higher than the overall average response rate of patients and healthcare professionals in surveys worldwide (30), and thus is generalizable based on the standard of many journals (31, 32). The primarily positive answers (65%) to the question: "Did you feel well prepared for the OSCE?" (Figure 3) indicate a well-realized implementation process and support the two-step approach. The negative answers (25%) show the need for further improvement such as the implementation of mock OSCEs (33–35). Students criticized the short amount of time allotted for each station, a comment also found in other OSCE reports (11, 36). From a practical point of

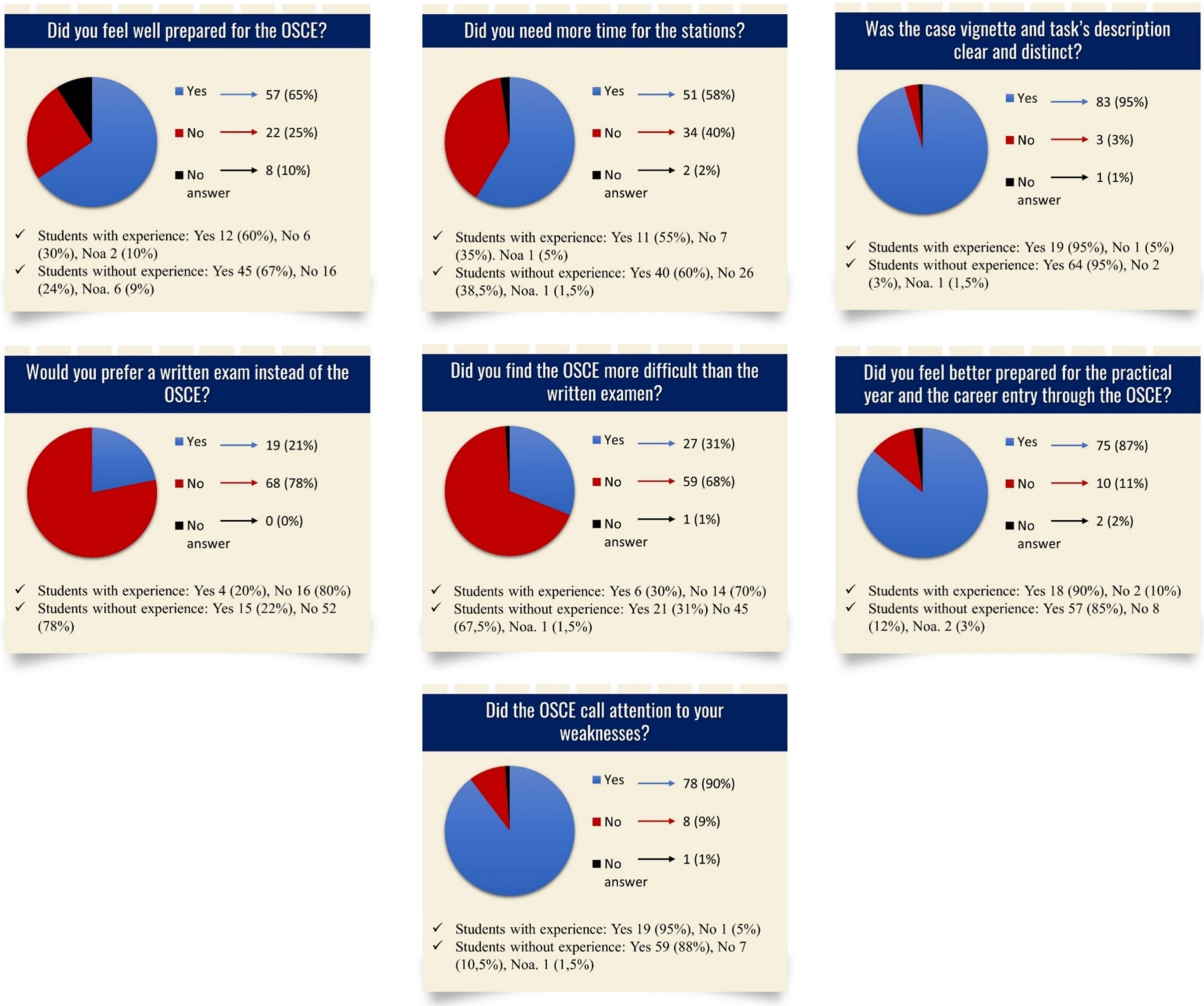


FIGURE 3
Questionnaire and results regarding the OSCE, Noa: No answer.

view, a short time period for each station allows for the completion of more and various stations in the same time frame, so, in this case, the time length of each station was chosen to last 5 min, similar to most described OSCEs and as recently published for Australian medical schools (37). From the examiner's point of view, the limited time demands that the student is well prepared and confident in performing the procedure. For an acceptable standard, the students must have practiced the task several times, comparable to the learning curve of surgical procedures [for instance, 80 cases are required for an acceptable standard in laparoscopic colorectal surgery (38)] (39). From the future physician's view, the limited time reflects the limited time available during a physician's daily routine. The case vignette and task descriptions of the stations were clear and distinct, which is often described by other studies but may be biased by the one-sided publication of good results (40–43). A total of 78% of the respondents prefer an OSCE, corresponding to studies in Ethiopia and Nigeria (44, 45). A reason for this may be that the students find the OSCE less difficult than written examinations, although this stands in contrast to the fact that published OSCE results show a wide range, some with better and some with worse results in comparison to written assessment (46, 47). On the other hand, 21% of the respondents prefer a written examination, which may be caused by higher stress and nervousness in an OSCE than in a written examination (48). A total of 87% of the respondents feel better prepared for their practical year and their career entry through the OSCE. Dental students described the OSCE as a superior method for evaluating their clinical skills and for presenting questions applicable to real-life clinical situations (43, 49). Furthermore, students attending an OSCE in Oakland show an improvement in long-term knowledge, supporting the benefit of an OSCE for their future careers (10). The students answered that the OSCE as an assessment method called attention to their weaknesses even before receiving the official results. In India, a similar rate of students reported that they could immediately identify their weak spots (50). (3) In the feedback questionnaire, several students reflected the use of models in the OSCE. Because of the ethical difficulties of performing genital examinations on actors, the examinations were realized using models, despite their known limitations. However, the availability of gynecological

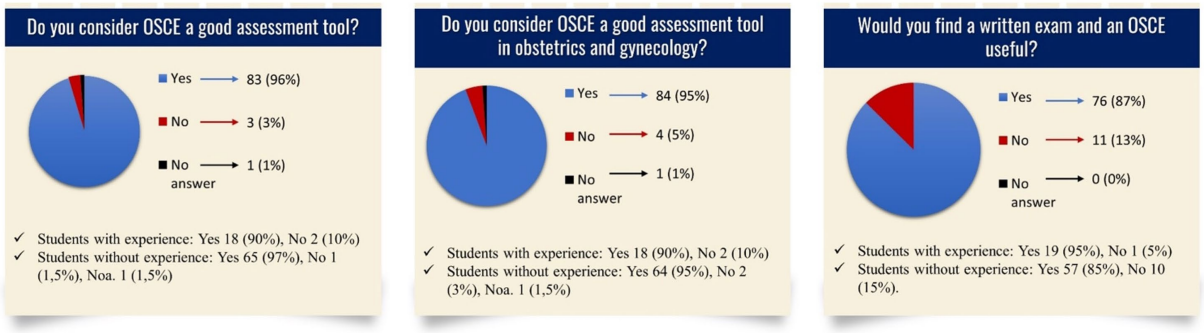


FIGURE 4
Questionnaire and results regarding OSCE in general, Noa: No answer.

teaching associates as live models to help teach gynecologic and obstetric examinations in the Netherlands and the United States of America (51–53), seems to be superior to the use of inanimate models (54) as well as provide a better cost–benefit ratio (55).

The establishment of OSCEs as an assessment method in the study of medicine, as required in the National Competence-Based Catalog of Learning Objectives in Medicine, German Catalog of Learning Objectives in Medicine, and the "Master Plan for Medical Studies 2020" (14, 15, 21), seems to be welcomed by the students because questions regarding their perception of OSCE as an assessment tool are mainly answered positively (Figure 4). In others countries, students and teachers have a similarly positive perception of the OSCE (42, 49, 56). This positive perception correlates with known evaluations, in which the students describe OSCE as an accurate measurement method to reflect their knowledge and skills (43, 57, 58). Therefore, the students' approval rate of OSCE as an assessment tool in obstetrics and gynecology is on a similar level as the approval rate of OSCE as an assessment tool in general (Figure 4) proving the possibility of implementing an OSCE in various medical disciplines (5–7). However, the students attending this OSCE find the combination of written examinations and an OSCE useful. The students may see the need to incorporate both assessment methods to reflect their behavioral and cognitive skills and to be trained in both skills (4, 59), thus supporting the two-step approach. The combination of both assessment methods has the strongest predictive validity of the subsequent performance, correlating with the students' wishes (60).

One limitation of this study is its design as a single-center study so that the students' perception of the OSCE is solely based on their learning experience at the University Hospital Bonn. Furthermore, the students' opinions may be influenced by the OSCE performed shortly before answering the poll. Although 113 students completed the OSCE, only 87 students (76.99%) answered the poll, so possibly, only students with a particular attitude answered the poll. A further limitation of this study is that the poll lacks an analysis of the students' experiences during the clerkship and the small group teaching. This could be relevant as the students surely had different experiences due to the daily fluctuation of cases during the clerkship and the varying teaching talents of the supervisors, leading to different levels of learning success. This may explain why 25% of students claimed to miss adequate preparation for the OSCE (Figure 3), whereas 75% of students described their preparation as adequate. Furthermore, the conception of the OSCE prevents an analysis regarding internal consistency such as Cronbach's alpha (61) because only a low number of stations are passed by all students. The realization of more stations for the assessment of the students is hindered by the limited possibility of dispensing physicians from their daily routines. The allocation of three stations from five possible stations allows more aspects of obstetrics and gynecology to be covered despite the shortage of staff.

OSCEs in general and specifically in obstetrics and gynecology are perceived positively by students. The students prefer the OSCE as an additional assessment method, thereby welcoming the changes regarding the implementation of more OSCEs in the study of medicine and favoring the establishment of a two-step learning approach in Germany.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Commission of University Hospital Bonn. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because the study is based on a voluntary evaluation of teaching methods without biomedical experiments on humans.

Author contributions

RP: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. AA: Conceptualization, Writing – review & editing. EE: Conceptualization, Data curation, Writing – review &

editing. AM: Resources, Writing – review & editing. NS: Resources, Writing – review & editing, Data curation. HP: Writing – review & editing, Validation, Writing – original draft. EW: Resources, Writing – review & editing. UG: Resources, Writing – review & editing. AW: Data curation, Writing – review & editing. BS: Data curation, Resources, Writing – review & editing. FR: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. *BMJ*. (1975) 1:447–51. doi: 10.1136/bmj.1.5955.447

2. Rushforth HE. Objective structured clinical examination (OSCE): review of literature and implications for nursing education. *Nurse Educ Today*. (2007) 27:481–90. doi: 10.1016/j.nedt.2006.08.009

3. Müller S, Settmacher U, Koch I, Dahmen U. A pilot survey of student perceptions on the benefit of the OSCE and MCQ modalities. *GMS J Med Educ*. (2018) 35:Doc51. doi: 10.3205/ZMA001197

4. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. (1990) 65:S63–7. doi: 10.1097/00001888-199009000-00045

5. Manogue M, Brown G. Developing and implementing an OSCE in dentistry. *Eur J Dent Educ*. (1998) 2:51–7. doi: 10.1111/j.1600-0579.1998.tb00039.x

6. Brighton R, Mackay M, Brown RA, Jans C, Antoniou C. Introduction of undergraduate nursing students to an objective structured clinical examination. *J Nurs Educ*. (2017) 56:231–4. doi: 10.3928/01484834-20170323-08

7. Smith V, Muldoon K, Biesty L. The objective structured clinical examination (OSCE) as a strategy for assessing clinical competence in midwifery education in Ireland: a critical review. *Nurse Educ Pract*. (2012) 12:242–7. doi: 10.1016/j.nepr.2012.04.012

8. Gilson G. Assessing clinical competence of medical students in women's health care: use of the objective structured clinical examination. *Obstet Gynecol*. (1998) 92:1038–43. doi: 10.1016/S0029-7844(98)00329-9

9. Elzubeir MA, Rizk DE. Assessing confidence and competence of senior medical students in an obstetrics and gynaecology clerkship using an OSCE. *Educ Health (Abingdon)*. (2001) 14:373–82. doi: 10.1080/13576280110082231

10. Berg J, Ogunyemi D, Venuti J, Ferrari T. Using Interprofessional OSCE-based simulation collaboration to teach obstetrics to second-year medical students. *Obstet Gynecol*. (2017) 130:53S–S. doi: 10.1097/01.AOG.0000525784.90852.91

11. McFaul PB, Taylor DJ, Howie PW. The assessment of clinical competence in obstetrics and gynaecology in two medical schools by an objective structured clinical examination. *Br J Obstet Gynaecol*. (1993) 100:842–6. doi: 10.1111/j.1471-0528.1993.tb14311.x

12. Nalesnik SW, Mills CS, Olsen CH, Haffner WHJ, Zahn CM. Creating an ideal objective structured clinical exam for an obstetrics and gynecology medical student clerkship. *Am J Obstet Gynecol*. (2005) 193:1544–50. doi: 10.1016/j.ajog.2005.06.067

13. Prasad L, Hockstein S, Safdieh JE, Harvey K, Christos PJ, Kang Y. An objective structured clinical exam on breaking bad news for clerkship students: in-person versus remote standardized patient approach. *MedEdPORTAL*. (2023) 19:11323. doi: 10.15766/mep_2374-8265.11323

14. Craig C, Kasana N, Modi A. Virtual OSCE delivery: the way of the future? *Med Educ*. (2020) 54:1185–6. doi: 10.1111/medu.14286

15. Masterplan 2020. (2022). Available at: https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/masterplan-medizinstudium-2020.html (Accessed Jun 17, 2022).

16. LOOOP-Netzwerk der Charite. Nationaler Kompetenzbasierter Lernzielkatalog Medizin, Version 2.0. Nationaler Kompetenzbasierter Lernzielkatalog Medizn. Available at: https://nklm.de/zend/menu. (Accessed Jul 5, 2022).

17. Recker F, Dohmen S, Riedel F, Egger E, Weiss M, Stope M, et al. Implementierung kompetenzbasierter Lehre und Prüfungen in die Frauenheilkunde im Rahmen der neuen ärztlichen Approbationsordnung 2020. *Gynäkologie*. (2021) 54:515–20. doi: 10.1007/s00129-021-04813-5

18. Kern DE, Thomas PA, Hughes MT eds. *Curriculum development for medical education: a six-step approach*. *2nd* ed. Baltimore, Md: Johns Hopkins University Press (2009).

19. Plöger R, Abramian A, Egger E-K, Gembruch U, Mustea A, Sänger N, et al. Development and feasibility of interdisciplinary objective structured clinical examination (OSCE) in obstetrics and gynecology. *Arch Gynecol Obstet*. (2023) 307:1265–8. doi: 10.1007/s00404-023-06909-w

20. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. (2016) 38:655–62. doi: 10.1007/s11096-016-0257-x

21. Institut für medizinische und pharmazeutische Prüfungsfragen. (2021). Available at: http://www.impp.de/files/PDF/Gegenstandskataloge/Medizin/gk2-2021-Auflage05_1.pdf (Accessed Jun 17, 2022).

22. Khan KZ, Gaunt K, Ramachandran S, Pushkar P. The objective structured clinical examination (OSCE): AMEE guide no. 81. Part II: Organisation & Administration. *Med Teach*. (2013) 35:e1447–63. doi: 10.3109/0142159X.2013.818635

23. Turner JL, Dankoski ME. Objective structured clinical exams: a critical review. *Fam Med*. (2008) 40:574–8.

24. Wagner MG, Fischer MR, Scaglione M, Linsenmaier U, Schueller G, Berger FH, et al. Subspecialisation in emergency radiology: proposal for a harmonised European curriculum. GMS journal for. *Med Educ*. (2017) 34:Doc61. doi: 10.3205/ZMA001138

25. Rajkumar Honest PC, David KV, Rahman SPMF, Sankarapandian V. Designing a curriculum for postgraduate training in family medicine in India. *J Family Med Prim Care*. (2021) 10:3453–8. doi: 10.4103/jfmpc.jfmpc_2484_20

26. Scala JJ, Braun NJ, Shamardani K, Rashes ER, Wang W, Mediratta RP. Applying Kern's six steps to the development of a community-engaged, just-in-time, interdisciplinary COVID-19 curriculum. *J Med Educ Curric Dev*. (2022) 9:238212052210963. doi: 10.1177/23821205221096370

27. Deis N, Koch C, Dreimüller N, Gaitzsch E, Weißkircher J, Jünger J, et al. Development, implementation, and evaluation of a curriculum for medical students on conflicts of interest and communicating risk. *GMS J Med Educ*. (2020) 37:Doc3. doi: 10.3205/ZMA001296

28. Khanna P, Roberts C, Lane AS. Designing health professional education curricula using systems thinking perspectives. *BMC Med Educ*. (2021) 21:20. doi: 10.1186/s12909-020-02442-5

29. McKimm J, Jones PK. Twelve tips for applying change models to curriculum design, development and delivery. *Med Teach*. (2018) 40:520–6. doi: 10.1080/0142159X.2017.1391377

30. Meyer VM, Benjamens S, Moumni ME, Lange JFM, Pol RA. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. *Ann Surg*. (2022) 275:e75–81. doi: 10.1097/SLA.0000000000004078

31. Fincham JE. Response rates and responsiveness for surveys, standards, and the journal. *Am J Pharm Educ*. (2008) 72:43. doi: 10.5688/aj720243

32. Carley-Baxter LR, Hill CA, Roe DJ, Twiddy SE, Baxter RK, Ruppenkamp J. Does response rate matter? Journal editors use of survey quality measures in manuscript publication decisions. *Surv Pract*. (2009) 2:1–7. doi: 10.29115/SP-2009-0033

33. Bevan J, Russell B, Marshall B. A new approach to OSCE preparation – PrOSCEs. *BMC Med Educ*. (2019) 19:126. doi: 10.1186/s12909-019-1571-5

34. Robinson P, Morton L, Haran H, Manton RChichester Medical Education Centre, St Richard's Hospital, Chichester, UK. Mock OSCEs improve medical students' confidence and reduce anxiety related to summative examinations. *Educ Med J*. (2017) 9:41–5. doi: 10.21315/eimj2017.9.2.4

35. Young I, Montgomery K, Kearns P, Hayward S, Mellanby E. The benefits of a peer-assisted mock OSCE. *Clin Teach*. (2014) 11:214–8. doi: 10.1111/tct.12112

36. Pierre RB, Wierenga A, Barton M, Branday JM, Christie CD. Student evaluation of an OSCE in paediatrics at the University of the West Indies, Jamaica. *BMC Med Educ*. (2004) 4:22. doi: 10.1186/1472-6920-4-22

37. Heal C, D'Souza K, Banks J, Malau-Aduli BS, Turner R, Smith J, et al. A snapshot of current objective structured clinical examination (OSCE) practice at Australian medical schools. *Med Teach*. (2019) 41:441–7. doi: 10.1080/0142159X.2018.1487547

38. Tekkis PP, Fazio VW, Lavery IC, Remzi FH, Senagore AJ, Wu JS, et al. Evaluation of the learning curve in Ileal pouch–anal anastomosis surgery. *Ann Surg*. (2005) 241:262–8. doi: 10.1097/01.sla.0000152018.99541.f1

39. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J*. (2007) 83:777–9. doi: 10.1136/pgmj.2007.057190

40. Yusuf L. Objective structured clinical examination? How students perceive their learning after OSCE. Pak. *J Med Sci*. (2021) 37:37. doi: 10.12669/pjms.37.4.4005

41. Skrzypek A, Szeliga M, Stalmach-Przygoda A, Górski S, Kowalska B, Kocurek A, et al. The objective structured clinical examination (OSCE) from the perspective of 3rd year's medical students – a pilot study. *Folia Med Cracov*. (2017) 57:67–75.

42. Majumder MAA, Kumar A, Krishnamurthy K, Ojeh N, Adams OP, Sa B. An evaluative study of objective structured clinical examination (OSCE): students and examiners perspectives. *Adv Med Educ Pract*. (2019) 10:387–97. doi: 10.2147/AMEP.S197275

43. Hammad M, Oweis Y, Taha S, Hattar S, Madarati A, Kadim F. Students' opinions and attitudes after performing a dental OSCE for the first time: a Jordanian experience. *J Dent Educ*. (2013) 77:99–104. doi: 10.1002/j.0022-0337.2013.77.1.tb05450.x

44. Gelan EA, Essayas R, Gebressilase K. Perception of final year medical students about objective structured clinical examination in the department of general surgery. *Ethiop Med J*. (2015) 53:183–9.

45. Ukadike Okugbo S, Agbonrofo P, Irowa O. Five years after commencing the objective structured clinical examination: are we getting it right? Medical students' assessment as the measuring index. *Afr Health Sci*. (2020) 20:960–5. doi: 10.4314/ahs.v20i2.52

46. Lebdai S, Bouvard B, Martin L, Annweiler C, Lerolle N, Rineau E. Objective structured clinical examination versus traditional written examinations: a prospective observational study. *BMC Med Educ*. (2023) 23:69. doi: 10.1186/s12909-023-04050-5

47. Memon S, Shaikh S-H. Comparison of performance on written and OSCE assessment during end semester pediatric examination: comparison of domain based achievement during pediatric end semester examination. Pak. *J Med Sci*. (2020) 36:36. doi: 10.12669/pjms.36.4.2026

48. De MR FÉ, Pinto RZ, PMM A, ÉLM V, Teixeira AL, Ferreira FR, et al. Stress, anxiety, self-efficacy, and the meanings that physical therapy students attribute to their experience with an objective structured clinical examination. *BMC Med Educ*. (2020) 20:296. doi: 10.1186/s12909-020-02202-5

49. Graham R, Zubiaurre Bitzer LA, Mensah FM, Anderson OR. Dental student perceptions of the educational value of a comprehensive, multidisciplinary OSCE. *J Dent Educ*. (2014) 78:694–702. doi: 10.1002/j.0022-0337.2014.78.5.tb05721.x

50. Lele SM. A mini-OSCE for formative assessment of diagnostic and radiographic skills at a dental college in India. *J Dent Educ*. (2011) 75:1583–9. doi: 10.1002/j.0022-0337.2011.75.12.tb05218.x

51. Beckmann CRB, Barzansky BM, Sharf BF, Meyers K. Training gynaecological teaching associates. *Med Educ*. (1988) 22:124–31. doi: 10.1111/j.1365-2923.1988.tb00422.x

52. Pickard S. Can gynaecology teaching associates provide high quality effective training for medical students in the United Kingdom? Comparative study. *BMJ*. (2003) 327:1389–92. doi: 10.1136/bmj.327.7428.1389

53. de Klerk HW, Boere E, van Lunsen RH, Bakker JJH. Women's experiences with vaginal examinations during labor in the Netherlands. *J Psychosom Obstet Gynaecol*. (2018) 39:90–5. doi: 10.1080/0167482X.2017.1291623

54. Janjua A, Smith P, Chu J, Raut N, Malick S, Gallos I, et al. The effectiveness of gynaecology teaching associates in teaching pelvic examination to medical students: a randomised controlled trial. *Eur J Obstet Gynecol*. (2017) 210:58–63. doi: 10.1016/j.ejogrb.2016.10.006

55. Janjua A, Roberts T, Okeahialam N, Clark TJ. Cost-effective analysis of teaching pelvic examination skills using Gynaecology teaching associates (GTAs) compared with manikin models (the CEAT study). *BMJ Open*. (2018) 8:e015823. doi: 10.1136/bmjopen-2017-015823

56. Larsen T, Jeppe-Jensen D. The introduction and perception of an OSCE with an element of self- and peer-assessment. *Eur J Dent Educ*. (2008) 12:2–7. doi: 10.1111/j.1600-0579.2007.00449.x

57. Nasir AA, Yusuf AS, Abdur-Rahman LO, Babalola OM, Adeyeye AA, Popoola AA, et al. Medical students' perception of objective structured clinical examination: a feedback for process improvement. *J Surg Educ*. (2014) 71:701–6. doi: 10.1016/j.jsurg.2014.02.010

58. Awaisu A, Abd Rahman NS, Nik Mohamed MH, Bux Rahman Bux SH, Mohamed Nazar NI. Malaysian pharmacy students' assessment of an objective structured clinical examination (OSCE). *Am J Pharm Educ*. (2010) 74:34. doi: 10.5688/aj740234

59. Kirton SB, Kravitz L. Objective structured clinical examinations (OSCE) compared with traditional assessment methods. *Am J Pharm Educ*. (2011) 75:111. doi: 10.5688/ajpe756111

60. Wilkinson TJ, Frampton CM. Comprehensive undergraduate medical assessments improve prediction of clinical performance. *Med Educ*. (2004) 38:1111–6. doi: 10.1111/j.1365-2929.2004.01962.x

61. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. (2011) 2:53–5. doi: 10.5116/ijme.4dfb.8dfd

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 02 February 2024
DOI 10.3389/fmed.2024.1330443

OPEN ACCESS

EDITED BY
Fedde Scheele,
VU Amsterdam, Netherlands

REVIEWED BY
Florian Recker,
University of Bonn, Germany
Jessica Van Der Aa,
Academic Medical Center, Netherlands

*CORRESPONDENCE
Lise Brogaard
✉ lbrj@clin.au.dk

RECEIVED 30 October 2023
ACCEPTED 08 January 2024
PUBLISHED 02 February 2024

CITATION
Brogaard L, Hinshaw K, Kierkegaard O, Manser T, Uldbjerg N and Hvidman L (2024) Developing the TeamOBS-vacuum-assisted delivery checklist to assess clinical performance in a vacuum-assisted delivery: a Delphi study with initial validation. *Front. Med.* 11:1330443. doi: 10.3389/fmed.2024.1330443

Developing the TeamOBS-vacuum-assisted delivery checklist to assess clinical performance in a vacuum-assisted delivery: a Delphi study with initial validation

Lise Brogaard[1,2]*, Kim Hinshaw[3], Ole Kierkegaard[4], Tanja Manser[5], Niels Uldbjerg[1,2] and Lone Hvidman[2]

[1]Department of Obstetrics and Gynecology, Aarhus University Hospital, Aarhus, Denmark, [2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, [3]Department of Obstetrics and Gynecology, Sunderland Royal Hospital, Sunderland, United Kingdom, [4]Department of Obstetrics and Gynecology, Horsens Regional Hospital, Horsens, Denmark, [5]Fachhochschule Nordwestschweiz (FHNW) School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

Introduction: In Northern Europe, vacuum-assisted delivery (VAD) accounts for 6–15% of all deliveries; VAD is considered safe when conducted by adequately trained personnel. However, failed vacuum extraction can be harmful to both the mother and child. Therefore, the clinical performance in VAD must be assessed to guide learning, determine a performance benchmark, and evaluate the quality to achieve an overall high performance. We were unable to identify a pre-existing tool for evaluating the clinical performance in real-life vacuum-assisted births.

Objective: We aimed to develop and validate a checklist for assessing the clinical performance in VAD.

Methods: We conducted a Delphi process, described as an interactive process where experts answer questions until answers converge toward a "joint opinion" (consensus). We invited international experts as Delphi panelists and reached a consensus after four Delphi rounds, described as follows: (1) the panelists were asked to add, remove, or suggest corrections to the preliminary list of items essential for evaluating clinical performance in VAD; (2) the panelists applied weights of clinical importance on a Likert scale of 1–5 for each item; (3) each panelist revised their original scores after reviewing a summary of the other panelists' scores and arguments; and (4) the TeamOBS-VAD was tested using videos of real-life VADs, and the Delphi panel made final adjustments and approved the checklist.

Results: Twelve Delphi panelists from the UK (n = 3), Norway (n = 2), Sweden (n = 3), Denmark (n = 3), and Iceland (n = 1) were included. After four Delphi rounds, the Delphi panel reached a consensus on the checklist items and scores. The TeamOBS-VAD checklist was tested using 60 videos of real-life

vacuum extractions. The inter-rater agreement had an intraclass correlation coefficient (ICC) of 0.73; 95% confidence interval (95% CI) of [0.58, 0.83], and that for the average of two raters was ICC 0.84 95% CI [0.73, 0.91]. The TeamOBS-VAD score was not associated with difficulties in delivery, such as the number of contractions during vacuum extraction delivery, cephalic level, rotation, and position. Failed vacuum extraction occurred in 6% of the video deliveries, but none were associated with the teams with low clinical performance scores.

Conclusion: The TeamOBS-VAD checklist provides a valid and reliable evaluation of the clinical performance of vaginal-assisted vacuum extraction.

KEYWORDS

performance, emergency, obstetric, vacuum extraction, team, video, checklist

1 Introduction

In Northern Europe, 6–15% of women have had a vacuum-assisted delivery (VAD) (1). Notably, most births using a vacuum have good outcomes, and VAD is generally accepted as safe when performed by appropriately trained healthcare providers (2). When delivery is indicated in the second stage of labor, obstetricians must balance the differing risks of instrumental-assisted births with those of second-stage cesarean births. Women delivered successfully using a vacuum have a higher chance of uncomplicated births in subsequent pregnancies than those delivered through a cesarean section. Furthermore, cesarean birth in the second stage can be challenging because of increased maternal and perinatal risks (3). Therefore, instrumental delivery remains a core obstetrical competence, and systematic clinical performance evaluation is crucial to ensure ongoing quality assessment and research to improve clinical training (4).

Available checklists for VAD have been designed to support procedural task execution, as cognitive aids or by evaluating by item-by-item feedback. These existing checklists in vacuum extraction use simple dichotomous items (not done/done) or Likert scales (5–7). However, a growing body of empirical evidence suggests that performance assessment should include weighted checklist items to differentiate between essential and less important actions. Furthermore, the inclusion of time frames helps to create a more refined assessment of performance (8, 9). The main advantage of these performance assessment tools is the production of an objective summative score that is valuable in quality assessment, benchmarking performance and research (10).

We could not identify an existing performance assessment tool for real-life vacuum-assisted births that fulfilled the abovementioned requirements; therefore, this study aimed to develop and validate a checklist for assessing the clinical performance of VAD.

Abbreviations: VAD, vacuum-assisted delivery; ICC, intraclass correlation coefficient; CI, confidence interval.

2 Materials and methods

2.1 Delphi method

We developed the TeamOBS-VAD checklist using a Delphi process to evaluate clinical performance (**Figure 1**). The Delphi method is an interactive process in which experts answer questions in four rounds until the answers reach a consensus. This is an internationally recognized method for solving research questions with different clinical approaches in practice and limited evidence (11–13). Our research aimed to develop a list of core items/tasks that a team should perform when conducting a VAD. The Delphi process was conducted online using Google survey tool forms.

Eighteen international obstetricians were invited to participate in the Delphi panel; 12 accepted and completed the Delphi process. The Delphi panelists were obstetric consultants from the UK (n = 3), Norway (n = 2), Sweden (n = 3), Denmark (n = 3), and Iceland (n = 1). The Delphi process was anonymous to ensure equal weight for all participants' arguments and suggestions. The Delphi steering committee (NU, LB, OK, and LH) drafted a preliminary list of items identified from the literature and international guidelines, each representing a core task in the VAD.

In the first round, the panel reviewed the preliminary list of items and was asked to add, remove, or suggest the wording of items and argue why. In the second round, panelists weighted each item for clinical importance on a 5-point Likert scale, where "5" was highly important and "1" was least important. In round three, the panelists reassessed the weights after reviewing a summary of the other panelists' scores and arguments. Consensus was defined when 90% of the panelists' scores fell within three neighboring categories using the Likert scale range of 1–5.

The TeamOBS-VAD checklist was designed using a predefined blueprint (10). The checklist resulted in a total score calculated using a "*weighted score*" of 0–100% as a percentage of the highest possible points and the assessors' subjective global rating, the "*patient safety score*," ranging from 0–100% (100% served as a goal for others). The "*patient safety score*" represented the assessors' subjective global rating of all treatment actions and offered the opportunity to evaluate aspects of performance that were not

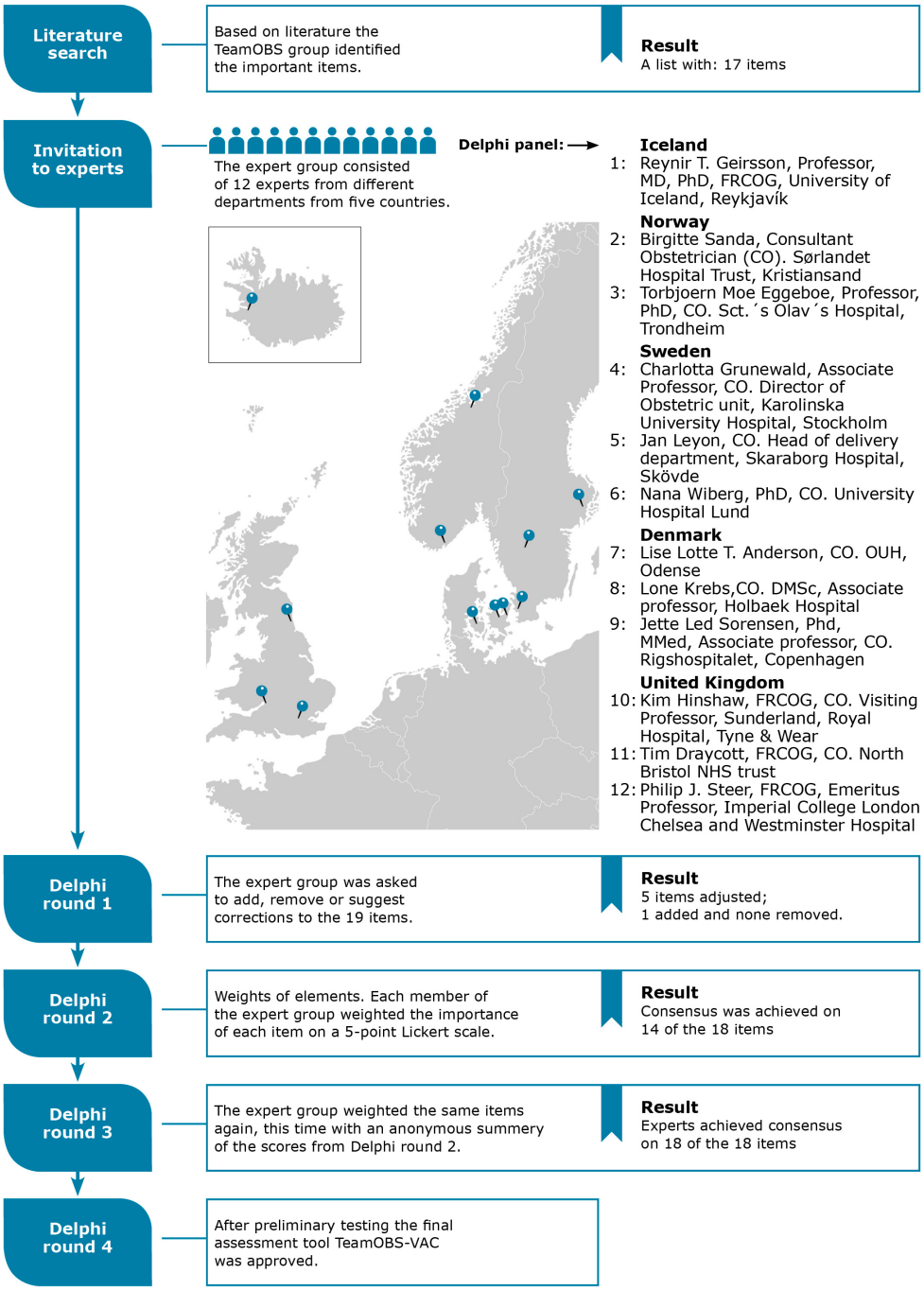


FIGURE 1
The Delphi process for the TeamOBS-VAD checklist.

captured by the 18 items. The TeamOBS-VAD score was calculated as follows: (*weighted score* + *patient safety score*)/2.

2.2 Validity and reliability testing

We used the conceptual definitions and arguments for validity described by Cook et al. (14, 15) to test validity and reliability. We used video recordings of real-life VADs, collected with informed consent from all individuals present (patients and staff) in the videos from two Danish hospitals: Aarhus University and Horsens Regional Hospitals. Aarhus University Hospital, with approximately 5,000 deliveries per year, provides level III maternal care (16) and Horsens Regional Hospital, with approximately 2,000 deliveries per year, provides level II maternal care. All 17 birthing suites at the two hospitals were equipped with two or three high-definition minidome surveillance cameras and a microphone attached to the ceiling, allowing for a comprehensive view of the room. As previously described, the recordings were automatically activated using Bluetooth (17). The instruments used for vacuum

extraction were a Malmström or Bird metal size 5–6 cm cup or a soft silicone cup. Videos were obtained with informed consent over 15 months between January 2015 and March 2016 and analyzed between 2018 and 2019.

Consultants LH and LA conducted tests for validity and reliability. They were experienced video raters from previous TeamOBS studies. In a session, they were trained as "raters" to develop familiarity with the checklist, followed by detailed discussions about the high and low score definitions for each item in the tool. The TeamOBS-VAD checklist was then independently applied to 60 real-life VAD video recordings. One month later, they reassessed 20% of the recordings (randomly selected) to evaluate both inter-rater and intra-rater agreements. Notably, all 60 videos were ranked based on their clinical performance scores. NU, LB, and LH reviewed the videos chronologically to determine where to set the low, acceptable, and high team performance levels.

2.3 Ethics

This study was approved by the Central Denmark Region's legal department, Danish Data Protection Agency (2012-58-006), and Central Denmark Region's Research Foundation (Case No. 1-16-02-257-14). All the participants (patients and staff) volunteered to participate and provided informed consent.

2.4 Statistical analysis

The clinical performance scores were analyzed on a logit-transformed scale to meet the criteria for normality and back-transformed using the inverse logit function (18). Rater agreement was described as summative using the intraclass correlation (19), Bland–Altman plots, and limits of agreement (20). STATA 17 (StataCorp LP, College Station, TX, USA) was used for statistical analysis.

3 Results

The Delphi panel achieved consensus in round one after adjusting for five items and adding one. Notably, all panelists weighted items according to importance on a Likert scale of 1–5; they reached a consensus on 14/18 items in the second round and a consensus on all 18 items in the next round. The TeamOBS-VAD checklist was developed and tested for usability in a simulation based on the Delphi method, and the raters found it easy to understand and use (**Figure 2**, **Supplementary material 1**).

The validity and reliability were tested by applying the TeamOBS-VAD checklist to 60 videos of real-life VADs. The inter-rater agreement for an individual rater had an intraclass correlation coefficient (ICC) of 0.73; 95% confidence interval (95% CI) of [0.58, 0.83], and that for the average of two raters ICC 0.84; 95% CI [0.73, 0.91]. The intra-rater agreement was tested as raters re-evaluated 11 videos, and the agreements had an ICC 0.74; 95% CI [0.24, 0.92] for rater one and ICC 0.90; 95% CI [0.67, 0.97] for rater 2. Agreement was described using the Bland–Altman plot. The limits used to indicate low, acceptable, and high clinical performance were a score of <60%, 60–84%, and 85–100%, respectively; **Figure 3**. The arguments for validity and reliability are presented in **Table 1**.

4 Discussion

4.1 Main findings

We used an international Delphi process to develop the TeamOBS-VAD checklist to evaluate the clinical performance of VAD. The checklist allowed the calculation of the total performance score. The validity and reliability of the TeamOBS-VAD checklist were high when applied by two raters. However, they were still acceptable when used by one rater.

4.2 Strengths and limitations

A crucial strength of the study is the international, diverse Delphi panel as this ensured a sensible construct along with clinical applicability and increased the possibilities for international adoption and acceptability (21). The panel's feedback and commentaries rounds guaranteed substantiated adaptations in the described tasks as well of the framework of the checklist (13). We recognize that the inclusion of a larger number of experts and inclusion of other professional groups such as midwives in expert panels could have been of additional benefit. Furthermore, as all experts in the Delphi are based Northern Europe countries, we recognize that the checklist will apply primarily in these countries. However, with minor modifications the checklist may be useful in other countries as well (22).

The use of videos of real-life vacuum deliveries was also a significant strength to the validity and reliability (15). Informed consent was obtained for all included videos, fulfilling all Danish ethical and legal requirements. However, informed consent may introduce potential selection bias, as we cannot exclude the possibility that low-performing teams were less willing to provide consent (23). It was nevertheless reassuring that 95% of the obstetricians consented to include all their videos, and only two videos were deleted when the staff withdrew consent (24).

4.3 Interpretation

The Delphi process was valuable in ensuring the inclusion of different international perspectives in managing vacuum extraction. The respective panelists' national guidelines had similarities; however, some elements differed (25). Discussions in the Delphi panel included the acceptable number of pulls, when to abandon the attempt, checking the position with ultrasound (and it's weighting according to importance), and examining and suturing the perineum after delivery. The weighted score of pain relief was discussed, as some panelists thought the idea of delivery using a vacuum without considering further medical pain relief was inappropriate. However, others did not consider the need for additional analgesia. These different views may reflect the expectations of the panelists and the availability or use of epidural and spinal analgesia (26). Two Delphi panelists preferred

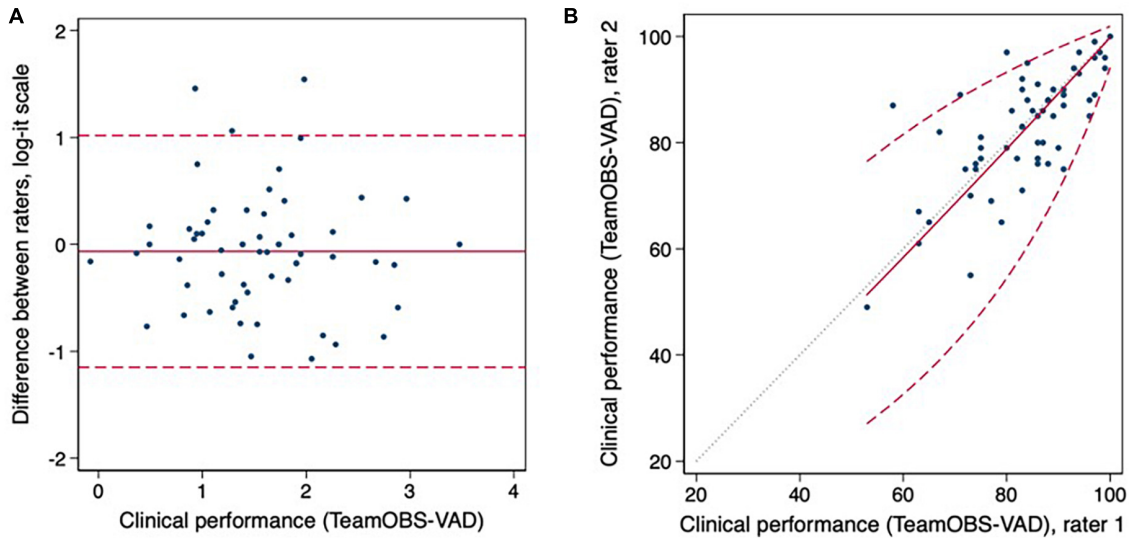


FIGURE 2
In graphs **(A,B)** the inter-rater agreement are visualized as Bland Altman plots and limits of agreement. Clinical performance data was analyzed on the log-it scale to meet the assumptions of constant mean, SD, and normality.

TEAM | OBS **Clinical Performance Checklist evaluating vaginal delivery by vacuum extraction**

Scoring: Put a cross in the box and give "not indicated", "cannot be assessd", "0", "1" or "2"

Stage of management Category	No. Item	Item description	Item weight	Not indicated 2 points	Cannot be assessed 2 points	Item weight x points	Not done or not considered 0 points	Partially or incorrectly done, or not done in a timely manner 1 point	Done correctly, completely and in a timely manner 2 points	Item weight x points
Preparation of the staff team	1-1	Appropriate staff trained in vaginal instrumental delivery	4	☐	☐		☐	Partially done – safe but could be improved ☐	Appropriate number and charge of staff present who are trained in vaginal instrumental delivery ☐	
	1-2	Team reviews background ('check in')	4	☐	☐		☐	Partially done – safe but could be improved ☐	Check-in includes: Parity of the mother; fetus: gestational age, weight and presentation; other risk factors ☐	
	1-3	Team reviews indication & urgency of instrumental delivery	4	☐	☐		☐	Partially done - safe but could be improved - timing not clearly specified ☐	The team states the indication (fetal or maternal) and time frame for delivery ☐	
Assessment	2-1	Considers whether contractions are adequate	3.5	☐	☐		☐	Partially done – safe but could be improved ☐	Assesses whether contractions are adequate (considers oxytocin - careful assessment if multgravid) ☐	
	2-2	Abdominal examination	3	☐	☐		☐	Partially done – safe but could be improved ☐	Examination regarding whether the head is not or is only partly palpable per abdomen (head should be no more than 1/5 palpable) ☐	
	2-3	Vaginal examination	4	☐	☐		☐	Partially done – safe but could be improved ☐	Examination regarding: Vertex presentation, cervical dilatation, rupture of the membranes, assessment of station, position and moulding ☐	
	2-4	Ultrasound assessment of fetal position and descent, if clinically unsure	2	☐	☐		☐	Partially done – safe but could be improved ☐	Ultrasound assessment of fetal position and descent undertaken for clinical uncertainty ☐	
Preparation of the mother	3-1	Ensures correct position	3.5	☐	☐		☐	Partially done – safe but could be improved ☐	Ensure correct position of the mother, buttocks to the edge of bed ☐	
	3-2	Informed consent	3	☐	☐		☐	Partially done – safe but could be improved ☐	Clear explanation of the procedure and informed consent given ☐	
	3-3	Considers analgesia	3.5	☐	☐		☐	Partially done – safe but could be improved ☐	Considers and ensures appropriate analgesia ☐	
	3-4	Empty bladder	3.5	☐	☐		☐	Considers whether the bladder is empty - no specific plan or action ☐	Team ensures the bladder is empty or bladder is emptied ☐	
Procedure	4-1	Choice of instrument	4	☐	☐		☐	Partially done – safe but could be improved ☐	The operator chooses the instrument most appropriate for the clinical situation and his/her level of skill ☐	
	4-2	Conducting the operative vaginal birth (application & initial traction)	4	☐	☐		☐	Partially done – safe but could be improved ☐	Instruments are placed and handled correctly according to guidelines ☐	
	4-3	Timing / traction	3.5	☐	☐		☐	Timing or length 'borderline' (based on local guidance). Traction safe but could be improved. ☐	The delivery is conducted within max 4 contractions and/or within 15-20 minutes (May be modified according to local guideline). Appropriate traction ☐	
	4-4	Ensuring fetal well-being	4	☐	☐		☐	Partially done - not done after *every* contraction ☐	Fetal heartrate is assessed after each contraction ☐	
	4-5	Relevant precautions taken against perineal tears	3.5	☐	☐		☐	Partially done – safe but could be improved ☐	Correct support during the delivery, "don´t push too hard", hands on perineum and limited use of episiotomy ☐	
Safety	5-1	Re-assessment if poor progress	4	☐	☐		☐	Done partially or not in a timely manner ☐	Reassesses appropriately: when the cup detaches two to three tiems or there is no evidence of progressive descent after 1-2 contractions. Re-checks position. ☐	
	5-2	Abandons attempted vaginal delivery when appropriate	4	☐	☐		☐	Done partially or not in a timely manner ☐	Abandons vaginal delivery when: there is no evidence of progressive descent with moderate traction, or where delivery is not imminent following three or four contractions (May be modified according to local guideline) ☐	
Patient safety score	6-1					sum				sum

0% 50% 100%
Patient safety is threatened/ Non-acceptable performance | Acceptable performance | Excellent

% = Patient safety score

Weighted score: (____ / 130 − ____) x 100 = ____
sum sum

Clinical performence score: ____

(Patient safety score + Weighted score total)/2: ____

Grafisk Service 8392

FIGURE 3
The TeamOBS-VAD checklist.

to include non-technical skills such as communication, leadership, and teamwork in the TeamOBS-VAD.

The steering committee had *a priori* decided not to include non-technical skills in the checklist, as validated obstetric teamwork assessment tools already exist for this purpose (27). From a methodological point of view tools for rating non-technical skills should be independent of the actual clinical problem while the clinical performance ratings need to address these specifics. Thus, the Delphi process and framework of the checklist did focus on the clinical performance (24, 28).

We did not identify any published checklists which are specifically designed to produce a summative score for VAD performance. Previously published checklists have been designed as cognitive aids, for supporting procedural task execution, or evaluating using item-by-item feedback (5–7). These classic evaluation checklists use simple dichotomous items (i.e., not

TABLE 1 Arguments for validity.

Arguments for TeamOBS-VAD validity					
Sources of validity		Validity question?	Data	Method	Results
1.	Content evidence	Measure what it was intended to?	Delphi process	Expert panel of 12 senior obstetricians from five countries	Consensus of items, weight of importance, and final checklist
			Blueprint	Used five-step approach	TeamOBS-VAD used a rating scale with five categories, weighed items, and a global rating scale
2.	Response process evidence	Easy to use and understand?	Rater handbook	Systematic feedback Usability testing	After 1 h of training, the obstetricians were comfortable with the checklist
3.	Internal structure evidence	Distinguish high from low performance?	60 videos of real-life deliveries by vacuum	Review videos in chronological order by clinical performance, open discussion	Performance Low < 60% Acceptable 60–84% High 85–100%
		Reproducibility?	60 videos of real-life deliveries by vacuum	Inter-rater agreement Individual Average of two raters	ICC (95% CI) 0.73 [0.58, 0.83] 0.84 [0.73, 0.91]
			11 videos of real-life (18%)	Intra-rater agreement Reevaluation >1 month	rater 1: ICC 0.72 [0.24, 0.92] rater 2: ICC 0.90 [0.67, 0.97]
			Across scenarios	Indication of vacuum	Obstericians found the checklist easy to use across scenarios
4.	Relations with other variables evidence	Is the performance associated with low-risk deliveries?	60 videos of real-life deliveries by vacuum	All videos were listed for the number of contractions, position (mid, lower, outlet), and failed or successful vaginal delivery	Clinical performance had no association with the number of contractions or positions. Failed vacuum extraction was listed in 6% of the videos; however, none of these teams achieved low clinical performance
5.	Consequences evidence	Disagreements between raters	60 videos of real-life deliveries by vacuum	Disagreement >15 points between raters	Six videos (10%) Recommended two reviewers for high precision

done/done); however, dichotomous items are often not sufficient for the assessment of more complex tasks. Therefore, tools for performance evaluation include more categories (e.g., not indicated/incorrectly performed/performed late/timely and correctly performed) (10). Other requirements of a performance checklist include weighted checklist items to differentiate between essential and less important actions and time frames to help create a more refined assessment of performance (8, 9). Therefore, we included both weighted items and more categories and these should be considered when developing future evaluation checklists.

The summative clinical performance score is useful in (a) assessing adherence to accepted guidelines, (b) supporting individual learning by mapping the learning curve, and (c) quality assessment within a labor and delivery ward (10). We developed the TeamOBS-VAD as a tool to produce a summative clinical performance score based on items weighted for importance. We ensured that difficult deliveries and deliveries in which the teams abandoned the vaginal delivery attempt did not automatically result in a low score. Conversely, simple outlet deliveries did not automatically result in high scores. The TeamOBS-VAD clinical performance score was not associated with the number of contractions if the number was acceptable, cephalic level, rotation, or position. Failed vacuum extraction was observed in 6% of the videos; none were teams with low clinical performance scores.

Educators and trainees have experienced difficulties developing and maintaining clinical competence in VAD because of reduced working hours and instrumental delivery rates (29). Therefore, we must rethink our learning path for vacuum extraction to improve and speed up trainees' learning. The first step could be to systematically assess performance, as improving current methods is difficult if we do not measure them objectively (30). Filming vaginal deliveries could be a second step in rethinking our learning path because a systematic assessment of the performance of trainees or departments allows us to investigate our team's performance and offer targeted training (31). Studies evaluating video use for educational purposes have reported high patient and staff acceptability and compliance (23). Notably, 30% of the staff found that filming provoked mild anxiety; however, they confirmed that the educational value outweighed it (32). Solving the ethical and legal issues associated with video recordings in emergency care may improve our knowledge and serve as a foundation for providing better patient care (33, 34).

External validity must be considered before applying the TeamOBS-VAD checklist in other settings. Our checklist reflects adherence to guidelines and accepted practices in Northern Europe. Thus, before applying the TeamOBS-VAD checklist in different settings, it may be necessary to agree with the present statements in the tool for "done correctly." Significant differences in opinion, the checklist, and item weights of importance will be imprecise. In addition, scoring an abandoned attempted vaginal delivery is meaningless if an urgent cesarean section is unavailable. Validity testing was conducted in two Danish hospitals, and it may be

necessary to re-evaluate the validity if the delivery guidance differs substantially.

5 Conclusion

The TeamOBS-VAD checklist we developed is valid, reliable, and easy to use in assessing clinical vacuum deliveries. It may help train individuals and evaluate team performance in a department.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was approved by the Central Denmark Region's legal department, Danish Data Protection Agency (2012-58-006), and Central Denmark Region's Research Foundation (Case No. 1-16-02-257-14). All the participants (patients and staff) volunteered to participate and provided informed consent. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Writing – original draft. KH: Formal analysis, Methodology, Supervision, Writing – review & editing. OK: Conceptualization, Data curation, Funding acquisition, Resources, Supervision, Validation, Writing – review & editing. TM: Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. NU: Conceptualization, Methodology, Supervision, Writing – review & editing. LH: Conceptualization, Data curation, Formal analysis, Investigation, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We thank the following organizations and departments for their financial support: Tryg Foundation (Trygfonden) (Grant ID no. 109507), Regional Hospital in Horsens, Department of Obstetrics and Gynecology, and Regional Postgraduate Medical Education Administration Office, North.

Acknowledgments

We thank the Delphi panel and all the patients who volunteered to participate in the study. We thank our colleagues who participated in this study and continued to strive for excellence. A special thanks to Consultant Lars Høj and Midwives Astrid Johansen and Lene Yding for assisting in testing the checklist. Our analysis was approved by Erik Parner, Professor of Statistics at Aarhus University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1330443/full#supplementary-material

References

1. Muraca G, Sabr Y, Lisonkova S, Skoll A, Brant R, Cundiff G, et al. Morbidity and mortality associated with forceps and vacuum delivery at outlet, low, and midpelvic station. *J Obstet Gynaecol Can.* (2019) 41:327–37. doi: 10.1016/j.jogc.2018.06.018

2. Murphy D, Strachan B, Bahl R. Assisted vaginal birth: green-top guideline no. 26. *BJOG.* (2020) 127:e70–112. doi: 10.1111/1471-0528.16092

3. Gei A. Prevention of the first cesarean delivery: the role of operative vaginal delivery. *Semin Perinatol.* (2012) 36:365–73. doi: 10.1053/j.semperi.2012.04.021

4. Merriam A, Ananth CV, Wright J, Siddiq Z, D'Alton M, Friedman A. Trends in operative vaginal delivery, 2005–2013: a population-based study. *BJOG.* (2017) 124:1365–72. doi: 10.1111/1471-0528.14553

5. Mannella P, Giordano M, Guevara M, Giannini A, Russo E, Pancetti F, et al. Simulation training program for vacuum application to improve technical skills in vacuum-assisted vaginal delivery. *BMC Preg Childbirth.* (2021) 21:338. doi: 10.1186/s12884-021-03829-y

6. Maagaard M, Oestergaard J, Johansen M, Andersen L, Ringsted C, Ottesen B, et al. Vacuum extraction: Development and test of a procedure-specific rating scale. *Acta Obstet Gynecol Scand.* (2012) 91:1453–9. doi: 10.1111/j.1600-0412.2012.01526.x

7. Staat B, Combs C. SMFM Special Statement: Operative vaginal delivery: checklists for performance and documentation. *Am J Obstet Gynecol.* (2020) 222:B15–21. doi: 10.1016/j.ajog.2020.02.011

8. Donoghue A, Nishisaki A, Sutton R, Hales R, Boulet J. Reliability and validity of a scoring instrument for clinical performance during Pediatric Advanced Life Support simulation scenarios. *Resuscitation.* (2010) 81:331–6. doi: 10.1016/j.resuscitation.2009.11.011

9. Donoghue A, Durbin D, Nadel F, Stryjewski G, Kost S, Nadkarni V. Effect of high-fidelity simulation on pediatric advanced life support training in pediatric house staff a randomized trial. *Pediatr Emerg Care.* (2009) 25:139–44.

10. Schmutz J, Eppich W, Hoffmann F, Heimberg E, Manser T. Five steps to develop checklists for evaluating clinical performance: An integrative approach. *Acad Med.* (2014) 89:996–1005. doi: 10.1097/ACM.0000000000000289

11. Wolf H, Schaap A, Bruinse H, Smolders-de Haas H, van Ertbruggen I, Treffers P. Vaginal delivery compared with caesarean section in early preterm breech delivery: a comparison of long term outcome. *Br J Obstet Gynaecol.* (1999) 106:486–91.

12. Morgan P, Lam-McCulloch J, Herold-McIlroy J, Tarshis J. Simulation performance checklist generation using the Delphi technique. *Can J Anaesth.* (2007) 54:992–7. doi: 10.1007/BF03016633

13. Hsu C, Sandford B. The Delphi technique: making sense of consensus. *Pract Assess Res Eval.* (2007) 12:1–8. doi: 10.1016/S0169-2070(99)00018-7

14. Cook D, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ.* (2015) 49:560–75. doi: 10.1111/medu.12678

15. Cook D, Zendejas B, Hamstra S, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ Theory Pract.* (2014) 19:233–50. doi: 10.1007/s10459-013-9458-4

16. Menard M, Kilpatrick S, Saade G, Hollier L, Joseph G, Barfield W, et al. Levels of maternal care This document was developed jointly by the. *Am J Obstet Gynecol.* (2015) 212:259–71. doi: 10.1016/j.ajog.2014.12.030

17. Brogaard L, Hvidman L, Hinshaw K, Kierkegaard O, Manser T, Musaeus P, et al. Development of the TeamOBS-PPH – targeting clinical performance in postpartum hemorrhage. *Acta Obstet Gynecol Scand.* (2018) 97:13336. doi: 10.1111/aogs.13336

18. Carstensen B. *Comparing Clinical Measurement Methods: A Practical Guide*. Hoboken, NJ: Wiley (2010). doi: 10.1136/pgmj.59.687.72-a

19. Koo T, Li M. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012

20. Bland J, Altman D. Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol.* (2003) 22:85–93. doi: 10.1002/uog.122

21. Barrett D, Heale R. What are Delphi studies? *Evid Based Nurs.* (2020) 23:68–9. doi: 10.1136/ebnurs-2020-103303

22. Schaap T, Bloemenkamp K, Deneux-Tharaux C, Knight M, Langhoff-Roos J, Sullivan E, et al. Defining definitions: a Delphi study to develop a core outcome set for conditions of severe maternal morbidity. *BJOG.* (2019) 126:394–401. doi: 10.1111/1471-0528.14833

23. Brogaard L, Uldbjerg N. Filming for auditing of real-life emergency teams: a systematic review. *BMJ Open Qual.* (2019) 8:e000588. doi: 10.1136/bmjoq-2018-000588

24. Brogaard L, Kierkegaard O, Hvidman L, Jensen K, Musaeus P, Uldbjerg N, et al. The importance of non-technical performance for teams managing postpartum haemorrhage: video review of 99 obstetric teams. *BJOG.* (2019) 126:15655. doi: 10.1111/1471-0528.15655

25. Tsakiridis I, Giouleka S, Mamopoulos A, Athanasiadis A, Daniilidis A, Dagklis T. Operative vaginal delivery: A review of four national guidelines. *J Perinat Med.* (2020) 48:189–98. doi: 10.1515/jpm-2019-0433

26. King T. Epidural anesthesia in labor. Benefits versus risks. *J Nurse Midwifery.* (1997) 42:377–88.

27. Onwochei D, Halpern S, Balki M. Teamwork assessment tools in obstetric emergencies: A systematic review. *Simul Healthc.* (2017) 12:165–76. doi: 10.1097/SIH.0000000000000210

28. Marks M, Mathieu J, Zaccaro S. A temporally based framework and taxonomy of team processes. *Acad Manage Rev.* (2001) 26:356–76.

29. Baskett T. Operative vaginal delivery – An historical perspective. *Best Pract Res Clin Obstet Gynaecol.* (2019) 56:3–10. doi: 10.1016/j.bpobgyn.2018.08.002

30. Manser T, Brösterhaus M, Hammer A. You can't improve what you don't measure: Safety climate measures available in the German-speaking countries to support safety culture development in healthcare. *Z Evid Fortbild Qual Gesundhwes.* (2016) 114:58–71. doi: 10.1016/j.zefq.2016.07.003

31. Bahl R, Murphy D, Strachan B. Non-technical skills for obstetricians conducting forceps and vacuum deliveries: qualitative analysis by interviews and video recordings. *Eur J Obstet Gynecol Reprod Biol.* (2010) 150:147–51. doi: 10.1016/j.ejogrb.2010.03.004

32. Davis L, Johnson L, Allen S, Kim P, Sims C, Pascual J, et al. Practitioner perceptions of trauma video review. *J Trauma Nurs.* (2013) 20:150–4. doi: 10.1097/JTN.0b013e3182a172b6

33. Townsend R, Clark R, Ramenofsky M, Diamond DL. ATLS-based videotape trauma resuscitation review: education and outcome. *J Trauma.* (1993) 34:133–8. doi: 10.1097/00005373-199301000-00025

34. Cheng A, Brown L, Duff J, Davidson J, Overly F, Tofil N, et al. Improving cardiopulmonary resuscitation with a CPR feedback device and refresher simulations (CPR cares study) a randomized clinical trial. *JAMA Pediatr.* (2015) 169:137–44. doi: 10.1001/jamapediatrics.2014.2616

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 22 February 2024
DOI 10.3389/fmed.2024.1339857

OPEN ACCESS

EDITED BY
Erich Brenner,
Innsbruck Medical University, Austria

REVIEWED BY
Florian Recker,
University of Bonn, Germany
Michael J. Wolyniak,
Hampden–Sydney College, United States
Ligia Garcia-Bejar,
Panamerican University, Mexico

*CORRESPONDENCE
Merel H. de Heer
✉ merelhdeheer@gmail.com

RECEIVED 16 November 2023
ACCEPTED 31 January 2024
PUBLISHED 22 February 2024

CITATION
de Heer MH, Driessen EW, Teunissen PW and Scheele F (2024) Lessons learned spanning 17 years of experience with three consecutive nationwide competency based medical education training plans.
Front. Med. 11:1339857.
doi: 10.3389/fmed.2024.1339857

Lessons learned spanning 17 years of experience with three consecutive nationwide competency based medical education training plans

Merel H. de Heer[1]*, Erik W. Driessen[2], Pim W. Teunissen[2,3] and Fedde Scheele[1,4]

[1]Amsterdam UMC Location Vrije Universiteit Amsterdam, Research in Education, Amsterdam, Netherlands, [2]School of Health Professions Education (SHE), Faculty of Health Medicine and Life Sciences (FHML), Maastricht University, Maastricht, Netherlands, [3]Department of Obstetrics and Gynecology, Maastricht University Medical Center (MUMC+), Maastricht, Netherlands, [4]Athena Institute, Faculty of Science, VU, Amsterdam, Netherlands

Introduction: Curricula for postgraduate medical education have transformed since the introduction of competency based medical education (CBME). Postgraduate training plans offer broader training with different competencies and an outcome-based approach, in addition to the medical technical aspects of training. However, CBME also has its challenges. Over the past years, critical views have been shared on the potential drawbacks of CBME, such as assessment burden and conflicts with practicality in the workplace. Recent studies identified a need for a better understanding of how the evolving concept of CBME has been translated to curriculum design and implemented in the practice of postgraduate training. The aim of this study was to describe the development of CBME translations to curriculum design, based on three consecutive postgraduate training programs spanning 17 years.

Method: We performed a document analysis of three consecutive Dutch gynecology and obstetrics training plans that were implemented in 2005, 2013, and 2021. We used template analysis to identify changes over time.

Results: Over time, CBME-based curriculum design changed in several domains. Assessment changed from a model with a focus on summative decision to one with an emphasis on formative, low-stakes assessments aimed at supporting learning. The training plans evolved in parallel to evolving educational insights, e.g., by placing increasing emphasis on personal development. The curricula focused on a competency-based concept by introducing training modules and personalized authorization based on feedback rather than on a set duration of internships. There was increasing freedom in personalized training trajectories in the training plans, together with increasing trust towards the resident.

Conclusion: The way CBME was translated into training plans has evolved in the course of 17 years of experience with CMBE-based education. The main areas of change were the structure of the training plans, which became increasingly open, the degree to which learning outcomes were mandatory or not, and the way these outcomes were assessed.

KEYWORDS

competency based medical education, postgraduate education, training plan, design, development, experience, implementation

1 Introduction

Competency based medical education (CBME) is an outcome based approach to education. The definition of CBME is highly variable in the literature (1). Frank et al. propose the following definition: CBME is an outcomes-based approach to the design, implementation, assessment, and evaluation of medical education programs, using an organizing framework of competencies (2). The outcome based approach is well aligned with workplace-based learning, such as postgraduate medical education, since all residents need to have a similar base of competence, mostly regarding medical accountability and patient safety. Already, in 1978 the WHO promoted the wide use of competency-based models in medical education (3). The popularity of CBME has increased ever since, as can be seen by increased implementation in postgraduate curricula (4) and the use of competency frameworks globally, such as the CanMEDS framework and ACGME Milestones (5, 6).

While the introduction of CBME has transformed training plans for postgraduate medical education, clinicians, educationalists and curriculum designers recognize the need for insight in how to use CBME (7, 8). There are few publications that share experiences and lessons learned from working with a CBME-based postgraduate training plan (9, 10). In this paper, we explore how the concept of CBME has evolved in training plans of postgraduate education for obstetrics and gynecology. We aim to identify lessons learned from 17 years of experience with CBME-based curriculum design in Dutch obstetrics and gynecology postgraduate education.

CBME has several characteristics, but the three core aspects of most CBME-based curricula are an outcome-based approach, workplace-based assessment and time-variable learning (7, 11–13). The outcome-based approach entails the translation of competency descriptions to practice in training plans. Workplace-based assessment covers how the residents' progress towards their intended outcomes is monitored and assessed (14). Competency-based time-variable learning focuses on attained competencies, independent of the time needed to achieve those competencies (15).

Although CBME has been implemented worldwide and is a popular educational model for medical education, the literature describes several challenges related to CBME in postgraduate education (8, 14, 16–22). The challenges mentioned in literature can be categorized in three main themes, which are elaborated below: *the practicality of CBME* (20–22), the *assessment burden* (8, 23) and also the *developments in the work field* (18, 24–26). *The practicality of CBME* covers the way the educational concept is used and implemented. Implementation in an existing health system is challenging (2, 27). The challenge is often debated as a balance between strict implementation and sufficient flexibility to ensure that the concept works in a local context. Strict implementation of CBME with a compulsory and fixed curriculum can serve the intended goal of CBME, but this compulsory and fixed curriculum appears to be not realistic in practice and not in line with implementation literature (7, 21, 28, 29). CBME can come across as rigid or strict (7, 21, 28). In addition, several authors described *the assessment burden* as a mental burden on residents due to the high frequency of observations and assessments. At the same time, the assessment burden is also described as a time-consuming practical burden for teachers because of the frequent moments of direct observation and the cumbersome administration of assessment outcomes (8, 23). Challenges related to *the developments in the work field* entail several aspects: training plans remain in need of change due to ongoing developments in health care systems, such as shared decision-making (18), but also due to educational insights and innovations resulting from ongoing research on the educational and clinical care outcomes of CBME (26, 30). Examples of those educational insights could be the growth mindset (ability can increase by focus, persistence and coaching) and mastery goal orientation (development of effective intrapersonal learning strategies for current-and lifelong learning). These educational approaches align with the tenets of CBME well, but also focus on competences reaching beyond the medical training itself (25, 26). Although it can be challenging to adjust set competencies and training plans, CBME should be continuously refined and informed by ongoing innovation and developments (18, 25).

CBME was first implemented in the Dutch obstetrics and gynecology training plan 17 years ago. In 2005 the first competency-based O&G training plan was launched, followed by a revision in 2013, and the most recent training plan revision for O&G in the Netherlands took place in 2021. Over those years, the design of the training plans evolved to cope with the previously mentioned challenges. The aim of this study was to describe the development of CBME translations to curriculum design, based on three consecutive postgraduate training plans spanning 17 years. By exploring how CBME has been used in successive O&G training plans, we aim to elucidate how CBME-based curriculum design has evolved and which lessons curriculum designers have learned. The outcomes of this study can be used to inform the development of future postgraduate training plans.

2 Methods

We conducted a qualitative document analysis, using three national Dutch obstetrics and gynecology (O&G) training plans spanning 17 years. We used template analysis, which is a form of thematic analysis (31). Thematic analysis is a method for analyzing qualitative data that allows for the identification of repeated patterns or themes across the dataset (32). Template analysis offers a description of a number of practical steps to be undertaken that starts with the description of with *a-priori* themes derived from pre-existing knowledge in literature (31).

2.1 Setting and data

This study was set in the Netherlands, where the O&G training plan covers 6 years of training. At least 2 years of the training take place in an academic hospital, which is the main base for the specialty training, and at least 2 years take place in a general hospital within the same training region. In total, there are 8 academic hospitals, which form the nucleus of a training region, and 48 general hospitals that participate in the postgraduate education of O&G residents, divided over the eight training regions.

O&G has a national training plan as well as local training plans based on competency-based medical education. The national training plan describes the content and structure of assessment, the intended learning outcomes, quality assurance, and implementation strategies

and provides guidelines for local training regions and hospitals. The national training plan also provides a theoretical background of the developments in the medical field and in educational concepts that are of influence on training plans. Local training plans effectuate the national training plan in the local context of the training hospital. For this study, we analyzed national training plans.

Until the year 2005, the training in Obstetrics and Gynecology in the Netherlands was semi-structured based on rotations over several clinical departments and a master-apprentice model. The higher numbers of trainees and the increasing complexity of the profession were the main reasons to develop a national training plan with a deliberate structure, both in content and in educational philosophy. This coincided with the advent of CBME and the publication of the first CanMEDS competency framework. The structure and clear outcome description offered by a CBME approach led the curriculum design committee to opt for this approach. The designers realized that although the most up to date educational knowledge was used for the training plans, experience gained by years of practice and developments in educational science, in the field of Obstetrics and Gynecology and in the health systems itself would urge a regular update.

Since the introduction of CBME, there have been 3 consecutive training plans (TP) for obstetrics and gynecology. These plans were published and subsequently implemented in 2005 (TP' 05), 2013 (TP '13), and 2021 (TP' 21).

TP '05 was the first competency-based Dutch national obstetrics and gynecology training plan. This training plan adopted the 1996 competency model of the Canadian Medical Education Directives for Specialists (CanMEDS).

TP '13, the second postgraduate training plan, contained a core training plan and newly introduced electives. The core training plan included all the general required learning outcomes of the specialty training and constituted the training program's core, which was equally compulsory for each resident. By contrast, the electives allowed for additional in-depth training in certain domains of obstetrics and gynecology. In addition, some forms of time-variable learning were implemented, which had become possible due to changes in legislation.

TP '21, the most recent training plan, introduced a new concept: in addition to a core training plan with electives, this new plan also featured themes, i.e., descriptions of topics in which residents can develop work-related competencies and pay attention to personal and professional development.

2.2 Data analysis

Data were analyzed between January and June 2023. The steps in this process, which are listed in Table 1, were based on the step guide for template analysis of King et al. (31). We started by defining *a priori* themes based on literature: CBME in practice, assessment, and time-based versus time-variable competency-based learning (11, 12). Next, the first author read the training plans, followed by initial coding. The first author shared the outcomes of initial coding with the research team for joint analysis. Based on this analysis, the team developed the template. If codes did not fit the *a priori* themes, we added new themes. Subsequently, the final coding process started, using iterative coding. The research team discussed the definitive template and codes. Data management and coding were performed with Atlas.ti.

TABLE 1 Steps in the analysis.

1. Definition of a priori themes
2. Familiarise oneself with the documents
3. Initial coding
4. Template development and final coding
5. Listing themes
6. Interpretation
7. Quality and reflexivity

2.3 Rigor and reflexivity

All the members of the research team are involved in medical education research. The team has expertise in education, medical practice and postgraduate training plan design. MH has clinical experience in obstetrics and gynecology. She performed the coding and used a logbook and reflective memos throughout the entire process of this study.

FS is a gynecologist and a professor of health systems innovation and education in Amsterdam. His area of expertise is health system innovation and education. He was involved in the development of the three training plans, and he is the author of several articles on the effects and difficulties of implementation strategies.

PT, also a gynecologist, has expertise in education development and medical education research and was involved in developing the 2021 training plan. As the chair of the revision committee, he used the growing body of research on work-based education and evaluations of CBME and work-based assessment as an inspiration for the revision of the O&G curriculum. As a researcher in the field, he critically examines how any educational intervention impacts practice and supports or, at times, hinders learning in and from practice. His involvement in the field as a gynecologist, a health professions education researcher and a curriculum designer shaped his interpretations of the data.

ED is an educationalist with expertise in medical education development and research. He has expertise in training plan development in several specialties. Since 2005, he has studied the implementation of CBME and has developed and led workplace-based faculty development sessions for his training region. The faculty development sessions made him aware of the frictions in CBME practice and of how trainees and teachers cope with these frictions. This background shaped ED's interpretations; he considered how each curriculum incorporated the theoretical developments and the experiences of the trainees and teachers.

3 Results

3.1 Themes

Our thematic analysis of the differences between the consecutive training plans revealed changes that could be summarized in three themes:

- Theme 1: from a fixed to an open training plan
- Theme 2: increasing degrees of freedom in personalized training trajectories
- Theme 3: assessment, from accountability to trust

The three themes will be described below, and we will use quotes from the three training plans to support the outcomes of the thematic analysis. The three main lessons learned from these themes are collected in Table 2. We also found some recurrent elements that were essential for all three training plans. We will present these recurrent elements at the end of the results section. The results are summarized in Table 3.

TABLE 2 The three main lessons learned.

From a fixed to an open training plan
1. In the Dutch context, a structured postgraduate training plan, with elaborate assessment and a fixed outcomes matrix developed towards a partially open training plan. Besides the fixed core of the specialty, there may be given room for both electives and themes with societal relevant themes to choose from. Not all learning outcomes of the specialty training were the same for every resident.
Increasing degrees in freedom in personalized training plans
2. There was an increase in the degree of freedom and personalization throughout the training plans. The balance shifted from a solid foundation with focus on the intended learning outcomes, towards room for individual development and attention for both professional and personal development. Providing opportunities for transformative learning.
Assessment, from accountability to trust
3. There can be a distinction in domains that need concrete assessment and domains that can be followed up differently. Parts of the training can also be guided with trust towards the residents, best effort obligation, self-reflection and a feedback culture.

3.2 Theme 1: from a fixed to an open training plan

In this section, we focus on the structure of the training plan. Our analysis showed a development from a training plan in 2005 that was fixed in terms of intended learning outcomes, assessment and competencies, to a training plan in 2021 that is partially open-ended. By a fixed versus an open training plan, we mean the following: in a fixed training plan, all the intended learning outcomes and the ways to reach those outcomes are set and the same for every resident. In an open training plan, not all the intended learning outcomes are fixed, they are not the same for every resident and the ways in which residents reach their end goals can differ as well.

3.2.1 A fixed training plan

TP '05 was the first plan that introduced a competency-based model into postgraduate medical education. It incorporated competencies and placed a strong emphasis on assessment and portfolios. The training plan comprised detailed instructions for how the plan should be executed in the training regions and hospitals. All the intended learning outcomes were the same for every resident. Competencies were elaborated into sub-competences and linked to

TABLE 3 Summary of results.

	2005	2013	2021
Theme 1 From a fixed to an open training plan	Fixed training plan ILOs* same for all Outcome-based education in line with CanMEDS terminology All ILO's captured in EPA's	Core training plan and electives ILOs may differ depending on electives, same within elective Outcome-based education in line with CanMEDS terminology All ILO's captured in EPA's	Core training plan, electives and themes Personalized ILOs for themes ILOs differ for electives Growth-and outcome-based education in line with workplace-based education concepts. ILOs covered by a combination of EPA's and themes.
Theme 2 Increasing degrees of freedom in personalized training trajectories	Area of interest: Residents could spend additional time in an area of interest, if the set ILO's were achieved	Choice of elective: 4 general years and 2 years of electives, in those 2 years there was freedom of choice within the elective options Emphasis on personal wellbeing	Choice of elective Freedom within themes: when, how and what within 5 developmental themes with personalized learning outcomes. Attention for professional and personal wellbeing and development
Theme 3 Assessment, from accountability to trust	Frequent assessment in various, mostly summative forms	Introduction of programmatic assessment combining assessment for formative and summative purposes	Discontinuation of target numbers in assessment Introduction of "best effort obligation" in themes with self-reflection Focus on trust in the learning process of the resident
Recurrent elements	workplace-based assessment	portfolio	faculty development	quality assurance	implementation strategy

*ILOs: intended learning outcomes.

entrustable professional activities (EPAs). The intended learning outcomes combined should result in a good gynecologist.

3.2.2 Keep up with expansion of medical knowledge

TP '13 placed greater emphasis on the implementation of competency-based learning and adjusted to changes in the medical field, such as the increasing expansion of medical knowledge. An important element was the introduction of electives in the training plan, in addition to a core training plan. While the core training plan, which covered the basis of the specialty training, was the same for every resident, the electives allowed residents to further deepen their knowledge in certain domains of the specialty (e.g., benign gynecology, oncology, and obstetrics). *"It has been established that developments within the field are happening so quickly that no one is able to keep up with the knowledge explosion that is taking place for all the electives. So there is a certain concentration (on the electives) after the broad foundation has been laid."* – TP '13. However, residents' training plans still had to fit within the clearly defined learning outcomes and assessment framework.

3.2.3 A more open training plan

TP '21 introduced a new concept based on an increasing body of research on how aspects of CBME and assessment work in practice. TP '21 continued the structure of a core curriculum combined with electives but with the addition of themes. TP '21 describes four themes that cover topics in which residents can develop their work-related competencies, knowledge and skills and pay attention to their personal and professional development: work-life balance and work enthusiasm, network health systems, organization of health, knowledge and innovation. *"Ongoing developments require healthcare providers to have an adaptive and creative view that extends beyond the boundaries of their own specialty and the hospital. One which increasingly uses insights from fields such as change management. … It means that it is essential that we learn together to deal with changes and disruptions in a more strategic and educated way."—TP'21.* This argument was one of the reasons underpinning the decision to create a partially open-ended training plan. The core training plan is still fixed in terms of the intended learning outcomes, but it is up to the residents to determine how they will achieve these outcomes and in which order. The training plan provides even more openness in structure in the electives and themes.

3.3 Theme 2: increasing degrees of freedom in personalized training trajectories

This second theme is the counterpart of the first theme, which focused on the changes in the structure of the training plans. With an increasing openness in the structure, the space for personalized learning trajectories increased as well. Each revision provided more freedom of choice regarding the focus of personal and professional development. TP '13 and TP '21 stipulated that residents should not all follow the same learning path. Although every resident should achieve similar core outcomes, both plans offered room for individualized development beyond the core objectives. Below, we will describe how this personalization has developed and how it was operationalized in the training plans.

3.3.1 Area of interest

TP '05 offered little room for personalization. Every resident would follow the same training plan with the same learning outcomes. There were no electives. However, there were some opportunities for personalization in learning trajectories. For example, residents could expand on an area of interest if they achieved the competencies more quickly than their peers. "*Based on the portfolio, the fast resident will be able to demonstrate that he or she has more time for the area of interest." – TP '05.* The area of interest could for example be a specific medical domain within O&G training or specific surgical skills. Furthermore, the training plan described fixed intended learning outcomes, but the resident and the local training institution could decide on the order in which the resident achieved the competencies of those intended learning outcomes.

3.3.2 Electives

TP '13 introduced electives in addition to a core training plan. With the increase in knowledge in the medical field and the increasing number of sub-specializations after specialty training, the training plan was adapted to the fact that it was no longer feasible for all residents to achieve the same intended learning outcomes. The former intended learning outcomes were revised and became partially optional in the form of electives. Residents could spend the time left after their core training on the electives of their choice. "*In addition to the core training plan of four years, TP '13 introduces a two-year elective period so that every gynecologist is broadly employable but can also be deployed in an area of interest. This has marked the beginning of a new phase" – TP '13* As well as providing room for electives, the training plan also offers room for personalization by paying attention to personal wellbeing. The plan indicates that during the annual reviews, program directors should not only evaluate the residents' professional performance, but also their personal well-being and their work-life balance, since these factors also influence professional functioning.

3.3.3 Development themes and electives

TP '21 explicitly describes personalization, in various forms: through electives, themes and attention to personal development. TP '21 preserved the electives and the core training plan of TP '13 and complemented them with four themes that focus on work-related competencies as well as on personal and professional development: work-life balance and work enthusiasm, network health systems, organization of health, and knowledge and innovation. The themes have no set outcomes; the residents are free to develop, albeit under best effort obligation, which means that they are expected to demonstrate their best efforts and growth within each of the themes. The development is monitored via self-reflection, progress conversations between residents and supervisors, and a portfolio. The authors of TP '21 state that developing a variety of skills and interests in combination with the core training plan will result in a good gynecologist: "*TP '21 is characterized by the confidence in residents and gynecologists that they carefully train future gynecologists… with the space to realize individualized training trajectories in addition to a solid foundation." TP '21.*

3.4 Theme 3: assessment: from accountability to trust

Our analysis of the three training plans showed a clear change in assessment strategy. TP '05 and TP '13 applied the pedagogical approach "assessment drives learning," which resulted in an assessment matrix with a high frequency of clinical feedback and skills assessments, clinical evaluations, knowledge exams, simulation exams and progress meetings, which were collected in a portfolio. By contrast, TP'21 applied a more development-oriented assessment approach, based on the assumption that it is not necessary to measure every developmental aspect. Rather than assessing whether a development took place, TP '21 introduced a combination of assessment and best effort obligation, with trust in the learning process. *"We do not see a competent gynecologist as the sum of ticked off competences."* – TP '21.

3.4.1 From numbers to narratives

TP '05 used frequent assessment in various, mainly summative, forms. "Target numbers can be useful to ensure sufficient experience for the resident. It is not easy to properly assess competency levels… it is safer not to abandon the principle of numbers, which guarantee experience."—TP '05. All the intended learning outcomes of the training plan were translated into EPAs, which were assessed throughout the training plan with various assessment forms and at different levels of independency. Each resident's progress was documented in a portfolio. It was thought that residents who had demonstrated their capacity to perform the required EPAs in these different forms of assessment would become good gynecologists. "The resident is entitled to sufficient participation in practical situations with professional coaching from the trainer. The trainer expects the resident to demonstrate, with the aid of EPAs and a portfolio, a growth of competences according to plan."—TP '05.

In the subsequent training plans, the assessments increasingly included narrative feedback in combination with attention to personal development. TP '13 maintained the assessment forms of TP '05 and extended those with narrative feedback options. For example, when trainers assessed an EPA, they could add narrative feedback to the level of independency. The same option was available for the clinical assessment of surgical skills or for the assessment of communication skills. Narrative feedback was collected in the portfolio, as were the other forms of assessment, such as the outcomes of knowledge exams and the independency levels achieved in various EPAs. The narrative feedback was used to monitor progress and could refer to different competency domains.

3.4.2 Towards a feedback culture

TP '21 stressed the importance of a feedback culture: "Feedback is one of the most important ways to help a resident learn. In order for a resident to effectively give and receive feedback, a supportive climate and feedback culture are required. This means that it should be considered normal to provide each other with explanations and feedback." – TP '21. The training plan was designed to find a balance between assessment of learning and assessment for learning. The plan placed less emphasis on summative assessment than the previous plans and reduced the frequency of assessment. The role of formative assessment in the form of narrative feedback was more pronounced.

The assessment of EPAs differed from the assessment of themes. EPAs were assessed at different independency levels and with specific criteria, whereas themes were evaluated through self-reflection based on best effort obligation and via narrative feedback in the portfolio. Since best effort obligation meant that there was no concrete assessment on the theme, learning outcomes were no longer all captured in EPAs. TP '21 introduced a distinction between criterion-based assessment and learning objectives with demonstrable development. Learning outcomes that need concrete assessment (for example, for the sake of patient safety) were criterion-based. "Some themes are mainly about providing guidance for development in areas for which we cannot establish universal end goals. Feedback on development within the themes will often require openness and self-reflection of the resident and the supervising trainer. To facilitate this conversation and create some form of transparency around the resident's development within the themes, the residents are asked to accept a best efforts obligation." TP '21.

3.5 Recurrent elements

Our data showed certain recurring elements, which were fundamental for the structure of the training plans and were developed alongside the content of the training plan.

3.5.1 Workplace-based assessment and portfolio

All training plans used workplace-based assessment, with a portfolio system to collect feedback and to show the residents' progress in earning EPAs and their development throughout the specialty training.

3.5.2 Emphasis on faculty development

TP '05 was the first training plan for residents that also introduced faculty development. The training plan stipulated that the staff of the hospitals should be trained in CBME through "teach the teacher" training. It also indicated that trainers should be trained in the theory of competency-based learning, in giving feedback, in assessing entrusted professional activities and in the use of the portfolio. The two subsequent training plans also strongly emphasized the importance of faculty development, which was described as a key factor for implementation in practice. *"Teach the teacher-training plans are the fundament for quality improvement of residency training"—TP '05.*

3.5.3 Quality assurance

All three training plans pay attention to quality assurance. They state that training regions should meet certain criteria of quality. Training centers are asked to perform their own quality assurance through a PDCA-cycle: plan-do-check-act, which should involve residents, trainers and other relevant stakeholders. TP '21 specifies that achieving improvements requires an open feedback culture and willingness to change and refers to the following quote concerning the need for an open feedback culture: *'Quality is cultivated by people learning together.'* (33)*– TP '21.*

3.5.4 Implementation strategy

All three training plans present an implementation strategy that indicates how the training plan could best be implemented in training

regions and hospitals. The implementation strategies included change management strategies, timelines, and planned visits to the training region's academic hospital and to the associated general hospitals. The strategies also included evaluations and collecting feedback from the work floor to improve the training plan and to improve the implementation. *"To implement the new training plan, it is not enough to publish this document. In order to succeed, there must be an actual implementation." TP '13.*

4 Discussion

In this study, we aimed to describe the development of CBME translations to curriculum design, based on three consecutive postgraduate training plans spanning 17 years. By exploring how CBME has been used in successive O&G training plans, we aimed to elucidate how CBME-based curriculum design has evolved and which lessons curriculum designers have learned.

The way CBME was translated into training plans has evolved in the course of 17 years of experience with CMBE-based education. The main areas of change were the structure of the training plans, which became increasingly open, the degree to which learning outcomes were mandatory or not, and the way these outcomes were assessed. The last training plan introduced certain learning outcomes that were not fixed and that could therefore be adapted very well to the local context and needs of training hospitals, as well as to the personal interests of the residents. These adaptations can take place without distracting from the core of CBME, which is that residents develop the competencies required for their profession while they work in practice. The residents can grow towards the point where they can be entrusted to perform professional activities. In addition, they can also acquire the competencies needed to cope with future developments.

4.1 Time-variable learning

While time-variable learning is often presented as one of the core components of CBME (12, 34), recent studies noted that time-variability is in conflict with practice, due to the high burden that can be imposed on the organization and the workforce (24). This burden is described as logistic chaos (2), calling into question how the theory can be reconciled with practice in both the educational frameworks and the existing healthcare systems (20, 24). In our analysis of the three training plans, time-variability did not emerge as a separate theme, which is in contrast to the outcomes of our literature search and our initial template. Instead, an important theme that emerged was room for personalization. Time-variable learning is compatible with a fixed training plan; the intended learning outcomes should be reached during the course of residency, but if they have not been reached at a certain moment, more time can be granted. A partially open training plan makes fully time-variable learning less relevant and provides room for personalization in achieving competence: in addition to the compulsory learning outcomes of the training plan, there are also preference-related learning opportunities. Fully time-variable learning could lead to an erosion of learning opportunities if learners were to move on to new departments once they obtained the compulsory learning targets because in that case, other, non-compulsory learning opportunities could disappear (35).

4.2 Placing our lessons in context: trialability, the assessment burden and transformative learning

Our analysis of the differences between the three consecutive training plans demonstrated the advancing insights into the capacity of a CBME-based curriculum to accommodate the challenges of CBME that were described in the introduction: the practicality of CBME (20, 21), the assessment burden (23) and the developments in the work field (18, 24). Figure 1 provides insight in the interpretation and extrapolation of trends of the evolving CBME landscape in postgraduate medical.

4.2.1 The practicality of CBME: trialability

Challenges regarding the practicality of CBME are often debated in terms of a balance between the strict implementation of CBME within a set framework and sufficient flexibility for a local context. Strict implementation is often considered to serve the intended goal of CBME, but a strict framework appears to be not realistic in practice or not in line with implementation literature (7, 21, 28, 29). As a strength, authors state that the potential of CBME lies in its clarity about requirements and in strictness of implementation (36, 37). However, CBME is an innovation that requires involvement and willingness of trainers, legislation adaptations and an augmented workforce (2, 38, 39). While different contexts may require a different balance between strictness and flexibility, our analysis of the way in which the three training plans were redesigned, showed that the concept of CBME is suitable for trialability, i.e., the degree to which an innovation may be experimented with on a limited basis (40, p. 16). According to Rogers, trialability is one of the most important preconditions for successful implementation of innovations. Trialability is positively correlated with the rate of adoption (40). In the implementation stage of the innovation-decision process, reinvention may occur during the trial of the innovation. Then, the innovation may be changed or modified by the potential adopter.

Dagnone et al. stated that CBME should be implemented under tactical guidance from CBME leaders, with committed stakeholder investment at local levels (41). On the one hand, the theory and the experts should be leading the way. However, in practice, there should be room for local diversity, shared leadership, re-invention and development to implement this complex change in medical education (41). Throughout our analysis of the three training plans, trialability was evident, including evaluations from the work floor. The consecutive training plans incorporated lessons learned from practice: maintaining the prior strengths while trying to overcome previous challenges and to implement new insights.

4.2.2 The assessment burden

The assessment of CMBE has been described as a burden on residents and teachers alike. For residents, it constituted a mental and practical burden due to the high frequency of observations and assessments, which teachers experienced as a time-consuming burden because of the frequent moments of direct observation and the cumbersome administration of assessment outcomes (23). A recent scoping review on CBME discourses describes the workload of assessment a challenge as well (8). They added uncertainty of the uniformity of quality of assessment to this conversation. TP '21 distinguished between learning outcomes that must be formally assessed (also for the sake of patient safety) and learning outcomes that can be evaluated by other forms of monitoring residents' progress. In

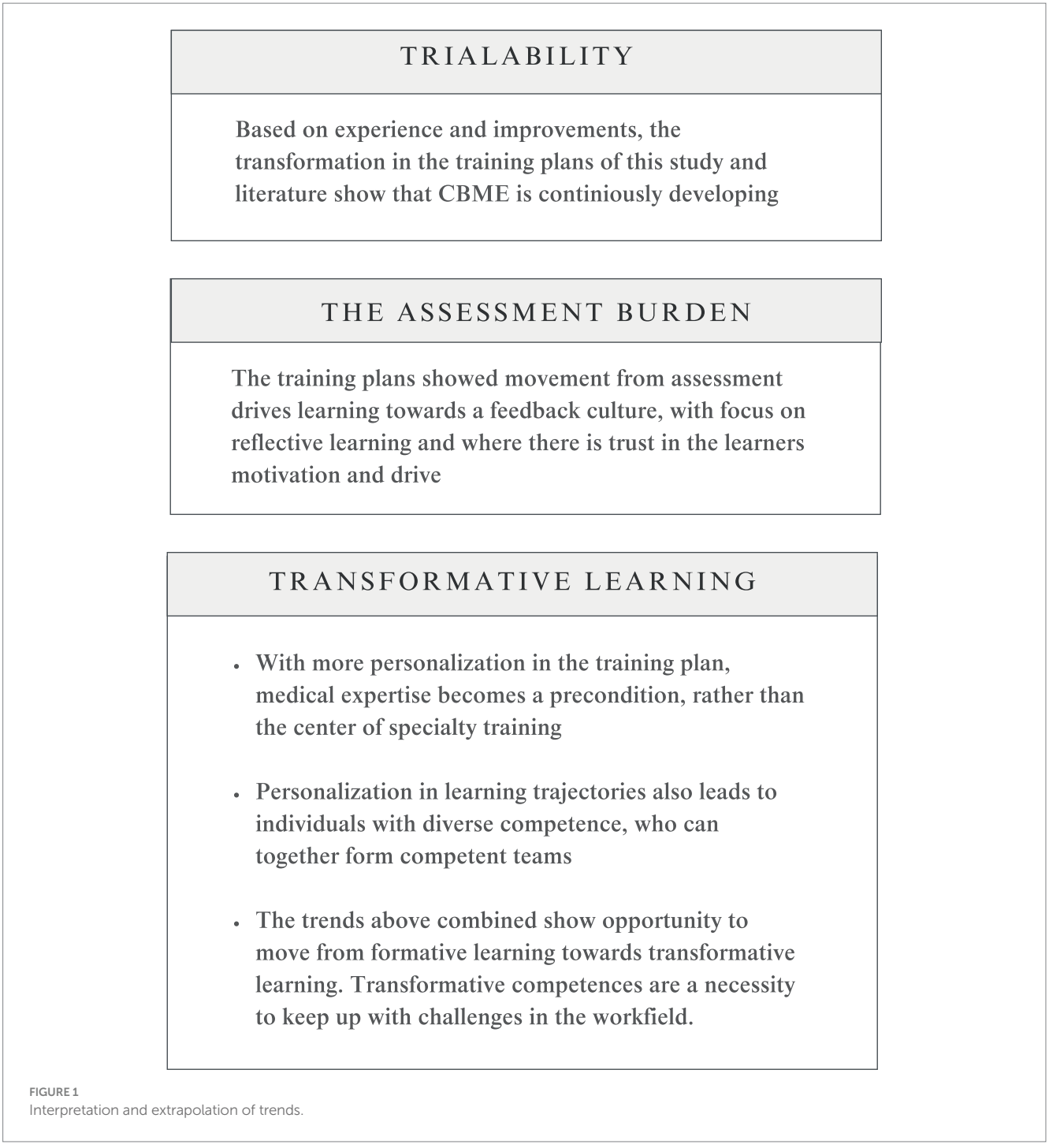


FIGURE 1
Interpretation and extrapolation of trends.

addition to reducing the frequency of observations and assessments, the '21 program describes a feedback culture: opening the dialogue between residents and faculty so that they can both learn from each other in a safe learning environment (42). This culture can foster trust in the learning process of the residents, allowing learning outcomes in certain domains to be monitored through reflection and narrative feedback.

4.2.3 The developments in the workfield: transformative learning

The three training plans stressed the importance of incorporating new educational insights, changes in the medical field and innovations in healthcare systems. This finding is in line with international literature (7, 16, 24, 43) and with contemporary educational insights regarding the importance of transformative learning (25). Our analysis showed that opportunities for transformative learning were provided by the partially open training plans of TP '13 and TP '21 and their increasing emphasis on narrative feedback and self-reflection. Transformative learning is the expansion of consciousness which enables individuals to question their own feelings, beliefs, and assumptions, and their perspective (44). Mezirow (44) and Frenk (43) have shown the importance of transformative learning for critical reflection and

for developing new perspectives, which are vital in achieving changes such as innovations in health care systems. A special interest group of the Dutch federation for medical education investigated the modernization of postgraduate medical education and they concluded that transformative learning is still in its early stages in postgraduate medical education, despite its importance in health care transitions (45).

The next step could be to actively implement transformative learning in training plans. The training of professionals and the education of 'enlightened change agents' for transformation in health care can become possible with transformative learning, more governance and support from academic leaders with a broader perspective on the future of health care (45). Further research is needed to gather the international perspectives of health professionals and educationalists on healthcare education and to integrate them in the practical implementation of transformative learning.

4.3 Limitations and future research

One of the limitations of this study is that it focuses on the CBME training plans developed in a single nation and for a single specialty. In different countries and regions, legislation and health care systems may be different, which may strongly influence the design of training plans. In this study we did not aim to generate generalizable results. However, despite the previously mentioned limitations, the shared experience and the developments in the training plans have potential value for curriculum designers in other medical specialties and in other countries as well. Our findings may transcend our specific research context in both CBME content that is used in similar ways elsewhere, as well as the link to the mentioned challenges, that are not unique to O&G or the Netherlands. Therefore, we expect our results to be helpful to other contexts, for example in the (re) design of a national curriculum.

Future research might analyze training plans designed in a different country or for different specialties. More data should become available on practical experiences and developments over time since CBME is used increasingly around the world and in different specialties. The practical experiences could focus on the outcomes and effects of CBME training plans on learners, patients, and faculty to bridge the gap between theoretical adaptations and real-world impact. We would also recommend investigating the experiences of learners within CBME frameworks, which could contribute to a more holistic understanding of the impact of these educational approaches on medical professionals (in training).

Future research should identify best practices for the implementation of CBME (9). Sharing experiences and lessons learned over time could improve the implementation and the practicality of CBME. We would also recommend researching the outcomes and learning effects of the partially open structure and the assessment approach used in TP '21.

In this paper, we did not investigate the enactment of the training plans nor the effects of the three CBME training plans on learners, patients and faculty. A previous study indicated that the effects of CBME on learners have not yet been explored (9). Future research could focus on learner experiences to underpin the theoretical foundation of CBME and to evaluate recent training plan innovations.

5 Conclusion

Our analysis of three Dutch training plans showed that the training plan designers chose to re-adjust the training plans with attention to alignment of theory and practice. In doing so, they balanced a solid foundation in a core training plan with a partially open training plan, with room for professional and personal development, expanded with themes that provide opportunities for acquiring competencies beyond the scope of current medical content itself.

To utilize the full potential of CBME while reducing the assessment burden and conflicts with practice, curricula could place greater emphasis on trust towards residents, creating learning opportunities for new competences.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

MH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. ED: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft. PT: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft. FS: Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer FR is currently organizing a Research Topic with the author FS.

Publisher's note

References

1. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. (2010) 32:631–7. doi: 10.3109/0142159X.2010.500898

2. Frank JR, Snell LS, Cate OT, Holmboe ES, Carraccio C, Swing SR, et al. Competency-based medical education: theory to practice. *Med Teach*. (2010) 32:638–45. doi: 10.3109/0142159X.2010.501190

3. McGaghie WC, Miller GE, Sajid AW, Telder TV. *Competency-based curriculum development in medical education–WHO*. Geneva: World Health Organization (1978).

4. Cate OT. Competency-based postgraduate medical education: past, present and future. *GMS J Med Educ*. (2017) 34:Doc69. doi: 10.3205/zma001146

5. Eno C, Correa R, Stewart NH, Lim J, Westerman ME, Holmboe ES, et al. *Milestones guidebook for residents and fellows*. USA: Accreditation Council for Graduate Medical Education (ACGME) (2020).

6. Frank JR, Snell L, Sherbino J. *Can MEDS 2015 physician competency framework* Ottawa: Royal College of Physicians and Surgeons of Canada (2015).

7. Chaney KP, Hodgson JL. Using the five Core components of competency-based medical education to support implementation of CBVE. *Front Vet Sci*. (2021) 8:689356. doi: 10.3389/fvets.2021.689356

8. Hamza DM, Hauer KE, Oswald A, van Melle E, Ladak Z, Zuna I, et al. Making sense of competency-based medical education (CBME) literary conversations: a BEME scoping review: BEME guide no. 78. *Med Teach*. (2023) 45:802–15. doi: 10.1080/0142159X.2023.2168525

9. Gruppen L, Frank JR, Lockyer J, Ross S, Bould MD, Harris P, et al. Toward a research agenda for competency-based medical education. *Med Teach*. (2017) 39:623–30. doi: 10.1080/0142159X.2017.1315065

10. Danilovich N, Kitto S, Price DW, Campbell C, Hodgson A, Hendry P. Implementing competency-based medical education in family medicine: a narrative review of current trends in assessment. *Fam Med*. (2021) 53:9–22. doi: 10.22454/FamMed.2021.453158

11. Spady W. Competency based education: a bandwagon in search of a definition. *Am Educ Res Ass*. (1977) 6:9–14. doi: 10.3102/0013189X006001009

12. Van Melle E, Frank JR, Holmboe ES, Dagnone D, Stockley D, Sherbino J, et al. A Core components framework for evaluating implementation of competency-based medical education programs. *Acad Med*. (2019) 94:1002–9. doi: 10.1097/ACM.0000000000002743

13. Carol Carraccio SW, Englander R, Ferentz K, Martin C. Shifting paradigms: from Flexner to competencies. *Acad Med*. (2002) 77:361–7. doi: 10.1097/00001888-200205000-00003

14. Scheele F, Teunissen P, Van Luijk S, Heineman E, Fluit L, Mulder H, et al. Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach*. (2008) 30:248–53. doi: 10.1080/01421590801993022

15. Ten Cate O, Gruppen LD, Kogan JR, Lingard LA, Teunissen PW. Time-variable training in medicine: theoretical considerations. *Acad Med*. (2018) 93:S6–S11 (3S Competency-Based, Time-Variable Education in the Health Professions). doi: 10.1097/ACM.0000000000002065

16. van der Aa JE, Aabakke AJM, Ristorp Andersen B, Settnes A, Hornnes P, Teunissen PW, et al. From prescription to guidance: a European framework for generic competencies. *Adv Health Sci Educ Theory Pract*. (2020) 25:173–87. doi: 10.1007/s10459-019-09910-8

17. Harris KA, Nousiainen MT, Reznick R. Competency-based resident education-the Canadian perspective. *Surgery*. (2020) 167:681–4. doi: 10.1016/j.surg.2019.06.033

18. Holmboe ES. The transformational path ahead: competency-based medical education in family medicine. *Fam Med*. (2021) 53:583–9. doi: 10.22454/FamMed.2021.296914

19. Torralba KD, Jose D, Katz JD. Competency-based medical education for the clinician-educator: the coming of milestones version 2. *Clin Rheumatol*. (2020) 39:1719–23. doi: 10.1007/s10067-020-04942-7

20. Cate OT, Scheele F. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. (2007) 82:542–7. doi: 10.1097/ACM.0b013e31805559c7

21. Boyd VA, Whitehead CR, Thille P, Ginsburg S, Brydges R, Kuper A. Competency-based medical education: the discourse of infallibility. *Med Educ*. (2018) 52:45–57. doi: 10.1111/medu.13467

22. Shrivastava SR, Shrivastava PS. Qualitative study to identify the perception and challenges faced by the faculty of community medicine in the implementation of competency-based medical education for postgraduate students. *Fam Med Community Health*. (2019) 7:e000043. doi: 10.1136/fmch-2018-000043

23. Szulewski A, Braund H, Dagnone DJ, McEwen L, Dalgarno N, Schultz KW, et al. The assessment burden in competency-based medical education: how programs are adapting. *Acad Med*. (2023) 98:1261–7. doi: 10.1097/ACM.0000000000005305

24. van Rossum TR, Scheele F, Sluiter HE, Paternotte E, Heyligers IC. Effects of implementing time-variable postgraduate training programmes on the organization of teaching hospital departments. *Med Teach*. (2018) 40:1036–41. doi: 10.1080/0142159X.2017.1418850

25. Ross S, Pirraglia C, Aquilina AM, Zulla R. Effective competency-based medical education requires learning environments that promote a mastery goal orientation: a narrative review. *Med Teach*. (2022) 44:527–34. doi: 10.1080/0142159X.2021.2004307

26. Richardson D, Kinnear B, Hauer KE, Turner TL, Warm EJ, Hall AK, et al. Growth mindset in competency-based medical education. *Med Teach*. (2021) 43:751–7. doi: 10.1080/0142159X.2021.1928036

27. Dauphinee WD, Boulet JR, Norcini JJ. Considerations that will determine if competency-based assessment is a sustainable innovation. *Adv Health Sci Educ Theory Pract*. (2019) 24:413–21. doi: 10.1007/s10459-018-9833-2

28. Van Melle E, Hall AK, Schumacher DJ, Kinnear B, Gruppen L, Thoma B, et al. Capturing outcomes of competency-based medical education: the call and the challenge. *Med Teach*. (2021) 43:794–800. doi: 10.1080/0142159X.2021.1925640

29. Wentzell DD, Chung H, Hanson C, Gooi P. Competency-based medical education in ophthalmology residency training: a review. *Can J Ophthalmol*. (2020) 55:12–9. doi: 10.1016/j.jcjo.2019.07.004

30. Adreak N, Srivaratharajah K, Mullen KA, Pike A, Mackay MH, Comber L, et al. Incorporating a Women's cardiovascular health curriculum into medical education. *CJC Open*. (2021) 3:S187–91. doi: 10.1016/j.cjco.2021.09.020

31. Brooks J. The utility of template analysis in qualitative psychology research. *Qual Res Psychol*. (2015) 12:202–22. doi: 10.1080/14780887.2014.955224

32. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE guide no. 131. *Med Teach*. (2020) 42:846–54. doi: 10.1080/0142159X.2020.1755030

33. Koksma J. J., Kremer J. A. Beyond the quality illusion: The learning era. *Academic Medicine*. (2019) 94:166–169.

34. Sonnadara RR, Mui C, McQueen S, Mironova P, Nousiainen M, Safir O, et al. Reflections on competency-based education and training for surgical residents. *J Surg Educ*. (2014) 71:151–8. doi: 10.1016/j.jsurg.2013.06.020

35. Teunissen PW, Kogan JR, Ten Cate O, Gruppen LD, Lingard LA. Learning in practice: a valuation of context in time-variable medical training. *Acad Med*. (2018) 93:S22–S6 (3S Competency-Based, Time-Variable Education in the Health Professions). doi: 10.1097/ACM.0000000000002070

36. Frank JR, Snell LS, Oswald A, Hauer KEInternational CC. Further on the journey in a complex adaptive system: elaborating CBME. *Med Teach*. (2021) 43:734–6. doi: 10.1080/0142159X.2021.1931083

37. Frenk J, Chen CL, Michaud C. Chapter 1, transformative learning in health education for a new century: Interdependence in the education of health professionals. Eds. J. Frenk, L. C. Chen and C. Michaud (2019).

38. Ten Cate O, Billett S. Competency-based medical education: origins, perspectives and potentialities. *Med Educ*. (2014) 48:325–32. doi: 10.1111/medu.12355

39. Cooney R, Chan T, Gottlieb M, Abraham M, Alden S, Mongelluzzo J, et al. Academic primer series: key papers about competency-based medical education. *Western J Emerg Med*. (2017) 18:713–20. doi: 10.5811/westjem.2017.3.33409

40. Rogers E. *Diffusion of innovations*. New York: The Free Press-Division of Simon and Schuster (1995).

41. Dagnone JD, Chan MK, Meschino D, Bandiera G, den Rooyen C, Matlow A, et al. Living in a world of change: bridging the gap from competency-based medical education theory to practice in Canada. *Acad Med*. (2020) 95:1643–6. doi: 10.1097/ACM.0000000000003216

42. Malik RF, Buljac-Samardzic M, Amajjar I, Hilders C, Scheele F. Open organisational culture: what does it entail? Healthcare stakeholders reaching consensus by means of a Delphi technique. *BMJ Open*. (2021) 11:e045515. doi: 10.1136/bmjopen-2020-045515

43. Frenk J, Chen L, Bhutta ZA, Cohen J, Crisp N, Evans T, et al. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *Lancet*. (2010) 376:1923–58. doi: 10.1016/S0140-6736(10)61854-5

44. Mezirow J. Transformative learning: theory to practice. New directions for adult and. *Contin Educ*. (1997) 5–12. doi: 10.1002/ace.7401

45. Scheele F, Van Luijk S, Mulder H, Baane C, Rooyen CD, De Hoog M, et al. Is the modernisation of postgraduate medical training in the Netherlands successful? Views of the NVMO special interest group on postgraduate medical education. *Med Teach*. (2014) 36:116–20. doi: 10.3109/0142159X.2013.849333

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 19 March 2024
DOI 10.3389/fmed.2024.1371075

OPEN ACCESS

EDITED BY
Fedde Scheele,
VU Amsterdam, Netherlands

REVIEWED BY
Bhavani Veasuvalingam,
International Medical University, Malaysia
Marja Dijksterhuis,
Amphia Ziekenhuis, Netherlands

*CORRESPONDENCE
Fionnuala M. McAuliffe
✉ fionnuala.mcauliffe@ucd.ie

RECEIVED 15 January 2024
ACCEPTED 07 March 2024
PUBLISHED 19 March 2024

CITATION
Dunlop K, Dillon G, McEvoy A, Kane D, Higgins S, Mangina E and McAuliffe FM (2024) The virtual reality classroom: a randomized control trial of medical student knowledge of postpartum hemorrhage emergency management.
Front. Med. 11:1371075.
doi: 10.3389/fmed.2024.1371075

The virtual reality classroom: a randomized control trial of medical student knowledge of postpartum hemorrhage emergency management

Kristyn Dunlop[1], Grace Dillon[1], Aoife McEvoy[1], Daniel Kane[1,2], Shane Higgins[1], Eleni Mangina[3] and Fionnuala M. McAuliffe[1]*

[1]UCD Perinatal Research Centre, University College Dublin, The National Maternity Hospital, Dublin, Ireland, [2]Department of Obstetrics & Gynecology, Royal College of Surgeons in Ireland, Rotunda Hospital, Dublin, Ireland, [3]School of Computer Science, University College Dublin, Dublin, Ireland

Objective: To investigate the use of a virtual reality learning environment (VRLE) to enhance medical student knowledge of postpartum hemorrhage (PPH) emergency management and insertion of a postpartum balloon.

Methods: A randomized control trial involving medical students from University College Dublin, Ireland. Participants were randomly allocated to the intervention group (VRLE tutorial) or control group (PowerPoint tutorial on the same topic). All participants completed pre-learning experience and post-learning experience surveys. Both groups were timed and assessed on postpartum balloon insertion technique on a model pelvis. The primary outcome was assessment of student knowledge. Secondary outcomes included confidence levels, time taken to complete the task, technique assessment, satisfaction with the learning environment, and side effects of VR.

Results: Both learning experiences significantly ($p < 0.001$) enhanced student performance on the post-learning experience multiple choice questionnaire, with no difference between the intervention and control groups. In the intervention group, time for task completion was significantly less compared to the control group (1–2 min vs. 2–3 min, $p = 0.039$). Both learning experiences significantly ($p < 0.001$) enhanced student confidence, with no significant difference between intervention and control groups. 100% of the students using the VRLE enjoyed the experience, and 82.4% were very likely to recommend use of VRLE in medical education. 94.1% of the students felt the VRLE was beneficial over didactic teaching.

Conclusion: Receiving formal instruction, regardless of format, enhances students' knowledge and confidence of the topic covered. Students who received instruction via the VRLE assembled the postpartum balloon faster than students who received didactic teaching. VR may be beneficial in teaching hands-on procedural skills in obstetrics and gynecology education.

KEYWORDS

medical education, obstetrics and gynecology, virtual reality learning environment, postpartum hemorrhage, balloon tamponade

1 Introduction

Postpartum hemorrhage (PPH) remains a leading cause of maternal mortality and morbidity worldwide (1). Prompt recognition and initiation of emergency management and resuscitation is crucial to reduce morbidity (1). The most common cause of PPH is uterine atony (1), and where uterotonic medications fail, balloon tamponade is a simple, effective, and potentially life-saving measure (2). Therefore, PPH emergency management is a critical concept in undergraduate obstetrics and gynecology teaching in order to prepare our future physicians for clinical practice.

Unfortunately, due to the sensitive nature of many aspects of clinical obstetrics and gynecology, medical students often find it difficult to get hands-on clinical experience in the specialty (3). Additionally, the centre in which students attend their clinical placements may have varying birth rates, making PPH a potentially rare event for students to encounter during the course of their clinical placement (4). During emergency situations, medical students are often silent observers, reducing their exposure to critical aspects of PPH emergency management (3). Therefore, simulated clinical environments are increasingly part of the medical school curriculum, in order to foster development of essential skills in prompt resuscitation, emergency management and communication in advance of clinical practice (5). Simulated clinical environments offer an opportunity to gain knowledge and confidence in emergency management skills, with the ultimate goal of preparing participants for prompt recognition and management of PPH, in order to improve patient outcomes (4). Additionally, simulated clinical environments provide students with an opportunity to practice hands-on skills which may not be commonly encountered (3), such as insertion of a postpartum balloon for uterine tamponade. These simulated clinical environments offer students a safe place to make mistakes, without risking patient safety in the clinical environment (4).

As technology has advanced and educational technology has become more accessible, virtual reality (VR) is increasingly being used to improve the educational experience (6). VR technology creates an immersive environment in which the participant can explore and manipulate multimedia sensory environments in real-time (7). VR has the potential to enhance the simulated clinical environment by creating a real-life multisensory clinical learning environment. Learning through simulation and VR utilises the constructivist educational theory, whereby learners construct knowledge through their interaction with the learning environment, rather than passively taking in information (8). VR learning tools are often self-directed, and promote active engagement of the student to navigate their own learning experience, thus resulting in the student forming their own knowledge through a self-regulated process (9). Additionally, VR has the benefit of supporting knowledge acquisition for procedural skills, where the user can repeat procedural steps hands-on through the VR headset as many times as required for the individual learner to feel confident with the procedure (10), facilitating practice without the risk of patient harm (11).

Therefore, the objective of our study was to assess the value of using VR in simulating training for medical students in management of PPH and insertion of a postpartum balloon for uterine tamponade. The topic of PPH was chosen as it is an emergency situation, and a critical concept to understand at the undergraduate level. Additionally, we were able to consolidate generalizable resuscitation skills, and introduce the procedural component of a postpartum balloon for uterine tamponade, which students' would rarely see during their clinical placements. We hypothesized that VR would enhance the student learning experience and improve the insertion technique of the postpartum balloon, compared with traditional didactic teaching methods.

2 Materials and methods

A randomized control trial (RCT) of students in the clinical years of the undergraduate (6-year program) and graduate entry (4-year program) medical degree programs at University College Dublin (UCD) was conducted over 5 days from October 15–19, 2023. Graduate entry medical students are those who have completed a primary degree, and are now undertaking medicine as a second degree. Ethical approval for this study was obtained from the University College Dublin Research Ethics Committee.

2.1 Participants

Similar to previous successful studies conducted at the UCD Perinatal Research Centre, medical students were invited to take part in the study via class announcements on the university's e-learning platform (12, 13), and announcements made during clinical placement. The announcements gave a brief overview of the study, without giving specific details of the content to be covered. Interested students contacted the study team via email, and an information leaflet and consent form was sent to each student for review. All students in the clinical years of study at UCD were eligible for participation, except those younger than 18 years of age, or with a medical condition including cardiac (e.g., pacemakers), binocular vision abnormalities, psychiatric disorders or epilepsy. Informed written consent from all participants was obtained on arrival to the teaching space. Participants were advised they could withdraw consent to participate at any time.

2.2 Randomization

Figure 1 depicts the randomization and allocation process for the study. Participants were randomly allocated to intervention or control groups through the use of sequentially numbered, opaque, sealed envelopes. The researcher conducting the study was not blinded to the group allocation given the nature of the study design, however, allocation was prospectively concealed to the students.

2.3 Intervention and control groups

The intervention learning experience involved an immersive VR tutorial lasting for fifteen minutes, using an Oculus Lens-2 VR head-mounted display, designed to teach background material regarding PPH management and step-wise procedural instruction of postpartum balloon insertion. The VR learning environment (VRLE) was designed in collaboration with the UCD School of Computer Science and Magos, a company based in Athens, Greece, specializing in creation of VR programmes. Students assigned to the intervention group used the Oculus Headsets to participate in an interactive tutorial on PPH background, emergency management, and postpartum balloon

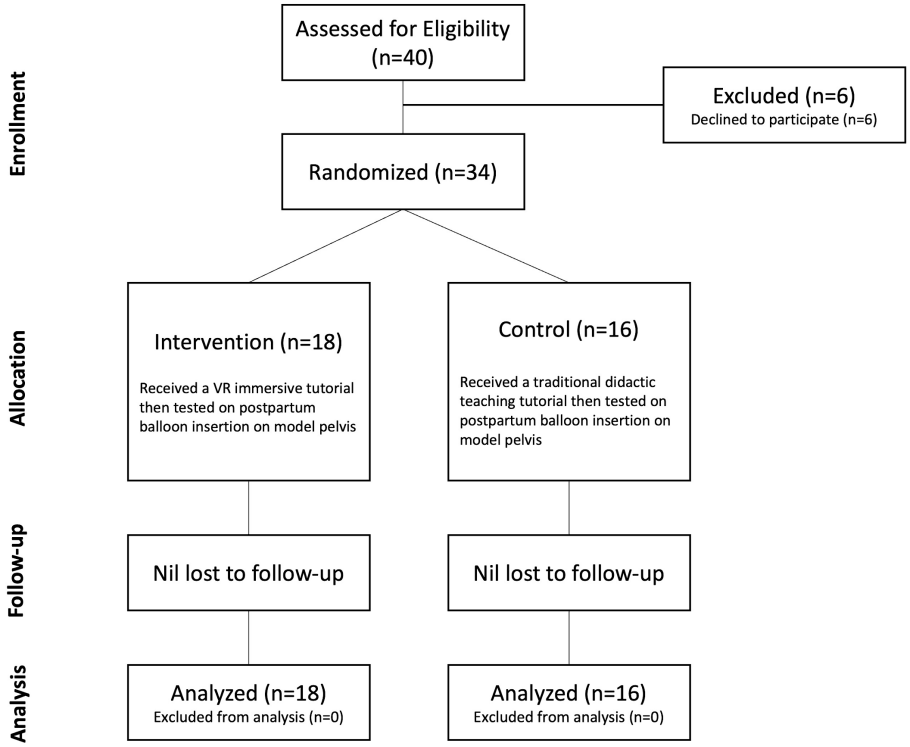


FIGURE 1
CONSORT (Consolidated Standards of Reporting Trials) diagram depicting the randomization and allocation process. VR – virtual reality.

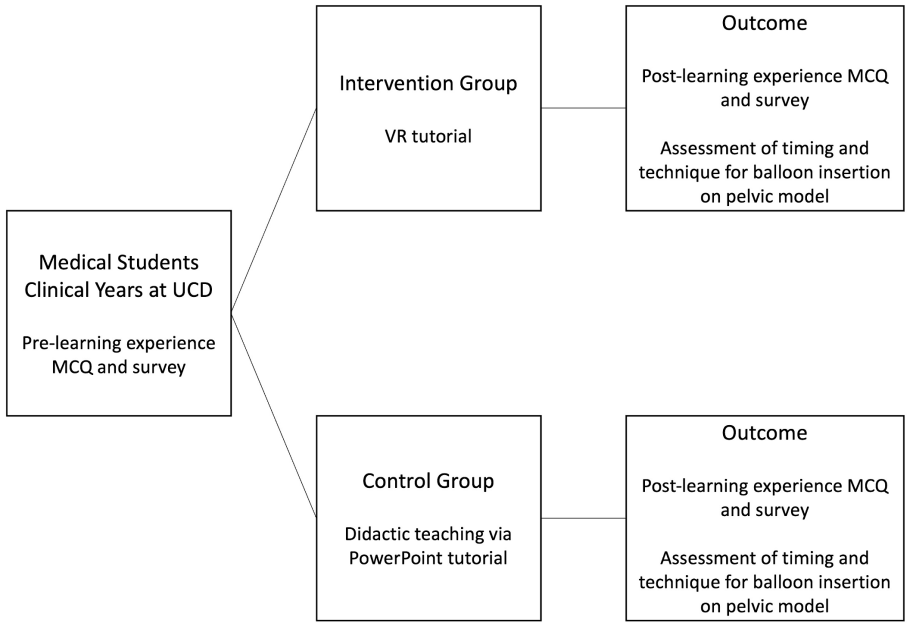


FIGURE 2
Study design. Pictorial representation of the study design and outcomes. UCD – University College Dublin, MCQ – multiple choice questionnaire, VR – virtual reality.

insertion (Figure 2). Participants then carried out postpartum balloon insertion on a model pelvis, and were marked on insertion technique and time taken to complete insertion. All participants completed pre-learning experience and post-learning experience questionnaires, including multiple-choice questions (MCQ) regarding PPH knowledge and management, prior experience of PPH and VR, satisfaction with the learning experience, and attitudes towards use of VR in medical education.

The control group underwent a traditional didactic learning experience consisting of PowerPoint presentation (Figure 2). The factual content of the traditional tutorial replicated the VRLE tutorial. The traditional tutorial was conducted by the same clinical tutor for the purpose of consistency in the teaching, and lasted for fifteen minutes. Once students had completed the didactic learning experience, they were asked to carry out postpartum balloon insertion on a model pelvis, and were marked on insertion technique and time taken to complete insertion. All participants completed pre-learning experience and post-learning experience questionnaires, including multiple-choice questions regarding PPH knowledge and management, prior experience of PPH and VR, satisfaction with the learning experience and attitudes towards the use of VR in medical education.

2.4 Outcomes

The primary outcome was student knowledge regarding PPH management, as measured by MCQ scores pre-learning experience compared to MCQ scores post-learning experience. Pre-specified secondary outcomes included time taken to complete the task, technique for postpartum balloon insertion, improvement in confidence levels, virtual reality-side effects and satisfaction with the learning experience. Each participant underwent objective, timed assessment of balloon insertion technique on the model pelvis to assess if there was a difference for either learning experience. Confidence levels of managing PPH and postpartum balloon insertion were assessed pre- and post-learning experience. The intervention group was asked to complete a questionnaire regarding side-effects experienced while using the VR headsets. All students were asked a series of questions regarding their attitudes towards the use of VR in medical education and satisfaction with the learning experience.

2.5 Statistical analysis

Data collected was entered and coded into Microsoft Excel. Statistical analysis was performed using SPSS software package (version 27; SPSS, Chicago, Illinois). Descriptive statistics were calculated as frequency and percentage for categorical data or mean and standard deviation for normally distributed variables. The paired t-test was used to compare mean scores for the MCQ and confidence levels pre- and post-learning experience. Independent t-test was used to compare the technique assessment between the intervention and control groups. Chi-square test of independence was used to assess the difference between categorical variables, including timing of insertion of the postpartum balloon on the model pelvis, as the timings were grouped categorically in increments of one minute on the assessment survey. Statistical significance was defined as a p-value <0.05.

3 Results

A total of 40 medical students expressed interest in participating in the study, and were screened for eligibility – 6 students ultimately declined to participate. 34 students were randomly allocated, 18 to the intervention (VRLE) group, and 16 to the control group receiving didactic teaching only (Figure 1). There were no participants lost to follow up and none excluded from the analysis.

Table 1 outlines the background demographics of the study participants. In the total cohort, 25 participants (73.5%) were female, and 9 participants (26.5%) were male. The majority of participants (52.9%) were between the ages of 18 and 24 years, with 38.2% of participants between the ages of 25 and 34 years. 19 participants were from the undergraduate medicine course, of which 13 (38.2% of the total cohort) were in the final year of the program. 15 participants were from the graduate medicine course, of which 13 (38.2% of the total cohort) were in the final year of the program. 61.8% of the total cohort had no prior experience using VR at the time of the study, with only 1 participant having used VR 15–20 times in the past. In terms of familiarity with PPH, 58.8% of the participants had no prior knowledge or experience of PPH, and 41.2% of the participants had only theoretical knowledge of PPH. Similar to the demographics of the overall cohort, there were significantly ($p = 0.038$) more females in the intervention group (Table 1). The remaining participant demographics were not significantly different between intervention and control groups.

With regards to the primary outcome testing student knowledge of PPH management, there was no significant difference in the mean MCQ score between the intervention or control groups when assessed before and after the learning experiences (Table 2). We noted a significant ($p < 0.001$) increase in mean MCQ score in both the intervention and control groups when comparing the pre- and post-learning experience MCQ scores (Table 3).

In terms of secondary outcomes, students were assessed on their ability to correctly assemble the components of a postpartum balloon and time taken to do so on a model pelvis. With regards the technique assessment, there was no difference in the mean score obtained between intervention or control groups (Table 4). We noted that students who participated in the intervention VRLE group assembled the postpartum balloon significantly ($p = 0.039$) faster [1.01–2 min vs. 2.01–3 min], compared with the control group (Table 4).

With regards to confidence levels, there was no significant difference in the mean confidence level between the intervention or control group when assessed before and after the learning experiences (Table 5). We noted a significant ($p < 0.001$) improvement in mean confidence level following the learning experience, in both the intervention and control groups (Table 6). Ultimately, 33 out of 34 participants (97%) felt that their overall confidence improved following training in PPH management and postpartum balloon insertion.

Table 7 describes the side effects experienced by students during the VRLE. 44.4% of those who used the VRLE experienced no side effects. Of those who did experience side effects, the most commonly reported symptoms were eye strain (27.8%), disorientation (27.8%), blurred/altered or double vision (16.7%), and nausea (16.7%).

Table 8 describes student satisfaction with elements common to both learning experiences. Overall, students were highly satisfied with both learning experiences, with mean ratings for each domain ranging from "agree" to "strongly agree" on the Likert-style scale. There was no difference in student satisfaction between the intervention and control group.

In relation to the students' views regarding the use of VR in medical education, 91.8% of participants agreed that VR technology could be useful as a learning tool in teaching obstetrics and gynecology topics to medical students. Of those who experienced the VRLE, 82.4% of participants were very likely to recommend its use in

10.3389/fmed.2024.1371075

TABLE 1 Baseline demographics of participants in intervention and control groups.

	Total cohort (n = 34)	Control (n = 16)	Intervention (n = 18)	p-value
Sex				0.038*
Male	9 (26.5%)	7 (43.8%)	2 (11.1%)	
Female	25 (73.5%)	9 (56.3%)	16 (88.9%)	
Age				0.772
18–24	18 (52.9%)	8 (50%)	10 (55.6%)	
25–34	13 (38.2%)	7 (43.8%)	6 (33.3%)	
35–45	3 (8.8%)	1 (6.3%)	2 (11.1%)	
Undergraduate entry medical students				0.221
Year 4	1 (2.9%)	-	1 (5.6%)	
Year 5	5 (14.7%)	2 (12.5%)	3 (16.7%)	
Year 6	13 (38.2%)	4 (25%)	9 (50%)	
Graduate entry medical students				0.221
Year 3	2 (5.9%)	2 (12.5%)	–	
Year 4	13 (38.2%)	8 (50%)	5 (27.8%)	
Any prior experience using VR?				0.934
Yes	13 (38.2%)	7 (38.9%)	6 (37.5%)	
No	21 (61.8%)	11 (61.1%)	10 (62.5%)	
Number of times VR used in the past:				0.302
0–5	33 (97.1%)	7 (100%)	5 (31.2%)	
5–10	–	–	–	
15–20	1 (2.9%)	–	1 (6.2%)	
Prior experience in management of a PPH/insertion of an intrauterine balloon:				0.774
No prior knowledge or experience	20 (58.8%)	11 (61.1%)	9 (56.3%)	
Theoretical knowledge only	14 (41.2%)	7 (38.9%)	7 (43.8%)	
Theoretical knowledge and some clinical experience	–	–	–	
Regular management of PPH but never seen insertion of postpartum balloon	–	–	–	
Regular management of PPH, seen insertion of postpartum balloon, but not confident	–	–	–	
Regular management of PPH, confident or have inserted postpartum balloon	–	–	–	

There were significantly more female participants in the overall cohort, with more females completing the intervention. The remaining baseline demographics were not significantly different between the two groups. Number of participants (n) and percentage of overall group size are reported. Differences between groups were assessed using Chi-square test of independence. A p-value < 0.05 was considered statistically significant.
VR – virtual reality, PPH – postpartum hemorrhage. Graduate entry medical students are those who have completed a primary degree, and are now undertaking medicine as a second degree.

TABLE 2 MCQ scores for intervention and control groups.

	n	MCQ mark Mean (SD)	Mean difference (95% CI)	p-value
Pre-learning experience				
Control	16	6.7 (2.0)	0.8 (−0.2, 1.9)	0.113
Intervention	18	7.7 (0.8)		
Post-learning experience				
Control	16	9.3 (1.1)	−0.1 (−0.8, 0.6)	0.770
Intervention	18	9.4 (0.9)		

MCQs were scored out of 10 points. Mean scores (standard deviation) are presented. MCQ scores were assessed using independent t-tests to evaluate mean score differences between control and intervention groups pre-learning experience and post-learning experience.
MCQ – multiple choice questionnaire, SD – standard deviation, CI – confidence interval.

TABLE 3 MCQ scores compared across time points.

	n	Pre-learning experience MCQ score Mean (SD)	Post-learning experience MCQ score Mean (SD)	Mean difference (95% CI)	*p*-value
Control	16	6.7 (2.0)	9.4 (0.9)	2.6 (1.6, 3.6)	<0.001
Intervention	18	7.7 (0.8)	9.3 (1.1)	1.6 (1.1, 2.1)	<0.001

MCQs were scored out of 10 points. Mean scores (standard deviation) are presented. MCQ scores were assessed using paired t-tests to evaluate the difference between the pre-learning and post-learning scores in each group. A p-value < 0.05 was considered statistically significant.
MCQ – multiple choice questionnaire, SD – standard deviation, CI – confidence interval.

TABLE 4 Student assessment post-learning experience.

	Control (n = 16)	Intervention (n = 18)	Mean difference (95% CI)	*p*-value
Technique assessment				
Mean score (SD)	7.5 (0.8)	7.6 (0.7)	0.09 (−0.5, 0.6)	0.743
Timing assessment				
Time to completion (minutes)	2.01–3.0	1.01–2.0		0.039*

The technique assessment was scored out of 10 based on manufacturer's guidelines. The timing assessment was marked categorically in 1 min increments. The technique assessment between groups was examined using independent t-tests. The timing assessment was evaluated using chi-square test of independence. A p-value < 0.05 was considered statistically significant.
SD – standard deviation, CI – confidence interval.

TABLE 5 Confidence levels for intervention and control groups.

Group	*n*	Confidence level Mean (SD)	Mean difference (95% CI)	*p*-value
Pre-learning experience				
Control	16	1.1 (0.5)	0.2 (−0.2, 0.5)	0.417
Intervention	18	1.3 (0.6)		
Post-learning experience				
Control	16	2.7 (0.7)	0.1 (−0.4, 0.6)	0.639
Intervention	18	2.8 (0.7)		

Confidence levels were assessed using Likert-type rating scales, with 1 representing no confidence and 5 as maximum score. Mean scores (standard deviation) are presented. Confidence levels were examined using independent t-tests to assess mean score differences between groups pre-learning experience and post-learning experience.
SD – standard deviation, CI – confidence interval.

TABLE 6 Confidence levels compared across time points.

Group	*n*	Pre-learning experience confidence level Mean (SD)	Post-learning experience confidence level Mean (SD)	Mean difference (95% CI)	*p*-value
Control	16	1.1 (0.5)	2.7 (0.7)	1.5 (1.2, 1.9)	<0.001
Intervention	18	1.3 (0.6)	2.8 (0.7)	1.5 (1.1, 1.8)	<0.001

Confidence levels were assessed using Likert-type rating scales, with 1 representing no confidence and 5 as maximum score. Mean scores (standard deviation) are presented. Confidence levels were examined using paired t-tests to examine the difference between the pre-learning and post-learning scores in each group. A p-value < 0.05 was considered statistically significant.
SD – standard deviation, CI – confidence interval.

university and training in obstetrics and gynecology. 94.1% of the participants felt that the VRLE was beneficial over didactic teaching, and 100% of the participants who experienced the VRLE enjoyed the learning experience.

4 Discussion

4.1 Main findings

The objective of our randomized control trial was to pilot the use of a VRLE to enhance medical students' knowledge of PPH management and insertion of a postpartum balloon for uterine tamponade. Receiving formal instruction, regardless of format, enhanced students' knowledge of PPH emergency management. Both learning experiences (VRLE and didactic teaching session) enhanced student performance on the post-learning experience MCQ. Additionally, students who participated in the VRLE were quicker at insertion of the postpartum balloon insertion on the model pelvis. Both learning experiences (VRLE and didactic teaching session) also improved student confidence with PPH management, as assessed on the post-learning experience survey. While 55.6% of VRLE participants did experience side-effects of VR, none of them were significant enough to cease the learning experience.

4.2 Comparison with existing literature

Similar to previous studies examining the use of VR in embryology, and anatomy, including fetal and pelvic anatomy (12, 14, 15), no significant differences were demonstrated in the post-learning experience knowledge scores when VRLE was compared to didactic teaching. Previous research has demonstrated that learners who use a VRLE complete tasks significantly quicker than those who used other forms of learning (12, 16). We also found that our students who used the VRLE completed the assessment on the model pelvis quicker than the control group.

Simulation training in obstetrics and gynecology has demonstrated positive effects on the knowledge and skills obtained by the participants and improves overall satisfaction and self-confidence with the content taught through simulation (3). Similar to simulation training, our results suggest that a VRLE has the potential to enhance the knowledge and skills of the students and improve their self-rated confidence levels with the material. It has been suggested that VR may function as an alternative to simulated clinical environments (17).

Medical students find immersive technology, such as VR, very enjoyable to use and adds to their overall learning experience (12, 13, 18). Previous research has demonstrated that students who are motivated and engaged by novel learning methods are more likely to retain knowledge, and may have improved learning outcomes (19). Similarly, we found high levels of satisfaction among students who used the VRLE, with 100% of the students enjoying the experience and 82.4% very likely to recommend the use of VRLE in medical education in the future.

TABLE 7 Side effects of VR.

Side-effect	*n* (%)
No side-effects	8 (44.4%)
Any side effect	10 (55.6%)
Dizziness	1 (5.6%)
Headache	1 (5.6%)
Blurred/altered or double vision	3 (16.7%)
Loss of awareness	–
Eye strain	5 (27.8%)
Eye or muscle twitching	–
Involuntary moving	–
Disorientation	5 (27.8%)
Impaired balance	–
Impaired hand-eye coordination	–
Excessive sweating	–
Increased salivation	–
Nausea	3 (16.7%)
Light-headedness	–
Discomfort or pain in the head or eyes	2 (11.1%)
Drowsiness	–
Fatigue	1 (5.6%)
Any symptoms similar to motion sickness	–

Side effects experienced by students while using the virtual reality learning environment. Participants completed a questionnaire regarding side effects experienced during the VRLE. Group ($n = 18$) and percentage of intervention group are reported for each side effect.

4.3 Implications for the future

Through this RCT, we have highlighted that a VRLE is an acceptable format in which to promote student engagement with concepts in obstetrics and gynecology. It is important to note that medical students are adult learners, and adults are more independent in their learning processes, striving for higher levels of autonomy and self-directed learning (20). In order to support this drive for self-directed learning, intrinsic motivation is required, which can be achieved when the student believes that the learning outcome is relevant to their practice (21). Therefore, the use of a VRLE specifically designed to achieve relevant learning outcomes may be beneficial to enhance student engagement and knowledge retention.

Additionally, VR has the potential to play a key role in creating a learning environment for acquisition of clinical skills in a safe

TABLE 8 Student satisfaction of the learning experiences.

	Control ($n = 16$) Mean (SD)	Intervention ($n = 18$) Mean (SD)	*p*-value
Easy to use/follow	2.1 (0.7)	2.1 (0.8)	0.871
Clear purpose and objectives	1.4 (0.5)	1.6 (0.6)	0.293
Support during the learning experience	1.4 (0.6)	1.6 (0.9)	0.433
Designed for my specific level of knowledge and skills	1.6 (0.6)	1.9 (0.6)	0.182
Opportunity to enhance understanding	1.3 (0.5)	1.2 (0.4)	0.491
Learning benefitted from the learning experience	1.5 (0.6)	1.4 (0.6)	0.723
Opportunity for feedback	1.7 (0.6)	1.8 (0.9)	0.534
The learning experience was very realistic	1.9 (0.7)	2.2 (1.3)	0.324
Recommendation of the learning experience	1.3 (0.5)	1.4 (0.8)	0.812

Students were asked to rate their learning experience on a Likert-style scale of 1 = strongly agree to 5 = strongly disagree, for the above domains. Scores are presented as mean (standard deviation), and comparisons between control and intervention groups was done using independent *t*-tests. No significant differences were noted. A p-value < 0.05 was considered statistically significant.

environment (7). Medical education relies heavily on continued practice of procedural skills necessary to begin a clinical career (7). However, in obstetrics and gynecology, this can be very difficult for medical students to achieve due to the sensitive and occasionally invasive nature of the specialty (3). Additionally, medical students require increased exposure to emergency situations in order to prepare them for their clinical practice in the future (3). Therefore, VR is uniquely positioned to create an immersive simulated clinical environment in which students can practice procedural skills as many times as necessary to enhance preparedness for their future clinical careers.

4.4 Strengths and Limitations

A strength of this study is that it is a randomized control trial assessing the use of a VRLE in enhancing medical student knowledge of an important emergency management topic, and assessing its use for teaching rarely encountered procedural skills. A limitation of the current study is the small sample size, which may have led us to be under-powered to discover differences in knowledge acquisition. Additionally, as in previous studies (12, 13, 15), side-effects due to cybersickness are acknowledged as a limitation to the use of VR technology. However, with advancement of VR technology, and repeated exposure to the technology, side effects are expected to decrease (22). Additionally, the cost of VR is acknowledged as a potential barrier to widespread use. Thus far, VR in medical education has not been widely available due to high costs and lack of evidence of its efficacy (23). Once the VRLE has been developed, it can be used repeatedly without further development costs, however it is acknowledged that regular maintenance and software updates are required to ensure function of the VR headsets. Of note, recent technological advancements have reduced the cost of VR headsets, making them more affordable (24). Future studies with larger cohorts of medical students could further explore the use of the VRLE as learning tools for medical students, especially for procedural skills.

5 Conclusion

Receiving formal instruction, regardless of format, enhances students' knowledge and confidence of the topic covered. Students who received instruction via the VRLE assembled the postpartum balloon faster than students who received didactic teaching. Use of VRLE in medical education represents an acceptable learning environment in which to engage students in order to enhance the overall learning experience. VR may be beneficial in teaching hands-on procedural skills in obstetrics and gynecology education.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by University College Dublin Research Ethics Committee. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KD: Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing, Project administration. GD: Data curation, Writing – original draft, Writing – review & editing, Formal analysis. AM: Conceptualization, Methodology, Project administration, Writing – review & editing. DK: Conceptualization, Software, Writing – review & editing, Methodology. SH: Funding acquisition, Supervision, Writing – review & editing, Conceptualization, Validation. EM: Software, Supervision, Writing – review & editing, Validation. FM: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing, Project administration, Validation.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by funding by the University College Dublin, Ireland and The National Maternity Hospital, Dublin, Ireland. Virtual reality headsets were previously purchased by University College Dublin, and the postpartum balloon insertion programme was created using funding from University College Dublin. Cook Medical provided the model pelvis for participant assessment, but did not contribute to the funding of this study.

Acknowledgments

The authors would like to sincerely thank all of the medical students for their participation in the study, the research centre, and the research team for their ongoing support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Mavrides EAS, Chandraharan E, Collins P, Green L, Hunt BJ, Riris S, et al. Prevention and Management of Postpartum Haemorrhage: Green-top guideline no. 52. *BJOG*. (2017) 124:e106–49. doi: 10.1111/1471-0528.14178

2. Purandare CN, Nazareth AK, Ryan G, Purandare NC. Role of balloon tamponade as a therapeutic non-surgical tool in controlling obstetric and gynecological hemorrhage in low-resource countries. *J Obstet Gynaecol India*. (2022) 72:285–90. doi: 10.1007/s13224-022-01662-7

3. Tauscher A, Stepan H, Todorow H, Rotzoll D. Interteam PERINAT—interprofessional team collaboration in undergraduate midwifery and medical education in the context of obstetric emergencies: presentation of simulation scenarios and empirical evaluation results. *GMS J Med Educ*. (2023) 40:Doc20. doi: 10.3205/zma001602

4. Minehart RD, Gallin H. Postpartum hemorrhage: the role of simulation. *Best Pract Res Clin Anaesthesiol*. (2022) 36:433–9. doi: 10.1016/j.bpa.2022.11.002

5. Shore EM, Davidson A, Arnason M, Kara H, Shah A, Shah R. Bridging the gap: incorporating simulation into obstetrics and Gynaecology undergraduate medical education. *J Obstet Gynaecol Can*. (2019) 41:191–196.e2. doi: 10.1016/j.jogc.2018.03.016

6. Jin J, Bridges SM. Educational technologies in problem-based learning in health sciences education: a systematic review. *J Med Internet Res*. (2014) 16:e251. doi: 10.2196/jmir.3240

7. Ryan GV, Callaghan S, Rafferty A, Higgins MF, Mangina E, McAuliffe F. Learning outcomes of immersive Technologies in Health Care Student Education: systematic review of the literature. *J Med Internet Res*. (2022) 24:e30082. doi: 10.2196/30082

8. McGaghie WC, Harris IB. Learning theory foundations of simulation-based mastery learning. *Simul Healthc*. (2018) 13:S15–20. doi: 10.1097/SIH.0000000000000279

9. Taylor DC, Hamdy H. Adult learning theories: implications for learning and teaching in medical education: AMEE guide no. 83. *Med Teach*. (2013) 35:e1561–72. doi: 10.3109/0142159X.2013.828153

10. Izard SG, Juanes JA, Garcia Penalvo FJ, Estella JMG, Ledesma MJS, Ruisoto P. Virtual reality as an educational and training tool for medicine. *J Med Syst*. (2018) 42:50. doi: 10.1007/s10916-018-0900-2

11. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. *Br J Surg*. (2012) 99:1322–30. doi: 10.1002/bjs.8819

12. Kane D, Ryan G, Mangina E, McAuliffe FM. A randomized control trial of a virtual reality learning environment in obstetric medical student teaching. *Int J Med Inform*. (2022) 168:104899. doi: 10.1016/j.ijmedinf.2022.104899

13. Ryan G, Rafferty A, Murphy J, Higgins MF, Mangina E, McAuliffe FM. Virtual reality learning: a randomized controlled trial assessing medical student knowledge of fetal development. *Int J Gynaecol Obstet*. (2023) 162:292–9. doi: 10.1002/ijgo.14684

14. Arents V, de Groot PCM, Struben VMD, van Stralen KJ. Use of 360 degrees virtual reality video in medical obstetrical education: a quasi-experimental design. *BMC Med Educ*. (2021) 21:202. doi: 10.1186/s12909-021-02628-5

15. Ryan G, Callaghan S, Rafferty A, Murphy J, Higgins M, Barry T, et al. Virtual reality in midwifery education: a mixed methods study to assess learning and understanding. *Nurse Educ Today*. (2022) 119:105573. doi: 10.1016/j.nedt.2022.105573

16. Nickel F, Brzoska JA, Gondan M, Rangnick HM, Chu J, Kenngott HG, et al. Virtual reality training versus blended learning of laparoscopic cholecystectomy: a randomized controlled trial with laparoscopic novices. *Medicine (Baltimore)*. (2015) 94:e764. doi: 10.1097/MD.0000000000000764

17. Birkham SL, Calogiuri G, Martinsen R. Advancing immersive virtual reality-based simulation practices: developing an evidence-based and theory-driven pedagogical framework for VR-based simulations of non-technical skills among healthcare professionals. *Interact Learn Environ*. (2023):1–13. doi: 10.1080/10494820.2023.2186896

18. Guetterman TC, Sakakibara R, Baireddy S, Kron FW, Scerbo MW, Cleary JF, et al. Medical Students' experiences and outcomes using a virtual human simulation to improve communication skills: mixed methods study. *J Med Internet Res*. (2019) 21:e15459. doi: 10.2196/15459

19. Huang H-M, Rauch U, Liaw S-S. Investigating learners' attitudes toward virtual reality learning environments: based on a constructivist approach. *Comput Educ*. (2010) 55:1171–82. doi: 10.1016/j.compedu.2010.05.014

20. Zigmont JJ, Kappus LJ, Sudikoff SN. Theoretical foundations of learning through simulation. *Semin Perinatol*. (2011) 35:47–51. doi: 10.1053/j.semperi.2011.01.002

21. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am Psychol*. (2000) 55:68–78. doi: 10.1037/0003-066X.55.1.68

22. Kim J, Luu W, Palmisano S. Multisensory integration and the experience of scene instability, presence and cybersickness in virtual environments. *Comput Hum Behav*. (2020) 113:106484:106484. doi: 10.1016/j.chb.2020.106484

23. Farra SL, Gneuhs M, Hodgson E, Kawosa B, Miller ET, Simon A, et al. Comparative cost of virtual reality training and live exercises for training Hospital Workers for Evacuation. *Comput Inform Nurs*. (2019) 37:446–54. doi: 10.1097/CIN.0000000000000540

24. Hamilton D, McKechnie J, Edgerton E, Wilson C. Immersive virtual reality as a pedagogical tool in education: a systematic literature review of quantitative learning outcomes and experimental design. *J Comput Educ*. (2021) 8:1–32. doi: 10.1007/s40692-020-00169-2

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 20 March 2024
DOI 10.3389/fmed.2024.1330457

OPEN ACCESS

EDITED BY
Fedde Scheele,
VU Amsterdam, Netherlands

REVIEWED BY
Joost Velzel,
Northwest Clinics, Netherlands
Leonie Van Rheenen,
Onze Lieve Vrouwe Gasthuis (OLVG),
Netherlands

*CORRESPONDENCE
L. Brogaard
lbrj@clin.au.dk

RECEIVED 30 October 2023
ACCEPTED 12 February 2024
PUBLISHED 20 March 2024

CITATION
Brogaard L, Rosvig L, Hjorth-Hansen KR, Hvidman L, Hinshaw K, Kierkegaard O, Uldbjerg N and Manser T (2024) Team performance during vacuum-assisted vaginal delivery: video review of obstetric multidisciplinary teams.
Front. Med. 11:1330457.
doi: 10.3389/fmed.2024.1330457

Team performance during vacuum-assisted vaginal delivery: video review of obstetric multidisciplinary teams

L. Brogaard[1,2]*, L. Rosvig[3], K. R. Hjorth-Hansen[4], L. Hvidman[2], K. Hinshaw[5], O. Kierkegaard[6], N. Uldbjerg[1,2] and T. Manser[7]

[1]Department of Obstetrics and Gynecology, Aarhus University Hospital, Aarhus, Denmark, [2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark, [3]Department of Obstetrics and Gynecology, Randers Hospital, Randers, Denmark, [4]Department of Oncology, Aarhus University Hospital, Aarhus, Denmark, [5]Department of Obstetrics and Gynecology, Sunderland Royal Hospital, Sunderland, United Kingdom, [6]Department of Obstetrics and Gynecology, Horsens Regional Hospital, Horsens, Denmark, [7]FHNW School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

Introduction: Vacuum extraction is generally considered an operator-dependent task, with most attention directed toward the obstetrician's technical abilities (1–3). Little is known about the effect of the team and non-technical skills on clinical outcomes in vacuum-assisted delivery. This study aimed to investigate whether the non-technical skills of obstetricians were correlated with their level of clinical performance via the analysis of video recordings of teams conducting actual vacuum extractions.

Methods: We installed between two or three video cameras in each delivery room at Aarhus University Hospital and Horsens Regional Hospital and obtained 60 videos of teams managing vacuum extraction. Appropriate consent was obtained. Two raters carefully reviewed the videos and assessed the teams' non-technical skills using the Assessment of Obstetric Team Performance (AOTP) checklist, rating all items on a Likert scale score from 1 to 5 (1 = poor; 3 = average; and 5 = excellent). This resulted in a total score ranging from 18 to 90. Two different raters independently assessed the teams' clinical performance (adherence to clinical guidelines) using the TeamOBS-Vacuum-Assisted Delivery (VAD) checklist, rating each item (0 = not done, 1 = done incorrectly; and 2 = done correctly). This resulted in a total score with the following ranges (low clinical performance: 0–59; average: 60–84; and high: 85–100). Interrater agreement was analyzed using intraclass correlation (ICC), and the risk of high or low clinical performance was analyzed on a logit scale to meet the assumption of normality.

Results: Teams that received excellent non-technical scores had an 81% probability of achieving high clinical performance, whereas this probability was only 12% among teams with average non-technical scores ($p < 0.001$). Teams with a high clinical performance often had excellent behavior in the non-technical items of "team interaction," "anticipation," "avoidance fixation," and "focused communication." Teams with a low or average clinical performance often neglected to consider analgesia, had delayed abandonment of the attempted vaginal delivery and insufficient use of appropriate fetal monitoring. Interrater reliability was high for both rater-teams, with an ICC for the non-technical skills of 0.83 (95% confidence interval [CI]: 0.71–0.88) and 0.84 for the clinical performance (95% CI: 0.74–0.90).

Conclusion: Although assisted vaginal delivery by vacuum extraction is generally considered to be an operator-dependent task, our findings suggest that teamwork and effective team interaction play crucial roles in achieving high clinical performance. Teamwork helped the consultant anticipate the next step, avoid fixation, ensure adequate analgesia, and maintain thorough fetal monitoring during delivery.

KEYWORDS

performance, obstetric, labor, vacuum extraction, video, teamwork, checklist

1 Introduction

Assisted vaginal delivery is employed in situations where the second stage of labor should be shortened due to maternal or fetal indications. The primary goal is to achieve expedited delivery while maintaining high-quality care. However, delivery wards worldwide have witnessed increased rates of cesarean sections in the second stage of labor. This trend may signify a decline in clinical skills or concerns regarding the risk of severe complications such as maternal perineal trauma, fetal scalp injuries, and even intracranial hemorrhages (1, 2). While vacuum extraction remains the most utilized method for assisted vaginal delivery, we must be vigilant in addressing and mitigating these risks.

Vacuum extraction is commonly perceived as an operator-dependent task, with emphasis placed on the technical skills of the obstetrician (1–3). However, there is a paucity of research exploring the importance of the non-technical skills employed by the entire delivery team (4–6). These non-technical skills have been categorized as cognitive, social, and personal resource skills. They include decision-making, situational awareness, communication, teamwork, leadership, and coping with stress and fatigue (7). These non-technical skills are important in acute emergency teams (8–10) and in obstetric teams managing postpartum hemorrhage (11), settings that clearly differ from delivery by vacuum extraction (12, 13).

In the context of vacuum extraction, all research on the non-technical skills has been confined to simulation settings (4, 5, 14, 15) which may not entirely capture the intricacies of real-word patient care. This underscores the imperative for a thorough evaluation of non-technical skills in the delivery suite (16). Therefore, this study analyzes videos of teams engaged in vacuum extractions during real-life deliveries, aiming to explore the non-technical skills and assess their potential correlation with the level of clinical performance.

2 Materials and methods

2.1 Study design and setting

Two Danish hospitals participated in the study: Aarhus University Hospital, delivering level 3 maternity care for 5,000 deliveries per year, with 50 physicians, 100 midwives, and 20 technicians; and Horsens Regional Hospital, delivering level 2 maternity care for 2,000 deliveries per year, with 25 physicians, 45 midwives, and 2 technicians. The delivery rooms of the hospitals are well-equipped with two or three high-definition mini-dome surveillance cameras and a ceiling microphone at the center of the room. To minimize the video recordings of normal deliveries, we designed a video system that was activated by Bluetooth on the obstetrician's phone. All cameras recorded 5-min loops until the obstetrician entered the room; the Bluetooth signal activated the cameras, and the preceding 5 min and subsequent time spent in the delivery room were recorded. Videos were only included if consent was obtained from all participants within 48 h; otherwise, videos were automatically deleted from the server.

Written consent was obtained from all participants appearing in the videos (staff, patients, and relatives). Information and collection of consent was a two-step procedure. First, all women received information about the research project in our outpatient clinic some weeks before labor, with records of this information noted in the medical electronic records. Subsequently, in the delivery ward, the midwife responsible for the delivery ensured written informed consent was obtained. The inclusion criterion encompassed all assisted deliveries conducted in our delivery suites, and the exclusion criteria included the absence of informed consent or technical errors in sound or video recording.

Vacuum extractions in Denmark are mainly conducted in the delivery room, with the opportunity to swiftly move to the theater to perform a cesarean section in cases of failed extraction. The incidence of cesareans is 20%, with approximately half of these being emergencies and half elective. The incidence of attempted vacuum extractions is 6–7% of all deliveries and failed vacuum extractions is approximately 9% of these. The hospitals guideline state that vacuum extraction should be discontinued if no descent occurs after two pulls or two pop-offs, or if delivery is not imminent following three pulls. The vacuum extraction is often conducted by a multidisciplinary team, where the consultant performs the extraction, a midwife supports the perineum, a second midwife ensures fetal monitoring, and a nurse assistant ensures all equipment for the extraction is available (for example, different-sized metal cups or silicone alternatives and vacuum machine). Forceps is rarely used for vertex presentation.

The video recordings of teams were recorded between November 2014 and July 2016 and analyzed during the summer of 2017.

Abbreviations: ANTS, Anesthetists' Nontechnical Skills; AOTP, Assessment of Team Performance; CI, Confidence interval; ICC, Intraclass correlation coefficient; VAD, Vacuum-assisted delivery.

2.2 Ethics approval

Ethical and legal approval for this study was obtained in May 2014 from the Central Denmark Region, the Danish Data Protection Agency (2012–58-006), and the Research Foundation of Central Denmark (Record No. 1–16–02-257-14). All videos were included with informed consent in conformity with the Danish code §264.

2.3 Clinical performance assessment by team OBS-VAD

We developed and validated a checklist to evaluate the clinical performance of assisted vaginal delivery via vacuum extraction. The checklist was the result of a Delphi process with 12 experts from the United Kingdom, Norway, Iceland, Sweden, and Denmark (17), and comprises 18 items, each with a weight of importance score: "Done correctly and in a timely manner" (2 points); "Done incorrectly or done correctly with delay" (1 point); "Not done" (0 points); "Cannot be assessed" (no value); or "Not indicated" (no value). The checklist results in a total score of 0–100% (100% = highest standard of care). The checklist also included a visual analog scale of patient safety ranging from 0 to 100%, where elements of management not included in the checklist were evaluated. The total clinical performance score was reported as: 'High' (85–100); 'Average' (60–84); or 'Low' (<60). Both raters (LH and LA) were senior obstetric consultants working in one of the two study hospitals. They were experienced in video review and received appropriate training, including an introduction to the TeamOBS-VAD checklist and a discussion of examples of high and low performance. Subsequently, the raters independently assessed all videos for clinical performance, and were blinded to each other's score.

2.4 Non-technical skills assessment by assessment of obstetric team performance

We used the AOTP (18) validated tool to assess the team's non-technical score. It comprises 18 items grouped into six categories: "Communication with the patient," "Task management," "Teamwork," "Situation awareness," "Communication with the team members," and "Environment of the room." Each item was scored on a Likert scale (1–5; 1, poor; 5, excellent), resulting in a total score between 18 and 90 (list of all items in Supplementary Table S3). Both raters (LB and KH) were physicians who were experienced in video review and AOTP from a previous study (11). The raters assessed all the videos independently and were blinded to each other's scores.

2.5 Statistical analysis

We assessed the following criteria: (1) the interrater agreement by intraclass correlation (ICC) and Bland–Altman analysis; (2) the agreement on each item by percentage agreement and weighed Cohen's kappa; (3) the association between the non-technical score and clinical performance by a restricted cubic spline regression analysis with three knots at 2.5, 3, and 3.5; and (4) the mean difference in clinical performance between the lowest (2) and the highest non-technical score (5) by spline regression analysis (19, 20). The regression models were checked using the diagnostic plots of the residuals. The confidence intervals for risk analysis were computed using a non-parametric percentile bootstrap. The potential confounding factors of parity, indication for vacuum (prolonged second stage or fetal compromise), classification (mid, low, and outlet), level of training (junior 1–5 years and senior >5 years), time of event, and hospital type were assessed using multiple linear regression analysis. We used STATA 15 (StataCorp LP, College Station, TX) for all statistical analysis.

3 Results

3.1 Included videos

Women expecting to give birth in one of the two study hospitals were informed about the project, and videos between November 2014 and July 2016 were included. The main reason for exclusion was the absence of consent after 48 h (28 cases), often because the mother had been discharged from the hospital before consent was obtained. Consent was declined by five healthcare providers, six mothers, or relatives. We included 60 videos, with 48% eligible cases (Figure 1). The 60 teams comprised 178 different healthcare providers, 60 different team combinations, and different levels of difficulty with vaginal delivery (Table 1).

3.2 Interrater agreement

The interrater agreement was high, as raters assessing the non-technical skills had an ICC of 0.83 (95% CI: 0.71–0.88) and consultants assessing clinical performance had an ICC of 0.84 (95% CI: 0.74–0.90). Agreement among raters for assessing the specific non-technical skills was 0.69–0.94 weighted kappa and agreement was visualized using Bland–Altman Plots and limits of agreement (Table 2; Supplementary Tables S1, S2). Four videos were discussed and reassessed by the two raters as their clinical performance assessments differed by >15%, which had been defined a priority as the maximum acceptable difference. This new consensus score was used in the risk analyses, but not in the ICC calculations. No videos were reassessed for non-technical scores.

3.3 Clinical performance and non-technical skills

Most teams achieved high clinical performance scores in ensuring the correct position of the mother (94%), appropriate number of staff present (91%), and delivery conducted within maximum four pulls (88%). The greatest challenge posed in clinical skill was considering analgesia, as 39% of the teams neither discussed nor mentioned analgesia as an option.

Clinical performance and non-technical skills correlated with a 'dose–response-like' association (Figure 2). Teams with excellent non-technical scores had an 81% probability of high clinical performance, whereas this probability was only 12% among teams with average non-technical scores ($p < 0.001$; Table 3). Clinically high-performing teams demonstrated excellent behavior in the

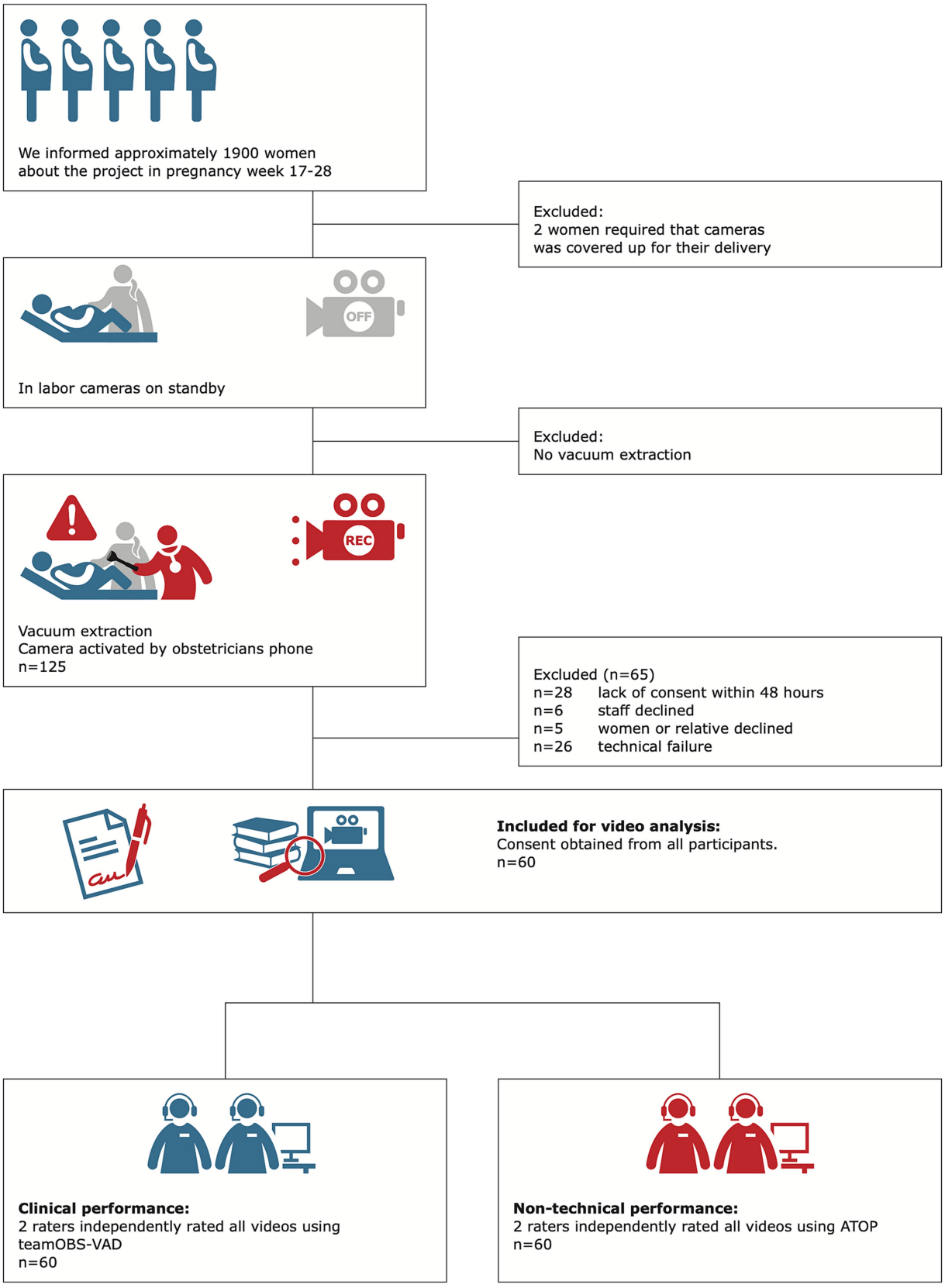


FIGURE 1
Inclusion of teams.

TABLE 1 Description of included teams.

		n	Mean/median/proportion
Team size, mean (SD)		60	4.5 (0.9)
Women previous (prior) vaginal delivery			
	1. No, n (%)	39	65%
	2. Yes, n (%)	5	8%
	3. Information not verbalized in video, n (%)	16	27%
Indication of vacuum extraction			
	1. Prolonged second stage of labor or for maternal benefit, n (%)	30	50%
	2. Suspicion of immediate or potential fetal compromise, n (%)	30	50%
Classification of assisted vaginal delivery*			
	1. Mid, n (%)	15	25%
	2. Low, n (%)	20	33%
	3. Outlet, n (%)	11	19%
	4. Information not verbalized in video, n (%)	14	23%
Choice of vacuum extractor			
	1. Soft vacuum cup, n (%)	15	25%
	2. Metal vacuum cup, n (%)	39	65%
	3. Replacing soft cup to metal, n (%)	6	10%
Procedure of the vacuum extraction			
	1. Delivery conducted in ≤3 contractions	48	80%
	2. Delivery conducted in >3 contractions	12	20%
	3. Delivery with cup detachment, n (%)	10	17%
	4. Duration of vacuum applied (min), median (IQR)	60	4.5 (2–9)
	5. Failed vacuum extraction, n (%)	4	6%
Experience of operator/ level of training			
	1. Resident, registrar, 1–5 years' experience, n (%)	25	42%
	2. Consultant, >5 years' experience, n (%)	35	58%
Time of event			
	1. Day (7:00–14:59), n (%)	18	30%
	2. Evening (15:00–23:59), n (%)	23	38%
	3. Night (00:00–06:59), n (%)	19	32%
Hospital type			
	1. University hospital, maternal care level 3, n (%)	32	53%
	2. Regional hospital, maternal care level 2, n (%)	28	47%

*The American College of Obstetricians and Gynecologist classification of types of operative vaginal delivery (ref guideline ACOG).

non-technical skills categories of team interaction, anticipation, avoidance of fixation, and focused communication (Supplementary Table S3). These results were robust in terms of the following confounders: hospital, team size, time of day, and number of pulls during the delivery (Supplementary Table S4).

4 Discussion

Assisted vaginal delivery by vacuum extraction has been considered an operator-dependent task; however, our findings indicate that the non-technical skills of teams play an important role in achieving high clinical performance. Thus, teams with an excellent non-technical score had an 81% chance of achieving high clinical performance, whereas this probability was only 12% among teams with average non-technical scores. Teamwork helped the obstetrician to anticipate the next steps, avoid fixation, ensure analgesia, and ensure sufficient fetal monitoring during delivery.

The main strength of this study is the automatic inclusion of videos using Bluetooth in the obstetrician's telephone, which activated the video cameras whenever the obstetrician entered the delivery room (21). This potentially reduced selection bias. The sample size is another strength, as the sample included 60 different teams and 178 healthcare providers from two hospitals, operators of the vacuum at all levels of experience, all degrees of difficulty, every day of the week, and any time of day. Finally, the two pairs of raters used a validated

TABLE 2 Interrater agreement for clinical performance and non-technical performance.

	Descriptive		ICC (95% CI)			
	Mean	Range	Individual rater*		Average of two raters**	
Clinical performance: TeamOBS-VAD score	83.2	(49–100)	0.73	(0.58–0.82)	0.84	(0.74–0.90)
Non-technical performance: ATOP average score	70.5	(39–87)	0.71	(0.55–0.82)	0.83	(0.71–0.88)

*Intraclass correlation (ICC) between raters. **All 60 videos were analyzed by two raters. The average agreement represents the ICC between two raters and the other two raters.

TABLE 3 Risk/chance of low/high clinical performance to the level of non-technical performance.

Non-technical score	Clinical performance: risk of low score*		Clinical performance: chance of high score**	
3 (Average)	4.3%	(0.2–12.4%)	12.3%	(2.0–29.2%)
4 (Good)	0.3%	(0.0–0.9%)	43%	(34.5–54.0%)
5 (Excellent)	0.01%	(≤0.01–0.01%)	80.6%	(63.7–94.1%)

These figures are based on 60 recordings from real-life teams managing major vaginal-assisted deliveries via vacuum extraction and risk (95% CI). *TeamOBS-VAD score below the minimum pass (score < 60). **TeamOBS-VAD score above high performance (score > 85).

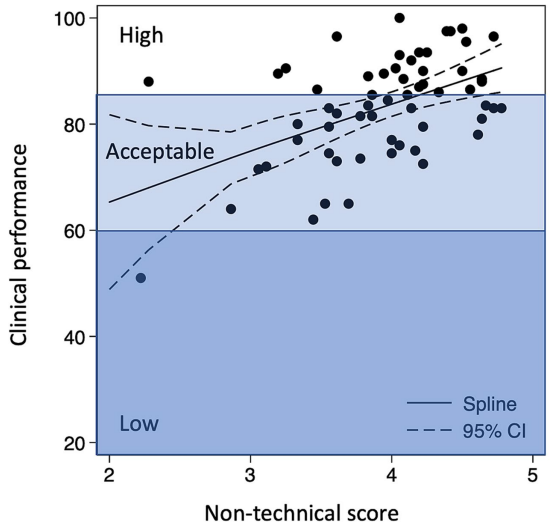


FIGURE 2
Clinical performance and non-technical skills.

checklist for systematic assessment and the interrater agreement was high.

Our study had certain limitations. We cannot exclude selection bias, as low-performing teams may have been less willing to provide their consent. However, it is reassuring that 95% of the healthcare providers gave consent to include their videos and that only two videos were deleted after staff withdrew their consent. We also acknowledge that our setup did not include all vacuum extractions, as the camera was activated by the consultant on call. There is a risk of assessment bias as AOTP raters may have been influenced by the actual clinical performance and vice versa. All teams were aware of the project, and this may have changed their behavior, a form of "Hawthorne effect" which might affect the observed association (22). Furthermore, despite having all the delivery rooms equipped with two or three cameras in the ceiling, the video could not capture all the details of the technical skills, and raters often missed tactile information gained during vaginal examination before the operator applied the cup (for example, molding, level, and position). Furthermore, our legal permission was restricted to analyze the videos; hence, we did not have access to any medical journal with maternal or neonatal outcomes. Finally, while our study found that clinical performance and non-technical scores were strongly correlated with a dose–response like association; this is no proof of causality.

The opportunity to review video recordings of teams managing vacuum extractions in real-life was a special experience that changed the way vacuum extractions were handled in our department (23). Epidural as pain relief in labor is available for every laboring woman in Denmark; however, uptake is driven by maternal request. Overall, approximately 27% of nulliparous women in labor request an epidural. We considered whether the high sound level on the recording reflected real life. Therefore, we tested our audio source with the help of a sound engineer and found that the quality was high and that the dB level in the room was reflected by the dB level in the video (24). These findings highlight the importance of improving our practice of offering women analgesia during vacuum extraction. Based on our study, we recommend prioritizing the wider use of pudendal block in cases of vacuum extraction and ensuring that women are adequately informed about assisted delivery preferable in the antenatal period (25).

Before this study, non-technical skills for vacuum extraction have only been investigated in simulated settings (6, 15, 26–28), but did suggest that non-technical skills contribute to successful delivery assisted by vacuum extraction (4, 14, 29). However, these studies did not describe the specific aspects of behavior or the specific non-technical skills that support teams to achieve high clinical performance. In this study, we observed how enacting specific non-technical skills helped obstetricians to anticipate the next steps, such as offering analgesia to women without an epidural or calling the pediatric team to the room when the indication was suspected fetal distress. Moreover, the team helped the consultant to keep track of the time and the number of contractions, thereby helping the consultant decide when to abandon the attempted vaginal delivery in a timely manner. This calls for rethinking our learning path for vacuum extraction, ensuring that trainees learn to include effective teamwork behavior in training and practice. Thus, this study confirms the results on surgical (30), trauma (31, 32), and resuscitation teams (33–35) regarding the importance of non-technical skills in providing high clinical performance.

High-performing teams in this study used specific behaviors that differed from those of high-performing teams managing postpartum hemorrhage (11). High-performing teams managing vacuum extraction demonstrated high scores in team interaction, anticipation, avoidance of fixation, and focused communication, whereas high-performing teams managing postpartum hemorrhage had high scores in vigilance, role assignment, stress management, and leadership (11). This difference may reflect the fact that most vacuum extractions are a 'semi-elective' procedure, whereas postpartum hemorrhage is usually an urgent, acute emergency. Another interesting disparity between the two settings was the importance of avoiding fixation, which we could only demonstrate in the vacuum extraction setting but not in the postpartum hemorrhage setting. This is related to the fact that the accoucheur who is conducting the vacuum delivery is 'task-focused' and therefore cannot maintain wider situational awareness. This requires a member of the team to 'step back' to monitor the whole situation (that is, 'situational leadership'). This tends to happen in teams managing acute emergencies, such as postpartum hemorrhage, where the situational leader steps back to co-ordinate and delegate tasks while ensuring that vigilance is maintained (7). Therefore, all team members should keep an eye on the overall situation while the accoucheur conducts the vacuum extraction and may need to consider whether an individual needs to 'step back' to maintain overall vigilance.

Video reviews offer a unique opportunity to review our performance, and future studies should investigate whether video reviews can help to improve our vacuum extraction performance (31, 34). Obstetric trainees primarily learn to perform vacuum extraction from senior colleagues by craft apprenticeship. In the last few decades, trainees have experienced difficulties in developing and maintaining obstetrical competencies, such as vacuum extractions, due to reduced working hours and a reduced rate of instrumental deliveries in obstetric care (33). Elective video review could be a valuable learning method to junior obstetricians to improve performance as they can revisit and reflect on their performance and learn from experienced colleagues (36). When they become consultants, they primarily work alone or with a junior trainee, and do not always have the same opportunity to benchmark their performance with several consultants; feedback may not always be ideal. Similar to other studies, we found that our colleagues had a genuine interest in learning about the outcomes of the video review, and we think video reviews offer us new learning opportunities for vacuum extraction to improve and maintain high performance in obstetric teams. Future studies are needed to explore how (37).

The external validity of our findings must be considered before they can be applied in other settings. In the labor and delivery wards, where the standard analgesia is epidural/spinal, our findings on further analgesia may not be relevant. However, we believe that the key findings of this study may be applicable in other settings where teams manage vacuum extractions, and we recommend including the specific behaviors of team interaction, anticipation, avoiding fixation, and focused communication to complement any technical training on vacuum extraction.

In conclusion, mastering vacuum extraction is a core competency in delivery wards and is important for the safety of women and children. Our study demonstrated how video reviews of real-life vacuum extraction can be used to evaluate the clinical and non-technical performance of teams to identify areas that require particular attention during training. Although vaginal vacuum extraction is generally accepted as an operator-dependent task, our findings indicate that the teams' non-technical skills play an important role in ensuring high clinical performance, as the accoucheur who is conducting the vacuum delivery is "task-focused" and cannot maintain wider situational awareness. Therefore, the team members are required to "step back" to monitor the whole situation, and the key non-technical skills associated with high clinical performance were team interaction, anticipation, avoidance of fixation, and focused communication.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethical and legal approval for this study was obtained in May 2014 from the Central Denmark Region, the Danish Data Protection Agency (2012-58-006), and the Research Foundation of Central Denmark (Record No. 1-16-02-257-14). All videos were included with informed consent in conformity with the Danish code §264. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

LB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. LR: Investigation, Writing – review & editing. KH-H: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing. LH: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Writing – review & editing. KH: Conceptualization, Methodology, Writing – review & editing. OK: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. NU: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing. TM: Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. We thank the following organizations and departments for their financial support: Tryg Foundation (Trygfonden; Grant ID no. 109507), Regional Hospital in Horsens, Department of Obstetrics and Gynecology, Regional Postgraduate Medical Education Administration Office, North.

Acknowledgments

We thank all the patients who volunteered to participate in the study and consultant Lars Høj for assisting in testing the checklist. Our analysis was approved by Erik Parner, Professor of Statistics, Aarhus University.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1330457/full#supplementary-material

References

1. Vacca A. Vacuum-assisted delivery: an analysis of traction force and maternal and neonatal outcomes. *Aust N Z J Obstet Gynaecol*. (2006) 46:124–7. doi: 10.1111/j.1479-828X.2006.00540.x

2. Sau A, Sau M, Ahmed H, Brown R. Vacuum extraction: is there any need to improve the current training in the UK? *Acta Obstet Gynecol Scand*. (2004) 83:466–70. doi: 10.1111/j.0001-6349.2004.0399.x

3. Bahl R, Murphy DJ, Strachan B. Qualitative analysis by interviews and video recordings to establish the components of a skilled rotational forceps delivery. *Eur J Obstet Gynecol Reprod Biol*. (2013) 170:341–7. doi: 10.1016/j.ejogrb.2013.06.034

4. Bahl R, Murphy DJ, Strachan B. Non-technical skills for obstetricians conducting forceps and vacuum deliveries: qualitative analysis by interviews and video recordings. *Eur J Obstet Gynecol Reprod Biol*. (2010) 150:147–51. doi: 10.1016/j.ejogrb.2010.03.004

5. Bracco F, Masini M, De Tonetti G, Brogioni F, Amidani A, Monichino S, et al. Adaptation of non-technical skills behavioural markers for delivery room simulation. *BMC Pregnancy Childbirth*. (2017) 17:89. doi: 10.1186/s12884-017-1274-z

6. Chang YS, Coxon K, Portela AG, Furuta M, Bick D. Interventions to support effective communication between maternity care staff and women in labour: a mixed-methods systematic review. *Midwifery*. (2018) 59:4–16. doi: 10.1016/j.midw.2017.12.014

7. Flin R, O'Connor P, Crichton M. *Safety at the sharp end: A guide to non-technical skills; chapter 2*. US: CRC Press. (2008);17–40.

8. Joint Commission on Accreditation of Healthcare Organizations, USA.. *Preventing maternal death*. Sentinel Event Alert. (2010) 26:1–4.

9. Guise JM, Segel S. Teamwork in obstetric critical care. *Best Pract Res Clin Obstet Gynaecol*. (2008) 22:937–51. doi: 10.1016/j.bpobgyn.2008.06.010

10. Andel C, Davidow SL, Hollander M, Moreno DA. The economics of health care quality and medical errors how big a problem is quality and patient safety? *J Health Care Finance*. (2012) 39:39–50.

11. Brogaard L, Kierkegaard O, Hvidman L, Jensen KR, Musaeus P, Uldbjerg N, et al. The importance of non-technical performance for teams managing postpartum haemorrhage: video review of 99 obstetric teams. *BJOG*. (2019) 126:1015–23. doi: 10.1111/1471-0528.15655

12. Leonard M, Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care*. (2004) 13:i85–90. doi: 10.1136/qshc.2004.010033

13. Patey R, Flin R, Flwtcher G, Nicola Mara RG. Developing a taxonomy of anesthetists' nontechnical skills (ANTS). *Advan Patient Safety*. (2005) 4:1–12.

14. Bahl R, Murphy DJ, Strachan B. Decision-making in operative vaginal delivery: when to intervene, where to deliver and which instrument to use? Qualitative analysis of expert clinical practice. *Eur J Obstet Gynecol Reprod Biol*. (2013) 170:333–40. doi: 10.1016/j.ejogrb.2013.06.033

15. Gruise JJ. *HHS Public Access*. (2008) 34:352–9.

16. Hinshaw K. Non-technical skills to improve obstetric practice In: SS Arulkumaran, editor. *Best practice in labour and delivery [internet]. 2nd* ed. Cambridge: Cambridge University Press (2016). 389–400.

17. Brogaard L, Hinshaw K, Kierkegaard O, Manser T, Uldbjerg N, and Hvidman L. Developing the TeamOBS-vacuum-assisted delivery checklist to assess clinical performance in a vacuum-assisted delivery: a Delphi study with initial validation. *Front Med (Lausanne)*. (2024) 11:1330443. doi: 10.3389/fmed.2024.1330443

18. Morgan PJ, Tregunno D, Pittini R, Tarshis J, Regehr G, Desousa S, et al. Determination of the psychometric properties of a behavioural marking system for obstetrical team training using high-fidelity simulation. *BMJ Qual Saf*. (2012) 21:78–82. doi: 10.1136/bmjqs-2011-000296

19. Carstensen B. *Comparing clinical measurement methods: A practical guide*. US: John Wiley & Sons, Ltd. (2010).

20. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies Ultrasound Obstet Gynecol. (2003) 22:85–93. doi: 10.1002/uog.122,

21. Brogaard L, Uldbjerg N. Filming for auditing of real-life emergency teams: a systematic review. *BMJ Open Qual*. (2019) 8:e000588. doi: 10.1136/bmjoq-2018-000588

22. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne effect: a randomised, controlled trial. *BMC Med Res Methodol*. (2007) 7:30. doi: 10.1186/1471-2288-7-30

23. RKKP. 2022 [cited 2023 Oct 18]. Dansk kvalitetsdatabase for fødsler. (2022). Available from: https://www.sundhed.dk/content/cms/66/4666_dkf-aarsrapport-2022_offentlig.pdf

24. Jensen KR, Hvidman L, Kierkegaard O, Gliese H, Manser T, Uldbjerg N, et al. Noise as a risk factor in the delivery room: a clinical study. *PloS One*. (2019) 14:e0221860. doi: 10.1371/journal.pone.0221860

25. Murphy DJ, Strachan BK, Bahl R. Assisted vaginal birth: green-top guideline no. 26. *BJOG*. (2020) 127:e70–e112. doi: 10.1111/1471-0528.16092

26. Siassakos D, Bristowe K, Draycott TJ, Angouri J, Hambly H, Winter C, et al. Clinical efficiency in a simulated emergency and relationship to team behaviours: a multisite cross-sectional study. *BJOG*. (2011) 118:596–607. doi: 10.1111/j.1471-0528.2010.02843.x

27. Cornthwaite K, Edwards S, Siassakos D. Reducing risk in maternity by optimising teamwork and leadership: an evidence-based approach to save mothers and babies. *Best Pract Res Clin Obstet Gynaecol*. (2013) 27:571–81. doi: 10.1016/j.bpobgyn.2013.04.004

28. Siassakos D, Draycott T, Montague I, Harris M. Content analysis of team communication in an obstetric emergency scenario. *J Obstet Gynaecol (Lahore)*. (2009) 29:499–503. doi: 10.1080/01443610903039153

29. Edozien LC. Towards safe practice in instrumental vaginal delivery. *Best Pract Res Clin Obstet Gynaecol*. (2007) 21:639–55. doi: 10.1016/j.bpobgyn.2007.03.006

30. Parker SH, Flin R, McKinley A, Yule S. Factors influencing surgeons' intraoperative leadership: video analysis of unanticipated events in the operating room. *World J Surg*. (2014) 38:4–10. doi: 10.1007/s00268-013-2241-0

31. Hoyt DB, Shackford SR, Fridland PH, Mackersie RC, Hansbrough JF, Wachtel TL, et al. Video recording trauma resuscitations: an effective teaching technique. *J Trauma*. (1988) 28:435–40. doi: 10.1097/00005373-198804000-00003

32. Townsend RN, Clark R, Ramenofsky ML, Diamond DL. ATLS-based videotape trauma resuscitation review: education and outcome. *J Trauma*. (1993) 34:133–8. doi: 10.1097/00005373-199301000-00025

33. Finer NN, Rich W. Neonatal resuscitation: toward improved performance. *Resuscitation*. (2002) 53:47–51. doi: 10.1016/S0300-9572(01)00494-4

34. Jiang C, Zhao Y, Chen Z, Chen S, Yang X. Improving cardiopulmonary resuscitation in the emergency department by real-time video recording and regular feedback learning. *Resuscitation*. (2010) 81:1664–9. doi: 10.1016/j.resuscitation.2010.06.023

35. Brogaard L, Hvidman L, Esberg G, Finer N, Hjorth-Hansen KR, Manser T, et al. Teamwork and adherence to guideline on newborn resuscitation-video review of neonatal interdisciplinary teams. *Front Pediatr*. (2022) 10:828297. doi: 10.3389/fped.2022.828297

36. Lesch H, Johnson E, Peters J, Cendán JC. VR simulation leads to enhanced procedural confidence for surgical trainees. *J Surg Educ*. (2020) 77:213–8. doi: 10.1016/j.jsurg.2019.08.008

37. Davis L, Johnson L, Allen SR, Kim PK, Sims CA, Pascual JL, et al. Practitioner perceptions of trauma video review. *J Trauma Nurs*. (2013) 20:150–4. doi: 10.1097/JTN.0b013e3182a172b6

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 20 March 2024
DOI 10.3389/fmed.2024.1282421

OPEN ACCESS

EDITED BY
Fedde Scheele,
VU Amsterdam, Netherlands

REVIEWED BY
Sagrario Gomez-Cantarino,
University of Castilla La Mancha, Spain
Etelka Moll,
Onze Lieve Vrouwe Gasthuis (OLVG), Netherlands

*CORRESPONDENCE
Rong Huang
huangrong_1986@hotmail.com

†These authors share first authorship

Received 24 August 2023
ACCEPTED 07 March 2024
PUBLISHED 20 March 2024

CITATION
Wu N, Li W, Huang R and Jiang H (2024) Effect of simulation-based training workshop on obstetric emergency team collaboration and communication: a mixed study. *Front. Med.* 11:1282421. doi: 10.3389/fmed.2024.1282421

Effect of simulation-based training workshop on obstetric emergency team collaboration and communication: a mixed study

Na Wu†, Wei Li†, Rong Huang* and Hui Jiang

Nursing Department of Shanghai Key Laboratory of Maternal Fetal Medicine, Shanghai Institute of Maternal-Fetal Medicine and Gynecologic Oncology, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China

Aims and objectives: To explore the effects of simulation-based midwife training workshops and determine whether such a program can improve team collaboration and communication.

Background: Simulation training improves communication, team cooperation, critical thinking, and situational awareness.

Design: This mixed study was conducted September 15–18, 2021.

Methods: Participants included 23 obstetricians and midwives who completed 2 days of simulation training, including communication, skills, teamwork, single technical operation, and scene running. The Clinical Teamwork Scale was used before and after the comparison, and the data were analyzed using a phenomenological analytic process.

Results: The total team cooperation, transparent thinking, closed-loop communication, overall decision-making, clear responsibility, and leadership scores of the trainees were significantly higher after than before the training. The experience of attending a simulated training workshop can be divided into two themes: innovative ways of offering training and active learning. Three key themes emerged from each category: education combined with recreation; full participation in interactions; and teamwork and communication. (1) application of knowledge (2) dissemination, and (3) sublimation of knowledge.

Conclusion: This study's findings indicated a good experience and higher team cooperation score among midwives participating in simulation-based training in China, the value of our work is to show that the researched teaching methods, although published in other contexts, are also valuable in the Chinese context, suggesting that they will pass on the methods and concepts of the simulated training to others and change the current status of classroom teaching, which is its most meaningful practical training effect.

Relevance to clinical practice: These results imply that simulation-based midwife training for obstetric emergencies is required to improve the comprehensive ability of midwives to address obstetric emergencies, thereby improving maternal clinical outcomes.

No patient or public contribution: Neither patients nor the public were involved in this study, and the midwives and obstetricians voluntarily participated.

KEYWORDS

midwife, simulation-based training, experience, team cooperation, obstetric emergency situations, clinical teamwork scale

1 Introduction

The cognitive load (CL) theory was proposed by van Merriënboer and Sweller (1) and consists of the following (2): internal and extraneous loads of learning tasks and related loads of the learning process. In the CL theory, self-reported commitment and effort, stress levels, and dual-task performance can be estimated in different ways. However, intrinsic loads can be managed from simple to complex learning tasks and from low- to high-fidelity scenarios (1), which can be improved through simulation.

Simulation is considered an effective way to assess nurses (3), especially team performance. The goals of simulation-based training especially for teamwork are addressing roles, situational awareness, leadership, communication, and responsibilities (4). The modern model of surgical simulation training, also called simulation-based training (SBT), which allows surgeons to obtain clinical experience or operating room skills in the simulation environment, is increasingly emerging with technological development (5). Moreover, the Accreditation Council for Graduate Medical Education has approved and integrated SBT into surgical curricula (6).

In the light of medical errors being the third leading cause of death in America, after heart disease and cancer (7), we know that supporting others, solving problems, exchanging information, and leadership can reduce the effects of human error during the simulation training process (8). Simulation training is superior to lecture-based teaching of technical and nontechnical skills (9) in terms of competence, confidence, and communication. Adult education in the field of obstetrics and gynecology has gradually shifted from traditional approaches, including lectures and teaching in clinical settings, to simulation-enhanced education. Learning about pelvic examinations through simulation-enhanced education brings a significant benefit to medical and midwifery students, thereby improving their technical skills, comfort, and patient communication (10). The number of articles on simulations has rapidly increased and simulation-based multidisciplinary team training has become popular over the past few years. A study from Pakistan showed that team performance significantly improved after training (11), highlighting the need to teach communication and leadership skills through simulation workshops. Similarly, Grogan et al. (12) showed that training in crisis resource management generated a positive attitude among trainees regarding coping with exhaustion, team building, communication with each other, and situational awareness.

The use of simulation training has also spread rapidly in China over the past 3 years, especially in the obstetrics community. Currently, in Shanghai, almost every midwife and nursing training program is held in our simulation training center; thus, we must understand the effects of simulation training and receive participant feedback.

The launch of the three-child policy by the Chinese government has raised concerns about a shortage of midwives; in addition, there will be a higher proportion of older mothers with potential complications. The "three delays," including in seeking appropriate care, obtaining timely care, and providing care, contribute to the pooled quality of services (13). Failures in communication, event reporting, and leadership are the root causes of patient safety incidents (14). These factors challenge the entire medical system and threaten maternal and child safety. Fostering trained midwives to adapt to different situations is important, and training may play an important role (15).

Few studies have explored midwives' team collaboration and communication or their feelings regarding attending training, especially in China. Our simulation workshop emphasizes team building, communication skills (such as closed-loop communication), leadership skills, rescue processes, and situational awareness. Therefore, here we sought to understand the midwives' thoughts and determine whether there was a difference between simulation- and lecture-based training. The nursing department of our hospital organized a national continuing medical education program called The Simulation-Based Training Workshop for Midwives. This study aimed to evaluate the effects of SBT programs on obstetric emergencies.

2 Materials and methods

2.1 Study design and settings

This study utilized a quasi-experimental and comparison group research design. A workshop that enrolled 24 participants was held at the simulation center of our hospital, which has two campuses and manages nearly 30,000 births annually.

2.2 Study participants

Midwives and obstetricians were eligible to attend the simulated training until the quota of applications was full upon meeting the following criteria: (1) worked in the labor room; and (2) spoke Mandarin. Table 1 shows the participants' baseline data.

One obstetrician was unable to participate due to work conflicts. Thus, a total of 23 participants, including 20 midwives and three obstetricians, participated in the simulated training. Sixteen midwives were interviewed because no new subject words appeared and the interview data were saturated. None had previously participated in SBT. This study was approved by the Ethics Committee of the Hospital, and the number was KS2026. All participants provided written informed consent before participating.

2.3 Simulation workshop

The SBT workshop was conducted September 15–18, 2021. The study was conducted at the simulation center of the hospital, which featured high-tech simulation equipment such as a SimMother, New B of Laerdal, and a Noelle™ birthing simulator (Gaumard, Miami, FL, United States).

TABLE 1 Professional and educational characteristics of study participants.

Age, years	28.69 (21–42)
Sex	
Female	21
Male	2
Profession	
Midwife	20
Obstetrics resident	3
Hospital	
Tertiary general hospital	7 (30.4%)
Secondary general hospital	7 (30.4%)
Tertiary specialized hospital	7 (30.4%)
Private hospital	2 (8.0%)

Values are shown as median (interquartile range), *n*, or *n* (%) as appropriate.

High-fidelity simulators produce respiratory sounds, heart sounds, and palpable pulses and are equipped with a monitor that displays the electrocardiograph, fetal heart rate, blood oxygen saturation, blood pressure, arterial waveform, pulmonary artery waveform, and anesthetic gas. Moreover, these simulators can be combined with computer simulations to conduct multidisciplinary team crisis management training.

The workshop included training interventions using computerized simulators and high-tech mannequins. The participants were separated into groups of 4–6 according to different scenarios. For example, the postpartum hemorrhage scenario featured six roles; thus, the participants were separated into four teams, and we ran the postpartum hemorrhage scenario four times. The participants in each team were separated by hospital level and working age until the groups were balanced. Each was an active participant and took turns in different scenarios, including the following: gentle birth, episiotomy model of the perineum (Figures 1A,B), shoulder dystocia, neonatal asphyxia resuscitation, emergency cesarean section, postpartum hemorrhage, and preeclampsia with seizures. The trainers introduced the environment and equipment before starting each scenario.

Before the simulation workshop began, all programmed scenarios were tested at least three times. The bedside monitor displayed the vital signs during the simulated scenarios. The team members must focus on and act on physiological data from the monitor, which was operated by an experienced faculty member in the tiring-room. Each scenario lasted for 15 min and was followed by a 20 min debriefing session. Topics included teamwork, communication skills, and leadership, all of which are important elements of an effective team in obstetric emergencies. During the debriefings, the video was suspended when the facilitators asked the participants to discuss and reflect on their performances.

We hypothesized that SBT would improve teamwork performance, especially communication skills, and enhance the training experience (see Figure 2).

2.4 Data collection procedure

The Clinical Teamwork Scale (CTS) was used to assess team performance in a simulation-based obstetric emergency context (16).

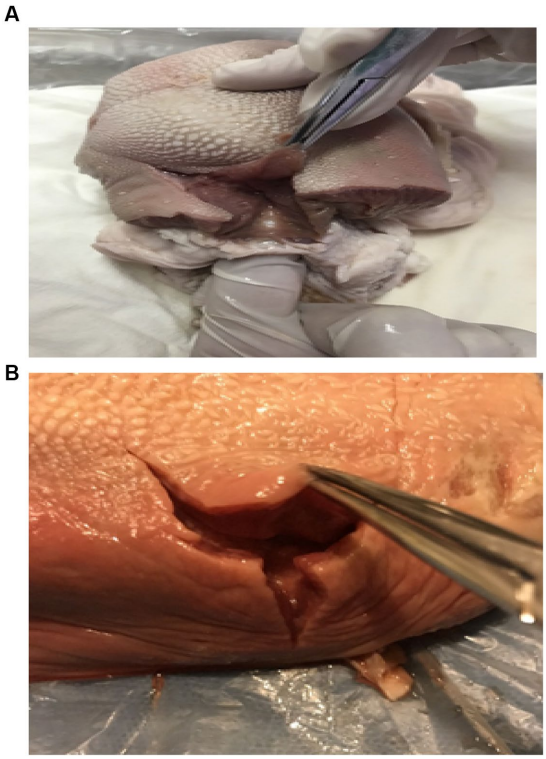


FIGURE 1
(A, B) Perineal wound using a cow's tongue.

The CTS has reported substantial agreement (kappa, 0.78) and high inter-rater reliability (interclass correlation coefficient, 0.98) (17) and contains teamwork behaviors that are applicable to any resuscitation environment. The CTS includes an overall/global teamwork rating as well as ratings for elements within five teamwork domains: communication, role responsibility, situational awareness, decision-making, and patient friendliness (all scored on a 10-point Likert scale). Item 9 is target fixation, which refers to the fact that in a simulated scene, the trainees are subjected to tunnel vision, which causes obstacles to management of the entire scene (18). This scale evaluates team performance in obstetric emergency simulations. The operational scenarios in our training workshop covered obstetric emergencies and dangerous situations; therefore, this scale was used to evaluate team cooperation.

2.5 Data management and analysis

A blinded review of the participants' performance in these scenarios was conducted by two trained nurses, one with expertise in emergency obstetrics and the other with experience in simulated education. An audit of 25% of the 32 scenarios demonstrated 99% consensus between the two reviewers.

2.6 Statistical analyses

The data were analyzed using Statistical Package for Social Sciences (version 26.0; IBM Corp., Armonk, NY, United States). CTS

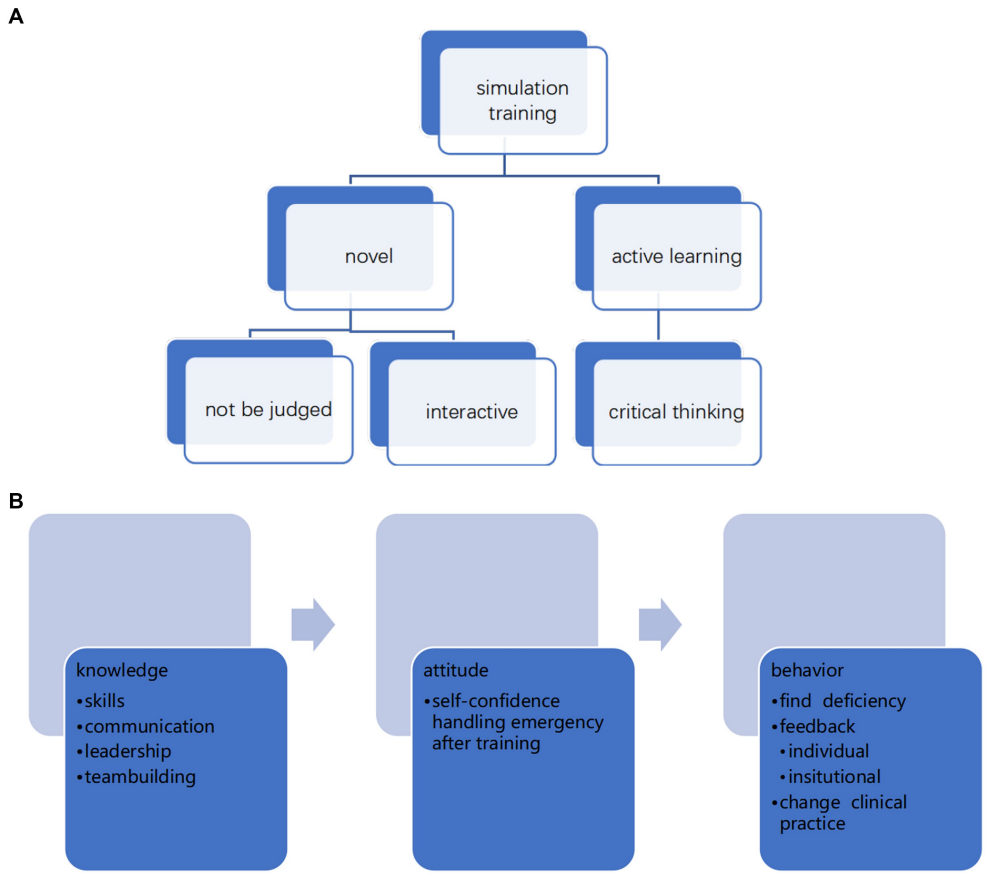


FIGURE 2
(A, B) Midwives' experiences.

TABLE 2 Semi-structured interview guide.

Questions
1. Have you heard of simulation training before?
2. Have you participated in obstetric emergency training before?
3. Why are you attending the simulation training?
4. Can you talk about your feelings about the simulation training?
5. Is there a gap with your expectations? If so, how?
6. What is the biggest impact of participating in simulation training on you?
7. How will your work change after this simulation?
8. Is there anything else you would like to share with us?

scores are described as median (range). To compare the differences in team performance between the pre- and post-training groups, the total median and median score of each CTS item were calculated, and a signed-rank test was conducted to investigate the difference in CTS scores. Statistical significance was accepted at a two-sided value of $p < 0.05$.

After the workshop, two researchers (NW and HJ) conducted in-depth interviews with 16 midwives that began with open-ended questions. An in-depth semi-structured interview outline was used (Table 2). The participants' behaviors and emotions were also noted. Each interview, which was conducted in a separate room, lasted 15–20 min. Body language was also noted to facilitate a later data analysis.

Colaizzi's (19) seven-step framework was used for the data analysis. To divide the experiences into categories, a passage reading was first performed within 24 h after the interview, and each participant's experience was read and assigned a keyword. The keywords were collected in tabular form. At the same time, we assigned every experience a number, so we could note the number and keyword of the experience in a second table and retrace how each experience was coded later. To combine the collected codes into supercategories, it was necessary to consider which keywords were similar or could be condensed and attributed to a similar behavior. In the final step, it was possible to assign all experiences to different categories. The results were cross-reviewed by two investigators and any disagreement was resolved by discussion with a third investigator. The transcribed data and extracted themes and sub themes were sent back to the participants and to ensure that the findings reflected their actual perspectives rather than the investigators' understanding of the phenomenon, a follow-up conversation was conducted.

After the workshop, the SBT was evaluated using a self-designed questionnaire, of which the content validity was mainly calculated using five expert scores, and the content validity index was 1. The questionnaire consisted of 17 workshop items that were answered using a five-point Likert-type scale (Table 3).

TABLE 3 Questionnaire including 17 workshop-related items.

N = 23	Item	Mean ± SD
Training objectives	1. The teaching goals are clear, reasonable, and practical	4.67 ± 0.48
	2. The objectives meet the needs of clinical work	4.67 ± 0.48
Simulation environment	3. The preparation matches the scenario description	4.67 ± 0.48
	4. The layout of the scene is reasonable	4.70 ± 0.46
Training content	5. There is plenty of information about the simulation scene	4.75 ± 0.44
	6. The scenarios are instructive and scientific	4.70 ± 0.46
	7. The theory is linked with practice	4.75 ± 0.44
	8. The simulation increased the knowledge from books	4.70 ± 0.46
Training methodology	9. The method is novel and interactive	4.79 ± 0.41
	10. The scenarios are scheduled for a reasonable time	4.79 ± 0.41
	11. Simulation is more useful to me than traditional lectures	4.79 ± 0.41
Training tutor	12. The tutors make me feel relaxed	4.88 ± 0.34
	13. The tutors are instructive	4.92 ± 0.28
Training application	14. The training content can be applied to my future work	5.00 ± 0.00
	15. The simulation training will be helpful to my work	4.96 ± 0.20
	16. Simulation training will be helpful to my teamwork abilities	4.96 ± 0.20
	17. Simulation training will be helpful to my communication abilities	4.96 ± 0.20

TABLE 4 Comparison of team performance pre- and post-simulation training assessed by the Clinical Teamwork Scale.

Item	Post-training	Pre-training	Statistic	*p*-value
1. Overall score	7.0 (4–9)	5.0 (4–6)	3.115	0.002
2. Overall communication	7.0 (5–9)	5.0 (3–6)	3.370	0.001
3. SBAR	7.0 (5–9)	4.0 (3–6)	3.783	0.000
4. Transparent thinking	7.0 (5–9)	6.0 (4–8)	2.410	0.016
5. Directed communication	7.0 (5–9)	5.0 (2–6)	3.313	0.001
6. Closed loop	7.0 (5–9)	5.0 (2–6)	3.552	0.000
7. Overall situational awareness	7.0 (5–8)	6.0 (4–8)	2.401	0.016
8. Resource allocation	6.0 (5–7)	6.0 (3–8)	0.722	0.470
10. Overall decision making	6.0 (5–8)	6.0 (4–8)	0.741	0.458
11. Prioritizing	6.0 (5–8)	6.0 (3–8)	1.384	0.166
12. Overall role responsibility	6.0 (5–8)	6.0 (4–8)	1.904	0.071
13. Role clarity	6.0 (5–9)	6.0 (4–8)	1.709	0.088
14. Perform as a leader/helper	7.0 (5–9)	5.0 (4–7)	3.438	0.001
15. Patient friendly	7.0 (4–9)	6.0 (4–8)	1.897	0.058

SBAR, situation, background, assessment, and recommendation.

3 Results

3.1 Team training was associated with higher median CTS

The total median score of all items was significantly higher in the post-training group (median, 7.0 range, 4.0–9.0) than in the pre-training group (median, 5.0; range, 4.0–6.0; $p = 0.002$). A comparison of the median CTS scores for communication, transparent thinking, overall situational awareness, and leadership showed a statistically significant difference between the pre- and post-training groups (Tables 4, 5).

3.2 General characteristics of interviewed study participants

In our study, the 16 midwives were 22–46 years of age, they came from different provinces of China, and their education levels ranged

TABLE 5 Comparison of presence of target fixation (item 9 of Clinical Teamwork Scale) pre- versus post-simulation training.

Item 9	Post-training (*n*)	Pre-training (*n*)	Total
No target fixation	10	7	17
Target fixation	2	4	6
Total	12	11	23

TABLE 6 Characteristics of interviewed midwives.

Code	Age	Education level	Hospital	Work experience, years
1	36	Bachelor's degree	Tertiary general hospital	11
2	25	Bachelor's degree	Secondary general hospital	3
3	28	Bachelor's degree	Tertiary specialized hospital	2
4	29	Bachelor's degree	Tertiary specialized hospital	2
5	25	Associate's degree	Tertiary general hospital	4
6	28	Associate's degree	Tertiary general hospital	4
7	23	Bachelor's degree	Tertiary specialized hospital	2
8	22	Bachelor's degree	Tertiary general hospital	6
9	22	Associate's degree	Tertiary general hospital	2
10	32	Bachelor's degree	Secondary general hospital	6
11	31	Bachelor's degree	Secondary general hospital	10
12	21	Bachelor's degree	Tertiary general hospital	1
13	33	Associate's degree	Tertiary general hospital	3
14	33	Associate's degree	Tertiary general hospital	2
15	26	Bachelor's degree	Tertiary general hospital	1
16	27	Associate's degree	Tertiary general hospital	3

from junior college to university. Table 6 summarizes the characteristics of each sample.

3.3 Effectiveness evaluation by midwife staff after training with face-to-face interviews

Satisfactory reactions emerged from the in-depth interviews. The experiences of midwives participating in the SBT can be summarized into two themes: novel training methods and active learning processes.

3.3.1 Midwives' feelings about novel training methods

In traditional Chinese teaching and training, students are often judged based on their grades, and trainees are frequently criticized. Under huge psychological pressure, most are unwilling to participate in training; rather, they passively accept the offered knowledge. SBT aims to identify and solve problems and improve capacity rather than criticize or blame. Therefore, midwives who participated in this SBT experienced new feelings.

3.3.2 Simulation training combines teaching and enjoyment

During the scenarios, the trainees interact with each other and the trainer provided a visual display of a standardized patient, especially when "family members created trouble," which made the training impressive.

Case 1

The training mode is very new. I only learned about this learning mode from textbooks before. I didn't think I would have the opportunity to experience such training after work.

Case 4

I didn't care much about patient's family psychology in the past, but the standardized patient gave me a significant impression of the patient's perception of care. When we tried to save the women, her husband continued to ask what happened to his wife. When you participate in a lecture, the teacher would just tell you, "Comfort the family," but then you may forget; however, after this simulation training, I would not.

3.3.2.1 Simulation training combines teaching with interaction

Traditional teaching involves teachers lecturing and students listening. For adult teaching, it may be difficult to stimulate a student's motivation to participate in the learning process, and the effect of training remains to be discussed.

Case 2

This model is vivid, which allows us to participate. It is more memorable than the traditional teaching model.

Case 3

I felt that all of the tutors were very professional and enthusiastic with close interactions. I learned a lot.

Case 6

I would listen to the tutors because they show how to do something and why. I didn't get bored and doze off.

Case 9

I learned "press the pubic symphysis" from a lecture and textbook about shoulder dystocia, but I didn't have the chance to operate the procedure. In the simulation training, I performed it, and the tutor provided advice. We interacted and I was given feedback.

Simulation training focuses more on student participation, arouses their enthusiasm, and encourages them to perform more work. Teachers play a guiding role and complete the teaching objectives by inspiring learning.

3.3.3 Simulation training emphasizes teamwork and communication

Workers in the labor room are not individuals; rather, they comprise a multidisciplinary team. Students further understood the importance of teamwork and communication through obstetric emergency simulation training.

Case 15

In the simulation scenario, we are required not only to form a group but also to communicate and cooperate with each other, which we usually would not be taught using books and lectures.

Case 7

Members conduct necessary action coordination, and a leader developed the team's ability to solve problems by relying on teamwork and communication.

Case 11

We just kept silent during the first scenario and did our own work, which was strange …After this workshop, I know we are all small building blocks in our work. Only by cohesion can we build a wonderful world.

3.3.4 Midwives' feelings about active learning process

3.3.4.1 Simulation training learning and application of knowledge

After the simulation training, students can take initiative to compare the results with their previous work styles and improve. These areas range from skills to theory to teamwork.

Case 2

I will bring your new ideas to our hospital so that we can help the women.

Case 4

When I was a doula, I wanted the women to give birth as soon as possible. I didn't pay attention to their psychology or comfort, which I will pay more attention to now.

Case 8

I used to aim to do my own work well. Now (after training), I know to cooperate and communicate with others as a team.

3.3.4.2 Simulation training learning and spread of knowledge

After the training, students not only changed the way they worked but also shared valuable lessons with their colleagues. This reflects an active learning process.

Case 3

I feel that your equipment and concepts are very advanced. I would like to give our leader some feedback after returning to work. Perhaps the hardware was not the same as yours, but we can change the concept step by step (laughter).

Case 5

A doctor in our hospital performed a cesarean section when the fetal heart slowed rather than attempting intrauterine fetal resuscitation. I would like to provide feedback when I return.

3.3.4.3 Simulation training turns knowledge into value

The trainees recognized the interesting methods and effects of simulation training and expected to obtain further learning and more training opportunities to constantly improve their abilities. Some were even planning to pursue teaching as part of their career path.

Case 10

I hope to learn more obstetric emergency training, such as cardiac arrest and amniotic fluid embolism.

Case 15

I hope to have the opportunity to participate in training frequently so that when I am old (working years and seniority promotion), I can also train new midwives (laughter).

Case 16

Your training method impressed me so much that I will try it with our younger (junior) midwives after I return.

3.4 Evaluation of satisfaction degree of midwife staff after training using self-assessed questionnaire

The mean scores for training objectives, simulation environment, training content, training methodology, training tutors, and training

applications are listed in Table 3. All interviewees enjoyed the simulation-based midwifery training workshop and hoped to participate in it in the future. They believed that team building and cooperation, closed-loop communication, and situational awareness could be learned through simulation training.

All participants were interested in the workshop and gained new knowledge. As Pacheco Granda and Salik (3) demonstrated, our course design ensured that the materials and equipment in these scenarios were realistic and appropriate for clinical practice with sufficient physiological complexity.

4 Discussion

This study aimed to describe midwives' attitudes, experiences, and team performance in SBT workshops in China. We found that simulation training allowed students to fully participate in operations, which can improve their skill levels, knowledge, and team cooperation scores compared with purely theoretical teaching. These findings contribute to the knowledge that trainees, especially midwives, have low seniority. Many changes in their perceptions have improved their ability to think critically and transparently. This is in line with previous evidence that, during simulated scenarios, teams could practice, reflect, and provide solutions to potential obstacles to medical care (20). Simulation training has been shown in improving knowledge acquisition (3). Moreover, participants could perform these operations on the simulator, which provided tools that allowed them to bridge theory and practice (21). This study also indicated a good experience with SBT, as participants could practice repeatedly in a familiar and safe environment without patient harm (4).

Our study showed a higher team cooperation score in the post-training group compared to the median scores in the pre-training group. Because the simulation-based team training emphasized effective communication, team cooperation, leadership, and situational awareness. This was also in line with Chang's study (22), which reported a significant difference between simulation and lecture training in total Situation Awareness Global Assessment Technique scores, and situational awareness increased mainly because of the improvement in perception abilities.

Simulation programs reportedly improve safety and communication skills (23). Thus, simulation training made the participants feel safe and they would not be judged to the same criticism as they would in traditional training. According to the CL theory, in low- and high-fidelity environments, intrinsic loads can be managed, relaxing participants and allowing them not to worry about the task itself. When adults see the value and relevance of what they learn, they spend more time learning the subject. Otherwise, they need opportunities for feedback and reflection to ensure self-improvement (24), which could be reduced in simulation training. As previous studies (25, 26) demonstrated, simulations change midwives' behaviors.

During SBT, debriefing, which is defined as feedback and a reflective response in a scenario in which participants discuss and reflect on their performance under tutor guidance, is considered the "heart and soul" link (27). First, participants were asked to briefly summarize what they had learned after running the scenario and then focus primarily on what had worked well and what could be improved rather than criticizing those things that had not gone well. Video playback played an important role in the simulations. During debriefing, the participants could see how they performed rather than how they thought they performed (28) or how things should be done according to routine rules.

Simulation training aims to identify and solve problems rather than criticize or place blame. Scores are the only criteria used in China's traditional teaching and training programs, which trainees are more likely to be criticized and are under great psychological pressure. Most are unwilling to participate in training; therefore, a safe environment should be provided (29). Trainers should create a "safe" training environment free from psychological burdens that allows trainees to actively and voluntarily participate in simulated training.

Through simulation training, midwives, especially those with low seniority, can master basic theories, knowledge, and skills in clinical work along with good communication and emergency handling abilities. These midwives can provide standardized services for patients to prevent and manage obstetric emergencies (30) and improve the clinical outcomes of pregnant women.

Most of the findings are well known in the world literature, however, the value of our work is to show that the researched teaching methods, although published in other contexts, are also valuable in the Chinese context. This enhances the validity of the methods. Our workshop was more like "training the trainers," when they returned to their respective units, students not only changed the way they worked but also shared some valuable with colleagues. They passed on the methods and concepts of simulated training and changed the current status of classroom teaching, which was the most meaningful practical training effect.

4.1 Study limitations

This study was conducted at the largest obstetrics and gynecology simulation center in Shanghai, China. Participants from the simulation training workshop comprised different levels of midwives from different hospitals throughout the country. As a qualitative study, there were very few thoughts from doctors, and we must increase the number of obstetricians and assess their views on simulated training and the training effect in future studies. They established temporary teams to protect women from postpartum hemorrhage, shoulder dystocia, etc. The team members were unfamiliar with each other, although they completed a 2 days partnership. This may have affected team performance, particularly on the first day.

4.2 Relevance to clinical practice

These results suggest that simulated training can improve the training effect more than traditional lectures. The experience of attending simulated training includes innovative ways of offering training and active learning. The simulation training made the participants feel safe and not judged as by traditional teachers. This reduces the cognitive load. Thus, the environment was relaxed, the experience was new and innovative, and the scenario setting and teaching methods demonstrated by the tutors were full of heuristic. According to the knowledge-attitude-behavior pattern, good experience (knowledge gained in an interesting way) changed traditional learning attitudes and prompted the participants to reflect and make changes to their clinical practice. Our research explored

only the experience of simulation training and team building, satisfaction with training, and improvement in trainee knowledge due to learning. In the future, we will study high-level simulation training, such as level three, referring to the application of learned behaviors and skills in clinical obstetrics practice. The fourth level refers to the effects of training on measurable clinical outcomes. Emergency obstetrics training must be performed using simulations that are safe for both healthcare providers and patients. The participants could perform repetitive operations, especially those related to rare emergencies, on a simulator without harming actual patients.

5 Conclusion

This study indicated a good experience and a higher team cooperation score of midwives participating in SBT in China. The experience completing the simulated training included innovative ways of offering training and active learning, passing the methods and concepts of simulated training on to colleagues, and changing the current status of classroom teaching, which is the most meaningful practical training effect. These findings imply that simulated training for obstetrics emergencies can be used to improve performance of both technical and nontechnical skills of midwives.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Shanghai First Maternity and Infant Hospital (Approval number: KS2026, Chairperson, Ye Luo). The patients/participants provided their written informed consent to participate in this study.

Author contributions

NW: Writing – original draft, Formal analysis, Data curation, Conceptualization. WL: Writing – original draft, Supervision, Software, Conceptualization. RH: Writing – review & editing, Software, Methodology. HJ: Writing – review & editing, Supervision.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors acknowledge the contributions of the midwives and obstetricians who participated in this study. This study is the first to report the results of simulation-based training in our simulation training center. These results provide an incentive and power to design future training plans for more midwives, which may improve overall clinical practice.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. (2010) 44:85–93. doi: 10.1111/j.1365-2923.2009.03498.x

2. Andersen SAW, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills: a randomized controlled trial. *Adv Simul*. (2016) 1:20. doi: 10.1186/s41077-016-0022-1

3. Pacheco Granda FA, Salik I. *Simulation training and skill assessment in critical care*. Treasure Island, FL: StatPearls (2022).

4. Merién AE, van de Ven J, Mol BW, Houterman S, Oei SG. Multidisciplinary team training in a simulation setting for acute obstetric emergencies: a systematic review. *Obstet Gynecol*. (2010) 115:1021–31. doi: 10.1097/AOG.0b013e3181d9f4cd

5. Chen G, Jin S, Xia Q, Wang Z, Shi Z, Chen G, et al. Insight into the history and trends of surgical simulation training in education: a bibliometric analysis. *Int J Surg*. (2023) 109:2204–13. doi: 10.1097/js9.0000000000000468

6. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA*. (2011) 306:978–88. doi: 10.1001/jama.2011.1234

7. Makary MA, Daniel M. Medical error-the third leading cause of death in the us. *BMJ*. (2016) 353:i2139. doi: 10.1136/bmj.i2139

8. Berg TA, Hebert SH, Chyka D, Nidiffer S, Springer C. Use of simulation to measure the effects of just-in-time information to prevent nursing medication errors: a randomized controlled study. *Simul Healthc*. (2021) 16:e136–41. doi: 10.1097/sih.0000000000000529

9. Noyes JA, Carbonneau KJ, Matthew SM. Comparative effectiveness of training with simulators versus traditional instruction in veterinary education: meta-analysis and systematic review. *J Vet Med Educ*. (2022) 49:25–38. doi: 10.3138/jvme-2020-0026

10. Le Lous M, Dion L, Le Ray C. Simulation training for pelvic examination: a systematic review. *J Gynecol Obstet Hum Reprod*. (2023) 52:102666. doi: 10.1016/j.jogoh.2023.102666

11. Zubair U, Zubair Z. Surgical resident training in Pakistan and benefits of simulation based training. *J Pak Med Assoc*. (2020) 70:904–8. doi: 10.5455/jpma.282116

12. Grogan EL, Stiles RA, France DJ, Speroff T, Morris JA Jr, Nixon B, et al. The impact of aviation-based teamwork training on the attitudes of health-care professionals. *J Am Coll Surg*. (2004) 199:843–8. doi: 10.1016/j.jamcollsurg.2004.08.021

13. Mohammed MM, El Gelany S, Eladwy AR, Ali EI, Gadelrab MT, Ibrahim EM, et al. A ten year analysis of maternal deaths in a tertiary hospital using the three delays model. *BMC Pregnancy Childbirth*. (2020) 20:585. doi: 10.1186/s12884-020-03262-7

14. Thomas AN, MacDonald JJ. Patient safety incidents associated with failures in communication reported from critical care units in the north west of England between 2009 and 2014. *J Intensive Care Soc*. (2016) 17:129–35. doi: 10.1177/1751143715626938

15. Kaldheim HKA, Fossum M, Munday J, Johnsen KMF, Slettebø Å. A qualitative study of perioperative nursing students' experiences of interprofessional simulation-based learning. *J Clin Nurs*. (2021) 30:174–87. doi: 10.1111/jocn.15535

16. Guise JM, Deering SH, Kanki BG, Osterweil P, Li H, Mori M, et al. Validation of a tool to measure and promote clinical teamwork. *Simul Healthc*. (2008) 3:217–23. doi: 10.1097/SIH.0b013e31816fdd0a

17. Fransen AF, van de Ven J, Merién AE, de Wit-Zuurendonk LD, Houterman S, Mol BW, et al. Effect of obstetric team training on team performance and medical technical skills: a randomised controlled trial. *BJOG*. (2012) 119:1387–93. doi: 10.1111/j.1471-0528.2012.03436.x

18. Green B, Parry D, Oeppen RS, Plint S, Dale T, Brennan PA. Situational awareness—what it means for clinicians, its recognition and importance in patient safety. *Oral Dis*. (2017) 23:721–5. doi: 10.1111/odi.12547

19. Wirihana L, Welch A, Williamson M, Christensen M, Bakon S, Craft J. Using Colaizzi's method of data analysis to explore the experiences of nurse academics teaching on satellite campuses. *Nurse Res*. (2018) 25:30–4. doi: 10.7748/nr.2018.e1516

20. Craft-Blacksheare M, Frencher Y. Using high fidelity simulation to increase nursing students' clinical postpartum and newborn assessment proficiency: a mixed-methods research study. *Nurse Educ Today*. (2018) 71:198–204. doi: 10.1016/j.nedt.2018.09.031

21. Walker DM, Holme F, Zelek ST, Olvera-García M, Montoya-Rodríguez 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) 15:117. doi: 10.1186/s12909-015-0401-7

22. Lee Chang A, Dym AA, Venegas-Borsellino C, Bangar M, Kazzi M, Lisenenkov D, et al. Comparison between simulation-based training and lecture-based education in teaching situation awareness. A randomized controlled study. *Ann Am Thorac Soc*. (2017) 14:529–35. doi: 10.1513/AnnalsATS.201612-950OC

23. Kiessling A, Amiri C, Arhammar J, Lundbäck M, Wallingstam C, Wikner J, et al. Interprofessional simulation-based team-training and self-efficacy in emergency medicine situations. *J Interprof Care*. (2022) 36:873–81. doi: 10.1080/13561820.2022.2038103

24. Ahn S-E, Rimpiläinen S. Maintaining sofia—or how to reach the intended learning outcomes during a medical simulation training. *Int J Learn Technol*. (2018) 13:115–29. doi: 10.1504/IJLT.2018.092095

25. Kumar A, Wallace EM, Smith C, Nestel D. Effect of an *in-situ* simulation workshop on home birth practice in Australia. *Women Birth*. (2019) 32:346–55. doi: 10.1016/j.wombi.2018.08.172

26. Kurup V, Matei V, Ray J. Role of *in-situ* simulation for training in healthcare: opportunities and challenges. *Curr Opin Anaesthesiol*. (2017) 30:755–60. doi: 10.1097/aco.0000000000000514

27. Johansson E, Lindwall O, Rystedt H. Experiences, appearances, and interprofessional training: the instructional use of video in post-simulation debriefings. *Int J Comput-Support Collab Learn*. (2017) 12:91–112. doi: 10.1007/s11412-017-9252-z

28. Salik I, Paige JT. *Debriefing the interprofessional team in medical simulation*. Treasure Island, FL: StatPearls (2022).

29. Coyne E, Calleja P, Forster E, Lin F. A review of virtual-simulation for assessing healthcare students' clinical competency. *Nurse Educ Today*. (2020) 96:104623. doi: 10.1016/j.nedt.2020.104623

30. Cooper N, O'Brien S, Siassakos D. Training health workers to prevent and manage post-partum haemorrhage (Pph). *Best Pract Res Clin Obstet Gynaecol*. (2019) 61:121–9. doi: 10.1016/j.bpobgyn.2019.05.008

frontiers | Frontiers in Medicine

TYPE Curriculum, Instruction, and Pedagogy
PUBLISHED 25 March 2024
DOI 10.3389/fmed.2024.1304417

OPEN ACCESS

EDITED BY
Florian Recker,
University of Bonn, Germany

REVIEWED BY
Michael J. Wolyniak,
Hampden–Sydney College, United States
Julia Radosa,
Saarland University Hospital, Germany

*CORRESPONDENCE
William Atiomo
✉ William.atiomo@mbru.ac.ae

RECEIVED 29 September 2023
ACCEPTED 04 March 2024
PUBLISHED 25 March 2024

CITATION
Atiomo W, Ennab F, Stanley A and Ezimokhai M (2024) Evaluating an obstetrics and gynecology teaching program for medical students incorporating simulation-based education underpinned by cognitive load theory.
Front. Med. 11:1304417.
doi: 10.3389/fmed.2024.1304417

Evaluating an obstetrics and gynecology teaching program for medical students incorporating simulation-based education underpinned by cognitive load theory

William Atiomo*, Farah Ennab, Adrian Stanley and Mutairu Ezimokhai

College of Medicine, Dubai Healthcare City, Mohammed Bin Rashid University (MBRU) of Medicine and Health Sciences, Dubai, United Arab Emirates

Although there have been previous publications on curriculum innovations in teaching O&G to medical students, especially utilizing simulation-based education, there have been none, as far as we know, incorporating and evaluating the outcomes using cognitive load theory. The aim of this article was to describe the introduction, implementation, and evaluation of an innovative teaching program in O&G, incorporating simulation-based education, underpinned by cognitive load theory. Cognitive load is defined as the amount of information a working memory can hold at any one time and incorporates three types of cognitive load—intrinsic, extraneous, and germane. To optimize learning, educators are encouraged to manage intrinsic cognitive load, minimize extraneous cognitive load, and promote germane cognitive load. In these sessions, students were encouraged to prepare in advance of each session with recommended reading materials; to limit intrinsic cognitive load and promote germane cognitive load, faculty were advised ahead of each session to manage intrinsic cognitive load, an open-book MCQ practice session aimed to reduce anxiety, promote psychological safety, and minimize extraneous cognitive load. For the simulation sessions, the faculty initially demonstrated the role-play situation or clinical skill first, to manage intrinsic cognitive load and reduce extraneous cognitive load. The results of the evaluation showed that the students perceived that they invested relatively low mental effort in understanding the topics, theories, concepts, and definitions discussed during the sessions. There was a low extraneous cognitive load. Measures of germane cognitive load or self-perceived learning were high. The primary message is that we believe this teaching program is a model that other medical schools globally might want to consider adopting, to evaluate and justify innovations in the teaching of O&G to medical students. The secondary message is that evaluation of innovations to teaching and facilitation of learning using cognitive load theory is one way to contribute to the high-quality training of competent future healthcare workers required to provide the highest standard of care to women who are crucial to the overall health and wellbeing of a nation.

KEYWORDS

obstetrics, gynecology, teaching, medical students, simulation, cognitive load theory

1 Introduction

The reproductive health of women determines a nation's health and the health of its future generations. The Global Strategy for Women, Children, and Adolescents' Health, agreed in 2016, emphasizes that all women have the right to the highest attainable standard of health and wellbeing, including the physical, mental, and social aspects of health (1). Nonetheless, in many parts of the world, women continue to experience a range of poor health outcomes including maternal death during pregnancy and childbirth, with 800 women dying each day globally in 2020 (2). High-quality training and continuing professional development of competent future healthcare workers are therefore crucial to the overall health and wellbeing of a nation, including its economic prosperity. Medical schools play a pivotal role in providing this required high-quality training to their medical students, to improve women's health, particularly during the Obstetrics and Gynaecology (O&G) clerkship.

Unfortunately, several challenges in teaching O&G to medical students continue to limit the ability of medical schools to provide this required high-quality training in many parts of the world. These include a decrease in delivery suite experience (3), a decline in the number of opportunities for students to learn how to perform a pelvic examination (4) and male students are also more likely to experience gender bias from patients during their O&G clerkships, therefore limiting their learning opportunities (5). Variable clinical exposure in O&G depending on the available learning opportunities at differing hospital sites can result in a variation in student clinical skill acquisition. Therefore, there is a need for medical schools to develop innovative teaching programs in O&G, to address some of these challenges, and specifically to ensure that they can continue providing the required high-quality training to their students. Some potential solutions to the challenges of teaching O&G include simulation-based education (including surgical skills training), simulated patients, and inter-professional education (6). However, it is important that curriculum innovations are based on educational principles that have been proven to promote, rather than inhibit learning.

One such educational theory is the cognitive load theory (7). Human memory, which is key to learning, consists of two key aspects: short-term (or working) memory and long-term memory. Short-term memory is finite in capacity and duration, and only able to hold up to four to seven chunks of information at any one time (8). On the other hand, long-term memory is believed to be infinite in capacity (9). During the learning process, new information is first transferred into short-term memory from where it is processed and transferred into long-term memory for future retrieval. Cognitive load is defined as the amount of information working memory can hold at any one time. Cognitive load theory incorporates three types of cognitive load—intrinsic, extraneous, and germane. Intrinsic cognitive load is the mental effort required to process new information that is directly relevant to learning. Extraneous cognitive load includes distractors that take up space in short-term memory (noise, irrelevant material on PowerPoint slides, and negative emotions in the learner such as anxiety). Germane cognitive load is the mental effort put into the acquisition/development of schemas of information to be held in long-term memory. To optimize learning, educators are encouraged to manage intrinsic cognitive load, minimize extraneous cognitive load, and promote germane cognitive load (10). High scores on the germane cognitive load (self-perceived learning) scale have been reported to be predictive of higher academic performance (11). A recent publication (2023) highlights the growing recognition of cognitive load theory as an effective theory of instruction. This is underpinned by 40 years of research and is designed to advance what is known about how students learn and how instructional methods should be designed to promote learning effectively (12).

One example of how cognitive load theory has been used to improve educational outcomes is a study (13) which found that medical students randomly assigned to practice intravenous venous line insertion using progressive training from low- to mid- to high-fidelity simulators had a higher rating on global clinical performance, communication, and technical skills than those students who trained with either a low-fidelity or high-fidelity simulator alone. It is thought (14) that the complex cognitive processes involved in consolidating, retrieving, and transferring knowledge might not have been possible if the initial working memory processing was hindered by cognitive overload. Progressive learning was thought to result in better learning outcomes because by gradual knowledge building in a low- to high-complexity sequence, students were able to increase their knowledge stored in long-term memory and were ultimately able to tackle the highest complexity situations and gain the necessary exposure to the highest complexity. In another example (15), emergency medicine resident training in the classification of orthopedic fractures was investigated by randomizing the learners into two groups. One group involved active learning with the classification chart provided after each diagnostic answer submission. However, the other group were guided by forming a diagnosis by providing the classification chart with each diagnostic question. The latter optimized germane cognitive load, and this group had higher test scores and lower perceived overall cognitive load scores. With regard to simulation-based education, four other studies also demonstrate positive benefits from applying the principles of cognitive load theory. In the first study (16), researchers investigated the effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills. They found that novice medical students who received cognitive load theory-based instructions had significantly increased cognitive load during post-training procedures compared to those who received standard instructions. This increased cognitive load was reflected in their performance, as the intervention group had a significantly lower final-product score than the control group. The second study (17) focused on ultrasound-guided internal jugular catheterization training. Using cognitive load theory principles, the researchers developed a curriculum incorporating progressive part practice in a simulation laboratory. They compared the technical proficiency of residents trained with this curriculum to those trained with a single simulation session. After three sessions, the experimental group showed significantly better hand motion and completion time scores compared to the control group. Even when assessed for retention at a later date, the experimental group still performed significantly better than the control group. In the third study (18), researchers used cognitive load theory principles to design a low-fidelity simulation (LFS) for the assessment and management of deteriorating patients. They measured the self-rated ability of undergraduate nurses in pre- and post-tests and found that their ability significantly increased after participating in the LFS. The fourth study (19) focused on using preparatory e-learning modules to improve performance in simulation-based education. The researchers developed online modules based on cognitive load theory and simulation-based

education principles and assessed their impact on cognitive load and performance. Participants who received the online modules had higher intrinsic and germane cognitive load, and lower extraneous cognitive load during the course component compared to those who did not (control). During the simulation-based objective structured clinical examination, the online modules group performed significantly better than the control group. Overall, these studies highlight the importance of considering cognitive load theory principles in the design of simulation training. In theory, simulations, such as illustrative diagrams and video clips, are assumed to promote germane cognitive load though they may increase intrinsic cognitive load.

Although there have been previous publications on curriculum innovations in teaching O&G to medical students (20–23), especially utilizing simulation-based education, there have been none, as far as we know, incorporating and evaluating the cognitive load in O&G. The aim of this article was to describe the introduction, implementation, of an innovative teaching program in O&G that incorporates simulation-based education, aimed at addressing some of the current challenges in teaching O&G to medical students and evaluation of its cognitive loads and outcomes. We believe it is a framework that other medical schools might want to consider and to provide the high-quality training required to improve women's health outcomes globally.

2 Methods

2.1 Learning environment and needs assessment

The Bachelor of Medicine Bachelor of Surgery (MBBS) program at Mohammed Bin Rashid University, Dubai, UAE (MBRU), is a 6-year undergraduate medical program divided into three phases. Phases 1 (Year 1) and 2 (Years 2 & 3) consist of basic sciences organ system courses to prepare the students for clinical clerkships in Phase 3 (Years 4–6) of the program. All students undertake a Year 3 three-credit Human Reproduction course, while the 8-week O&G clerkship takes place in Year 5.

The aim of the 8-week O&G clerkship is to familiarize students with the signs and symptoms of normal and abnormal reproductive function and to teach the basic examinations in O&G. This is achieved using a blended learning approach of face-to-face teaching on placements at both government and private hospitals, simulation sessions on the MBRU campus, and online resources, which include a study guide, range of videos, and revision material. The course aims to emphasize and reinforce skills for taking an appropriate history, performing a physical and pelvic examination, formulating a differential diagnosis as well as a treatment plan, and effectively managing patients. Students also undertake a 4-week O&G placement in the final (6th) year of the program. This takes the form of an apprenticeship which provides the students an opportunity to consolidate their knowledge, skills, and professional competencies in O&G, before graduating as doctors.

The first cohort of students were admitted in 2016. In early 2022, a review of the O&G clerkship was undertaken after the first cohort of students had completed their Year 5 clerkship. This involved several analyses. Mostly, a quantitative analysis was undertaken of formal student feedback scores for the first cohort. In addition, a qualitative (thematic) analysis was undertaken of free-text comments obtained from the students who were currently in Year 5 and had completed their O&G placements at the time of the review and a qualitative (thematic) analysis of free-text comments from a Year 6 student, reflecting on their experience of the O&G clerkship, which they completed in the 2020/2021 academic year. The quantitative analysis revealed that the key areas of strength were that 'the clerkship objectives were clearly communicated at the beginning of the placement' and that 'teaching materials were provided in advance (when appropriate)'. However, the areas identified for improvement were that 'formal teaching could be more relevant to the course objectives, feedback provided on students' clinical performance could be timely and informative' and 'ensuring that the teaching in Phases 1 and 2 of the MBBS programs better prepared the students for the clinical clerkship'.

From qualitative analysis, the top positive theme was 'good teaching and support of learning from the adjunct faculty' and the top recommendation for improvement was the need for 'centralized' MBRU-based teaching (in contrast to teaching at individual hospital placement sites to fragmented clerkship sub-sets). Informal feedback from students and faculty identified a variation in opportunities for medical students to learn clinical skills relevant to O&G, such as pelvic and obstetric examination and delivery of a baby, depending on the location of the hospital placement and the students' gender. A curriculum mapping exercise was also undertaken to identify knowledge and skills gaps in the Year 5 O&G clerkship: Core topics listed in the MBRU Year O&G 5 study guide (2021–2022) and faculty-recorded lectures on the MBRU learning management system were mapped to the Royal College of Obstetrics and Gynecology (RCOG) undergraduate curriculum available at the time of the exercise (24).

2.2 Pedagogical framework and format

The outcome of the evaluation exercise was communicated to the Dean of the medical school and other senior faculty with a proposal to address the issues identified. The proposal included the introduction of four 2- to 3-h MBRU-based centralized (for all students in the clerkship group) teaching sessions on weeks 1, 3, 5, and 7 of each 8-week clerkship rotation. The sessions were held on a Monday morning (aligning to the over-arching Year 5 curriculum delivery) and included an induction/introduction session on week 1 and sessions on labor and delivery (week 3), obstetric emergencies (week 5), and gynecology emergencies (week 7). The chosen topics addressed some of the issues highlighted in the student's feedback and were core topics important to the curriculum of all O&G clerkships globally (25). Two revision sessions before each biannual (December and May) examination (1 for each half of the cohort) were also introduced. Table 1 shows a summary of a typical day of instruction for students in the new curriculum as compared to the old.

Figure 1 illustrates a typical session and how the principles of cognitive load theory were applied. Students were encouraged to prepare in advance of each session where relevant, with recommended reading material communicated ahead of the sessions: This was designed to limit intrinsic cognitive load on the day of the session and promote germane cognitive load (8). Faculty were also advised ahead of each session to manage intrinsic cognitive load. The usual structure of the sessions included starting with a brief introduction and

TABLE 1 Summary of a typical day of instruction for students in the new curriculum as compared to the old.

	Old curriculum				New curriculum
	Students in hospital site 1	Students in hospital site 2	Students in hospital site 3	Students in hospital site 4	All students
TIME/DAY	Monday	Monday	Monday	Monday	Monday
0:730–08:30	SELF STUDY	Ward Rounds/Handover	Handover with Hospitalist	Handover in Ward	*Scheduled Discipline Teaching at MBRU
08:30–12:00	Community Obstetrics and Gynecology Clinic	Join Registrar (resident) in ward round	Ward work tutorials	Ward (Student 2 and student 5), Antenatal clinic (Student 3), Operating theater (Student 1 and Student 4)	
13:00–16:30		Tutorial in hospital from 12:00–13:00 then MBRU longitudinal theme teaching afternoon from 14:00 to 17:00	MBRU longitudinal theme teaching afternoon from 14:00 to 17:00	MBRU longitudinal theme teaching afternoon from 14:00 to 17:00	MBRU longitudinal theme teaching afternoon from 14:00 to 17:00

*All students. From 09:00 to 12:00 h. MBRU, Mohammed Bin Rashid University.

Title: LABOR & DELIVERY	CLT principles
Pre-reading material (Text and video).	ICL and GCL
Program.	
1- 10-minute overview of normal labour and delivered to activate prior knowledge and promote germane cognitive load. Pitched at the level of medical students to manage intrinsic cognitive load.	ICL and GCL
2- Interactive tutorial by invited faculty.	ICL
3- Open book MCQ practice	ECL
4- 15-minute break	
5- OSCE scenario of students performing normal vaginal delivery on a simulator.	ICL and ECL
6- Feedback and evaluation.	Leppink CL scale

ICL= Intrinsic Cognitive Load, ECL = Extraneous cognitive load, GCL = Germane Cognitive Load, CL= Cognitive Load.

FIGURE 1
Description of the program and the application of the principles of cognitive load theory. ICL, Intrinsic cognitive load, ECL, extraneous cognitive load, GCL, germane cognitive load, CL, cognitive load.

overview of the topic, followed by a 25-min interactive 'flipped classroom' discussion with the faculty. This was followed by a 20-min open-book MCQ practice session: These formative questions were relevant to the topic and written in a similar format to those used in their end-of-block examinations. The purpose of the open-book MCQ format, in contrast to closed-book MCQs, was to reduce anxiety, promote psychological safety, and minimize extraneous cognitive load (26). Students were permitted to discuss possible solutions in groups of two or three. This was followed by a 15-min break, after which the students engaged in a 45-min simulation scenario relevant to the topic of the day. The induction day included an introduction to history taking, pelvic examination, obstetric examination, and revision of key clinical O&G conditions which the students had been taught in their Year 3 Human Reproduction course. The scenarios for the induction session in week 1 included practice on task trainers on pelvic and obstetric examinations. During the labor and delivery session on week 3, each student performed a normal vaginal delivery on a task trainer. The obstetric emergency session on week 5 involved an OSCE scenario on the management of a patient with post-partum hemorrhage and the gynecology emergency session on week 7 involved an OSCE scenario of a simulated patient with an ectopic pregnancy. These scenarios start with the facilitating faculty, initially demonstrating the role-play situation or clinical skill being taught first, before asking the students to participate. This aimed to manage intrinsic cognitive load and reduce extraneous cognitive load (8), (27). The sessions concluded with a 10-min feedback and evaluation session during which students completed an online structured questionnaire and provided verbal instant feedback on their perceptions of how the sessions went. The revision sessions held before each biannual (December and May) examination, which involved students rotating through three mock OSCE scenarios on pelvic examination, obstetric examination, and gynecology history taking, on task trainers and a simulated patient,

TABLE 2 Student feedback questionnaire used.

Student feedback questionnaire used		
1	The topics covered in the session were very complex.	IL
2	The session covered theories that I perceived as very complex	
3	The session covered concepts and definitions that I perceived as very complex.	IL
4	The instructions and/or explanations during the session were very unclear.	EL Ins
5	The instructions and/or explanations during the session were in terms of learning very ineffective.	EL Ins
6	The instructions and/or explanations during the session were full of unclear language	EL Ins
7	Low quality audio made the instructions hard to follow.	EL Ins
8	Noises in the environment made it difficult to focus on the learning content.	EL Noi
9	Distractions in the environment made learning ineffective.	EL Noi
10	Unrelated events occurring in the environment made it difficult to focus.	EL Noi
11	My activities on my phone/computer made it difficult to focus on the learning content.	EL Dev
12	Messages and notifications from my phone/computer made learning unclear.	EL Dev
13	Others' phone/computer use distracted me, making it hard to learn.	EL Dev
14	Technical issues made learning ineffective.	EL Dev
15	Problems with technology made it difficult to focus.	EL Dev
16	The session really enhanced my understanding of the topic(s) covered.	GL/SPL
17	The session really enhanced my knowledge and understanding of obstetrics and gynecology.	GL/SPL
18	The session really enhanced my understanding of the theories covered.	GL/SPL
19	The session really enhanced my understanding of concepts and definitions.	GL/SPL
20	Please feel free to include any free text comments below.	

IL, Intrinsic load; EL Ins, extraneous load instructions; EL Noi, extraneous load noises; EL Dev, extraneous load devices; GL, germane load; SPL, self-perceived learning.

conducted in a similar style to the actual examination. Feedback was provided to students on their performance and general expectations during the final OSCE examinations.

The practicalities of implementing the sessions included communicating the plans to key faculty and liaising with key staff at the simulation center at MBRU to support delivering the sessions and timetabling. A meeting was held with the faculty facilitating each session, or an email was sent, the week before the session, to discuss the plan, during which cognitive load theory was discussed. Students did not have a formal teaching session on cognitive load theory as part of their education. However, on several occasions across the academic year, the principles of the theory were informally discussed with the students as part of the advice given to support their study skills. MBRU also holds an annual medical education meeting for faculty participating in the medical student teaching program during which the principles of cognitive load theory are discussed. The sessions were delivered during the 2022–2023 academic year with a slight modification following feedback from the first cohort of students for a need to consider introducing a 10-min brief overview of the topic by an in-house faculty member of the day in a didactic format at the start of each session.

2.3 Evaluation

The sessions were evaluated by analyzing the results of data obtained after students had anonymously completed an online structured questionnaire (Supplementary Figure S1 and Table 2). The questionnaire had a quantitative and qualitative domain. The quantitative domain was based on an expanded cognitive load scale (28) initially developed by Leppink et al. (29) and is one of the most validated and widely used self-report measures of intrinsic load (IL), extraneous load (EL), and germane load (GL). The introductory paragraph to the questionnaire informed students that the inventory measured their cognitive load during the teaching or revision session. They were advised to read each of the questions carefully and mark their responses to each question, on a rating scale from 0 (not all the case) to 10 (completely the case). The three components measured included the intrinsic cognitive load (ICL) of the sessions, the extraneous cognitive load (ECL), and the germane cognitive load or self-perceived learning (SPL) (11, 30). The qualitative domain asked, 'Please feel free to include any free text comments below'.

The data were analyzed using Microsoft Excel. The means, standard deviations, median, and interquartile ranges were calculated for all the scores obtained in response to the questions relevant to the cognitive load domain. Specifically, all the scores in response to questions 1 to 3 in Table 2 ('The topics covered in the session were very complex', 'The session covered theories that I perceived as very complex', and 'The session covered concepts and definitions that I perceived as very complex') were used to calculate the mean and median intrinsic cognitive load (IL) imposed by the teaching sessions. The scores in response to questions 4–7 in Table 2 ('The instructions and/or explanations during the session were very unclear', 'The instructions and/or explanations during the session were in terms of learning very ineffective', 'The instructions and/or explanations during the session were full of unclear language', and 'Low quality audio made the instructions hard to follow') were used to calculate the mean and median extraneous cognitive load stemming from instructions (EL

Ins). The scores in response to questions 8–10 in Table 2 ('Noises in the environment made it difficult to focus on the learning content', 'Distractions in the environment made learning ineffective', and 'Unrelated events occurring in the environment made it difficult to focus') were used to calculate the mean and median extraneous cognitive load stemming from noises (EL Noi). The scores in response to questions 11–15 in Table 2 ('My activities on my phone/computer made it difficult to focus on the learning content', 'Messages and notifications from my phone/computer made learning unclear', 'Others' phone/computer use distracted me, making it hard to learn', 'Technical issues made learning ineffective', and 'Problems with technology made it difficult to focus') were used to calculate the mean and median extraneous cognitive load stemming from devices (EL Dev). Finally, the scores in response to questions 16–19 in Table 2 ('The session really enhanced my understanding of the topic(s) covered', 'The session really enhanced my knowledge and understanding of obstetrics and gynecology', 'The session really enhanced my understanding of the theories covered', and 'The session really enhanced my understanding of concepts and definitions') were used to calculate the mean and median germane cognitive load or self-perceived learning (GCL/SPL). The student's responses to the qualitative domain asking, 'Please feel free to include any free text comments below', were analyzed with a natural language processing program on 3 July 2023 (31) with the prompt being 'Provide a thematic analysis of the text in the following table', with the responses of free-text students copied and pasted from the table of results into the prompt box, after the prompt. The natural language processing program used was ChatGPT. Developed by OpenAI, ChatGPT is a language model that utilizes natural language processing within the field of artificial intelligence. Its purpose is to provide text-based responses to queries in a manner that simulates human conversation. ChatGPT is constructed using the Transformer deep learning architecture, allowing it to recognize language patterns and generate coherent and realistic text. Through training on extensive text data, it has garnered the ability to generate responses across various subjects, ranging from basic inquiries to intricate conversational topics.

2.4 Institutional review board considerations

Formal institutional review board approval was not sought for the study because it was a desk-based retrospective review of student feedback data that did not involve any direct patient contact. The Medical Research Council (MRC) Regulatory Support Center/UK NHS Health Research Authority (HRA) online decision support tool[1] does not classify the study as research. Nevertheless, the study was submitted to the chair of the MBRU institutional review board (IRB) committee for review, who provided written confirmation that the study was exempt and did not require their approval.

1 http://www.hra-decisiontools.org.uk/research/ accessed on 4 July 2023

3 Results

The new centralized MBRU-based O&G teaching sessions were delivered from 23 August 2022 to 24 April 2023. There were 18 sessions in total, with four sessions (induction to O&G, labor and delivery, obstetric emergencies, and gynecology emergencies) delivered on the Monday morning of weeks 1, 3, 5, and 7 for each of the four, 8-week O&G clerkships over the academic year (16 in total) and two revision sessions before each biannual (December and May) examination (one session for each half of the cohort). In total, 48 students were in this Year 5 cohort; thus, 12 students rotated through each 8-week O&G clerkship. The maximum possible number of responses to the online structured questionnaire that students completed at the end of each session was 240 (16 sessions × 12 students and 2 sessions × 24 students): 206 responses were submitted and thus a response rate of 86%.

3.1 Students self-reported scores of cognitive loads

Table 3 presents the results of the students' self-reported scores of cognitive load scales associated with the new O&G teaching sessions at MBRU. The mean (standard deviation (SD)) and median (interquartile range (IQ)) intrinsic cognitive scores from all sessions were 3.9 (2.9) and 3 (1-6) respectively. The mean and median scores of measures of extraneous cognitive load ranged from 1 to 1.5. The mean (SD) and median (IQ range) scores for measures of germane cognitive load or self-perceived learning on the cognitive load scale used to evaluate all the sessions were 9.5 (1.3) and 10 (10-10).

The sessions that imposed the highest intrinsic cognitive load were those on obstetric emergencies and labor and delivery with mean scores of 4.5 (3) and 4.4 (2.8), respectively, and median scores of 3 (2-7) and 4 (2-7), respectively. The session that imposed the lowest intrinsic cognitive load was on gynecological emergencies with mean and median scores of 3.4 (2.9) and 2 (1-6) respectively. The session on obstetric emergencies, which imposed the highest intrinsic cognitive load, also had the highest measures of germane cognitive load or self-perceived learning with mean (SD) and median (IQ range) scores of 9.7 (0.7) and 10 (10-10) respectively. However, the revision sessions which were associated with low mean and median scores of intrinsic cognitive load 3.5 (2.9) and 2 (1-5) were also associated with the highest measures of germane cognitive load or self-perceived learning with mean (SD) and median (IQ range) scores of 9.7 (0.6) and 10 (10-10), respectively.

3.2 Thematic analysis of student's free-text comments

The results of the thematic analysis of the student's free-text feedback are presented in Table 4. Overall, the thematic analysis found a positive response to the sessions, highlighting their efficacy in preparing participants for their rotations, reinforcing knowledge, and providing hands-on experience. Participants also provided valuable suggestions for improvement to enhance future sessions.

TABLE 3 Students self-reported scores of the different components of the cognitive load scale, associated with the O&G teaching sessions at MBRU.

	All sessions	Clerkship introduction	Labor & delivery	Obstetric emergencies	Gynecology emergencies	Revision
Mean ICL (SD)	3.9 (2.9)	3.5 (2.6)	4.4 (2.8)	4.5 (3)	3.4 (2.9)	3.5 (2.9)
Median ICL (IQ range)	3 (1–6)	2 (1–5)	4 (2–7)	3 (2–7)	2 (1–6)	2 (1–5)
Mean ECL INS (SD)	1.2 (1)	1.4 (1.6)	1.5 (0.5)	1 (0.8)	1.3 (1.3)	1.1 (0.3)
Median ECL INS (IQ range)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)
Mean ECL NOI (SD)	1.2 (1)	1.2 (1.4)	1.1 (0.6)	1.1 (0.4)	1.4 (1.40)	1.1 (0.2)
Median ECL NOI (IQ range)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)
Mean ECL DEV (SD)	1.1 (0.9)	1.3 (1.4)	1.0 (0.3)	1.1 (0.4)	1.3 (1.4)	1.1 (0.2)
Median ECL DEV (IQ range)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)	1 (1–1)
Mean GCL/SPL (SD)	9.5 (1.3)	9.3 (1.7)	9.4 (1.4)	9.7 (0.7)	9.4 (1.4)	9.7 (0.6)
Median GCL/SPL (IQ range)	10 (10–10)	10 (9–10)	10 (9.8–10)	10 (10–10)	10 (9–10)	10 (10–10)

4 Discussion

The results of the evaluation of this new teaching program in O&G described in our study found that the students perceived that they invested relatively low mental effort in understanding the topics, theories, concepts, and definitions discussed during the sessions. There was low extraneous cognitive load because of the nature of the instructions or distractions from noise, or electronic devices (phone, computer, or technical issues) and measures of germane cognitive load or self-perceived learning were high. Students expressed appreciation for the organized and practical nature of the sessions, as well as the guidance provided by instructors. Areas for improvement were also identified, such as incorporating additional teaching materials and allowing for more time for practice. These changes have already been implemented for the subsequent academic year (2023–2024).

Although no previous study had measured cognitive load during simulation teaching of O&G, one study (30) measured cognitive load in 41 Year 5 (final year) students undergoing simulation teaching in a medical and surgical scenario. The median intrinsic cognitive load was slightly lower (score of 3) in our O &G scenario study compared with the scores (3.7 to 4.2) in the medical and surgical scenario study (23). The median self-perceived learning scores in our O&G scenario study were higher (score of 10) than the scores (6 to 6.8) in the study of medical and surgical scenarios (30), but the median extraneous cognitive load scores were similar in both studies (1 and 0.9). As high scores on the self-perceived learning scale have been reported to be predictive of high academic performance (11), these results could be interpreted as better achievement of learning outcomes in our study. On the other hand, the different results could be due to different designs of simulation teaching or the context for the studies. It is possible that the The low intrinsic cognitive load scores observed in our study might be attributed to the fact that these sessions occurred in parallel with clinical placements. Thus, students interact with real patients and problems, which may have prepared them for learning, as well as receiving pre-reading material ahead of the sessions. The location of the sessions outside hospital placements and within a dedicated simulation learning environment may have reduced extraneous cognitive load alongside the other measures we adopted during the sessions (e.g., open-book MCQs) to reduce this. Effective learning, as reflected in the high self-perceived learning/germane cognitive load scores, might also have taken place because of the mindset and the multiple methods of teaching, including simulation.

The sessions on obstetric emergencies and labor and delivery imposed the highest intrinsic cognitive load (a higher mental effort invested in learning) as the students may have found these tasks 'complex'. These were OSCE-style scenarios simulating the clinical environment. The first was on a patient with post-partum hemorrhage and the second required the student to perform a vaginal delivery on a task trainer. These findings are consistent with previous research showing that simple tasks help students gain more self-confidence (32) but that complex simulated clinical environments impose greater intrinsic and extraneous cognitive load and stress on students. In contrast, the session that imposed the lowest intrinsic cognitive load was on a gynecological emergency, which involved a simulated OSCE scenario of a patient with an ectopic pregnancy. The latter was undertaken in week 7, and thus, students may have become more familiar with the style of teaching or the scenario was not perceived as complex as the earlier simulations.

Moreover, it was intriguing that there were high scores of self-perceived learning in the 'complex' sessions on obstetric emergencies and labor and delivery, which had imposed the highest intrinsic cognitive load. This does not quite fit what we would have expected; however, a recent randomized controlled trial in simulation-based teaching might provide some insight. This found that environmental complexity contributes to intrinsic cognitive load, but students seemed to strategically manage their own cognitive load and learn from these simulations (33). On the other hand, the associated drama, anxiety, and excitement during the simulation sessions on labor and delivery, could have made it a challenge to identify the specific learning outcomes. This infers that the relationship between intrinsic cognitive load, environmental complexity, and learning gained may not be straightforward. Thus, more research is required to clarify the link between task complexity, cognitive load, and learning in simulation-based teaching. It may be the case that increased intrinsic cognitive load is a 'price to pay' for acquiring germane cognitive load in simulation. However, from a practical perspective, good practice would be for task complexity to be adapted to the expertise level of the learners and increased progressively as they become more competent.

The expanded scale in our student evaluation used a multidimensional conceptualization of the extraneous load construct

TABLE 4 Thematic analysis of student's free-text comments:

1	Appreciation and gratitude: Many participants expressed their gratitude and appreciation for the sessions, citing how helpful and beneficial they were in preparing them for their rotations or exams. They thanked the organizers and instructors for their efforts.
2	Preparation and knowledge reinforcement: Participants mentioned how the sessions helped them prepare for their rotations, solidify their theoretical knowledge, and apply their learning in practical scenarios. They highlighted the importance of gaining confidence and being more comfortable in clinical settings.
3	Suggestions for improvement: Some participants offered suggestions for improvement, such as adding more time for practice, including visual aids or videos to enhance learning, and providing clearer guidelines or instructions.
4	Hands-on experience and interactive learning: Participants valued the hands-on experience and interactive nature of the sessions, which involved simulations and discussions. They appreciated the opportunity to practice different roles, receive feedback, and engage in case-based scenarios.
5	Constructive criticism: While most participants have positive feedback, a few provided constructive criticism regarding aspects such as visibility during demonstrations, the need for more formal teaching sessions, or the inclusion of specific visual aids to enhance the learning experience.
6	Positive impact on confidence and skills: Many participants expressed how the sessions enhanced their confidence, knowledge, and clinical skills. They mention feeling more prepared to assist during deliveries, handle emergencies, or approach certain conditions.
7	Organization and time management: Several participants appreciated the organization and structure of the sessions. They highlighted the importance of effective time management and proper planning, including the allocation of sufficient time for each activity.
8	Reinforcing theoretical concepts: Many participants emphasized the relevance of bridging theory and practice. They mention the benefits of discussing concepts, clarifying doubts, and reinforcing theoretical knowledge through practical application.

that was relevant to physical and online teaching environments. The expanded Leppink cognitive load scale includes items related to instructions/explanations with sub-dimensions, including extraneous load stemming from noises, and extraneous load stemming from both media and devices within the environment. The Leppink scale was used because of its wide use and its perceived ability to measure cognitive load in more realistic learning environments, consistent with the learning environment of our students. We were also mindful that teaching with clinical simulation can induce both emotional and cognitive overload (30) and were keen to objectively evaluate this.

The findings from the needs assessment prior to introducing the new program are consistent with previous research showing that core knowledge and competencies acquired during O&G clerkships vary widely depending on the medical school (25). The call for a need to ensure that teaching in the earlier phases of the MBBS program should better prepare them for the clinical clerkship is consistent with the finding that clinical reasoning requires knowledge in real or simulated clinical environments (34). This matter was addressed by introducing a formal induction session, which included an introduction to history taking and revision of core O&G clinical conditions, which the students had previously been taught during the Year 3 Human Reproduction course.

The challenge of gender bias against male students, which was one of the factors leading to the curricular reforms in this study, should hopefully be addressed by this new program. As these barriers might persist post-qualification, it is important that postgraduate educational programs (residency and continuing professional development) also explore innovative educational curricula to ensure high-quality training and professional development to address this challenge. It is interesting to note that in a systematic review of 15 studies (35), patients prioritized their physician's care, technical skills, compassion, and experience over gender when choosing their obstetricians and gynecologists. Therefore, barriers to learning because of gender bias may be less of a challenge for practicing doctors compared with medical students.

The strengths of this innovative program curriculum in teaching O&G to medical students are in the incorporation of cognitive load theory in the design, implementation, and evaluation. The process was also consistent with the recommended steps of curriculum development (36). The format used in the sessions was rated highly by students in a previous publication (20). To adapt this format further, we were, however, mindful of underpinning our sessions with principles of cognitive load theory (37).

There were some limitations. This included the lack of a control group, lack of student assessment data to objectively measure learning (examination results), response bias, and potentially limited generalizability.

It was difficult to develop an appropriate and ethical control group because of the nature of the study. Moreover, the complexity of the intervention would have made it a challenge to develop a control group; for example, the type of cognitive load addressed shifted with different aspects of the new curriculum. There were multiple interventions. Students were encouraged to prepare in advance of each session with recommended reading materials; this was aimed to limit intrinsic cognitive load and promote germane cognitive load, mindful that the teaching would involve a brief introduction followed by a 'flipped classroom' discussion. In addition, faculty were advised ahead of each session to manage intrinsic cognitive load during the 25-min tutorial. The format of the 20-min open-book MCQ practice session aimed to reduce anxiety, promote psychological safety, and minimize extraneous cognitive load. For the simulation sessions, the faculty initially demonstrated the role-play situation or clinical skill first, which aimed to manage intrinsic cognitive load and reduce extraneous cognitive load. It would therefore have been challenging to determine which specific component of the intervention worked, even with a control group. The feasibility and resource implications of a systematic variation in the various components of the intervention (e.g., types of cognitive load) to determine which component worked, in a study with a control group or groups would also have been challenging, and particularly so with a small cohort of less than 50 students. However, as proof of concept in our setting, this study has provided preliminary data to inform the design of future studies.

With respect to assessments, objective measures of learning outcomes, such as summative examination results, could provide more robust evidence of the effectiveness of the program. However, the student feedback in the study was collected anonymously and it was not possible to link responses from individual students to their examination results. The O&G summative examination results from both the cohort of 48 students who underwent the novel teaching program and the cohort in the previous academic year (34 students) were analyzed: There was no significant difference in either OSCE or theory results. This is not surprising. There are challenges with assessing the effectiveness of any new teaching program because of confounding variables. Some of these include test anxiety, variation in marking standards, student motivation, student social interactions outside of the classroom, and the student's independent study. Furthermore, the sessions only covered a part of the syllabus, whereas the O&G examination covered wider aspects. Longitudinal assessment data might have also provided insights into the sustainability of the effects of program on students' learning outcomes and clinical performance. However, this was not possible as the data were obtained from the students anonymously.

As the study was conducted at a specific institution, this limits the generalizability of the proposed teaching program to other medical schools. Resource limitations may also impact the feasibility and scalability of this innovative teaching program to other medical schools, especially those in low-resource settings, as building simulation centers with high-fidelity manikins require modern and expensive equipment (38). On the other hand, as long as the theoretical principles of cognitive load theory underpin the program, creative solutions (39) such as simulated patients and part-task trainers could provide a starting point in low-resource settings.

Feedback mechanisms might be biased toward students who are more vocal or have stronger opinions. However, our survey response bias was unlikely given the high response rate (86%) to the student evaluation questionnaires.

The order in which the questions were asked in the survey tool used may also have influenced the results, as it has been shown that asking learners about their intrinsic cognitive load (ICL) first, makes them give higher ICL ratings compared to asking them about their extraneous cognitive load (ECL) first (40). Finally, as far as we know, there is not a defined acceptable cutoff value for the different types of cognitive load we measured using the cognitive load measurement scale we used in our study. However, on a rating scale of 1–10, the mean and median scores of 9.5 and 10, of germane cognitive load or self-perceived learning, demonstrated in our study were consistent with positive learning outcomes.

In conclusion, the introduction, implementation, and evaluation of this innovative way of teaching O&G to medical students, underpinned by cognitive load theory, demonstrated positive outcomes. It is a model that other medical schools globally might want to consider, raise standards in teaching O&G (and other subjects) to medical students, and address some of the current educational challenges facing the teaching of O&G. The promotion of learning using cognitive load theory in the students who attended these sessions, should hopefully, contribute to the high-quality training of sufficient future healthcare workers required to provide the highest standard of care to women who are crucial to the overall health and wellbeing of a nation.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WA: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. FE: Project administration, Supervision, Writing – review & editing. AS: Writing – review & editing, Formal analysis. ME: Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. All authors are funded by the College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai Healthcare City, Dubai, United Arab Emirates.

Acknowledgments

The authors would like to thank colleagues at the Khalaf Ahmed Al Habtoor Medical Simulation Center, Mohammed Bin Rashid University Medicine and Health Sciences, for their support in delivering the sessions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1304417/full#supplementary-material

References

1. van den Broek N. Keep it simple – effective training in obstetrics for low- and middle-income countries. *Best Pract Res Clin Obstet Gynaecol*. (2021) 80:25–38. doi: 10.1016/j.bpobgyn.2021.10.007

2. Lawrence ER, Klein TJ, Beyuo TK. Maternal mortality in low and middle-income countries. *Obstet Gynecol Clin N Am*. (2022) 49:713–33. doi: 10.1016/j.ogc.2022.07.001

3. Hogan E, Woods C, Buttrose M, Abenthum L, Cheng HC, de Costa C. The changing birth suite experience for Australian medical students. *Aust N Z J Obstet Gynaecol*. (2016) 56:537–42. doi: 10.1111/ajo.12495

4. Janjua A, Roberts T, Okeahialam N, Clark TJ. Cost-effective analysis of teaching pelvic examination skills using Gynaecology teaching associates (GTAs) compared with manikin models (the CEAT study). *BMJ Open*. (2018) 8:e015823. doi: 10.1136/bmjopen-2017-015823

5. Chang JC, Odrobina MR, McIntyre-Seltman K. The effect of student gender on the obstetrics and gynecology clerkship experience. *J Womens Health*. (2010) 19:87–92. doi: 10.1089/jwh.2009.1357

6. Kumar A, Ameh C. Start here- principles of effective undergraduate training. *Best Pract Res Clin Obstet Gynaecol*. (2021) 80:114–25. doi: 10.1016/j.bpobgyn.2021.11.010

7. van Merriënboer JJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ*. (2010) 44:85–93. doi: 10.1111/j.1365-2923.2009.03498.x

8. Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE guide no. 86. *Med Teach*. (2014) 36:371–84. doi: 10.3109/0142159X.2014.889290

9. Sweller J, van Merriënboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. *Educ Psychol Rev*. (2019) 31:261–92. doi: 10.1007/s10648-019-09465-5

10. Ghanbari S, Haghani F, Barekatain M, Jamali A. A systematized review of cognitive load theory in health sciences education and a perspective from cognitive neuroscience. *J Educ Health Promot*. (2020) 9:176. doi: 10.4103/jehp.jehp_643_19

11. Bergman EM, de Bruin ABH, Vorstenbosch MATM, Kooloos JGM, Puts GCWM, Leppink J, et al. Effects of learning content in context on knowledge acquisition and recall: a pretest-posttest control group design. *BMC Med Educ*. (2015) 15:133. doi: 10.1186/s12909-015-0416-0

12. Hanham J, Castro-Alonso JC, Chen O. Integrating cognitive load theory with other theories, within and beyond educational psychology. *Br J Educ Psychol*. (2023) 93:239–50. doi: 10.1111/bjep.12612

13. Brydges R, Carnahan H, Rose D, Rose L, Dubrowski A. Coordinating progressive levels of simulation fidelity to maximize educational benefit. *Acad Med*. (2010) 85:806–12. doi: 10.1097/ACM.0b013e3181d7aabd

14. Fraser KL, Ayres P, Sweller J. Cognitive load theory for the Design of Medical Simulations. *Simul Healthc*. (2015) 10:295–307. doi: 10.1097/SIH.0000000000000097

15. Cramer N, Zuckerbraun NS, Puller J, Furtado AD, Deb A, Dorfsman ML, et al. Putting theory to practice: applying cognitive load theory to resident medical education. *Pediatr Emerg Care*. (2022) 38:e771–5. doi: 10.1097/PEC.0000000000002371

16. Andersen SAW, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. The effect of implementing cognitive load theory-based design principles in virtual reality simulation training of surgical skills: a randomized controlled trial. *Adv Simul*. (2016) 1:20. doi: 10.1186/s41077-016-0022-1

17. McGraw R, Chaplin T, Rocca N, Rang L, Jaeger M, Holden M, et al. Cognitive load theory as a framework for simulation-based, ultrasound-guided internal jugular catheterization training: once is not enough. *CJEM*. (2019) 21:141–8. doi: 10.1017/cem.2018.456

18. Say R, Visentin D, Betihavas V, Minutillo S. A cognitive load theory simulation design to assess and manage deteriorating patients. *Int J Nurs Educ Scholarsh*. (2019) 16:9. doi: 10.1515/ijnes-2019-0009

19. Gutierrez G, Lunsky IO, Van Heer S, Szulewski A, Chaplin T. Cognitive load theory in action: e-learning modules improve performance in simulation-based education. A pilot study. *Can J Emerg Med*. (2023) 25:893–901. doi: 10.1007/s43678-023-00586-z

20. Kevelighan EH, Duffy S, Walker FF. Innovations in teaching obstetrics and gynaecology–the theme afternoon. *Med Educ*. (1998) 32:517–21. doi: 10.1046/j.1365-2923.1998.00241.x

21. Paterson H, Kenrick K, Wilson D. Teaching the Y generation obstetrics and gynaecology skills: a survey of medical students' thoughts on a new program. *Aust N Z J Obstet Gynaecol*. (2012) 52:151–5. doi: 10.1111/j.1479-828X.2012.01415.x

22. Satin AJ. Simulation in obstetrics. *Obstet Gynecol*. (2018) 132:199–209. doi: 10.1097/AOG.0000000000002682

23. Everett EN, Forstein DA, Bliss S, Buery-Joyner SD, Craig LB, Graziano SC, et al. To the point: the expanding role of simulation in obstetrics and gynecology medical student education. *Am J Obstet Gynecol*. (2019) 220:129–41. doi: 10.1016/j.ajog.2018.10.029

24. RCOG. *Undergraduate training in O&G*. (2009). Available at:https://www.rcog.org.uk/careers-and-training/training/curriculum/undergraduate-training-in-og/.

25. Atiomo WU, Stanley AG, Ezimokhai MM. A literature review and proposed framework for a core curriculum in obstetrics and gynecology for medical students globally. *Int J Gynaecol Obstet*. (2023) 161:386–96. doi: 10.1002/ijgo.14575

26. Atiomo W, Mbaki Y, Bhardwaj K, Hagan P. Academic attainment in international medical students might be optimised by educating them about cognitive load theory. *Int J Med Educ*. (2022) 13:90–1. doi: 10.5116/ijme.6238.4dfd

27. Tremblay M-L, Lafleur A, Leppink J, Dolmans DHJM. The simulated clinical environment: cognitive and emotional impact among undergraduates. *Med Teach*. (2017) 39:181–7. doi: 10.1080/0142159X.2016.1246710

28. Andersen MS, Makransky G. The validation and further development of the multidimensional cognitive load scale for physical and online lectures (MCLS-POL). *Front Psychol*. (2021) 12:642084. doi: 10.3389/fpsyg.2021.642084

29. Leppink J, Paas F, Van der Vleuten CPM, Van Gog T, Van Merriënboer JJG. Development of an instrument for measuring different types of cognitive load. *Behav Res Methods*. (2013) 45:1058–72. doi: 10.3758/s13428-013-0334-1

30. Fredericks S, ElSayed M, Hammad M, Abumiddain O, Istwani L, Rabeea A, et al. Anxiety is associated with extraneous cognitive load during teaching using high-fidelity clinical simulation. *Med Educ Online*. (2021) 26:1994691. doi: 10.1080/10872981.2021.1994691

31. De Paoli S.. (2023). Can large language models emulate an inductive thematic analysis of semi-structured interviews? An exploration and provocation on the limits of the approach and the model. Available at:https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/.

32. Tremblay M-L, Leppink J, Leclerc G, Rethans J-J, Dolmans DHJM. Simulation-based education for novices: complex learning tasks promote reflective practice. *Med Educ*. (2019) 53:380–9. doi: 10.1111/medu.13748

33. Tremblay M-L, Rethans J-J, Dolmans D. Task complexity and cognitive load in simulation-based education: a randomised trial. *Med Educ*. (2023) 57:161–9. doi: 10.1111/medu.14941

34. Richmond A, Cooper N, Gay S, Atiomo W, Patel R. The student is key: a realist review of educational interventions to develop analytical and non-analytical clinical reasoning ability. *Med Educ*. (2020) 54:709–19. doi: 10.1111/medu.14137

35. Nguyen BT, Streeter LH, Reddy RA, Douglas CR. Gender bias in the medical education of obstetrician-gynaecologists in the United States: a systematic review. *Aust N Z J Obstet Gynaecol*. (2022) 62:349–57. doi: 10.1111/ajo.13511

36. Scala JJ, Braun NJ, Shamardani K, Rashes ER, Wang W, Mediratta RP. Applying Kern's six steps to the development of a community-engaged, just-in-time, interdisciplinary COVID-19 curriculum. *J Med Educat Curri Develop*. (2022) 9:23821205221096370. doi: 10.1177/23821205221096370

37. Meguerdichian M, Walker K, Bajaj K. Working memory is limited: improving knowledge transfer by optimising simulation through cognitive load theory. *BMJ Simul Technol Enhanc Learn*. (2016) 2:131–8. doi: 10.1136/bmjstel-2015-000098

38. Jeet G, Prinja S, Aggarwal A. Cost analysis of a simulation-based training for health workforce in India. *Indian J Public Health*. (2017) 61:92–8. doi: 10.4103/ijph.IJPH_189_15

39. How RS. How simulation-based medical education can be started in low resource settings. *J Ayub Med Coll Abbottabad*. (2019) 31:636–7.

40. Skulmowski A. Learners emphasize their intrinsic load if asked about it first: communicative aspects of cognitive load measurement. *Mind Brain Educ*. (2023) 17:165–9. doi: 10.1111/mbe.12369

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 27 March 2024
DOI 10.3389/fmed.2024.1363222

OPEN ACCESS

EDITED BY
Erich Brenner,
Innsbruck Medical University, Austria

REVIEWED BY
Matthias Siebeck,
LMU Munich University Hospital, Germany
Elena Jost,
University Hospital Bonn, Germany

*CORRESPONDENCE
Emma Paternotte
✉ emmapaternotte@gmail.com

RECEIVED 30 December 2023
ACCEPTED 27 February 2024
PUBLISHED 27 March 2024

CITATION
Paternotte E, Dijksterhuis M, Goverde A, Ezzat H and Scheele F (2024) Comparison of OBGYN postgraduate curricula and assessment methods between Canada and the Netherlands: an auto-ethnographic study. *Front. Med.* 11:1363222. doi: 10.3389/fmed.2024.1363222

Comparison of OBGYN postgraduate curricula and assessment methods between Canada and the Netherlands: an auto-ethnographic study

Emma Paternotte[1]*, Marja Dijksterhuis[2], Angelique Goverde[3], Hanna Ezzat[4] and Fedde Scheele[5]

[1]Department of Obstetrics and Gynaecology, Gelre Hospitals, Apeldoorn, Netherlands, [2]Department of Obstetrics and Gynaecology, Amphia Ziekenhuis, Breda, Netherlands, [3]Department of Obstetrics and Gynaecology, University Medical Center Utrecht, Utrecht, Netherlands, [4]Division of General Gynaecology and Obstetrics, University of British Columbia, Vancouver, BC, Canada, [5]Department of Obstetrics and Gynaecology, Onze Lieve Vrouwe Gasthuis (OLVG), Amsterdam, Netherlands

Introduction: Although the Dutch and the Canadian postgraduate Obstetrics and Gynecology (OBGYN) medical education systems are similar in their foundations [programmatic assessment, competency based, involving CanMED roles and EPAs (entrustable professional activities)] and comparable in healthcare outcome, their program structures and assessment methods considerably differ.

Materials and methods: We compared both countries' postgraduate educational blueprints and used an auto-ethnographic method to gain insight in the effects of training program structure and assessment methods on how trainees work. The research questions for this study are as follows: what are the differences in program structure and assessment program in Obstetrics and Gynecology postgraduate medical education in the Netherlands and Canada? And how does this impact the advancement to higher competency for the postgraduate trainee?

Results: We found four main differences. The first two differences are the duration of training and the number of EPAs defined in the curricula. However, the most significant difference is the way EPAs are entrusted. In Canada, supervision is given regardless of EPA competence, whereas in the Netherlands, being competent means being entrusted, resulting in meaningful and practical independence in the workplace. Another difference is that Canadian OBGYN trainees have to pass a summative written and oral exit examination. This difference in the assessment program is largely explained by cultural and legal aspects of postgraduate training, leading to differences in licensing practice.

Discussion: Despite the fact that programmatic assessment is the foundation for assessment in medical education in both Canada and the Netherlands, the significance of entrustment differs. Trainees struggle to differentiate between formative and summative assessments. The trainees experience both formative and summative forms of assessment as a judgement of their competence and progress. Based on this auto-ethnographic study, the potential for further harmonization of the OBGYN PGME in Canada and the Netherlands remains limited.

KEYWORDS

postgraduate medical education, assessment, EPA, comparison, auto-ethnographic

Introduction

It is widely acknowledged that well-structured postgraduate medical training is important to deliver well-trained medical specialists who can provide the best possible care. To certify that trainees develop the necessary skills and attitudes in addition to the knowledge required to practice, the competency framework CanMEDS was developed in Canada in 1996 (1). This competency-based framework described the competences that trainees should acquire during their training. In 2015, the CanMEDS framework was enhanced by incorporating competency milestones for each CanMED role (2).[1] This approach has been implemented in several medical education systems globally. For instance, the CanMEDS roles have been used in the official training blueprints of all postgraduate medical education (PGME) in the Netherlands (3).[2]

Further expansion of the CanMEDS has led to the development of EPAs. EPAs are described as Entrustable Professional Activities. The Netherlands was among the pioneers in implementing EPAs in their training programs (4–8). The Obstetrics and Gynecology PGME training in Canada also transitioned toward EPAs (see text footnote 1). EPAs are "units of professional practice, defined as tasks or responsibilities to be entrusted to unsupervised execution by a trainee once he or she has attained sufficient competence" (9).

The EPAs are designed to ensure a gradual level of entrustment, known as progressive independence. Entrustment is intended to mean trust, meaning that if the postgraduate trainee is entrusted to perform a certain activity, they would be expected to carry out the procedure without supervision (4). For instance, potential applications of EPAs include assessing competence, making entrustment decisions for independent practice, facilitating professional development, and informing curriculum development. It is known that the uptake of EPAs in the curricula varies among the different specialty training programs (10).

Although competency-based training, CanMeds roles, and EPAs have been implemented in postgraduate OBGYN-training in several countries, assessment methods considerably differ. Differences can be attributed to political, social, and cultural reasons (4, 11). It has been suggested that patient outcomes, notably major complications, are associated with the quality of training received. Interestingly, quality of training is not associated with licensing examination scores (12). Aabake et al. demonstrated that in Europe, examinations during OBGYN training were used in 89% of the included countries. All examinations were mandatory, but the study did not clarify whether they were formative or summative (11). In other words, formative assessments are often considered assessments *for* learning, while summative assessments are assessments *of* learning (13). It is interesting to find out why some postgraduate OBGYN training programs use certain assessment tools, and why others do not. For instance, despite Canada having implemented competency-based training and EPAs, the adoption of new assessment methods and the utilization of a training logbook are not yet widespread practices (14). As stated on the Royal College website,[3] new formal assessment methods have been implemented and formalized since Garofalo's findings in 2017.

Worldwide, the Netherlands and Canada hold leading positions in the advancements and modernization of medical education (14). Given the similar foundations of PGME in Canada and the Netherlands, it is interesting to discuss differences, particularly in competency-based education and programmatic assessment. Therefore, our research questions are "Which differences in formal assessment methods exist in Obstetrics and Gynecology postgraduate medical education and how does this impact the advancement to higher competency for the postgraduate trainee?" To answer these questions, we employed the auto-ethnographic observation method to gain insights in which differences exist and how these differences are perceived by the postgraduate trainees. We scrutinized the Obstetrics and Gynecology PGME blueprints of both countries to find similarities and differences. Highlighting the differences between both Canadian and the Dutch assessments systems with respect to cultural and political differences may provide insight into assessments that could enhance each training program. By harmonizing curricula, variability in training outcomes will decrease, which could improve the quality of training overall. Moreover, it would facilitate benchmarking and comparisons across centers and streamline the exchange of trainees between countries (14).

Materials and methods

This study is grounded in auto-ethnography, a qualitative research methodology in which the author uses writing and self-reflection to probe personal experiences, thereby deriving broader sociocultural meaning and understanding (15).

Auto-ethnography combines elements from autobiography and ethnography and represents both a scholarly "process" and "product." It moves away from traditional approaches to understanding and representing "culture" which has historically been rooted in rigid structures. Instead, auto-ethnography "acknowledges and accommodates subjectivity, emotionality, and the researcher's influence on research, rather than hiding from these matters or assuming they don't exist" (16).

Participants and setting

This is qualitative research includes an introduction of the enrolled researchers. This project involved five participants: EP, FS, MD, AG, and HE. The main auto-ethnographer was EP, a

Abbreviations: OBGYN, Obstetrics and Gynecology; EPA, Entrustable professional activities; PGME, Postgraduate medical education; EP, Emma Paternotte; FS, Fedde Scheele; MD, Marja Dijksterhuis; AG, Angelique Goverde; HE, Hanna Ezzat; LOGO, Landelijk Opleidingsplan Obstetrie en Gynaecologie (Formal national Education document of Obstetrics and Gynecology); OSATS, Objective Structured Assessment of Technical Skills; Mini-CEX, Mini-Clinical Evaluation Exercise; CbD, Case-based discussion; NOTSS, Non-Technical Skills for Surgeons; O-score, Ottawa Surgical Competency Operating Room Evaluation.

1 https://www.royalcollege.ca

2 https://www.knmg.nl/ik-ben-arts/cgs/regelgeving/specialismen

3 https://www.royalcollege.ca/en/credentials-exams/assessment-canadian-us.html

Dutch 6th-year postgraduate trainee in OBGYN who finished an elective in Vancouver, Canada. She is the subject of the performed auto-ethnographic study and received support and supervision throughout the project from the other four participants. FS is gynecologist and professor in Health Systems Innovation and Education in the Netherlands, where he supervised PhD students and conducted research in assessment methods. MD is a gynecologist in the Netherlands with a PhD in assessment methods in medical education. AG is gynecologist and program director of EP in postgraduate medical education of OBGYN in Utrecht, the Netherlands. HE is gynecologist and program director of the postgraduate residency program in British Colombia, Canada. The connection among our team members was forged serendipitously.

Data collection

The predetermined guiding theme for this auto-ethnography was how postgraduate trainees perceive assessment methods and how formal assessment methods are shaped throughout the course of residency training. EP's experiences as an elective OBGYN postgraduate trainee were discussed with the other researchers. As the process of reflection and discussion developed, the need to scrutinize the blueprints of PGME of Canada and the Netherlands became apparent. Differences and similarities were found. These findings were then discussed with gynecologists and postgraduate trainees of both the Netherlands and British Columbia, Canada. Extensive reflection and discussion between EP and these trainees and gynecologists yielded insights into how postgraduate trainees perceive their assessment methods and how does this impact the advancement toward higher competency. These reflective sessions were either audiotaped or noted.

TABLE 1 Overview of PGME in OBGYN focused on assessment.

	Canada	The Netherlands
Years of UGME/Med school	4	6
Years of PGME OBGYN	5	6
EPAs in PGME OBGYN	36	12
Working hours per week	No duty hour restrictions	Maximum 48 h
Feedback	Formative throughout the whole year	Formative throughout the whole year
Progress assessment	Every 6 months by competency committee	In the first year every 3 months, after first year at least annually, progress meeting by trainee and program director
Progress test	Yearly, formative, not required by Royal College	Yearly, formative, mandatory
Surgical examination	Year 2, summative	Mandatory courses
Final examination	Written and Oral	None
Postgraduate training program paid by:	Ministry of health. Money to cover resident salaries not given directly to hospitals, but to health authorities or larger bargain units. No funding for "disutility" provided.	Ministry of health, welfare and sports, executed by the Dutch healthcare Authority (NZA). Money is forwarded to the teaching hospital where the trainee is currently employed. It covers both costs for training (salary and education) and disutility.

Results

The results are presented in two sections. The first section provides a general overview based on the blueprints of the Canadian and the Dutch curricula, with a focus on assessment methods. In Table 1, an overview of postgraduate medical education in OBGYN in both Canada and the Netherlands is given. Table 2 describes the EPAs in the Netherlands, whereas Table 3 describes the EPAs in Canada.

The second part is a discussion of how these assessment methods manifest in the workplace. This discussion is based on the experiences of EP and is the auto-ethnographic part.

PGME and assessment of OBGYN in the Netherlands

Postgraduate training can be pursued directly after 6 years of undergraduate medical training. Typically, candidates undergo a period of work in non-training grade posts in order for the candidate to gain practical experience and increase their chances of successfully applying for a postgraduate training program. Entering a postgraduate training program is based on an interview, which may sometimes be combined with a personal assessment. The postgraduate training program in OBGYN lasts 6 years on full-time basis, although part-time placement is allowed to a certain degree. The program consists of hands-on practical training in delivering patient care, complemented by formal teaching, and academic study. Programs are completed in a structured learning environment in teaching hospitals supervised by faculty. Rotations are evenly divided between academic and district teaching hospitals.

The national training curriculum is based on EPAs (Table 2), and progress is registered in the personal electronic portfolio. The curriculum focusses on delivering competent general OBGYNs who share common basic skills but have specific areas of differentiation, depending on their area of interest. The curriculum of the OBGYN postgraduate training is captured in a national training curriculum for obstetrics and gynecology LOGO (Landelijk Opleidingsplan Obstetrie en Gynaecologie) (17), where a daily feedback culture is a basic requirement for the programs. One of the most salient features of current training is the increased number and variety of formative assessment moments: assessments that are aimed at providing feedback that can direct and stimulate learning. Additionally, an annual formative online progress test, which consists of 160 multiple choice questions, is mandatory. This test is designed to generate learning goals based on the test results. The low-stakes formative assessments

TABLE 2 Overview of EPAs of OBGYN in the Netherlands (17).

Core EPA # 1: Pregnancy and labor, low-risk
Core EPA # 2: Pregnancy and labor, high-risk
Core EPA # 3: Postpartum healthcare
Core EPA # 4: Abnormal uterine bleeding
Core EPA # 5: Early pregnancy
Core EPA # 6: Prolapse and pelvic floor complaints
Core EPA # 7: Abdominal pain (acute and chronic)
Core EPA # 8: Vulvar and vaginal abnormalities
Core EPA # 9: Healthcare for pre-malignant lesions
Core EPA # 10: Basic oncological care
Core EPA # 11: Fertility healthcare
Core EPA # 12: Lifecycle endocrinology

and learner self-reflection are pivotal to progression within the training program and not the formative annual national written progress test (8). There is no final oral or written examination. Moreover, postgraduate trainees may participate in mandatory or non-mandatory courses, such as an ultrasound course. All formative assessments, together with progress reports, proof of experience, feedback of supervisors [for example Objective structured assessment of technical competencies (OSATS) and mini clinical evaluation exercise (mini-CEX)], individual reflection, and the results of the annually formative progress test are collected in a personal electronic portfolio.

Assessment for entrustment of specific EPAs and the final conclusion on whether or not the postgraduate trainee is competent to deliver high standards of care is conducted by the educational team chaired by the program director and is based on the records in the electronic portfolio (all activities and development). If the EPA is entrusted, the trainee is allowed to perform this EPA independently.

PGME and assessment of OBGYN in Canada

Medical school students seeking to start postgraduate training in OBGYN apply to the Canadian residency matching program. This residency matching program includes a detailed portfolio, and the selection process includes a file review of medical school performance records, reference letters, and an interview. The specialty training requirements in OBGYN include a 5-year training program governed by the Royal College of Physicians and Surgeons of Canada. A new curriculum, called Competency by Design, was introduced in 2019 (16).[4] The curriculum is framed around 36 EPAs (Table 3).[5] These EPAs represent essential specialist

TABLE 3 Overview of EPAs of OBGYN in Canada* (see text footnote 4).

Transition to Discipline EPA #1: Performing initial assessments for uncomplicated obstetric patients.
Transition to Discipline EPA #2: Performing initial assessments for uncomplicated gynecological patients.
Foundations EPA #1: Providing routine prenatal care to a low-risk, healthy population.
Foundations EPA #2: Performing assessments of antenatal fetal well-being.
Foundations EPA #3: Assessing and providing initial management for patients with common obstetric presentations.
Foundations EPA #4: Managing labor and childbirth.
Foundations EPA #5: Performing uncomplicated cesarean sections with a skilled assistant.
Foundations EPA #6: Providing early postpartum care.
Foundations EPA #7: Providing consultation and initial management for patients with urgent and emergent gynecologic presentations.
Foundations EPA #8: Counseling and management for patients requiring family planning.
Foundations EPA #9: Providing consultation for patients with gynecological conditions.
Foundations EPA #10: Performing minor gynecologic operative procedures.
Foundations Special Assessment #1: Performing critical appraisal of health literature and initiating scholarly projects.
Core EPA #1: Providing preconception and antenatal care to women with high-risk pregnancies.
Core EPA #2: Managing patients with acute conditions presenting in the antenatal and perinatal periods.
Core EPA #3: Managing complex vaginal deliveries.
Core EPA #4: Performing complex cesarean sections.
Core EPA #5: Diagnosing and managing postpartum complications.
Core EPA #6: Performing obstetric and gynecologic ultrasound.
Core EPA #7: Providing definitive management for patients with acute gynecologic emergencies.
Core EPA #8: Providing care for patients with complex gynecologic conditions and/or medical comorbidities
Core EPA #9: Assessing and initiating management for patients with reproductive challenges
Core EPA #10: Diagnosing and managing pediatric and adolescent patients with common gynecologic conditions.
Core EPA #11: Providing care for patients with pelvic floor dysfunction.

(Continued)

4 https://www.royalcollege.ca/en/cbd/understanding-cbd.html

5 https://www.royalcollege.ca/content/dam/documents/accreditation/competence-by-design/non-resource-documents/epa-observation-templates-e.html

TABLE 3 (Continued)

Core EPA #12: Assessing, diagnosing, and managing patients with chronic pelvic pain and sexual health concerns.
Core EPA #13: Assessing and managing patients with gynecologic malignancies.
Core EPA #14: Performing advanced hysteroscopy.
Core EPA #15: Performing major vaginal and vulvar procedures.
Core EPA #16: Performing major laparoscopic gynecologic procedures.
Core EPA #17: Performing major open abdominal gynecologic procedures.
Core EPA #18: Managing patients with surgical complications.
Core EPA #19: Managing the birthing unit.
Transition to Practice EPA #1: Managing complex patients, including those requiring longitudinal care.
Transition to Practice EPA #2: Discussing difficult news.
Transition to Practice SA #1: Conducting scholarly work.
Transition to Practice SA #2: Teaching and managing learners.

*The Canadian system of EPAs is subdivided into stages. Stage 1 is transition to discipline, stage 2 is foundations, stage 3 is core, and stage 4 transition to practice. This structure is maintained across all postgraduate programs in all disciplines. The trainee must obtain all the EPAs for a certain stage before the competency committee promotes them to the next stage, and they can start working on the associated EPAs for that stage.

competencies, leading to optimal OBGYN care outcomes. The EPAs are designed to align with the stage the trainee is in. The pedagogical concept is that EPAs should not be time based, but based on whenever a competency has been demonstrated. However, as has been stated in *The Future of medical education in Canada*[6]: "this approach to using competencies to assess the performance of physicians in practice is still in its infancy."

Currently, the assessment philosophy of Canadian medical education is based on this competency by the design structure. This philosophy includes formative assessments throughout the whole year. The workplace-based assessments include the following, for instance: OSATS, mini-CEX, Case-based discussion (CbD), Team observation forms, bi-Annual Review of Competence Progression by the competency committee, and Non-Technical Skills for Surgeons (NOTSS). The concept is that continuous assessment drives and promotes learning. In year 2, the postgraduate trainee of surgical specialties, such as OBGYN, needs to pass a summative surgical foundations examination examination (see text footnote 3). In year 5, a summative written examination is taken, followed by the final summative oral examination 6 months later. If the postgraduate trainee fails one of these summative examinations, the trainee is required to retake the failed examinations. Additionally, there is an annual formative written progress test, which is not mandatory but is utilized to prepare trainees for the final examination (see text footnote 1) (14).

The formal methods of assessment are described in the "Standards of Accreditation for Residency Programs in Obstetrics and Gynecology" (see text footnote 1). The competency committee comprises supervisors (gynecologists) from various clinics, as well as the program director. Together, they decide whether an EPA is entrusted. Even after an EPA is entrusted, some degree of ongoing direct or indirect supervision remains the cultural, legal, and practically enforced norm.

Reflection

Postgraduate medical education in Canada and the Netherlands share the same philosophy, including programmatic assessment in competency based medical education. Medical education in Canada and the Netherlands are regarded as the highest standards worldwide (13). In both countries, development and improvement are key factors in medical education.

In both the Netherlands and Canada, EPAs and programmatic assessment have been implemented successfully and this has been confirmed by stakeholders. Despite these similarities, there are a few key differences in the curricula and assessment methods experienced by EP and discussed with the research team.

The first difference between OBGYN education is the total duration of medical training. In Canada, medical school (undergraduate training) and residency (postgraduate medical training) embrace a total of 9 years, whereas in the Netherlands, total duration is 12 years, usually longer due to junior doctors commonly working in non-training posts. Another key finding that aligns with the first difference is that non-training posts are not possible in Canada. In Canada, non-training posts for junior doctors who have finished medical school do not exist.

The second difference is the total amount EPAs. In the Netherlands, the 12 EPAs are used for the whole 6 years of postgraduate training, while in Canada, EPAs are organized into stages. The stages represent a stepwise progression of competence during the training period. Comparing the content of the EPAs informs us that there are no big differences in the end terms of the OBGYN training. However, it appears that the Canadian OBGYN EPAs are more fragmented than those in the Netherlands.

The third difference lies in how entrustment is applied by both countries. In the Netherlands, OBGYN has implemented this entrustment process in daily clinical practices; a trainee, entrusted to perform a cesarean section, is legally and culturally allowed to perform the procedure unsupervised. For the trainee, the entrustment granted during the training years contributes to an inner sense of confidence that they indeed are capable of working as a gynecologist (unsupervised). Most PGME programs in the Netherlands offer the trainees with rotations where they can experience the role of supervising younger trainees, which contributes to their confidence and professional growth toward independent practice. In the Netherlands, entrustment translates into a meaningful change in status. In contrast, in Canada,

6 https://www.afmc.ca/wp-content/uploads/2022/10/FMEC-CPD_Synthesized_EN_WEB.pdf

entrustment means that the resident is progressing as would be expected for their training. Canadian trainees are usually supervised during (surgical) procedures until their certification as a gynecologist. Both postgraduate trainees and supervisors describe this as the norm. A quote from one of the supervisors is as follows: "Competency by design is still a bit out of reality of daily clinical education."

An explanation for this significant difference in the meaning of entrustment might be found in how the levels of entrustment are interpreted. As explained on the website of the Royal College, the steps of assessment to full entrustment of an EPA are described by the use of the O-score. This score is a 9-item evaluation tool designed to assess technical competence in surgical trainees using behavioral anchors. This O-score seems to be written through the eyes of the supervisor (highest level: "I did not need to be there"), whereas in the Netherlands, this scoring system is through the eyes of the trainee (highest level: "I supervise this procedure or EPA"). The entrustment rating of an EPA by the Royal College is as follows: "Rating trainees as independent does not mean that they are now always allowed to independently perform that task. It means that they were independent on this occasion" (18). The entrustment rating of independence does not reflect the medico-legal reality, nor the expectations of patients and trainers of the presence of a certified specialist who is ultimately responsible, especially in surgery. The reason for this may be found in how the regulation and granting of postgraduate medical education is arranged in Canada as the Royal College is responsible for prescribing and assessing the learning standard, but not for regulating or granting a license to practice. Ultimately, all trainees, even until the last moment of training, have an educational license only, which means that they have no license to practice independently and everything they do must be supervised. The supervisor is the "most responsible physician" if something goes wrong. It is understandable why a faculty surgeon, who owns that responsibility, might hesitate to allow a resident, even one they believe is highly competent, to perform a cesarean section without supervision just because the competency committee (CC) said that they were competent to do so. However, the designation by the CC holds little weight outside the program.

A fourth difference is the formative and summative examination systems. Both countries have a yearly progress test, in the form of a written examination. This is obligatory for the Dutch resident, and its results are used formatively to identify areas for further study. However, in Canada, the yearly formative test is not mandatory; the Royal College does not require, endorse, or even suggest this test. Instead, the program directors and teachers want trainees to take the test as a necessity to prepare trainees for their final examination. This final examination has to be taken by postgraduate trainees in Canada but is not known in the Dutch postgraduate training program. Though competency by design in Canada preaches programmatic training of competencies and formative assessments, the final assessment is summative. Postgraduate trainees in Canada find it inconceivable to not have a final examination. The trainees described that supervisors treat them differently as soon as they passed this examination. Ending the PGME without this final examination is unimaginable for them, which gives them the feeling that they have met a certain standard and that this will allow them to get a job anywhere else in Canada. Legally the resident still does not have a license after passing the examination to act independently, but the examination certification is the key requirement to that license. In contrast, Dutch postgraduate trainees are satisfied with their assessment system without a summative exit examination, mainly because the Dutch trainee work totally independent at the end of their training and, therefore, a summative exit examination would not add anything to the growing process. Growing into being fully entrusted is an organic process that develops along the way. Dutch trainees view examinations as a snapshot of their performance and query how well any such examination would represent their competence. Interestingly, despite the formative intent of PGME in OBGYN in both countries, neither set of postgraduate trainees perceive their training as such. Many postgraduate trainees regard all the feedback moments as small summative assessments and feel continuously judged about their performance.

Discussion

To answer our first research questions ("which differences in formal assessment methods exist in Obstetrics and Gynecology of postgraduate medical education?"), we created an overview of the formal assessment methods. For the second research question ("how does this impact the advancement to higher competency for the postgraduate trainee?"), we observed four main differences in the curricula of PGME of OBGYN between Canada and the Netherlands. The most striking difference lies in the way that entrustment is interpreted and put into practice in both countries, since this is of consequence for the role of assessment in the entrustment process and even more trainees feeling adequately prepared to work as a gynecologist. However, the Royal College explains the entrustment of EPAs as the resident is progressing as would be expected for their stage, for the Dutch trainees entrustment reflects their professional development toward independent practice. Additionally, the number of EPAs differs enormously, which might also reflect the difference between wanting control and needing a "pass" on smaller parts, vs. believing in trust the knowledge that the whole is more than the total sum of small parts. In addition to this consideration, there is the requirement for passing a summative exit examination in Canada, whereas in the Netherlands, this requirement has been rendered redundant. In conclusion, a higher competence is reached more or less on a similar way in both Canada and the Netherlands. However, the meaning of this higher competence is interpreted differently.

To summarize, programmatic assessment in Canada appears to be based on the "assessment *for* learning" principle. The fact that there is a summative final examination also makes assessment feel more like an "assessment *of* learning" or perhaps it is a combination of the two. In contrast, programmatic assessment in the Netherlands tends more toward "assessment for learning" alone. However, in both countries, trainees tend to experience the assessments as more summative than formative, in general, and there is ongoing discussion about how to make assessments feel formative for postgraduate trainees. We propose that there is a role for both kinds of assessment, in line with the different levels of knowledge and skills as described in Miller's pyramid (19).

Formative assessment is the instrument for coaching a trainee from the "does" level in Miller's pyramid to the "shows" level, for instance in surgical procedures. Summative assessment is the instrument to actually ascertain and document whether the trainee has reached the higher "shows" level and to inform supervisor and trainee about the level of independence in specific tasks or procedures. Making the aim of the formative assessment explicit could help the trainee to focus more on the learning process, while clearly laying out the summative assessment moments gives the trainee a perspective on entrustment decisions (20).

Conclusion

There is no one-size-fits-all solution, and no specific programmatic assessment has been proven to be superior to others (21). A tip for both countries regarding their assessment methods is that clear communication about the purpose of assessment and the use of assessment outcomes needs to be formulated (22). All supervisors need to be trained to provide non-judgmental constructive formative feedback (20). Competency-based medical education requires an ongoing process to evaluate and improve the assessment methods (13). This difference in assessment program is largely explained by cultural and legal aspects of postgraduate training and consequent differences in licensing practice. Based on this auto-ethnographic study, the harmonization of the OBGYN PGME in Canada and the Netherlands appears limited since cultural, legal, and practical aspects of assessment and licensing predominate. In our opinion, this is a missed opportunity because harmonization could help to resolve labor shortages where needed. We recommend investigating whether there are options to facilitate the harmonization process between postgraduate trainees between the Netherlands and Canada.

Limitations

A limitation of this study is that not all perspectives could be captured. There may be opinions and experiences of trainees, teachers, and supervisors that were missed. Other limitations include the inherent subjectivity of the ethnography method and the limited sample size.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies involving humans because it is an auto ethnographic study which does not contain personal information but only interpretations of the writer based on humans. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements because no direct material of humans was used.

Author contributions

EP: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft, Investigation. MD: Conceptualization, Supervision, Writing – review & editing. AG: Supervision, Writing – review & editing. HE: Writing – review & editing, Data curation, Methodology. FS: Writing – review & editing, Data curation, Methodology, Conceptualization, Supervision.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Frank JR. *The Can MEDS 2005 Physician Competency Framework: Better Standards*. Ottawa: Better Physicians Better Care. (2005).

2. Frank J, Snell L, Sherbino J. *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada. (2015).

3. Scheele F, den Rooyen C, van Luijk S, van Loon K, van der Lee N. Better education for obstetrics and gynecology (BOEG). In: *Dutch National Competency Based Curriculum for Obstetrics & Gynaecology*. (2011).

4. ten Cate O, Scheele F. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. (2007) 82:542–7. doi: 10.1097/ACM.0b013e31805559c7

5. Van der Vleuten CPM, Schuwirth LWT, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clini Obstet Gynaecol*. (2010) 24:703–19. doi: 10.1016/j.bpobgyn.2010.04.001

6. Billett S. Learning through practice: models, traditions, orientations and approaches [Electronic version] In: Billett S, editor. *Professional and Practice-Based Learning*. Dordrecht: Springer. (2010), p. 1–20.

7. Ten Cate O. Managing risks and benefits: key issues in entrustment decisions. *Med Educ.* (2017) 51:879–81. doi: 10.1111/medu.13362

8. Caccia N, Nakajima A, Scheele F, Kent N. Competency-based medical education: developing a framework for obstetrics and gynaecology. *J Obstet Gynaecol Can*. (2015) 37:1104e12. doi: 10.1016/S1701-2163(16)30076-7

9. Ten Cate O. Nuts and bolts of entrustable professional activities. *J Grad Med Educ.* (2013) 5:157–8. doi: 10.4300/JGME-D-12-00380.1

10. van Loon KA, Bonnie LHA, van Dijk N, Scheele F. The influence of the workplace environment on the uptake of EPAs in EPA-based curricula. *Perspect Med Educ.* (2021) 10:200–6. doi: 10.1007/S40037-021-00658-9

11. Aabakke AJM, Kristufkova A, Boyon C, Bune LT. Workforce planning and training in Obstetrics and Gynaecology across Europe: a survey of national trainee societies. *Eur J Obstet Gynecol Reproduct Biol.* (2017) 214:156–61. doi: 10.1016/j.ejogrb.2017.05.007

12. Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA*. (2009) 302:1277e83. doi: 10.1001/jama.2009.1356

13. Lockyera J, Carracciob C, Chanc MK, Hartde D, Smeef D, Touchief C. Core principles of assessment in competency-based medical education. *Med Teach*. (2017) 36:609–16. doi: 10.1080/0142159X.2017.1315082

14. Garofalo M., Aggarwal R. Competency-based medical education and assessment of training: review of selected national obstetrics and gynaecology curricula. *J Obstet Gynaecol Can.* (2017) 39(7):534e544. doi: 10.1016/j.jogc.2017.01.024

15. Farrell L, Bourgeois-Law G, Regehr G, Ajjawi R. Auto-ethnography: introducing 'I' into medical education research. *Med Educ*. (2015) 49:974–82. doi: 10.1111/medu.12761

16. Ellis C, Adams TE, Bochner AP. Auto-ethnography: an overview. *Historical Soc Res/Historische Sozialforschung*. (2011) 36:273–90.

17. LOGO. *Landelijk Opleidingsplan Gynaecologie & Obstetrie*. (2022). Available online at: https://nvog-logo.nl (accessed Ferbruary 11, 2024).

18. Gofton W, Dudek N, Barton G, Bhanji F. *Workplace-Based Assessment Implementation Guide: Formative Tips for Medical Teaching Practice, 1st ed.* Ottawa: The Royal College of Physicians and Surgeons of Canada. (2017) p. 1–12. Available online at: http://www.royalcollege.ca/rcsite/documents/cbd/wba-implementation-guide-tips-medical-teaching-practice-e.pdf (accessed Ferbruary 11, 2024).

19. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med.* (1990) 1990:S63–7. doi: 10.1097/00001888-199009000-00045

20. PlP Brand and F Scheele. Feedback in de medische opleiding. *Ned Tijdschr Geneeskd*. (2022) 166:D6708.

21. Ross S, Hauer KE, Wycliffe-Jones K, Hall AK, Molgaard L, Richardson D, et al. Key considerations in planning and designing programmatic assessment in competency-based medical education. *Med Teach.* (2021) 43:758–64. doi: 10.1080/0142159X.2021.1925099

22. Sein AS, Rashid H, Meka J, Amiel J, Pluta W. Twelve tips for embedding assessment for and as learning practices in a programmatic assessment system. *Med Teach.* (2021) 43:300–6. doi: 10.1080/0142159X.2020.1789081

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 24 April 2024
DOI 10.3389/fmed.2024.1371141

OPEN ACCESS

EDITED BY
Simcha Yagel,
Hadassah Medical Center, Israel

REVIEWED BY
Lise Brogaard,
Aarhus University Hospital, Denmark
Guglielmo M. Trovato,
European Medical Association (EMA), Belgium

*CORRESPONDENCE
Johannes Weimer
✉ weimer@uni-mainz.de
Anna Dionysopoulou
✉ anna.dionysopoulou@unimedizin-mainz.de

RECEIVED 15 January 2024
ACCEPTED 26 March 2024
PUBLISHED 24 April 2024

CITATION
Weimer J, Recker F, Hasenburg A, Buggenhagen H, Karbach K, Beer L, Weimer A, Schiestl L, Lorenz L, Kloeckner R and Dionysopoulou A (2024) Development and evaluation of a "simulator-based" ultrasound training program for university teaching in obstetrics and gynecology–the prospective GynSim study. *Front. Med.* 11:1371141. doi: 10.3389/fmed.2024.1371141

Development and evaluation of a "simulator-based" ultrasound training program for university teaching in obstetrics and gynecology–the prospective GynSim study

Johannes Weimer [1]*, Florian Recker[2], Annette Hasenburg[3], Holger Buggenhagen[1], Karla Karbach[3], Lia Beer[1], Andreas Weimer[4], Lina Schiestl[3], Liv Lorenz[5], Roman Kloeckner[6] and Anna Dionysopoulou[3]*

[1]Rudolf Frey Learning Clinic, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany, [2]Department of Obstetrics and Prenatal Medicine, University Hospital Bonn, Bonn, Germany, [3]Department of Obstetrics and Gynecology, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany, [4]Center of Orthopedics, Trauma Surgery, and Spinal Cord Injury, University Hospital Heidelberg, Heidelberg, Germany, [5]Department of Radiation Oncology and Radiotherapy, University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany, [6]Institute of Interventional Radiology, University Hospital Schleswig-Holstein, Lübeck, Germany

Introduction: This study addresses the challenges of ultrasound education in obstetrics and gynecology, focusing on the potential benefits of simulation techniques in medical training. Aiming to evaluate the impact of a structured simulator-based training program, this prospective, randomized, interventional study examines its effects on educational outcomes for 5th year medical students.

Methods: A total of 153 medical students were randomized into two groups: one receiving both theoretical instruction and hands-on ultrasound simulator training (study group), and the other receiving only theoretical instruction (control group). The study assessed theoretical knowledge and practical skills at two time points: upon enrollment and at the end of the course. The practical skills were specifically evaluated using a dedicated test on the ultrasound simulator.

Results: Out of 153 students, 113 completed the study (study group n=59, control group n=54). The students in the study group demonstrated a greater improvement in theoretical test scores. They also achieved better results at the practical test, with regard to image quality, accuracy, and efficiency. Both groups showed an increase in self-confidence and competency in performing ultrasound examinations independently. Students expressed high satisfaction with the course and a positive attitude toward simulator-based training.

Discussion: Simulator-based training presents a valuable supplement to traditional clinical education methods in obstetrics and gynecology. This approach is particularly effective in overcoming the challenges posed by the sensitive nature of gynecological examinations in medical student training. The study highlights the benefits of integrating simulator-based methods into

medical curricula, improving both theoretical and practical ultrasound skills among students.

KEYWORDS

ultrasound training, simulator-based training, education, obstetrics, gynecology

1 Introduction

Ultrasound imaging, particularly transabdominal and transvaginal, is a critical diagnostic tool in Obstetrics and Gynecology (Ob/Gyn). Its non-invasive nature and excellent safety profile make it indispensable in clinical practice for differentiating between normal and pathological findings (1–3).

Gaining ultrasound imaging proficiency requires extensive training. Key medical bodies, including the EFSUMB and WFUMB, emphasize integrating ultrasound education into medical curricula (4, 5).

Traditionally, ultrasound training for medical students is conducted in clinical settings, supervised by experienced physicians. This method, though valuable, is fraught with challenges. It is time-consuming, requiring physicians to juggle their clinical responsibilities with teaching duties (6). Additionally, the patient-centered nature of this training can be stressful for students, especially during sensitive procedures like transvaginal sonography, and uncomfortable for patients who may be hesitant to be examined by inexperienced learners (7–9).

Theoretical ultrasound knowledge is gained from materials and courses, yet practical skills require hands-on practice. Present Ob/Gyn ultrasound training often fails to meet these needs, resulting in a practical skills gap among medical students (10). To address these challenges, peer-assisted learning has been proposed as an alternative. This approach involves students learning from and practicing with fellow students who have been trained as peer tutors (11, 12). While effective in certain areas like echocardiography or abdominal sonography, the intimate nature of Ob/Gyn examinations poses unique challenges for implementing peer-assisted learning with live models (13–15).

As a result, there is a growing consensus on the need for reform in ultrasound education in Ob/Gyn. This reform should align with the evolving needs of students and advancements in modern technology (16–19). Incorporating modern teaching materials and methods is crucial for developing practical specialist skills and understanding aspects of patient safety (4, 5, 20). In this regard, ultrasound simulators have emerged as a key component of innovative training concepts. They offer a risk-free environment for students to practice and hone their skills without the pressures and limitations of real patient interactions (21–23).

The use of simulation techniques in medical education is not new. Simulators have long been used for teaching anatomy, physiology, surgical techniques, and obstetric skills (24–26). Recent advancements in technology have led to the development of high-fidelity obstetric/gynecologic ultrasound simulators and even mobile ultrasound simulation applications for smartphones and tablets, enabling remote learning (27, 28).

Our study investigates how structured ultrasound training affects medical students' learning outcomes in Obstetrics/Gynecology (Ob/Gyn). It is a prospective, randomized study that assesses a program combining theory with simulator-based practice. The goal is to show that this approach significantly improves ultrasound skills in Ob/Gyn, measured by theoretical and practical simulator tests. We also look at secondary outcomes like increased self-confidence in performing real patient ultrasounds, satisfaction with the training, and interest in adopting this method in the curriculum. The results could significantly impact Ob/Gyn medical education, producing more skilled and confident practitioners.

2 Materials and methods

This is a prospective, single-center, randomized, interventional study (Figure 1). The study was designed in accordance with the CONSORT guidelines for reporting parallel group randomized trials and under the guidance for reporting intervention development studies in health research (GUIDED) (29, 30). The study was conducted from January 2022 to December 2022. It involved 5th year medical students that were randomized 1:1 to either receive a combination of theoretical teaching and hands-on training using an ultrasound simulator (study group) or to receive the theoretical lectures alone (control group). The contents of the course program were designed after taking into consideration the quality requirements for ultrasound examination in early and second-trimester pregnancy and the updated recommendations for the performance of basic gynecologic ultrasound examinations of the German Society for Ultrasound in Medicine (DEGUM) (1–3). The training sessions and evaluation tests took place at the Department of Obstetrics and Gynecology of the University Medical Center of the Johannes Gutenberg University Mainz, Germany. The study was designed in cooperation with the Rudolf-Frey Learning Clinic of the Johannes Gutenberg University Mainz and approved by the Ethics Committee of the Medical Association of Rhineland-Palatinate (Number: 2022-16372).

2.1 Selection and description of the ultrasound simulator

Prior to the study, we conducted a testing phase to determine which ultrasound simulator would best meet the needs of our study population. Therefore, we invited several ultrasound simulator companies to allow our medical team to test and evaluate their products. Three of them responded and agreed to take part to the this testing phase (Scantrainer 8:TAS/TVS OBGYN-Education

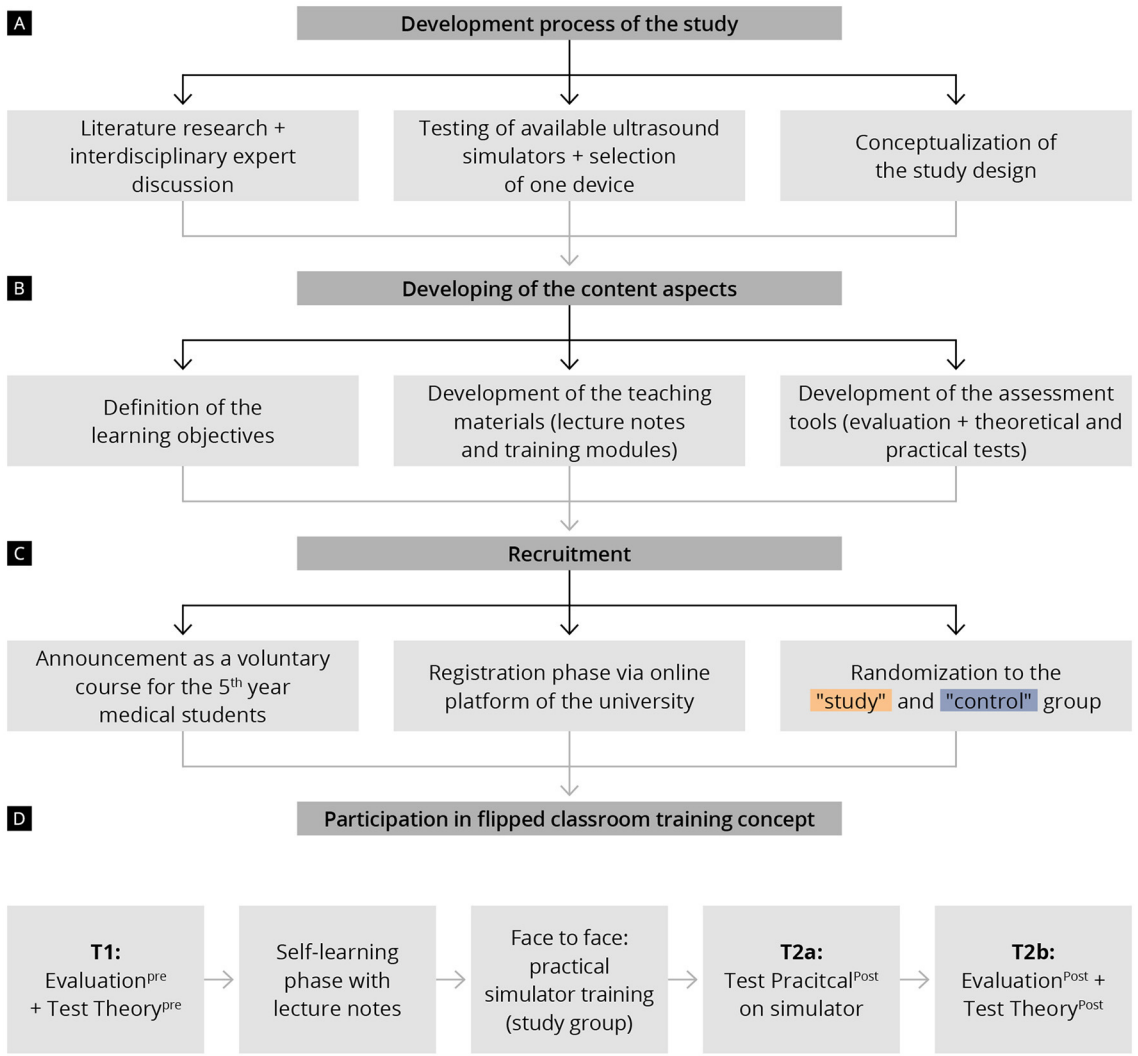


FIGURE 1
Chronological presentation of the study procedure and training course program, including data collection times (T1, T2a, and T2b) **(A)** Development process of the study; **(B)** Developing of the content aspects; **(C)** Recruitment; **(D)** Flipped classroom training concept.

pack), Skillsmed (Nuremberg, Germany), VirtaMed Portable GynoS™ OB/GYN ultrasound simulator (Zurich, Switzerland) and VIM-003 (Ob/Gyn) Simulator Base Unit, CAE Vimedix (Sarasota, United States). The three ultrasound simulators were installed in the Department of Obstetrics and Gynecology of the University Medical Center of the Johannes Gutenberg University Mainz and were available for testing for five working days. Twenty-six participants (14 medical students, nine residents, and three consultants in Obstetrics and Gynecology) took part in this testing phase. The technical characteristics and functions of the simulators, the training platform and courses included in the software, as well as aspects like self-learning and simulator feedback were evaluated. We decided to conduct our study using the Virtamed Portable GynoS™ OB/GYN ultrasound simulator, since our colleagues' evaluation showed a superiority of this simulator with regard to the above-mentioned aspects (Figure 2). The simulator includes transabdominal and transvaginal obstetric ultrasound modules with more than 100 cases available in the training platform. For an overview of all possible modules of the simulator, see Supplementary material S1.

2.2 Participant recruitment and eligibility criteria

The study was a voluntary part of the officially predetermined curriculum of the University Medical Center of the Johannes Gutenberg University Mainz, Germany. Being a 5th year medical student, consenting to participate in the study and completing the assessment tools (evaluation forms, theoretical test, and practical test) were defined as inclusion criteria. An invitation to participate describing the scope of the study was sent via email to all 5th year medical students of winter and summer semesters 2022. Informed written consent was obtained from all participating students.

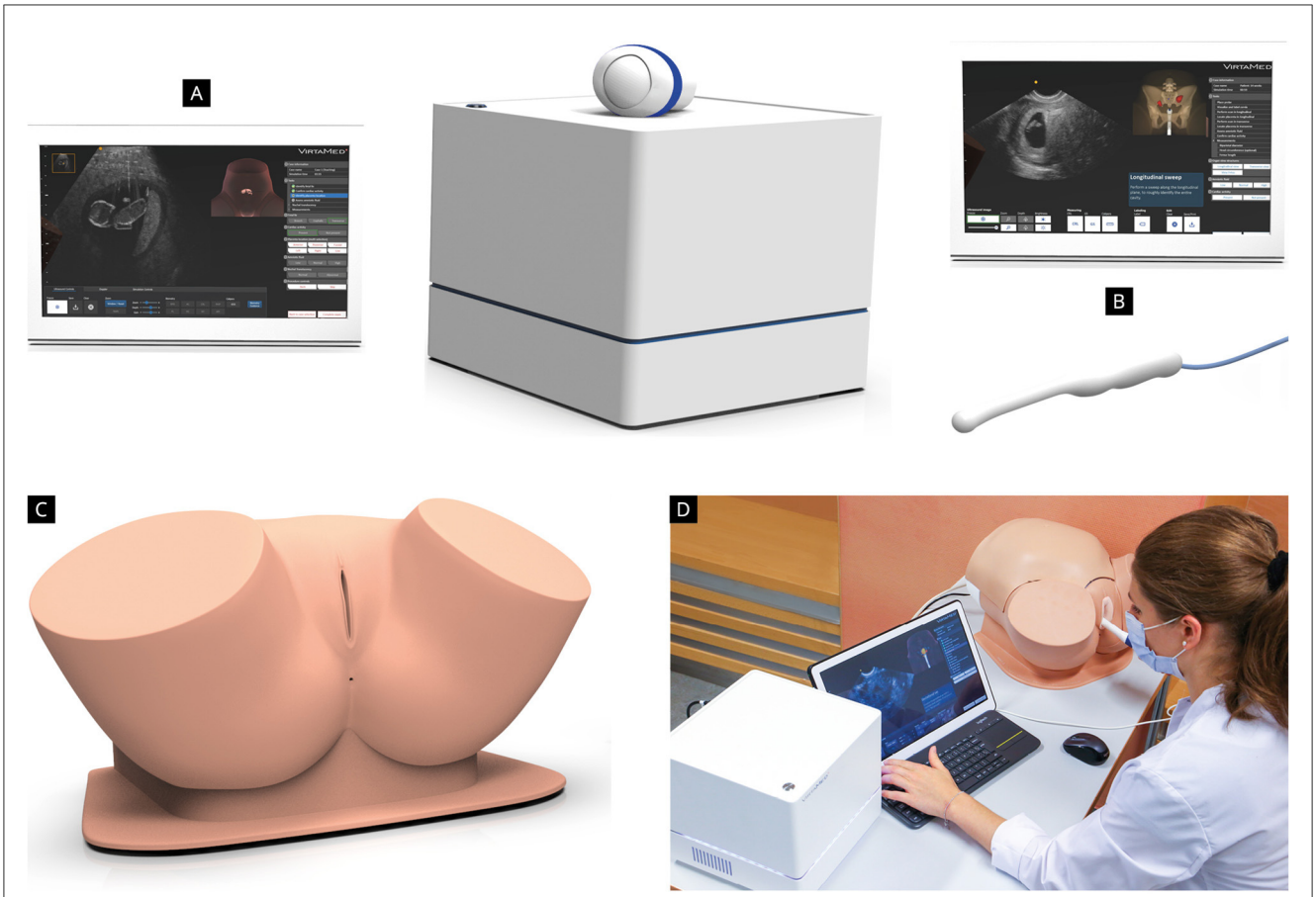


FIGURE 2
VirtaMed Portable GynoSTM™ OB/GYN ultrasound simulator (Zurich, Switzerland) consisting of an anatomic pelvic model equipped **(C, D)** with an anatomical uterus insert, two abdominal models (for fetuses younger than 18 weeks and for fetuses older than 18 weeks), as well as a transabdominal **(A)** and a transvaginal ultrasound **(B)** transducer replica.

2.3 Course program

The course program (Figure 1), based on the flipped classroom model, consisted of a structured preparation phase via lecture notes and one dedicated onsite course with face-to-face teaching on the ultrasound simulator (only for the study group).

2.3.1 Lecture notes

Both groups (study and control group) received lecture notes providing theoretical knowledge about the standard planes of a basic gynecologic and obstetric ultrasound examination (see Supplementary material S2). With the help of anatomical figures and images from real transabdominal and transvaginal ultrasound examinations, the students were guided through the most important normal and pathologic findings of Obstetrics and Gynecology. The following aspects of normal anatomy were covered: uterus in sagittal and transverse view, measurement of the endometrial thickness, normal appearance of the ovaries and pouch of Douglas, the appearance of a normal early pregnancy along with the measurement of the crown-rump length, basic biometrical measurements of the fetus in the second and third trimester, as well as the appearance and localization of the placenta and the evaluation of the cervical length and amniotic fluid volume. With regard to pathologic findings, the following topics were discussed: position variants of the uterus, ovarian cysts, early pregnancy loss, ectopic pregnancy and pregnancy of unknown location, as well as polyhydramnios, oligohydramnios and cervical shortening.

2.3.2 The onsite training course

Only students in the study group received the onsite training course on the ultrasound simulator. The control group received the aforementioned lecture notes only. All practical training sessions were identical and were performed by the same investigator after herself receiving dedicated training courses on the ultrasound simulator until she reached an expert level on the predefined training modules. The training course consisted of four modules and lasted about 75–90 min per student. The predefined learning objectives and practical tasks of each module are shown in Figure 3.

2.4 Evaluation forms and learning assessment

In order to measure obstetric/gynecologic ultrasound competency acquisition and the students' attitude toward the

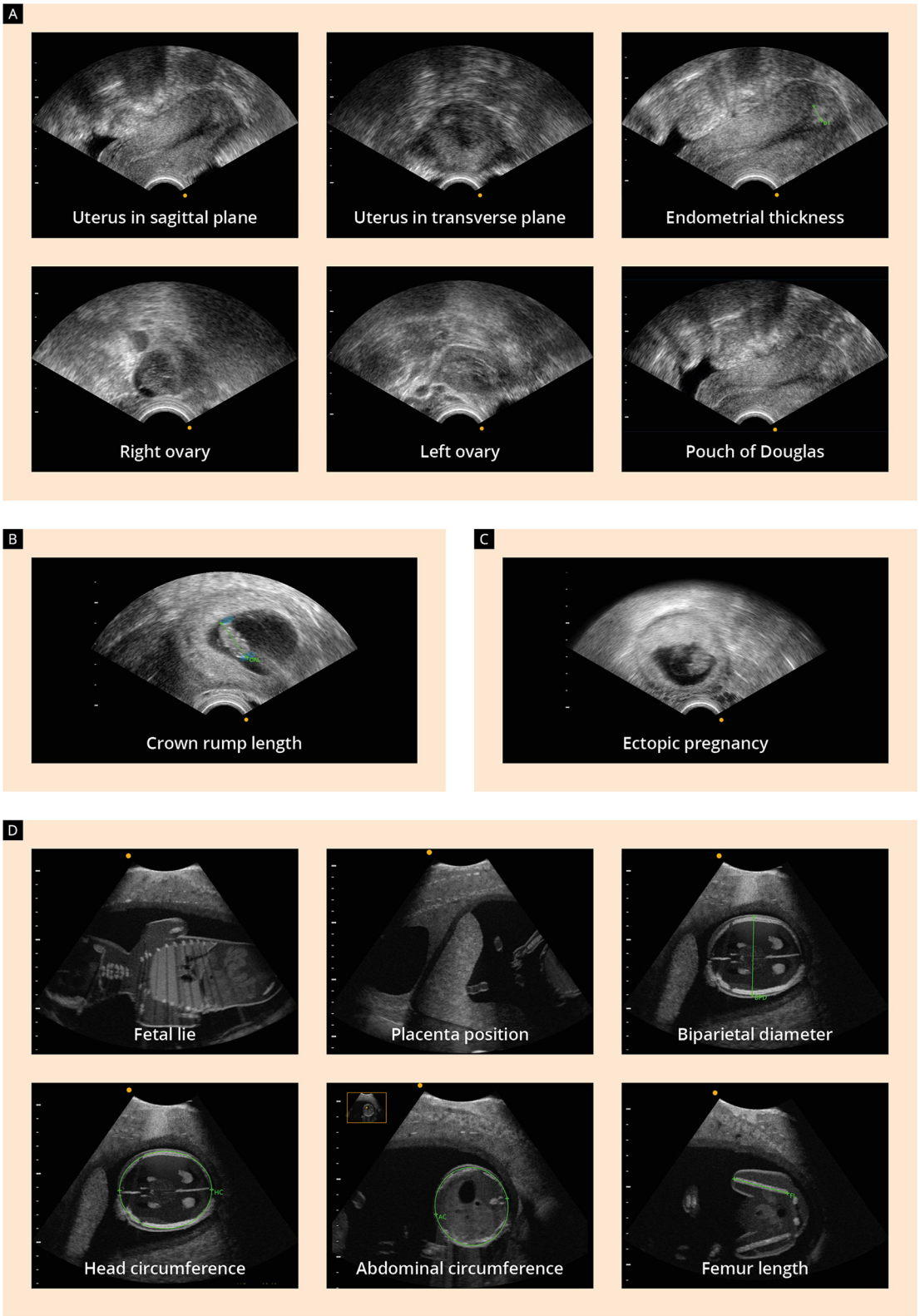


FIGURE 3
Presentation of the four modules of the onsite training course and learning objectives: **(A)** assessment of the uterus and endometrium on the sagittal and transverse plane, measurement of the endometrial thickness, identification of the ovaries and the pouch of Douglas; **(B)** identification of the embryo and measurement of the crown-rump length (CRL) on the 10th week of gestation; **(C)** visualization of an ectopic pregnancy; **(D)** second-trimester ultrasound examination, including identification of the fetal lie and placenta position and performing fetal biometric measurements.

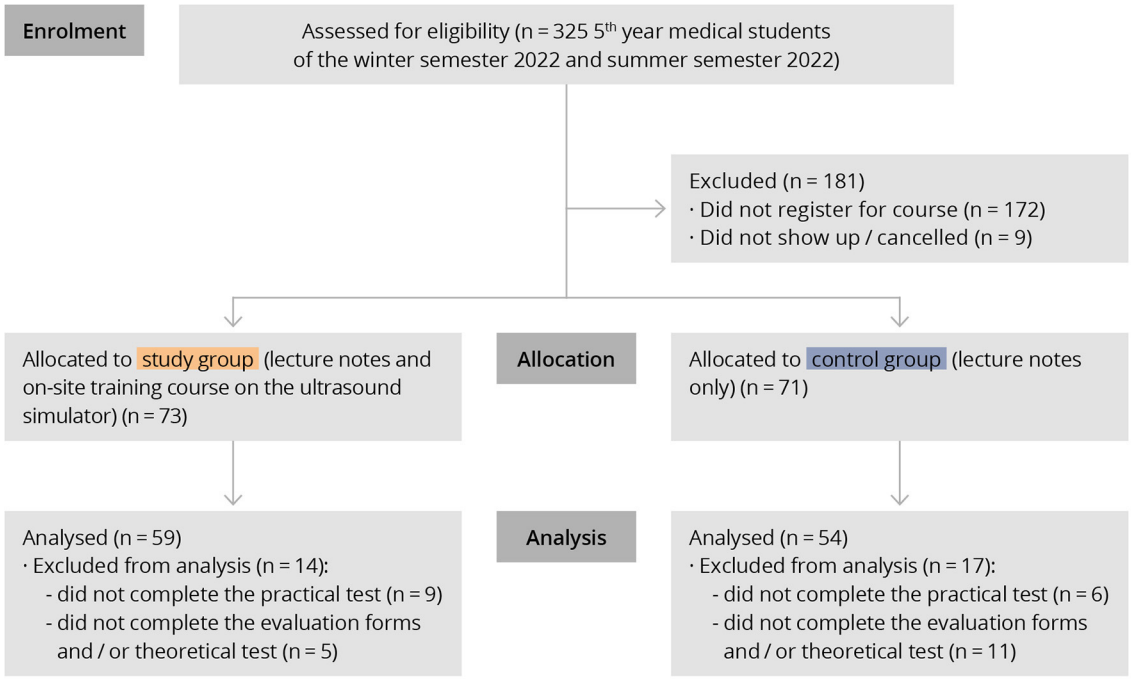


FIGURE 4
Flow Diagram showing participant recruitment and data analysis according to CONSORT guidelines.

course program, written evaluations and assessment of theoretical knowledge were performed at two time points: upon enrollment (T1: Evaluation^{pre} and Test Theory^{pre}) and at the end of the course program (T2b: Evaluation^{post} and Test Theory^{post}). Assessment of practical skills was performed in form of a practical test on the ultrasound simulator also at the end of the course program (T2a: Test Practical^{post}). All evaluation forms and learning assessment tools were designed by the authors of the study based on published assessment methods used to measure ultrasound competency in medical ultrasound education (31, 32).

2.4.1 Evaluation forms at time points T1 and T2b

The demographic characteristics of the participants as well as previous experience with ultrasound on real patients and/or on simulated settings were assessed upon enrollment (time point T1). The participants' subjective level of competency in transabdominal and transvaginal sonography, their level of self-confidence with regard to ultrasound examinations on real patients, their learning goals and motivation to practice on the simulated environment were assessed at time points T1 and T2b using questions based on a 7-point Likert scale (1 = strongly disagree with the statement; 7 = strongly agree with the statement). At time point T2b (Evaluation^{post}) the following aspects were also evaluated: the lecture notes, the properties of the ultrasound simulator, as well as the advantages, future perspectives and attitudes toward simulator-based ultrasound training in the field of obstetrics and gynecology.

2.4.2 Learning assessment: theoretical tests at time points T1 und T2b

Theoretical knowledge related to obstetric/gynecologic ultrasound was assessed at time points T1 (Test Theory^{pre}) and T2b (Test Theory^{post}). The contents of the theoretical tests were based on the predefined aforementioned learning goals. Each test included 21 single choice, multiple choice and free text questions (33) that were subdivided in the following four topics: (1) uterus and Douglas Pouch, (2) ovaries and ovarian pathology, (3) early pregnancy, (4) fetus and placenta (see Supplementary material S3). For the evaluation of the results of the theoretical tests a scoring system was developed. Correct answers scored 1 point and wrong answers zero points. We did not use negative scoring for wrong answers.

2.4.3 Learning assessment: test practical^{post} on the ultrasound simulator at time point T2a

In order to evaluate practical ultrasound skills, the participants underwent the Test Practical^{post} on the ultrasound simulator at time point T2a. Both groups performed the practical test independently, after receiving a short introduction into the settings and functions of the simulator. The practical test was supervised by medical experts who had previously received training on the ultrasound simulator. Technical assistance was provided whenever needed, but no feedback or instructions were given. The Test Practical^{post} (about 25 min) consisted of three cases, that were available within the simulator training platform (see Supplementary material S4).

The entire practical test was recorded for each participant. All images acquired were rated on a later time point by ultrasound experts qualitatively and quantitatively using predetermined rating criteria.

a) Qualitative assessment: The image quality was assessed. Outcome measures included the correct identification of the uterus, endometrium and adnexa, the visualization of a viable first-trimester pregnancy, the correct assessment of the fetal lie and heartbeat in the second trimester of pregnancy, the localization of the placenta and the establishment of the correct diagnosis for modules (a) and (b), using a pass/fail performance level (0: fail, 1: pass).
b) Quantitative assessment: The quantitative assessment was applied only if the students were able to identify and demonstrate correctly the required anatomical level and structure. The deviation of the students' following measurements from the reference value was calculated: measurement of the endometrial thickness, measurement of the crown rump length in the first trimester of pregnancy and biometric measurements (biparietal diameter, head circumference, abdominal circumference and femur length) of the fetus in the second trimester of pregnancy. The time needed to complete each module was also evaluated.

2.5 Statistic

In order to calculate the sample size required to detect a statistical significant effect, a power analysis was performed. Based on an expected effect size of 0.6, a significance level of 0.05 and a desired power of 0.80, the calculated sample size was set at 90 participants. The evaluations and theoretical tests were conducted digitally through an online questionnaire tool and were exported as an Excel spreadsheet. The results of the practical tests were reported manually in an Excel file. All data were manually evaluated using Microsoft Excel before analysis in R studio (RStudio Team [2020]. RStudio: Integrated Development for R. RStudio, PBC, http://www.rstudio.com, last accessed 06 01 2024) with R 4.0.3 (A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, http://www.R-project.org; last accessed 06 01 2024). Binary and categorical baseline parameters are expressed as absolute numbers and percentages. Continuous data are expressed as median and interquartile range (IQR) or as mean and standard deviation (SD). Categorical parameters were compared using Fisher's exact test and continuous parameters using the Mann-Whitney test. The results of the theory test were given as a percentage. In addition, pairwise correlations of metric variables were obtained, and the correlation effect sizes and significances were calculated. Furthermore, Mann-Whitney tests were constructed to compare the influence of individual factors on the results of the theoretical and practical tests. Finally, a multivariate linear regression model was produced to compare the influence of individual factors. $P < 0.05$ were considered statistically significant.

3 Results

3.1 Participants

A total of 325 5th year medical students of the winter and summer semesters 2022 were invited to participate in the study. In total, $n = 153$ students registered for the study. Nine of them canceled the course because of illness and other personal reasons. The final analysis included 113 students (study group $n = 59$, control group $n = 54$) who completed all assessment tools (Figure 4).

Participants' baseline characteristics are presented in Supplementary Table S5. There were no significant differences between the groups regarding age (study: 27.7 ± 3.3 years vs. control: 27.5 ± 4.1 years; $p = 0.70$) and the gender distribution. The majority of the students in both groups reported having previous experience with ultrasound examinations in general (study: 96.6% vs. control: 100%; $p = 0.52$), but most of them never have had contact with ultrasound simulators (study: 98.3% vs. control: 98.1%; $p = 1.00$). The only differences between the groups were that the participants in the control group stated that they had independently performed slightly more transvaginal ultrasound examinations that the ones in the study group and that within the study group, significantly more participants stated that they had already completed an apprenticeship (study: 66.1% vs. control: 40.7%; $p = 0.01$).

3.2 Subjective assessment of competency and self-confidence

Both groups were able to increase their ultrasound skills significantly ($p < 0.01$) over the course period. However, the study group achieved a significantly higher increase in the subjective level of skills competence (study: Δ 1.7 ± 1.0 vs. control: Δ 1.2 ±1.1; $p = 0.03$). Corresponding to that the level of self-confidence in performing ultrasound examinations independently increased to a larger extent in the study group over the study period (study: Δ 4.4 ±1.5 vs. control: Δ 3.3 ±1.5; $p < 0.001$). For further information, see Supplementary material S6.

3.3 Advantages of simulator-based ultrasound training, motivation and learning goals

Both groups rated the" Advantages of simulator-based ultrasound training" with high scale points (T1: study: 6.2 ± 0.80 vs. control: 6.0 ± 0.80; $p = 0.09$; T2b: study: 6.03 ± 0.81 vs. control: 5.7 ± 0.92; $p = 0.07$. At the end of the course, both groups rated the theme complexes "Motivation" and Learning goals" at a similarly high level, although the study group tended to rate both topic complexes with higher scale points. For further information, see Supplementary Table S7.

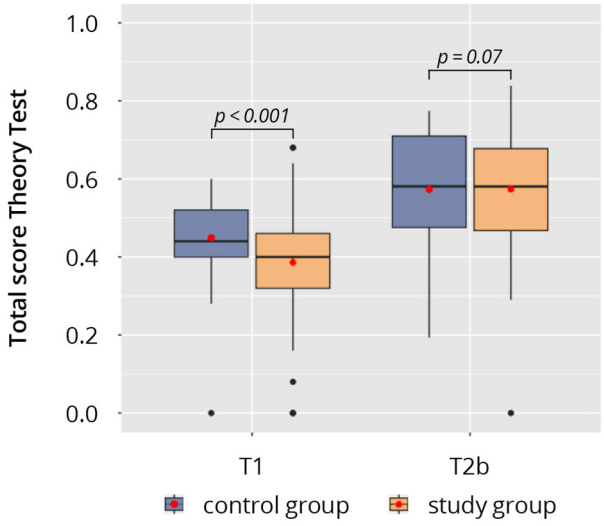


FIGURE 5
Results of the theoretical tests at time points T1 (Test Theory^{pre}) and T2b (Test Theory^{post}).

3.4 Evaluation of the teaching materials, the properties of the ultrasound simulator and future perspectives and attitudes toward simulator-based ultrasound training

Both the lecture notes (study: 5.5 ± 1.2 vs. control: 5.4 ± 1.3; $p = 0.55$) and the properties of the ultrasound simulator (study: 6.3 ± 0.8 vs. control: 6.2 ± 0.9; $p = 0.64$) were rated with similarly high scale points from both groups. The statements of the theme complex "Future perspectives and attitudes toward simulator-based ultrasound training in Obstetrics and Gynecology" were rated with significantly higher scale points from the participants of the study group. For further information, see Supplementary Table S8.

3.5 Learning assessment: theoretical tests and practical test on the ultrasound simulator

3.5.1 Theoretical tests: test theory^{pre} and test theory^{post}

The results of the theoretical tests are presented in Figure 5. Upon enrollment (time point T1, Test Theory^{pre}) the students in the control group achieved significantly higher scores in the theory test than the students in the study group (study: 0.39 ± 0.15 vs. control: 0.45 ± 0.10; $p = 0.01$). Both groups managed to improve their ultrasound knowledge during the course period and achieved significantly higher scores at the theoretical test at time point T2b (Test Theory^{post}) ($p < 0.001$). The study group however achieved a higher increase at the total score between time points T1 and T2b (study: Δ 0.19 ± 0.17 vs. control: Δ 0.12 ± 0.17; $p = 0.08$), even though this did not reach statistical significance.

3.5.2 Practical test on the ultrasound simulator (test practical^{post})

The results of the practical tests on the ultrasound simulator are shown in Figure 6 and Supplementary Table S9. Qualitative and quantitative assessment was performed. Quantitative assessment was, however, only applied if the students were able to identify and demonstrate correctly the required anatomical level and structures. Because the students in the study group were significantly better at doing so (study: 0.95 ± 0.07% vs. control: 0.85 ± 0.10%; $p < 0.001$), more datasets from the study group could be taken into account in the quantitative assessment analysis. Regarding the deviation from the reference values for the measurement of the endometrial thickness, the CRL and the biometric measurements of the fetus, the study group tended to be closer to the reference values than the control group (deviation study: 14.65 ± 13.4 vs. deviation control: 18.83 ± 12.90; $p = 0.30$). In addition, the study group required on average less time to complete the tasks (study: 1,480 ± 301 s vs. control: 1,551 ± 296 s; $p = 0.22$) and was significantly better at correctly interpreting the ultrasound findings than the control group (study: 0.88 ± 0.22% vs. control: 0.72 ± 0.25%; $p < 0.001$).

3.6 Regression analysis and correlations

Multivariable linear regression was performed to identify influential factors and potential confounders (see Supplementary Table S10). Only the subjective level of competency in obstetric/gynecologic ultrasound at time point T1 had a significant influence on the results of the Test Theory^{pre} ($\beta = 0.042$; $p = 0.01$). The participants' reported level of self-confidence with regard to obstetric/gynecologic ultrasound examinations, correlates with better results at the Test Theory^{pre}. This means that the participants, who were feeling more confident (subjective), were also better in the objective assessment of theoretical ultrasound knowledge as well.

With regard to the results of the Test Theory^{post} the variables: "total score on the Test Theory^{pre}" ($\beta = 0.366$; $p < 0.01$) and "studied the lecture notes" ($\beta = 0.02$; $p < 0.01$) were found to have significant influence on the results. Similar to the results of the regression analysis for the Test Theory^{pre}, only the subjective level of competency in obstetric/gynecologic ultrasound at time point T1 had significant influence on the results of the practical test on the ultrasound simulator ($\beta = 0.06$; $p = 0.02$).

Correlation analysis between subjective and objective data are provided in Supplementary Table S11. Significant correlations ($p < 0.05$) were found between the levels of subjective and objective competencies, as well as the participants' attitude and motivation between time points T1 and T2 (T2a and T2b) with a medium/strong effect size ($0.30 \leq r \leq 0.95$).

4 Discussion

This prospective, single-center, randomized study offers a thorough exploration into the evolving landscape of medical education, with a specific focus on the training of obstetric and gynecologic ultrasound skills using high-fidelity simulation technology. This study highlights the critical role

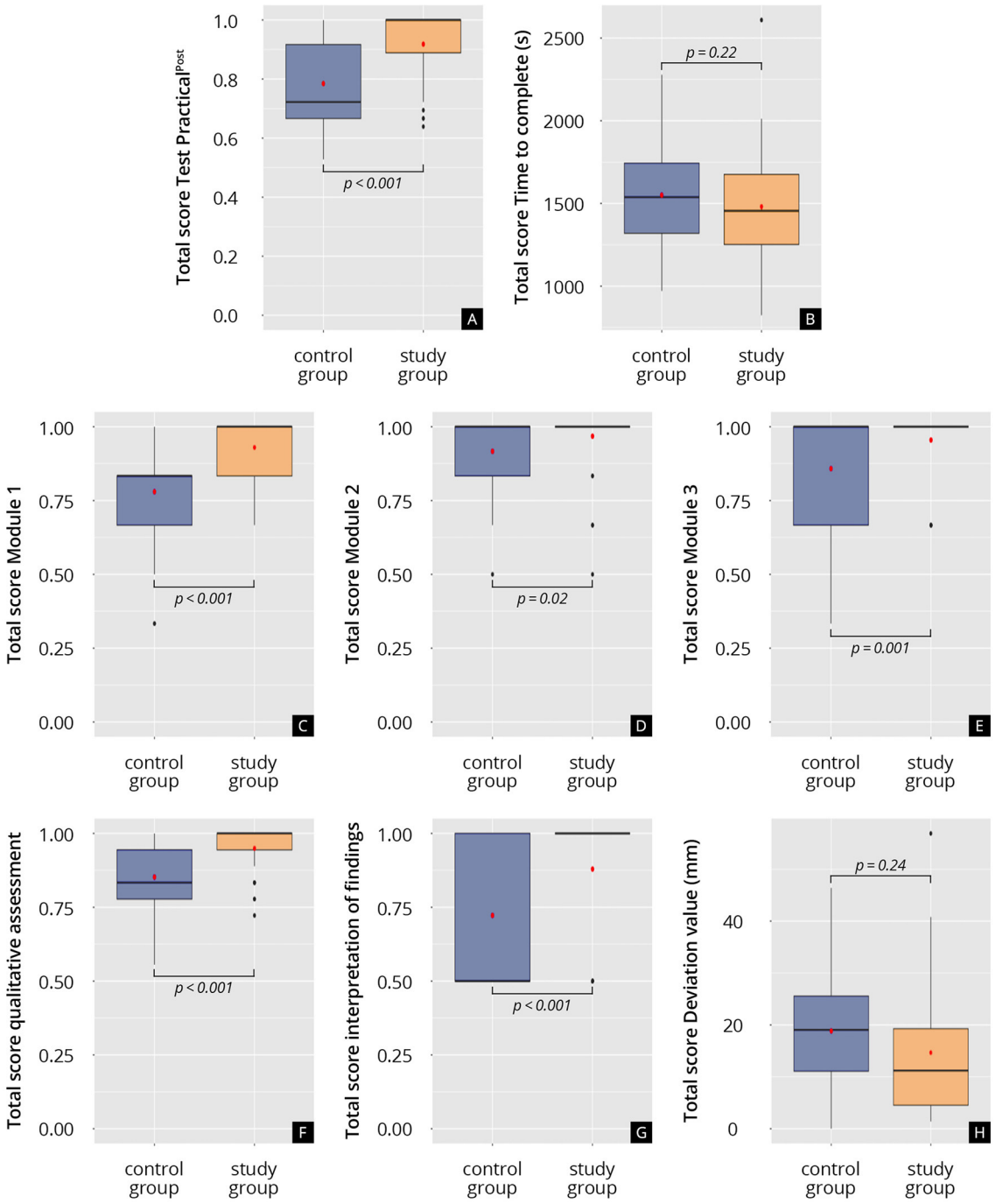


FIGURE 6
Results of the practical test on the ultrasound simulator (Test Practical^{post}). **(A)** Total score on the practical test, test Practical^{post}. **(B)** Time to complete the practical test, s: seconds. **(C)** Total score on Module 1 of the practical test. **(D)** Total score on Module 2 of the practical test. **(E)** Total score on Module 3 of the practical test. **(F)** Total score of the qualitative assessment analysis. **(G)** Total score of the interpretation of findings. **(H)** Total deviation of the measurements from the reference values, mm: millimeter.

of ultrasound in modern medicine, particularly in obstetrics and gynecology, and adopts a scientifically rigorous method to improve medical education through innovative teaching. Its design as a prospective, randomized trial following CONSORT and GUIDED guidelines reflects a commitment to high-quality research. The use of the VirtaMed Portable GynoS™ OB/GYN ultrasound simulator demonstrates a balance between advanced technology and practical use. Findings indicate that both groups improved their ultrasound knowledge, with the study group showing greater improvement and better practical test performance, underscoring the value of extra simulator training. One could argue, that, the group that receives an additional training intervention, is expected to also perform better in the evaluation test, than the group that receives no

intervention. However, this is only partially true, since the impact on skills also depends on the quality of the intervention itself. The better scores of the study group militate in favor of the quality of the training concept. Finally, both groups showed great satisfaction with the training course and rated the teaching materials and properties of the ultrasound simulator with very high scores. The students' evaluations show clearly their positive attitude toward simulator-based ultrasound medical education and their motivation to continue training on the simulated environment.

Our results are in accordance with previous published data regarding simulator-based ultrasound training for medical students. In a multicenter randomized trial, Etienne et al. examined the effect of the addition of a simulation course on the usual training in transvaginal ultrasound for medical students that were trained in an emergency gynecological unit. The course was found to be beneficial and the students were highly satisfied with the session as an initial training method (34). Similar results reported Cook *et al.* after a one-hour simulation training session in obstetric/gynecologic ultrasound for third-year medical students (35).

Several studies have examined the use of obstetric/gynecologic ultrasound simulators in training and evaluating residents, specialist doctors and medical students. Theoretical and practical simulator training as well as simulator-based trainees' evaluation techniques have been shown to be comparable with such that are patient-based (36). Moreover, simulator-based ultrasound training has been shown to be effective not only with regard to practical skills acquisition, but also in other areas of performance, like interpretation of findings, documentation and medical decision-making (37). This study's advocacy for the integration of simulator-based training in obstetrics/gynecology curricula could potentially lead to significant advancements in medical education. This perspective is supported by recent research, which suggests that the incorporation of technology-driven, interactive learning tools can enhance the educational experience and better prepare students for clinical practice (22, 38, 39).

Simulator-based medical training cannot of course replace training on real patients (40). It can, however, be used as an adjunct to traditional clinical methods of education, especially when instruction time and patient availability is limited. This is particularly important in the field of obstetrics and gynecology where the intimate character of the gynecological examination itself poses a great challenge to medical training. Simulation provides the possibility to train practical skills and acquire competencies in a safe, stress-free environment without the fear of committing errors or the risk of harming patients.

4.1 Strengths and limitations of the study

The study's strengths include its randomized design, consistent teaching methods, and standardized evaluation criteria, ensuring uniform training and objective grading. However, limitations include its voluntary nature, possibly attracting more ultrasound-interested students, and varying motivation levels between groups. Additionally, the study didn't assess impacts on patient safety or care quality, though previous research suggests that simulator training improves care efficiency and reduces patient discomfort and the need for repeat examinations and trainee supervision (41).

The study's focus on the immediate impacts of simulator-based training does not address the long-term retention of skills. Understanding how well these competencies are retained over time is crucial to evaluating the effectiveness of the training method. Additionally, while simulators provide a safe and controlled environment, they lack the unpredictability and complexity of real patient interactions. This may limit the preparedness of students for real-life clinical scenarios, which often involve direct patient communication and managing unexpected findings.

There is also a concern that students might become overly dependent on the simulator environment, which might not always be replicated in actual clinical settings where such technology may not be available. The high costs and limited accessibility of high-fidelity simulators further complicate the potential for widespread implementation, particularly in resource-limited settings (42). A specific cost-benefit economic analysis was not performed in our study. Even though we recognize that this is a possible limitation of the study, there are effects of simulator-based ultrasound training, like students 'preparation for clinical practice, effects on patient care and safety, as well as the satisfaction with the educational experience, that are impossible to measure with money. The ultrasound simulator is still available in our clinic for medical students to practice voluntarily and following projects are already planned.

The study predominantly focuses on quantitative assessments of knowledge and skill, possibly overlooking qualitative aspects such as learning style preferences or the subjective experience of learning with simulators. Additionally, the control group in the study received only theoretical teaching without practical hands-on experience, not considering other practical learning methods that do not involve high-fidelity simulators. This design choice might limit the scope of the study's conclusions.

4.2 Conclusion

The "GynSim" study marks a significant step forward in medical education for obstetrics and gynecology ultrasound training. This study shows how structured simulator-based training positively affects 5th-year medical students' educational outcomes, enhancing their theoretical knowledge, practical skills, and confidence. The findings support the effectiveness of simulation-based training, noted for its methodological strength, including randomization, a large sample size, consistent teaching, and objective evaluation. These aspects ensure the study's reliability and contribute insights into the benefits of simulator-based training, highlighting its role in preparing students for clinical practice and advocating for further research on its long-term benefits and curriculum integration to improve patient care and safety.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of the Medical Association of Rhineland-Palatinate. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

JW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing. FR: Methodology, Validation, Writing—review & editing. AH: Conceptualization, Investigation, Resources, Supervision, Writing—review & editing. HB: Conceptualization, Investigation, Resources, Supervision, Validation, Writing—review & editing. KK: Conceptualization, Data curation, Investigation, Methodology, Writing—review & editing. LB: Data curation, Methodology, Writing—review & editing. AW: Data curation, Formal analysis, Writing—review & editing. LS: Data curation, Visualization, Writing—review & editing. LL: Data curation, Visualization, Writing—review & editing. RK: Supervision, Writing—review & editing. AD: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

This study includes parts of the doctoral thesis of one of the authors (KK). We thank all participating students and lecturers for supporting our study. We would like to express our gratitude to Kay Stankov for his contributions to this publication. His dedicated efforts in consulting, supervising, and meticulously reviewing all statistical aspects have been instrumental in ensuring the rigor and accuracy of our research findings. We would like to also thank C. Christe for her help in revising the figures.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1371141/full#supplementary-material

References

1. Grab D, Merz E, Eichhorn KH, Tutschek B, Kagan KO, Heling KS, et al. Basic gynecologic ultrasound examination (level I): DEGUM, ÖGUM, and SGUM recommendations. *Ultraschall in der Medizin*. (2023) 44:151–61. doi: 10.1055/a-1851-5157

2. Rempen A, Chaoui R, Häusler M, Kagan KO, Kozlowski P, von Kaisenberg C, et al. Quality requirements for ultrasound examination in early pregnancy (DEGUM Level I) between 4+0 and 13+6 weeks of gestation. *Ultraschall in der Medizin*. (2016) 37:579–83. doi: 10.1055/s-0042-115581

3. Kähler C, Schramm T, Bald R, Gembruch U, Merz E, Eichhorn KH. Updated DEGUM quality requirements for the basic prenatal screening ultrasound examination (DEGUM Level I) between 18 + 0 and 21 + 6 weeks of gestation. *Ultraschall in der Medizin*. (2020) 41:499–503. doi: 10.1055/a-1018-1752

4. Cantisani V, Dietrich CF, Badea R, Dudea S, Prosch H, Cerezo E, et al. EFSUMB statement on medical student education in ultrasound [long version]. *Ultrasound Int Open*. (2016) 2:E2–7. doi: 10.1055/s-0035-1569413

5. Hoffmann B, Blaivas M, Abramowicz J, Bachmann M, Badea R, Braden B, et al. Medical student ultrasound education, a WFUMB position paper, part II. A consensus statement of ultrasound societies. *Med Ultrason*. (2020) 22:220–9. doi: 10.11152/mu-2599

6. Salvesen KA, Lees C, Tutschek B. Basic European ultrasound training in obstetrics and gynecology: where are we and where do we go from here? *Ultrasound Obstet Gynecol*. (2010) 36:525–9. doi: 10.1002/uog.8851

7. Tolsgaard MG, Rasmussen MB, Tappert C, Sundler M, Sorensen JL, et al. Which factors are associated with trainees' confidence in performing obstetric and gynecological ultrasound examinations? *Ultrasound Obstet Gynecol*. (2014) 43:444–51. doi: 10.1002/uog.13211

8. Santen SA, Hemphill RR, Spanier CM, Fletcher ND. 'Sorry, it's my first time!' Will patients consent to medical students learning procedures? *Med Educ*. (2005) 39:365–9. doi: 10.1111/j.1365-2929.2005.02113.x

9. Graber MA, Wyatt C, Kasparek L, Xu Y. Does simulator training for medical students change patient opinions and attitudes toward medical student procedures in the emergency department? *Acad Emerg Med*. (2005) 12:635–9. doi: 10.1111/j.1553-2712.2005.tb00920.x

10. Nitsche JF, Brost BC. Obstetric ultrasound simulation. *Semin Perinatol*. (2013) 37:199–204. doi: 10.1053/j.semperi.2013.02.012

11. Herrmann-Werner A, Gramer R, Erschens R, Nikendei C, Wosnik A, Griewatz J, et al. Peer-assisted learning (PAL) in undergraduate medical education: an overview. *Z Evid Fortbild Qual Gesundhwes*. (2017) 121:74–81. doi: 10.1016/j.zefq.2017.01.001

12. Dickerson J, Paul K, Vila P, Whiticar R. The role for peer-assisted ultrasound teaching in medical school. *Clin Teach*. (2017) 14:170–4. doi: 10.1111/tct.12541

13. Weimer J, Rolef P, Müller L, Bellhäuser H, Göbel S, Buggenhagen H, et al. FoCUS cardiac ultrasound training for undergraduates based on current national guidelines: a prospective, controlled, single-center study on transferability. *BMC Med Educ*. (2023) 23:80. doi: 10.1186/s12909-023-04062-1

14. Weimer J, Dionysopoulou A, Strelow KU, Buggenhagen H, Weinmann-Menke J, Dirks K, et al. Undergraduate ultrasound training: prospective comparison of two different peer assisted course models on national standards. *BMC Med Educ.* (2023) 23:513. doi: 10.1186/s12909-023-04511-x

15. Weimer JM, Widmer N, Strelow K-U, Hopf P, Buggenhagen H, Dirks K, et al. Long-term effectiveness and sustainability of integrating peer-assisted ultrasound courses into medical school - a prospective study. *Tomography.* (2023) 9:1315–28. doi: 10.3390/tomography9040104

16. Wolf R, Geuthel N, Gnatzy F, Rotzoll D. Undergraduate ultrasound education at German-speaking medical faculties: a survey. *GMS J Med Educ.* (2019) 36:4. doi: 10.3205/zma001242

17. Prosch H, Radzina M, Dietrich CF, Nielsen MB, Baumann S, Ewertsen C, et al. Ultrasound curricula of student education in europe: summary of the experience. *Ultrasound Int Open.* (2020) 6:E25–e33. doi: 10.1055/a-1183-3009

18. Bahner DP, Goldman E, Way D, Royall NA, Liu YT. The state of ultrasound education in U.S. medical schools: results of a national survey. *Acad Med.* (2014) 89:12. doi: 10.1097/ACM.0000000000000414

19. Tarique U, Tang B, Singh M, Kulasegaram KM, Ailon J. Ultrasound curricula in undergraduate medical education: a scoping review. *J Ultrasound Med.* (2018) 37:69–82. doi: 10.1002/jum.14333

20. Ziv A, Wolpe PR, Small SD, Glick S. Simulation-based medical education: an ethical imperative. *Acad Med.* (2003) 78:783–8. doi: 10.1097/00001888-200308000-00006

21. Chalouhi GE, Bernardi V, Ville Y. Ultrasound simulators in obstetrics and gynecology: state of the art. *Ultrasound Obstet Gynecol.* (2015) 46:255–63. doi: 10.1002/uog.14707

22. Taksoe-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:E42–54. doi: 10.1055/a-1300-1680

23. Dietrich CF, Lucius C, Nielsen MB, Burmester E, Westerway SC, Chu CY, et al. The ultrasound use of simulators, current view, and perspectives: Requirements and technical aspects (WFUMB state of the art paper). *Endosc Ultrasound.* (2023) 12:38–49. doi: 10.4103/EUS-D-22-00197

24. Owen H. Early use of simulation in medical education. *Simulat Healthc.* (2012) 7:102–16. doi: 10.1097/SIH.0b013e3182415a91

25. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA.* (2011) 306:978–88. doi: 10.1001/jama.2011.1234

26. Issenberg SB, McGaghie WC, Hart IR, Mayer JW, Felner JM, Petrusa ER, et al. Simulation technology for health care professional skills training and assessment. *JAMA.* (1999) 282:861–6. doi: 10.1001/jama.282.9.861

27. Østergaard ML, Ewertsen C, Konge L, Albrecht-Beste E, Bachmann Nielsen M. Simulation-based abdominal ultrasound training - a systematic review. *Ultraschall in der Medizin.* (2016) 37:253–61. doi: 10.1055/s-0042-100452

28. Hartmann TJ, Friebe-Hoffmann U, de Gregorio N, de Gregorio A, Lato C, Hüner B, et al. Novel and flexible ultrasound simulation with smartphones and tablets in fetal echocardiography. *Arch Gynecol Obstet.* (2022) 305:19–29. doi: 10.1007/s00404-021-06102-x

29. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ.* (2010) 340:c869. doi: 10.1136/bmj.c869

30. Duncan E, O'Cathain A, Rousseau N, Croot L, Sworn K, Turner KM, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. *BMJ Open.* (2020) 10:e033516. doi: 10.1136/bmjopen-2019-033516

31. Le Lous M, Klein M, Tesson C, Berthelemy J, Lavoue V, Jannin P. Metrics used to evaluate obstetric ultrasound skills on simulators: a systematic review. *Eur J Obstet Gynecol Reprod Biol.* (2021) 258:16–22. doi: 10.1016/j.ejogrb.2020.12.034

32. Höhne E, Recker F, Dietrich CF, Schäfer VS. Assessment methods in medical ultrasound education. *Front Med (Lausanne).* (2022) 9:871957. doi: 10.3389/fmed.2022.871957

33. Puthiaparampil T, Rahman MM. Very short answer questions: a viable alternative to multiple choice questions. *BMC Med Educ.* (2020) 20:141. doi: 10.1186/s12909-020-02057-w

34. Etienne M, Gabay L, Levaillant JM, Vivanti A, Dommergues M, Fernandez H, et al. Benefits of using a simulator in the initial training for transvaginal ultrasound examination in gynecologic emergency unit. *J Gynecol Obstet Human Reprod.* (2021) 50:101938. doi: 10.1016/j.jogoh.2020.101938

35. Cook J, Rao VV, Bell F, Durkin M, Cone J, Lane-Cordova A, et al. Simulation-based clinical learning for the third year medical student: effectiveness of transabdominal and transvaginal ultrasound for elucidation of OB/GYN scenarios. *J Clin Ultrasound.* (2020) 48:457–61. doi: 10.1002/jcu.22888

36. 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) 214:525.e1–e8. doi: 10.1016/j.ajog.2015.10.932

37. Tolsgaard MG, Ringsted C, Dreisler E, Norgaard LN, Petersen JH, Madsen ME, et al. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. *Ultrasound Obstet Gynecol.* (2015) 46:312–8. doi: 10.1002/uog.14780

38. 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. *Simulat Healthc.* (2016) 11:340–4. doi: 10.1097/SIH.0000000000000176

39. Hani S, Chalouhi G, Lakissian Z, Sharara-Chami R. Introduction of ultrasound simulation in medical education: exploratory study. *JMIR Med Educ.* (2019) 5:e13568. doi: 10.2196/13568

40. Moak JH, Larese SR, Riordan JP, Sudhir A, Yan GF. Training in transvaginal sonography using pelvic ultrasound simulators versus live models: a randomized controlled trial. *Acad Med.* (2014) 89:1063–8. doi: 10.1097/ACM.0000000000000294

41. Tolsgaard MG, Ringsted C, Rosthøj S, Nørgaard L, Møller L, Freiesleben NC, et al. The effects of simulation-based transvaginal ultrasound training on quality and efficiency of care: a multicenter single-blind randomized trial. *Ann Surg.* (2017) 265:630–7. doi: 10.1097/SLA.0000000000001656

42. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: the missing outcome in simulation-based medical education research: a systematic review. *Surgery.* (2013) 153:160–76. doi: 10.1016/j.surg.2012.06.025

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 02 May 2024
DOI 10.3389/fmed.2024.1380940

OPEN ACCESS

EDITED BY
Florian Recker,
University of Bonn, Germany

REVIEWED BY
Maximilian Riedel,
Technical University of Munich, Germany
Valeria Sebri,
European Institute of Oncology (IEO), Italy

*CORRESPONDENCE
Sebastian Griewing
✉ s.griewing@uni-marburg.de

RECEIVED 02 February 2024
ACCEPTED 24 April 2024
PUBLISHED 02 May 2024

CITATION
Griewing S, Knitza J, Gremke N, Wallwiener M, Wagner U, Lingenfelder M and Kuhn S (2024) Awareness and intention-to-use of digital health applications, artificial intelligence and blockchain technology in breast cancer care.
Front. Med. 11:1380940.
doi: 10.3389/fmed.2024.1380940

Awareness and intention-to-use of digital health applications, artificial intelligence and blockchain technology in breast cancer care

Sebastian Griewing[1,2,3,4]*, Johannes Knitza[2], Niklas Gremke[3], Markus Wallwiener[4,5], Uwe Wagner[3,4], Michael Lingenfelder[1] and Sebastian Kuhn[2]

[1]Institute for Healthcare Management, Chair of General Business Administration, Philipps-University Marburg, Marburg, Germany, [2]Institute for Digital Medicine, University Hospital Marburg, Philipps-University Marburg, Marburg, Germany, [3]Department of Gynecology and Obstetrics, University Hospital Marburg, Philipps-University Marburg, Marburg, Germany, [4]Commission for Digital Medicine, German Society for Gynecology and Obstetrics, Berlin, Germany, [5]Department of Gynecology and Obstetrics, Martin-Luther University Halle-Wittenberg, Halle, Germany

Emerging digital technologies promise to improve breast cancer care, however lack of awareness among clinicians often prevents timely adoption. This study aims to investigate current awareness and intention-to-use of three technologies among breast cancer healthcare professionals (HCP): (1) digital health applications (DHA), (2) artificial intelligence (AI), and (3) blockchain technology (BC). A 22-item questionnaire was designed and administered before and after a 30min educational presentation highlighting technology implementation examples. Technology awareness and intention-to-use were measured using 7-point Likert scales. Correlations between demographics, technology awareness, intention-to-use, and eHealth literacy (GR-eHEALS scale) were analyzed. 45 HCP completed the questionnaire, of whom 26 (57.8%) were female. Age ranged from 24 to 67 {mean age (SD): 44.93±12.62}. Awareness was highest for DHA (68.9%) followed by AI (66.7%) and BC (24.4%). The presentation led to a non-significant increase of intention-to-use AI {5.37 (±1.81) to 5.83 (±1.64)}. HCPs´ intention-to-use BC after the presentation increased significantly {4.30 (±2.04) to 5.90 (±1.67), $p<0.01$}. Mean accumulated score for GR-eHEALS averaged 33.04 (± 6.61). HCPs´ intended use of AI significantly correlated with eHealth literacy ($\rho=0.383$; $p<0.01$), intention-to-use BC ($\rho=0.591$; $p<0.01$) and participants´ age ($\rho=-0.438$; $p<0.01$). This study demonstrates the effect that even a short practical presentation can have on HCPs´ intention-to-use emerging digital technologies. Training potential professional users should be addressed alongside the development of new information technologies and is crucial to increase HCPs´ corresponding awareness and intended use.

KEYWORDS

artificial intelligence, blockchain, digital health application, gynecology, breast cancer, oncology

1 Introduction

Breast cancer is the most prevalent oncological entity affecting women in Germany, with over 70,000 new cases emerging each year (1). The adoption of digital technology in breast cancer care is becoming ever more essential due to the increasingly challenging care reality. While there was a short-lived decrease in cases in Germany thanks to improved screening processes, an aging population is likely to reverse this trend, resulting in more cases over time and increased demand for treatment. This challenge is intensified by the particularly favorable survival rate, which resulted from intensive research into innovative treatment options in the past decades, but simultaneously requires long-term follow-up adjuvant treatment to monitor recurrence and side effects, which demands ongoing commitment to patients (2, 3). On the other hand, persistent scientific efforts continue to uncover novel breast cancer treatment modalities. As a result, therapy options to treating breast cancer are rapidly evolving due to breakthroughs in diagnostic and treatment technologies. Recent diffusion of diagnostic tools, i.e., advanced genetic sequencing and the introduction of precision-targeted therapies including antibody-drug conjugates, are paving the way to the personalized cancer treatment approach (4, 5). This progress is accompanied with an overwhelming amount of complex data and information, which increasingly overwhelm practitioners in terms of complexity (6, 7).

In the meanwhile, the system of breast cancer treatment is confronted with an increasingly challenging care infrastructure (8). Sharing vital patient information, which is essential for making informed decisions in cancer treatment, is hindered by missing data infrastructure, interoperability and privacy issues (9). At the same time, the healthcare workforce is burdened by excessive documentation requirements and is shrinking due to demographic aging (10).

Innovative digital technologies can handle large-scale health data, uphold the corresponding privacy of patient data, and alleviate economic strains caused by increased patient numbers (11–13). According to a national study by the Commission for Digital Medicine of the German Society of Gynecology and Obstetrics (DGGG), the majority of gynecology specialists remain optimistic that digital advancements will alleviate respective challenges, enhance patient care, and foresee the integration of sophisticated algorithms (14).

These algorithms are commonly grounded in artificial intelligence (AI) techniques, including machine learning, deep learning, natural language processing, or neural networks. Although AI has become an integral part of daily life, it often stays unnoticed by its users. A recent UK study revealed that merely 17% of adults are cognizant of their engagement with AI technologies (15). In the realm of healthcare, AI applications have seen significant expansion in various domains such as image analysis, automated diagnostic procedures, intelligent drug delivery systems, and personalized treatment approaches (16). Specifically, AI technologies in breast cancer-related imaging, pathology, and supportive care not only alleviate the workload of healthcare professionals but also improve the accuracy and effectiveness in diagnosing and treating breast cancer (17).

Public awareness and understanding of emerging digital technologies vary. While over a third of the Germans is familiar and more than half feel confident in explaining AI, 38% have never heard of Blockchain (18).

Blockchain technology (BC), initially linked with the launch of the cryptocurrency Bitcoin in 2007, has rapidly evolved to encompass application areas extending beyond decentralized finance (19). Conceptually distinct from its role in payment networks, a BC functions as a distributed database, where data are stored across participating ledgers, eliminating the need for centralized storage (20). The adaptation of BC in healthcare has demonstrated its efficacy in providing efficient, decentralized data management, addressing existing data silos and high operational costs (21, 22). This approach was implemented in Estonia's blockchain-based healthcare system infrastructure in 2008, showcasing the potential advantages in terms of cost and efficiency and has been transferred to diverse care settings, including breast cancer treatment (10, 22, 23). Drawing from these developments, the European Commission initiated regulatory actions in 2022 that aim to establish a European health data space, leveraging the full potential of decentralized data management for healthcare advancements in Europe (21).

In contrast to the more sophisticated functionality of AI and BC technology, digital health applications (DHA) have already found their way into German healthcare renumeration system. To leverage quick access to the benefits of digital medicine on clinical care, the "app on prescription" was introduced with the passing of the Digital Healthcare Act (Digitale-Versorgungs-Gesetz, DVG) in 2019, entitling over 70 million people with statutory health insurance receive reimbursement for DHA, called Digitale Gesundheitsanwendungen (DiGA). Up to now, 59 DHA have mastered the regulatory hurdles including three applications for breast cancer.

Nevertheless, in Germany, a cumbersome digitization process prevails and leads to an increasing gap between these evidenced clinical benefits of emerging digital technologies and their actual use, also in gynecological oncology (14, 24, 25). Their integration into healthcare settings rely on the attitudes and characteristics of both clinicians and patients for sustained awareness and intention-to-use (26, 27). Previous studies, i.e., on the integration of virtual reality in breast cancer treatment processes, underscore the importance of considering physicians' psychological attitudes towards new technologies and exploring the transformative potential of interventions tailored to patients' needs and outcomes (28, 29). To the best of our knowledge, the perception of emerging digital technologies, such as AI and BC, has not been investigated comprehensively within the context gyne-oncology.

This study seeks to evaluate healthcare professionals' awareness and intention-to-use of these innovative technologies, and to determine if enhancing professional education could increase their intended use in breast cancer care. Furthermore, this research delves into the under-explored area of eHealth literacy among health professionals in gyne-oncology and its relation to both factors. As emerging technologies like AI and BC offer evidenced benefits for clinical care, it is essential to explore how professional education can bridge the gap between the potential and actual usage.

2 Materials and methods

2.1 Study design

The study was conducted as part of the regional gynecological specialist training days on breast cancer care (GYN-Fortbildungstage

Mammakarzinom) in Marburg, Germany on December 6, 2023. The event encompassed presentations about current developments in breast cancer care, i.e., targeted therapies for breast cancer including antibody-drug conjugates or changes in reimbursable disease management programs for breast cancer patients. Only caregivers which were actively involved in patient care of the regional breast cancer care network of the county of Marburg-Biedenkopf did attend. Recruitment for the study was based on voluntary participation in the survey by the attendees of the mentioned event. An ethics vote was waived by the Research Ethics Committee of Philipps-University Marburg on November 13, 2023 (23-279 ANZ) with reference to the necessity of primary data anonymization, neglecting the option for follow-up survey. Participants who did not want to take part in the survey were able to leave the room beforehand. No one took advantage of this option, all attending participants in the event took part. The requirement for inclusion in the study was the active participation in the regional breast cancer care network and patient care. Exclusion was performed for participants who do not have a medical profession (e.g., administrative employees). The comprehensive survey was initially shared with attendees via a QR-code, with responses gathered directly through Google Forms (Google LLC, Mountain View, United States). Following the survey, participants underwent a 30 min educational presentation about the application of artificial intelligence and blockchain in managing breast cancer. The presentation included the demonstration of previously publicized concepts on combined application of artificial intelligence, i.e., the use of large language models as an adjunct for tumor boards, as well as blockchain technology, i.e., as technological framework for decentralized interoperable data management in care networks (8, 10, 30). After the presentation, two questions concerning the intended use of these technologies were posed once more.

2.2 Study population

The study encompassed physician and non-physician healthcare professionals involved in breast cancer care within the regional network under investigation. The preliminary set of questions aimed to gather fundamental demographic data, including age, gender, and details concerning their practice in either outpatient or inpatient settings, in addition to their level of professional training.

2.3 Questionnaire

The initial survey consisted of 22 items (see Supplementary material S1 for original German version), beginning with four questions to capture basic demographic details of the participants as described in 2.2. This was followed by a series of 10 questions evaluating the participants' awareness and intention-to-use of three technologies (HCP): (1) digital health applications (DHA), (2) artificial intelligence (AI), and (3) blockchain technology (BC). Technology awareness was measured in binominal manner (yes or no) while intention-to-use was assessed using a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). The survey then addressed the components of the German version of the eHealth Literacy Scale (GR-eHEALS 1–8) (31, 32). Table 1 is provided to summarize the questionnaire and the abbreviations of its items.

2.4 eHealth literacy scale

The eHealth Literacy Scale (eHEALS), originally introduced by Norman and Skinner in 2006, has since been adapted into various languages and validated across multiple settings (32). The 8-item scale is used for assessing electronic health literacy in research populations, employing a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with total scores spanning from 8 to 40. Higher scores denote greater literacy levels. In our study, we utilized the validated German version of the eHEALS (GR-eHEALS) by Marsall et al. (31). This study marks the first application of GR-eHEALS among health professionals in the field of gynecology, prompting us to examine reliability and validity in this novel context in line with Norman and Skinner. Internal consistency was measured through Cronbach's alpha. Validity was examined using exploratory factor analysis. We employed the maximum likelihood method, adopting factors with an eigenvalue exceeding 1, in line with the Kaiser criterion. The decision on the number of factors to retain was based on the analysis of a scree plot.

2.5 Data analysis

Data analysis was conducted in IBM SPSS Statistics (Version 29.0.2.0 (20); IBM Corporation, Armonk, United States). We computed Spearman's rho to establish nonparametric correlations for the variables of age, ITU_AI, ITU_BC, and eHEALS_Total, and these were then evaluated for their statistical two-tailed significance. Additionally, we carried out Mann–Whitney Test to compare ITU_AI and ITU_BC scores across binary categories of gender and sector.

2.6 Pre-and post-comparison of intention-to-use

Following the educational session, the intention-to-use query was repeated for the investigated technologies (ITU_AI; ITU_BC). Differences in pre-and post-comparison were assessed by conducting a two-tailed Sign test on the two variable pairings (ITU_AI_PRE → ITU_AI_POST; ITU_BC_PRE → ITU_BC_POST).

3 Results

3.1 Demographics

The study population is characterized by a gender distribution of 57.8% female {mean age (SD): 48.26 (±13.96)} and 42.2% male {42.50 (±11.20)} for 45 participants {44.93 (±12.62), minimum 24 to maximum 67 years}. In terms of sectoral engagement, 35.6% are predominantly involved in outpatient care, while a majority of 64.4% operates within an inpatient setting. With respect to professional qualifications, the cohort comprises one-third specialists, accompanied by 24.4% of doctors in residency and 17.8% serving as attending physicians (see Figure 1).

TABLE 1 Questionnaire items.

Questionnaire part	Item	Question	Measurement
Demographics of study population	Age	How old are you?	Age in years
	Gender	Which gender do you consider yourself to be?	Single choice (Female, Male or Diverse)
	Sector	Do you work primarily in inpatient or outpatient care?	Single choice (Inpatient or Outpatient)
	Profession	In which position do you work?	Single choice (Other Medical Professional (No Physician), Final Year Medical Student, Resident Physician, Specialist Physician, Attending Physician, Chief Physician)
Digital health applications	DiGA1	I know digital health applications (DHA).	Single choice (yes or no)
	DiGA2	As a doctor, I have already prescribed digital health applications (DHA).	Single choice (yes or no)
Artificial intelligence	AI_1	I am aware of applications of artificial intelligence.	Single choice (yes or no)
	AI_2	I am aware of applications of artificial intelligence in healthcare.	Single choice (yes or no)
	AI_3	I am aware of applications of artificial intelligence breast cancer care.	Single choice (yes or no)
	ITU_AI	I would use applications of artificial intelligence in breast cancer care.	7-point Likert scale
Blockchain technology	BC_1	I am aware of applications of blockchain technology.	Single choice (yes or no)
	BC_2	I am aware of applications of blockchain technology in healthcare.	Single choice (yes or no)
	BC_3	I am aware of applications of blockchain technology in breast cancer care.	Single choice (yes or no)
	ITU_BC	I would use applications of blockchain technology in breast cancer care.	7-point Likert scale
GR-eHEALS	GR_eHEALS_1	I know how to find helpful health resources on the Internet.	5-point Likert scale
	GR_eHEALS_2	I know how to use the Internet to answer my health questions.	5-point Likert scale
	GR_eHEALS_3	I know what health resources are available on the Internet.	5-point Likert scale
	GR_eHEALS_4	I know where to find helpful health resources on the Internet.	5-point Likert scale
	GR_eHEALS_5	I know how to use the health information I find on the Internet to help me.	5-point Likert scale
	GR_eHEALS_6	I have the skills I need to evaluate the health resources I find on the Internet.	5-point Likert scale
	GR_eHEALS_7	I can tell high quality from low quality health resources on the Internet.	5-point Likert scale
	GR_eHEALS_8	I feel confident in using information from the Internet to make health decisions.	5-point Likert scale

3.2 Technology awareness

Figure 2 illustrates the reported awareness that participants have with the technologies being examined. DHA (Digitale Gesundheitsanwendungen, DiGAs) were the most recognized, with 68.9% awareness, followed closely by artificial intelligence at 66.7%, while blockchain technology as the least known, with only 24.4% of participants being aware of it. Regarding the specific utilization of these technologies in the healthcare sector or in the context of breast cancer treatment, the awareness rates were 51.1 and 26.7% for AI, and for blockchain technology, the rates were 17.8 and 13.3%, respectively.

3.3 Intention-to-use

The initial intention-to-use, as measured by a 7-point Likert scale, yielded a mean of 5.42 (SD; ±1.82) for artificial intelligence and 4.16 (±2.04) for blockchain technology. Figure 3 depicts the distribution frequencies corresponding to the respective scores and types of

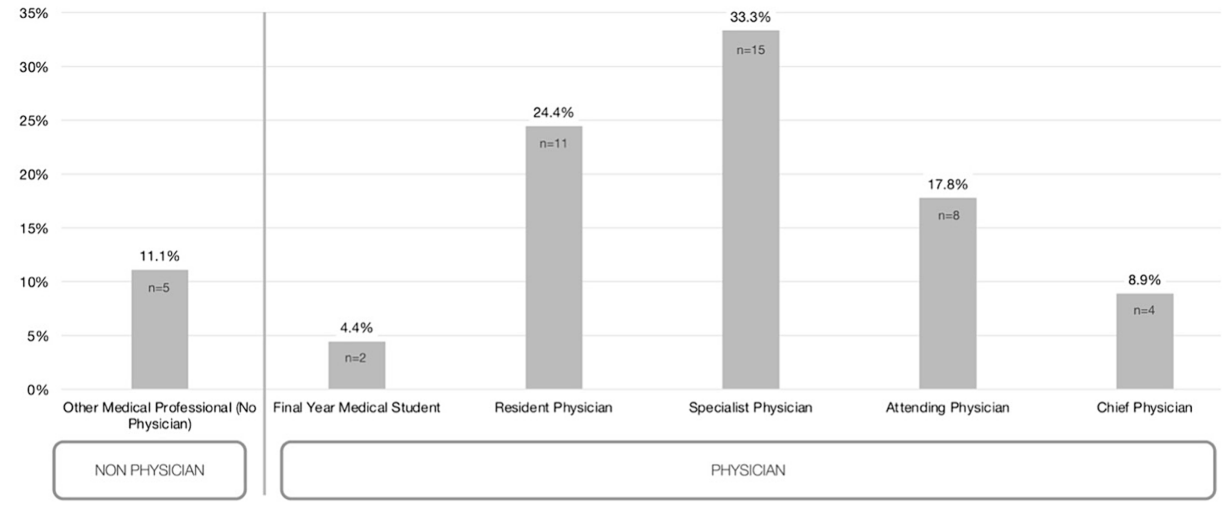


FIGURE 1
level of professional training of the study population.

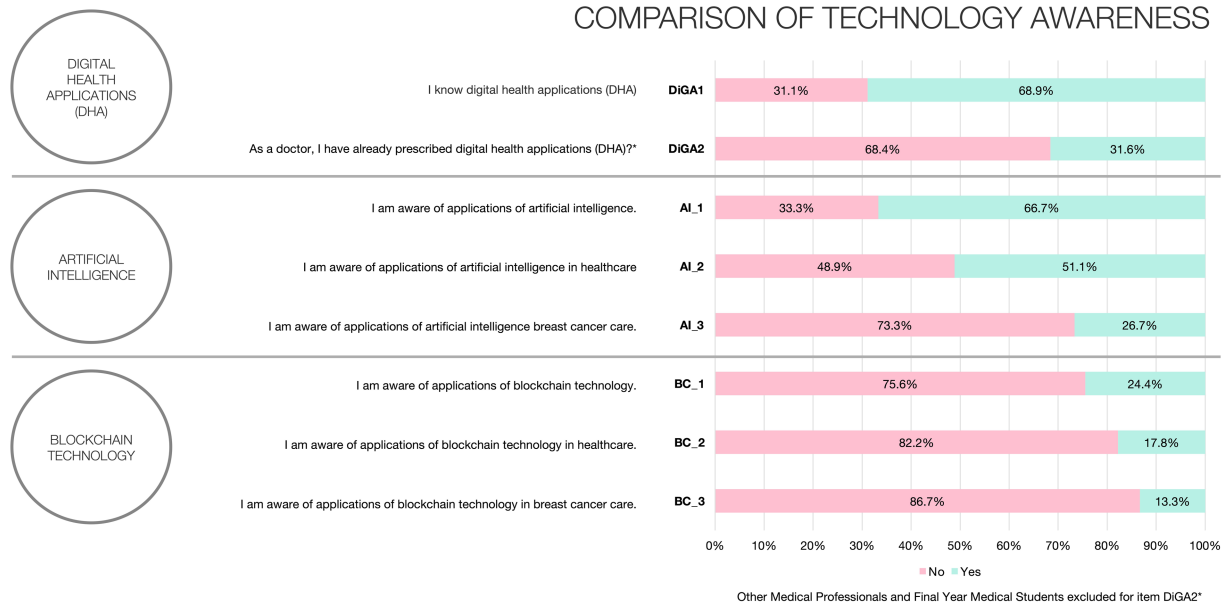


FIGURE 2
Technology awareness.

technology. Moreover, the analysis did not reveal any statistically significant variances in the intention-to-use when comparing across the binary classifications of gender and sector for either technology.

3.4 Correlation matrix

Spearman's rank correlation coefficient (ρ) was employed to evaluate the significance of associations among the variables: age, intention-to-use artificial intelligence (ITU_AI), intention-to-use blockchain technology (ITU_BC), and the aggregate scores of the German eHealth Literacy Scale (GR_eHEALS_Total). Consistent with the thresholds defined by Cohen et al., a robust positive correlation was identified between ITU_AI and ITU_BC ($\rho = 0.591$; $p < 0.01$) (33). Concomitantly, an intermediate-level negative correlation was discerned between participant age and ITU_AI ($\rho = -0.438$; $p < 0.01$), while an intermediate-level positive correlation was found between ITU_AI and GR_eHEALS_Total ($\rho = 0.383$; $p < 0.01$) (Table 2).

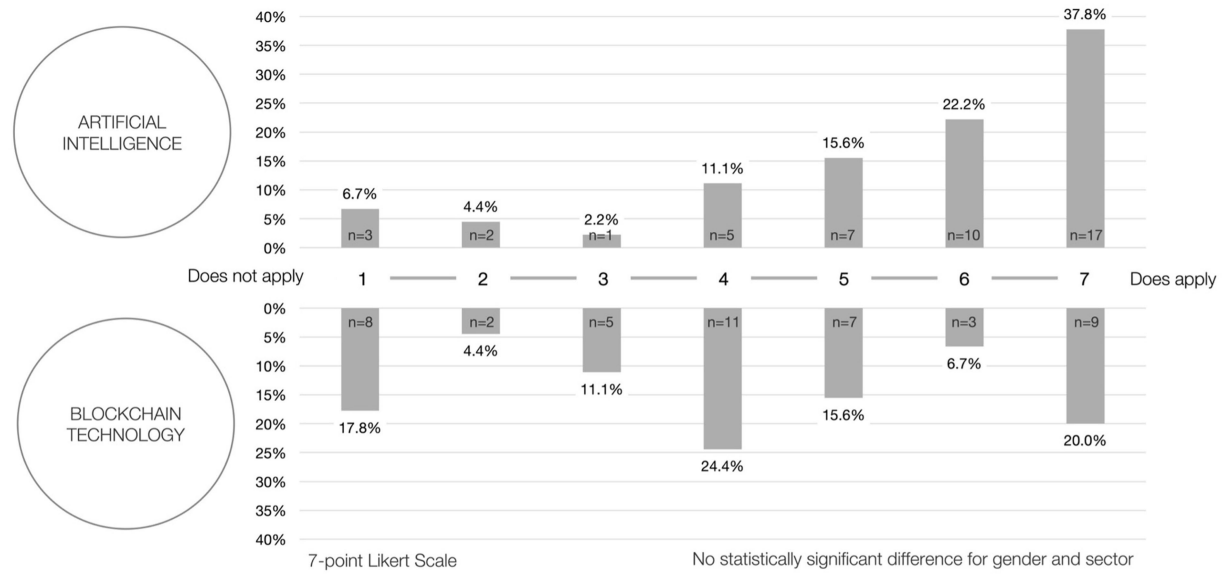


FIGURE 3
Intention-to-use in breast cancer care.

TABLE 2 Correlation matrix.

Correlation Matrix (Spearman's rho, ρ)				
	Age	ITU_AI	ITU_BC	GR_eHEALS_Total
Age	1.000	**−0.438****	−0.099	−0.125
ITU_AI	**−0.438****	1.000	**0.591****	**0.383****
ITU_BC	−0.099	**0.591****	1.000	0.206
GR_eHEALS_Total	−0.125	**0.383****	0.206	1.000

** Correlation is significant at the 0.01 level (2-tailed). Bold values represents significant correlation coefficients.

3.5 eHealth literacy scale

For the exploratory factor analysis, significant findings from Bartlett's test of sphericity (χ 2 = 304.000, $p < 0.001$) supported the factorability of the correlation matrix. The high value of Kaiser-Meyer-Olkin test (0.848) showed adequate sampling. Maximum likelihood method confirmed a single factor, which emerged from an initial eigenvalue of 5.47 and explained 68.37% of the total variance explained for. The factor loadings for this model were between 0.619 and 0.906. The data yielded a mean accumulated score for GR-eHEALS of 33.04 (SD ± 6.61; 95% CI {31.06, 35.03}) with a calculated Cronbach's alpha amounting to 0.928. Detailed results of the exploratory factor analysis and the corresponding scree plot are presented in Supplementary material S2, S3.

3.6 Change in intention-to-use after educational presentation

With regard to the post-education survey conducted, 30 respondents repeated the question on intention-to-use AI or BC on a 7-point Likert scale. In this sub-group, the mean intention-to-use AI was 5.37 (±1.81) before the session and 5.83 (±1.64) afterwards, which did not show a statistically significant change. The intended usage of BC did significantly increase from an initial average of 4.30 (±2.04) to 5.90 (±1.67), marking an improvement of +1.6 ($p < 0.01$) (Figure 4).

4 Discussion

Germany's healthcare system is characterized by a mandatory health insurance model, with the majority of citizens (89%) covered under statutory health insurance and a smaller fraction (11%) opting out for private health insurance (34). This system ensures extensive health coverage and contributes to the country's high healthcare spending of 12.9% of its Gross Domestic Product (GDP) in 2021, surpassing the European Union (EU) average by two percentage points (34).

The nation offers a substantial hospital sector, evident in its high number of hospital beds, ranking second in the EU behind Bulgaria and considerably above the EU average. Germany's physician and nurse density has increased at higher pace and surpasses European standards, reflecting a commitment to accessible medical care (34). According to the EU Statistics on Income and Living Conditions

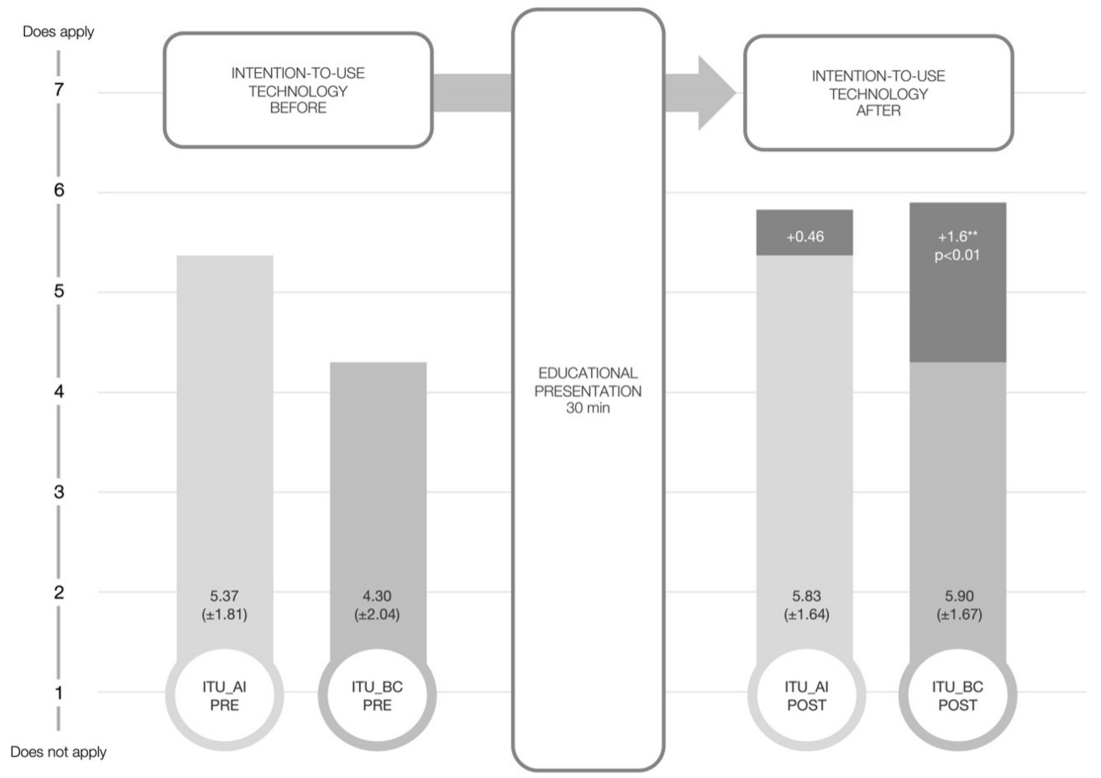


FIGURE 4
Change in intention-to-use before and after educational presentation.

(EU-SILC) survey, the proportion of reported unmet medical needs in Germany due to costs, waiting times or distances is among the lowest in the EU without differences between income groups (35). The country exceeds OECD averages in life expectancy, preventable mortality, and healthcare coverage, with a notable 85% of the population expressing satisfaction with healthcare quality (34).

Despite these strong indicators, the country reveals a rather low performance in the adoption and implementation of digital health solutions (36). Despite the assured reimbursement of telemedicine, Germany, together with its neighboring country France, has the lowest proportion of remote medical consultations in the EU (34). The OECD Digital Health at a Glance 2023 report explores the concept of digital health readiness by evaluating their member countries´ policy, analytical, technical and social environment in terms of twelve digital health readiness indicators, which should enable the successful use of digital health. Denmark distinguishes itself as a consistent leader, topping seven out of twelve indicators. Other countries like Finland, Korea, Sweden, Japan, the United States, and the Netherlands also showcase significant achievements. In comparison, Germany clearly lags behind, only performing well with respect to two OECD indicators, namely by offering a national strategy for digital health and high levels of digital security (36).

4.1 Principal findings

Public awareness and understanding of emerging information technologies vary and their application often stays unnoticed (15, 18, 37). Despite evidenced benefits for clinical care, a gap between the digital technologies´ potential and actual implementation persists (24). The dissemination of innovative technologies in healthcare necessitates an examination of how professional training can help to close this gap. A study conducted by the Commission for Digital Medicine of the German Society for Gynecology and Obstetrics underlines a significant incongruity between the perceived benefits of digital medicine and its practical application in gynecological care in Germany (14). Despite 78.4% of gynecologists recognizing the ability of digital medicine to reduce their increasing workload, only 13.5% acknowledge receiving institutional training. The research identifies critical obstacles to the adoption of digital medical technologies, primarily grounded in the scarcity of time and a deficiency in knowledge. This study illustrates that even a short training session of thirty minutes can significantly increase the intention-to-use less-known technologies like blockchain, elevating its intended use to the same level as that of more familiar technologies such as artificial intelligence.

4.2 eHealth literacy of gyne-oncological health professionals

eHealth literacy refers to the capacity of individuals to use emerging information and communications technologies for the enhancement or facilitation of health and healthcare services (38). Previous research has established the eHEALS as a reliable and valid

measure of eHealth literacy across different contexts (32). This study represents the first utilization of the German version of eHEALS (GR-eHEALS) to delineate the eHealth literacy of healthcare professionals in gynecological oncology, affirming the tool's applicability in this context (31). Additionally, the positive correlation between eHealth literacy scores and the intended use of AI reinforces the tool's validity in assessing competencies relevant to digital medicine. Our findings align with past research, which found a positive correlation of eHEALS with internet usage and reported computer knowledge (38, 39). The study confirms a sufficient degree of eHealth literacy among the investigated gyne-oncological care professionals (31, 32). This suggests that healthcare professionals in breast cancer care possess substantial digital skills to use emerging information technologies.

4.3 Differences in technology awareness

Despite demonstrating a satisfactory level of e Health literacy, adoption rates of emerging technologies often fall short of expectations (14, 24). In a survey conducted by the German Society of Gynecology and Obstetrics, gynecological professionals were asked about their actual use of DHA (Digitale Gesundheitsanwendung, DiGA) and a mere 10.2% of gynecologists had ever prescribed an app (14). Nearly half of the questioned professionals were not even aware that these applications were available for prescription, despite them being reimbursable in Germany since 2019.

This study suggests that the prescription rate of DHA among physicians has increased and provides further insights on less common emerging information technologies of AI and BC. As such their awareness in healthcare and breast cancer care fall behind the DHA. Despite the growing body of evidence underscoring their clinical applicability, lower levels of awareness among health professionals regarding their use in the management of breast cancer persist.

4.4 Differences in intention-to-use

These differences are not only observable in the context of technology awareness but are particularly pronounced in their delineated intended use for breast cancer care. AI is more widely recognized and has a higher intended use for breast cancer care in comparison to BC. Furthermore, the analysis reveals that younger individuals or those with higher levels of eHealth literacy show a greater inclination towards using AI – an observation that seems to be intuitively logic at first glance. Nevertheless, these observations are not consistently applicable to the less known technology of Blockchain. Professional education of 30 min did boost the intention-to-use both AI and BC, with the impact being particularly noteworthy and significant in the case of BC. As a result, after the educational presentation the intended adoption of BC equalizes with the level of intended adoption of AI in breast cancer care. The incorporation of AI in healthcare helps with diagnostic support, patient interaction and treatment customization. The effectiveness of diagnosis and treatment is usually the focus of scientific analysis. However, for a successful integration of the technology into clinical practice, psychological and social issues such as trust in AI, dependency risks and changes in the doctor-patient relationship must also be considered as suggested by Triberti et al. (40). The research by Strika et al. and Sebri et al. highlights the importance of understanding the attitudes of healthcare professionals and patients towards AI, an aspect that is often overlooked in research (27, 29). They argue that future studies should examine the broader impact of AI on the social and organizational aspects of healthcare, not just its effectiveness in diagnosis and treatment.

4.5 Limitations and research perspective

The research utilized a single-center approach, limiting the wider applicability and relevance of its findings beyond the context of the German healthcare system and its peculiarities in regulatory and socioeconomic dimensions. Unlike Germany, which has established the DiGA process to manage the prerequisites for DHAs´ reimbursability, other European nations do not offer standardized frameworks. As such, the equivalent term of DiGA for a DHA in Germany, needs to be interpreted in a different light compared to other European partner countries.

In a national context, the study's focus on a single region could be expanded to include multiple and diverse regions across Germany to enhance the national relevance of the findings. Nevertheless, this study, conducted during regional training on breast cancer care in Hesse, ensures a representative sample by interviewing a balanced mix of participants. The study's demographic balance across gender and professional training, coupled with the high centralization of care and the socioeconomically balanced, predominantly rural area of Marburg-Biedenkopf county, mirrors German standards well enough to provide a solid foundation for this kind of health services research. Previous epidemiological studies of the care network have proven to capture German standards to a sufficient degree (8, 41).

Nevertheless, to improve the external validity and broader applicability of the results, future research should adopt a national multi-network and international approach for comparison. This would enable the study's findings to be more transferable and applicable to various health systems and care settings. Furthermore, the present study explored the awareness and intention-to-use on conceptional applications of AI and BC, rather than a singular digital application or therapeutic intervention, i.e., an mHealth application. Future research on the adoption of digital health tools should incorporate a study design with follow-up survey to provide more insight into the long-term impact of the educational intervention on physicians and patients perceptions and understanding the factors that hinder or promote adherence to prevent discontinuation of their use.

5 Conclusion

The study emphasizes the role of professional education in bridging the gap between the evidenced clinical benefits and actual use of emerging health technologies in breast cancer care. It indicates that healthcare professionals in gynecological oncology are open to digital tools and show a sufficient degree of eHealth literacy. The adoption of emerging technologies nevertheless lags behind, often due to a lack of awareness, time and training. The study demonstrates that even brief educational interventions show the potential to increase the intention-to-use emerging, less-known technologies, suggesting that focused educational programs could significantly enhance the integration of emerging technologies into clinical practice.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This is an observational study with primary data anonymization. The Philipps-University Marburg Research Ethics Committee has confirmed that no ethical approval is required (23-279 ANZ).

Author contributions

SG: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JK: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. NG: Conceptualization, Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. MW: Supervision, Writing – review & editing. UW: Resources, Supervision, Writing – review & editing. ML: Project administration, Resources, Supervision, Writing – review & editing. SK: Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. SG and NG were supported by the Clinician Scientist program (SUCCESS-program) of Philipps-University of Marburg and the University Hospital of Giessen and Marburg. Open access funding was provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

Acknowledgments

The authors would like to thank Julia Greenfield for valuable comments and insights about the statistical analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1380940/full#supplementary-material

References

1. Robert Koch Institute. Brustdrüse – C 50 In: Robert Koch Institut, editor. *Krebs in Deutschland für 2019/2020. 14th* ed. Berlin: Robert Koch Institute and Gesellschaft der epidemiologischen Krebsregister in Deutschland e.V. (2023). 78–81.

2. European Commission. Europe's beating Cancer plan. (2021). Available at: https://health.ec.europa.eu/system/files/2022-02/eu_cancer-plan_en_0.pdf (Accessed February 2, 2024).

3. German Guideline Program in Oncology (German Cancer Society, German Cancer Ais, AWMF. Interdisciplinary evidence-based pratice guideline for early detecion, diagnosis, treatment and follow-up of breast cancer long version 4.4 AWMF registration number: 032/045OL. (2021). Available at: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/S3_Guideline_Breast_Cancer.pdf (Accessed February 2, 2024)

4. Tarawneh TS, Rodepeter FR, Teply-Szymanski J, Ross P, Koch V, Thölken C, et al. Combined focused next-generation sequencing assays to guide precision oncology in solid tumors: a retrospective analysis from an institutional molecular tumor board. *Cancers (Basel)*. (2022) 14:4430. doi: 10.3390/cancers14184430

5. Santa-Maria CA, Wolff AC. Antibody-drug conjugates in breast cancer: searching for magic bullets. *J Clin Oncol*. (2023) 41:732–5. doi: 10.1200/JCO.22.02217

6. Bhattacharya T, Brettin T, Doroshow JH, Evrard YA, Greenspan EJ, Gryshuk AL, et al. AI meets exascale computing: advancing cancer research with large-scale high performance computing. *Front Oncol*. (2019) 9:984. doi: 10.3389/fonc.2019.00984

7. Barker AD, Lee JSH. Translating "big data" in oncology for clinical benefit: progress or paralysis. *Cancer Res*. (2022) 82:2072–5. doi: 10.1158/0008-5472.CAN-22-0100

8. Griewing S, Gremke N, Lingenfelder M, Wagner U, Keil C. Resilience of gynecological and obstetric inpatient Care in Central Germany in times of repetitive socioeconomic shocks—an epidemiological study assessing standardized health services indicators and economic status according to the aG-DRG catalog. *Healthcare (Switzerland)*. (2023) 11:1683. doi: 10.3390/healthcare11121683

9. Dubovitskaya A, Novotny P, Xu Z, Wang F. Applications of Blockchain Technology for Data-Sharing in oncology: results from a systematic literature review. *Oncology (Switzerland)*. (2020) 98:403–11. doi: 10.1159/000504325

10. Griewing S, Lingenfelder M, Wagner U, Gremke N. Use case evaluation and digital workflow of breast Cancer care by artificial intelligence and Blockchain technology application. *Healthcare (Switzerland)*. (2022) 10:2100. doi: 10.3390/healthcare10102100

11. Poon H. Multimodal generative AI for precision health. *NEJM AI*. (2023) 1:1–3. doi: 10.1056/AI-S2300233

12. Goldberg C. Patient portal. *NEJM AI*. (2023) 1:AIp2300189. doi: 10.1056/AIp2300189

13. pwc. Sherlock in health: how artificial intelligence may improve quality and efficiency, whilst reducing healthcare costs in Europe. (2017). Available at: https://www.pwc.com.tr/en/sektorler/saglik/yayinlar/sherlock-saglik-sektorunu-arastiriyor.html (Accessed February 2, 2024).

14. Pfob A, Hillen C, Seitz K, Griewing S, Becker S, Bayer C, et al. Status quo and future directions of digitalization in gynecology and obstetrics in Germany: a survey of the commission digital medicine of the German Society for Gynecology and Obstetrics. *Arch Gynecol Obstet*. (2023) 309:195–204. doi: 10.1007/s00404-023-07222-2

15. UK Office for National Statictics (ONS). Public awareness, opinions and expectations about artificial intelligence. (2023). Available at: https://www.ons.gov.uk/businessindustryandtrade/itandinternetindustry/articles/publicawarenessopinionsandexpectationsaboutartificialintelligence/julytooctober2023#cite-this-article (Accessed February 2, 2024)

16. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med*. (2023) 388:1201–8. doi: 10.1056/NEJMra2302038

17. Yan S, Li J, Wu W. Artificial intelligence in breast cancer: application and future perspectives. *J Cancer Res Clin Oncol*. (2023) 149:16179–90. doi: 10.1007/s00432-023-05337-2

18. Statista. Artificial intelligence perspective of German society. (2023). Available at: https://de.statista.com/statistik/studie/id/112085/dokument/kuenstliche-intelligenz-in-der-deutschen-gesellschaft/ (Accessed February 2, 2024)

19. Sunny FA, Hajek P, Munk M, Abedin MZ, Satu MS, Efat MIA, et al. A systematic review of Blockchain applications. *IEEE Access*. (2022) 10:59155–77. doi: 10.1109/ACCESS.2022.3179690

20. Dubovitskaya A, Novotny P, Thiebes S, Sunyaev A, Schumacher M, Xu Z, et al. Intelligent health care data management using Blockchain: current limitation and future research agenda. *Lect Notes Comput Sci*. (2019) 11721 LNCS:277–88. doi: 10.1007/978-3-030-33752-0_20

21. Raab R, Küderle A, Zakreuskaya A, Stern AD, Klucken J, Kaissis G, et al. Federated electronic health records for the European health data space. *Lancet Digit Health*. (2023) 5:e840–7. doi: 10.1016/S2589-7500(23)00156-5

22. Yaqoob I, Salah K, Jayaraman R, al-Hammadi Y. Blockchain for healthcare data management: opportunities, challenges, and future recommendations. *Neural Comput & Applic*. (2022) 34:11475–90. doi: 10.1007/s00521-020-05519-w

23. Hirano T, Motohashi T, Okumura K, Takajo K, Kuroki T, Ichikawa D, et al. Data validation and verification using blockchain in a clinical trial for breast cancer: regulatory sandbox. *J Med Internet Res*. (2020) 22:e18938. doi: 10.2196/18938

24. Rainer Thiel A, Deimel L, Schmidtmann D, Piesche K., Hüsing T., Rennoch J, et al (2018) Gesundheitssystem-Vergleich Fokus Digitalisierung #SmartHealthSystems Digitalisierungsstrategien im internationalen Vergleich. Available at: https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Der_digitale_Patient/VV_SHS-Gesamtstudie_dt.pdf (Accessed February 2, 2024).

25. Pfob A, Griewing S, Seitz K, Hillen C, Becker S, Bayer C, et al. Current landscape of hospital information systems in gynecology and obstetrics in Germany: a survey of the commission digital medicine of the German Society for Gynecology and Obstetrics. *Arch Gynecol Obstet*. (2023) 308:1823–30. doi: 10.1007/s00404-023-07223-1

26. Papastergiou D, Kokaridas D, Bonotis K, Digelidis N, Patsiaouras A. Intervention effect of supportive group therapy and physical exercise on the quality of life of cancer patients. *Cent Eur J Sport Sci Med*. (2019) 25:5–13. doi: 10.18276/cej.2019.1-01

27. Sebri V, Durosini I, Mazzoni D, Pravettoni G. Breast cancer survivors' motivation to participate in a tailored physical and psychological intervention: a qualitative thematic analysis. *Behav Sci*. (2022) 12:271. doi: 10.3390/bs12080271

28. Reynolds LM, Cavadino A, Chin S, Little Z, Akroyd A, Tennant G, et al. The benefits and acceptability of virtual reality interventions for women with metastatic breast cancer in their homes; a pilot randomised trial. *BMC Cancer*. (2022) 22:360. doi: 10.1186/s12885-021-09081-z

29. Strika M, Durosini I, Sebri V, Pravettoni G. Healthcare Professionals' attitudes towards virtual reality for Cancer patients. *Annu Rev Cyberther Telemed*. (2023) 21:180.

30. Griewing S, Gremke N, Wagner U, Lingenfelder M, Kuhn S, Boekhoff J. Challenging ChatGPT 3.5 in Senology—an assessment of concordance with breast Cancer tumor board decision making. *J Pers Med*. (2023) 13:1502. doi: 10.3390/jpm13101502

31. Marsall M, Engelmann G, Skoda EM, Teufel M, Bäuerle A. Measuring electronic health literacy: development, validation, and test of measurement invariance of a revised German version of the eHealth literacy scale. *J Med Internet Res*. (2022) 24:e28252. doi: 10.2196/28252

32. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. *J Med Internet Res*. (2006) 8:e27. doi: 10.2196/jmir.8.4.e27

33. Cohen J. *Statistical power analysis for the behavioral sciences second edition*. New York: Routledge (1988).

34. OECD/European Observatory on Health Systems and Policies. *Germany: country health profile 2023, State of Health in the EU*. Paris: OECD Publishing (2023).

35. Eurostat. Unmet need for medical examinations in 2022. (2023). Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20231211-1 (Accessed March 24, 2024).

36. OECD. *Health at a glance 2023: OECD indicators*. Paris: OECD Publishing (2023).

37. Krusche M, Klemm P, Grahammer M, Mucke J, Vossen D, Kleyer A, et al. Acceptance, usage, and barriers of electronic patient-reported outcomes among German rheumatologists: survey study. *JMIR Mhealth Uhealth*. (2020) 8:e18117. doi: 10.2196/18117

38. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. *J Med Internet Res*. (2012) 14:e19. doi: 10.2196/jmir.1619

39. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. *J Med Internet Res*. (2006) 8:e9. doi: 10.2196/jmir.8.2.e9

40. Triberti S, Durosini I, Lin J, La Torre D, Ruiz GM. Editorial: on the "human" in human-artificial intelligence interaction. *Front Psychol*. (2021) 12:808995. doi: 10.3389/fpsyg.2021.808995

41. Griewing S, Wagner U, Lingenfelder M, Heinis S, Schieffer B, Markus B. Impact of the COVID-19 pandemic on delivery of gynecology and obstetrics Services at a Maximum Care University Hospital in Germany. *Geburtshilfe Frauenheilkd*. (2022) 82:427–40. doi: 10.1055/a-1687-9674

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 05 June 2024
DOI 10.3389/fmed.2024.1370836

OPEN ACCESS

EDITED BY
Fedde Scheele,
VU Amsterdam, Netherlands

REVIEWED BY
Martina Arcieri,
Ospedale Santa Maria della Misericordia di Udine, Italy
Andrea Giannini,
Umberto 1 Hospital, Italy
Paul Van Kesteren,
Onze Lieve Vrouwe Gasthuis (OLVG), Netherlands

*CORRESPONDENCE
Thomas Ind
✉ Thomas.Ind@rmh.nhs.uk

RECEIVED 15 January 2024
ACCEPTED 06 May 2024
PUBLISHED 05 June 2024

CITATION
Sickinghe A, Nobbenhuis M, Nelissen E, Heath O and Ind T (2024) Proficiency-based progression training in robot-assisted laparoscopy for endometrial cancer: peri-operative and survival outcomes from an observational cohort study.
Front. Med. 11:1370836.
doi: 10.3389/fmed.2024.1370836

Proficiency-based progression training in robot-assisted laparoscopy for endometrial cancer: peri-operative and survival outcomes from an observational cohort study

Ariane Sickinghe[1,2], Marielle Nobbenhuis[1], Ellen Nelissen[3], Owen Heath[1] and Thomas Ind[1]*

[1]Department of Gynecological Oncology, Royal Marsden Hospital, London, United Kingdom, [2]Faculty of Medicine, University Medical Centre Utrecht, Utrecht University, Utrecht, Netherlands, [3]Department of Gynecological Oncology, Royal United Hospitals, Bath, United Kingdom

Introduction: Over the last decade there has been a transition from traditional laparoscopy to robotic surgery for the treatment of endometrial cancer. A number of gynecological oncology surgical fellowship programmes have adopted robot-assisted laparoscopy, but the effect of training on complications and survival has not been evaluated. Our aim was to assess the impact of a proficiency-based progression training curriculum in robot-assisted laparoscopy on peri-operative and survival outcomes for endometrial cancer.

Methods: This is an observational cohort study performed in a tertiary referral and subspecialty training center. Women with primary endometrial cancer treated with robot-assisted laparoscopic surgery between 2015 and 2022 were included. Surgery would normally include a hysterectomy and salpingo-oophorectomy with some form of pelvic lymph node dissection (sentinel lymph nodes or lymphadenectomy). Training was provided according to a training curriculum which involves step-wise progression of the trainee based on proficiency to perform a certain surgical technique. Training cases were identified pre-operatively by consultant surgeons based on clinical factors. Case complexity matched the experience of the trainee. Main outcome measures were intra- and post-operative complications, blood transfusions, readmissions $<$ 30 days, return to theater rates and 5-year disease-free and disease-specific survival for training versus non-training cases. Mann–Witney U, Pearson's chi-squared, multivariable regression, Kaplan-Meier and Cox proportional hazard analyses were performed to assess the effect of proficiency-based progression training on peri-operative and survival outcomes.

Results: Training cases had a lower BMI than non-training cases (30 versus 32 kg/m^2, p = 0.013), but were comparable in age, performance status and comorbidities. Training had no influence on intra- and post-operative complications, blood transfusions, readmissions $<$ 30 days, return to theater rates and median 5-year disease-free and disease-specific survival. Operating time was longer in training cases (161 versus 137 min, p = $<$ 0.001). The range

of estimated blood loss was smaller in training cases. Conversion rates, critical care unit-admissions and lymphoedema rates were comparable.

Discussion: Proficiency-based progression training can be used safely to teach robot-assisted laparoscopic surgery for women with endometrial cancer. Prospective trails are needed to further investigate the influence of distinct parts of robot-assisted laparoscopic surgery performed by a trainee on endometrial cancer outcomes.

KEYWORDS

endometrial cancer, uterus cancer, robot-assisted laparoscopy, minimally-invasive surgery, training, survival, complications

1 Introduction

The use of minimally-invasive surgery for endometrial cancer has become widespread since the LACE and GOG LAP2 trials established non-inferiority of laparoscopic versus laparotomic surgery for disease-free and overall survival in endometrial cancer (1–3). Robot-assisted laparoscopic (RAL) surgery was introduced in gynecological surgery in 2005 (4) and provides more precision, better views, reduced patient morbidity and improved surgeon ergonomics compared to conventional laparoscopy enabling the surgeon to perform more complex surgery (5–10). These advantages are especially beneficial in obese patients undergoing open or laparoscopic hysterectomy as they are more prone to post-operative morbidity compared to non-obese patients (11–13). Obesity is the main risk factor for endometrial cancer and since its incidence is rising (14), the preferred approach in minimally-invasive surgery has shifted from straight-stick to robotic (5, 15, 16).

An increasing number of gynecological oncology surgical fellowship programs are embedding RAL surgery. The introduction of a new surgical technique is accompanied with a learning curve, which also has been assessed in conventional laparoscopic and robotic surgery (17–20). This underpins the need for a training curriculum.

Urologists were the first to develop a proficiency-based progression (PBP) training curriculum for robotic surgery (21). The Society of European Robotic Gynecological Surgery and British and Irish Association of Robotic Gynecological Surgeons followed by providing training a curriculum for robot-assisted gynecological surgery (22, 23). Previous studies have evaluated the effect of a single-surgeon learning curve on peri-operative outcomes (11, 13, 14). However, the general effect of PBP training on peri-operative and survival outcomes in endometrial cancer patients undergoing RAL surgery has not been evaluated.

The Royal Marsden Hospital is a tertiary cancer center in the United Kingdom treating high-risk endometrial cancer patients. It was the first center in the United Kingdom to adopt RAL surgery for gynecological cancer in 2007 and have used PBP training in RAL surgery for trainees subspecialising in gynecological oncology since 2015.

The objective of this study was to assess the impact of PBP training on peri-operative and survival outcomes in endometrial cancer patients undergoing RAL surgery.

2 Materials and methods

This project received institutional review board approval from the Royal Marsden Committee on Clinical Research on 17-11-2022. Project number SE1234.

2.1 Design

An observational cohort study was performed between 2015 and 2022. All patients intended to undergo RAL surgery for endometrial cancer as part of routine care at the Royal Marsden Hospital were included. This included a small group of patients who were operated in any other hospital of the Southwest Thames Gynecological Cancer Centre our surgical team operated in due to capacity constraints. The Southwest Thames Gynecological Cancer Centre is a consortium of hospitals that closely work together and share facilities. It includes the Royal Marsden Hospital, St George's Hospital, The London Clinic, Lister Hospital and the Princess Grace Hospital. All surgeries were performed by three robot-trained gynecological oncology surgeons on three generations of Da Vinci robots (S, Si, Xi).

The inclusion criteria consisted of women diagnosed with primary endometrial cancer and the intention of undergoing RAL hysterectomy, bilateral salpingo-oophorectomy and/or any lymph node dissection. All subtypes of endometrial cancer were included. Patients with any additional cancer, e.g., simultaneous ovarian, sigmoid or other type of cancer, were excluded from analysis. Patients who underwent RAL hysterectomy, bilateral salpingo-oophorectomy and/or any lymph node dissection for a non-endometrial type of cancer, e.g., cervical cancer, were excluded. Also, conforming to European Society of Gynecological Oncology (ESGO) guidelines (24) patients with advanced disease where cytoreduction was considered infeasible as judged by a multi-disciplinary team were excluded.

2.2 Data collection

Data was collected prospectively by two surgeons (TI and MN) from 2015 to 2022 and was stored in an encrypted and

secure database. Missing data was completed retrospectively by independent researcher (AS) in 2022 using information on the hospital's electronic patient record.

2.3 Identification of training cases

Trainees subspecialising in gynecological oncology were consultant surgeons subspecialising in gynecological oncologic surgery or gynecological registrars following the training program at the end of general gynecology training. Trainees followed a PBP training curriculum during 1–2 years provided by The Society of European Robotic Gynecological Surgery and British and Irish Association of Robotic Gynecological Surgeons (22, 23). Recommendations for this training curriculum were formulated by experienced gynecological robotic surgeons who performed The Delphi process (25, 26).

In PBP training, trainees follow a structured and standardized training with pre-set learning goals. Modules of training lead from e-learning, to virtual training, to model training, to procedural training. All clinical procedures are performed under the guidance of expert tutors and trainees can only progress to the next step of training if they are proficient in the previous steps as judged by supervising consultant surgeons in concordance with the SERGS training curriculum (22). Trainees start procedural training with vault suturing and end with performing a hysterectomy and lymph node dissection independently.

A case was marked as a "training case" if the trainee performed a part of the surgery on the console. A case was not marked as a "training case" if the trainee only performed first or second assistant tasks like robot docking, skin suturing or bringing in the uterus manipulator. PBP training cases were identified pre-operatively by consultant surgeons based on clinical factors, such as BMI and comorbidities. Patients with a BMI $>$ 50 or many comorbidities were not selected as training cases. Training case complexity was subjectively matched by the supervising surgeon to the trainee's proficiency and progression through the training program in concordance with the SERGS training curriculum (22).

All operations were performed at the Southwest Thames Gynecological Cancer Centre under direct supervision of one of two consultant surgeons. Consultant surgeons had extensive experience in robotic surgery (over 300 robotic cases performed per surgeon) and were trained by Lapco to provide training in minimally invasive surgery in a similar and certified manner (27).

2.4 Outcomes

Primary outcomes included intra- and post-operative complications before and after 30 days, blood transfusions, readmissions $<$ 30 days, return to theater, and 5-year disease-free and disease-specific survival. Intra-operative complications were defined as any type of surgical complication occurring during the operation. Post-operative complications within 30 days were graded according to the Clavien-Dindo classification (28).

Secondary outcomes included estimated blood loss, operating duration, rate of conversions, critical care unit (CCU)-admissions, length of stay (LOS) longer than one day and lymphoedema. A conversion was defined as the need to convert to laparotomy after docking of the robot due to an intra-operative complication or impossibility to complete robotically.

Prognostic risk groups according to ESGO/The European Society for Radiotherapy and Oncology (ESTRO)/The European Society of Pathology (ESP) were assessed ("ESGO risk groups" in short) (24). These recently developed guidelines for risk group determination incorporate clinicopathologic and molecular parameters and effectively predict survival in endometrial cancer (29).

2.5 Statistical analysis

Analyses were performed with the Statistical Package for the Social Sciences (SPSS) 28.01.1. Missing data analysis revealed missing data $>$ 5% for American Society of Anesthesiologists physical (ASA) status classification and World Health Organization (WHO) performance status. Imputation of missing data was done in SPSS using the median of nearby points for the variables ASA status classification and WHO performance status (30).

Mann–Witney U testing was used to assess differences in median values. Pearson's chi-squared testing was performed to assess the correlation between categorical dependent variables and the independent variable (training case yes/no). Multivariable logistic regression analysis was performed to assess the correlation between continuous clinical variables and the independent variable. Multivariable logistic regression analysis was performed to assess the correlation between training and peri-operative outcomes. Cox-regression analysis was performed for 5-year disease-free and disease-specific survival.

The effect of PBP training on post-operative and survival outcomes are expressed as odds and hazard ratios. Effect sizes were corrected for age, stage ($<$ 2/$\geq$ 2) and grade (low/high) analysis because these variables render clinical relevance for disease-free and disease-specific survival.

Statistical tests were two-sided with significance set at $p < 0.05$, with confidence intervals (CI) at the 95% level. Post-hoc testing according to Bonferroni was performed if Pearson's chi-squared testing rendered group differences (31). Bonferroni-corrected p-values are marked with an "*".

3 Results

3.1 Patient characteristics

In total 594 endometrial cancer cases were analyzed: 294 (49.4%) training cases and 300 (50.6%) non-training cases. Thirteen cases (1.9%) were excluded due to non-endometrial primary histology or any additional cancer. Eighteen gynecological oncology trainees were trained in a PBP manner with a mean number of 16 cases performed per trainee (range: 4–58).

10.3389/fmed.2024.1370836

TABLE 1 Patient characteristics.

	Training case (n = 294)	Non-training case (n = 300)	p-value
Age (years), median (range)	66 (31–91)	67 (34–93)	0.980
BMI (kg/m2), median (range)	30 (16–57)	32 (17–69)	0.013
ASA score, median (range)	2 (0–3)	2 (1–3)	0.655
ASA score, n (%)			
0	1 (0.4)	0	
1	29 (10.5)	28 (9.8)	
2	152 (54.9)	156 (54.5)	
3	95 (34.3)	102 (35.7)	
WHO performance status, median (range)	1 (0–3)	1 (0–3)	0.589
WHO performance status, n (%)			
0	77 (28.4)	100 (35.2)	
1	154 (56.8)	128 (45.4)	
2	30 (11.1)	32 (11.3)	
3	10 (3.7)	22 (7.7)	
Charlson Comorbidity Index 10-year survival estimate, median (range)	21.4% (0–90.1%)	21.4% (0–98.3%)	0.259
FIGO stage, median (range)	1 (1–4)	1 (1–4)	0.224
FIGO stage, n (%)			
1	190 (64.6)	219 (73.0)	
2	25 (8.5)	20 (6.7)	
3	68 (23.1)	52 (17.3)	
4	11 (3.7)	9 (3.0)	
Histology, n (%)			
Endometrioid	167 (56.8)	208 (69.3)	0.002*
Serous	69 (23.5)	54 (18.0)	0.101*
Clear cell carcinoma	14 (4.8)	9 (3.0)	0.267*
Carcinosarcoma	26 (8.8)	19 (6.3)	0.246*
Other	18 (6.1)	10 (3.3)	0.110*
Grade, median (range)	3 (1–3)	2 (1–3)	0.004
Grade, n (%)			
1	100 (34.0)	136 (45.3)	0.004*
2	46 (15.6)	45 (15.0)	0.826*
3	148 (50.3)	119 (39.7)	0.009*
ESGO Risk, *n* (%)			0.069
Low	70 (24.3)	95 (31.9)	
Intermediate	51 (17.7)	60 (20.1)	
High-Intermediate	29 (10.1)	34 (11.4)	
High	131 (45.5)	101 (33.9)	
Advanced	7 (2.4)	8 (2.7)	
Type of LN dissection, n (%)			
Sentinel LN dissection	227 (77.2)	203 (67.7)	0.009
Pelvic lymphadenectomy	74 (25.2)	88 (29.3)	0.255
Para-aortic lymphadenectomy	1 (0.3)	4 (1.3)	0.185
LN harvested, median (range)	3 (0–37)	3 (0–39)	0.337
(Neo-)adjuvant treatment, n (%)			
Neo-adjuvant treatment	15 (5.1)	11 (3.7)	0.253
Adjuvant treatment	197 (69.9)	170 (59.4)	0.031
Follow-up duration (months), median (range)	25 (0–60)	28 (0–60)	0.148

ASA, American Society of Anesthesiologists; BMI, body mass index; ESGO, the European Society of Gynecological Oncology; FIGO, International Federation of Gynecology and Obstetrics; LN, lymph node; WHO, World Health Organization. *Post-hoc analyses were performed for WHO performance status, histology and grade and p-values were adjusted following the Bonferroni method (25).

Table 1 shows the baseline characteristics for training and non-training cases. Groups were similar in age (66 versus 67 years, $p = 0.095$), median ASA physical status score (2 versus 2, $p = 0.655$), median WHO performance status (1 versus 1, $p = 0.589$) and Carlson Comorbidity Index 10-year median survival estimates (21.4 versus 21.4%, $p = 0.259$). Training cases had a lower BMI than non-training cases (30 versus 32 kg/m^2, $p = 0.013$).

No differences were found in median Fédération Internationale de Gynécologie et d'Obstétrique (FIGO) stage between groups (1 versus 1, $p = 0.224$), but training cases had patients with a higher median histopathological grade (3 versus 2, $p = 0.004$). Post-hoc testing showed a lower rate of grade 1 (34.0 versus 45.3%, $p = 0.004$*) and a higher rate of grade 3 tumors (50.3 versus 39.7%, $p = 0.009$*) in training cases. Training cases had a lower rate of endometrioid tumors (56.8 versus 69.3%, $p = 0.002$*) and a higher percentage of adjuvant treatment in training cases (69.9 versus 59.4%, $p = 0.031$). The distribution of European Society of Gynecological Oncology (ESGO) risk scores (29) did not differ between groups ($p = 0.069$). More sentinel lymph node dissections (77.2 versus 67.7%, $p = 0.009$) were performed in training cases. The median number of harvested lymph nodes (3 versus 3, $p = 0.337$), rates of pelvic (25.2 versus 29.3%, $p = 0.255$) and para-aortic lymphadenectomies (0.3 versus 1.3%, $p = 0.185$) were comparable. Median follow-up was comparable between groups (25 versus 28 months, $p = 0.148$).

TABLE 2 Peri-operative and survival outcomes in training and non-training cases.

	Training case (n = 294)	Non-training case (n = 300)	p-value
EBL (ml), median (range)	100 (0–1,200)	100 (0–2,700)	0.005
Conversion, n (%)	7 (2.4)	6 (2.0)	0.749
Blood transfusion, n (%)	6 (2.0)	14 (4.7)	0.076
Operating time (min), mean (range)	160 (25–308)	137 (20–385)	< 0.001
CCU-admission, n (%)	7 (2.5)	10 (3.6)	0.991
Length of stay (days), median (range)	1 (0–77)	2 (1–30)	0.007
Intra-operative complication, n (%)	14 (4.8)	23 (7.7)	0.143
Post-operative complication, n (%)	84 (30.4)	86 (30.4)	0.991
Clavien-Dindo grade, n (%)			0.665
1	30 (10.9)	21 (7.4)	
2	48 (17.4)	59 (20.8)	
3a	2 (0.7)	1 (0.4)	
3b	2 (0.7)	3 (1.1)	
4a	2 (0.7)	2 (0.7)	
Readmission < 30 days, n (%)	11 (3.8)	18 (6.0)	0.205
Return to theater, n (%)	3 (1.0)	5 (1.7)	0.492
Lymphoedema, n (%)	25 (8.4)	43 (11.3)	0.192
Disease-specific survival (months), median (range)	21 (0–60)	26 (0–60)	0.004
Disease-free survival (months), median (range)	25 (0–60)	28 (0–60)	0.182
Disease-specific death, n (%)	38 (12.8)	36 (9.1)	0.178
Recurrence, n (%)	66 (24.2)	62 (25.1)	0.825

CCU, critical care unit; EBL, estimated blood loss. Data are presented as n (% of training cases/non-training cases, rounded to one decimal).

3.2 Primary outcomes

Primary outcomes are displayed in Tables 2, 3 and Figures 1, 2.

3.2.1 Peri-operative outcomes

Intra- and post-operative complications were comparable across groups (4.8 versus 7.7%, p = 0.146; 30.4 versus 30.4%, p = 0.991, respectively). There was no difference in the distribution of Clavien-Dindo complication grades between groups (p = 0.665).

TABLE 3 Multivariable logistic regression on the effect of training on post-operative outcomes and cox proportional hazard ratios for disease-free and disease-specific survival.

	Odds ratio (95%-CI)	p-value
Intra-operative complication	0.7 (0.4–1.4)	0.364
Clavien-Dindo complication	0.9 (0.6–1.2)	0.436
Any post-operative complication	0.9 (0.6–1.2)	0.385
Readmission < 30 days	0.7 (0.3–1.5)	0.332
Return to theater	0.7 (0.2–2.8)	0.579
LOS > 1 day	0.6 (0.5–0.9)	0.003
Lymphoedema	0.6 (0.4–1.1)	0.101
	Hazard ratio (95%-CI)	**p-value**
Disease-free survival	0.9 (0.6–1.4)	0.892
Disease-specific survival	1.5 (0.8–2.6)	0.172

CI, confidence interval; LOS, length of stay. Odds and hazard ratios are corrected for stage, grade and age. Stage was divided in stage 1 or ≥ 2. Grade was divided in low (grade 1 and 2) and high grade (grade 3). Age was categorized per five years.

Readmissions < 30 days (3.8 versus 6.0%, p = 0.205), return to theater (1.0 versus 1.7%, p = 0.492) and blood transfusions (2.0 versus 4.7%, p = 0.076) did not differ (see Table 2).

The effect of PBP training on intra- and post-operative outcomes are expressed as odds and hazard ratios and shown in Table 3. Training did not increase odds ratios for intra-operative complications (0.6, p = 0.154), post-operative complications (1, p = 0.994), Clavien-Dindo complications, (0.9, p = 0.583), readmissions < 30 days (0.6, p = 0.202) and return to theater (0.6, p = 0.553).

3.2.2 Survival outcomes

Table 2 displays the rates of recurrences (24.2 versus 25.1%, p = 0.825) and deaths of disease (12.8 versus 9.1%, p = 0.178), which did not differ between groups. Median disease-free survival differed significantly between training and non-training cases (21 versus 26 months, p = 0.004). Median disease-specific survival (25 versus 28 months, p = 0.182) was comparable between groups. After correction for the confounders age, stage and grade (see methods) the hazard ratio for disease-free survival for training cases compared to non-training cases is 0.9 (95%-CI: 0.6–1.4, p = 0.892). For disease-specific survival the hazard ratio is 1.5 (95%-CI: 0.8–2.6, p = 0.172) after correction (see Table 3).

Figures 1, 2 show the Kaplan-Meier curves for 5-year disease-free and disease-specific survival. The estimated 5-year disease-free survival is 66.6% (95%-CI: 59.1–73.0%) for non-training cases and 68.5% (95%-CI: 61.3–74.5%) for training cases. The estimated 5-year disease-specific survival is 86.1% (95%-CI: 80.0–90.6%) for non-training and 79.1% (95%-CI: 72.6–84.2%) for training cases.

3.3 Secondary outcomes

Secondary outcomes are displayed in Tables 2, 3. A difference was found in the range of estimated blood loss (0–2,700 ml versus

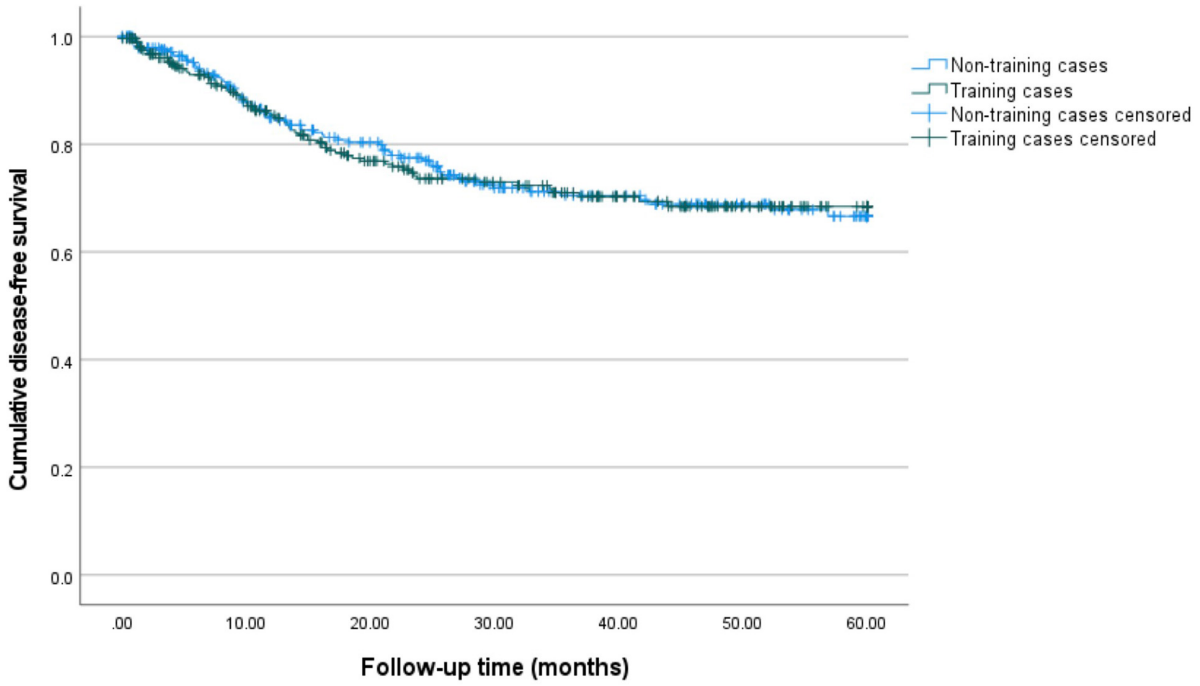


FIGURE 1
5-year disease-free survival for training and non-training cases. Kaplan–Meier curves for 60 months of follow-up are presented. Training cases are depicted in green and non-training cases are depicted in blue. The estimated 5-year disease-free survival is 66.6% (95%-CI: 59.1–73.0%) for non-training cases and 68.5% (95%-CI: 61.3–74.5%) for training cases.

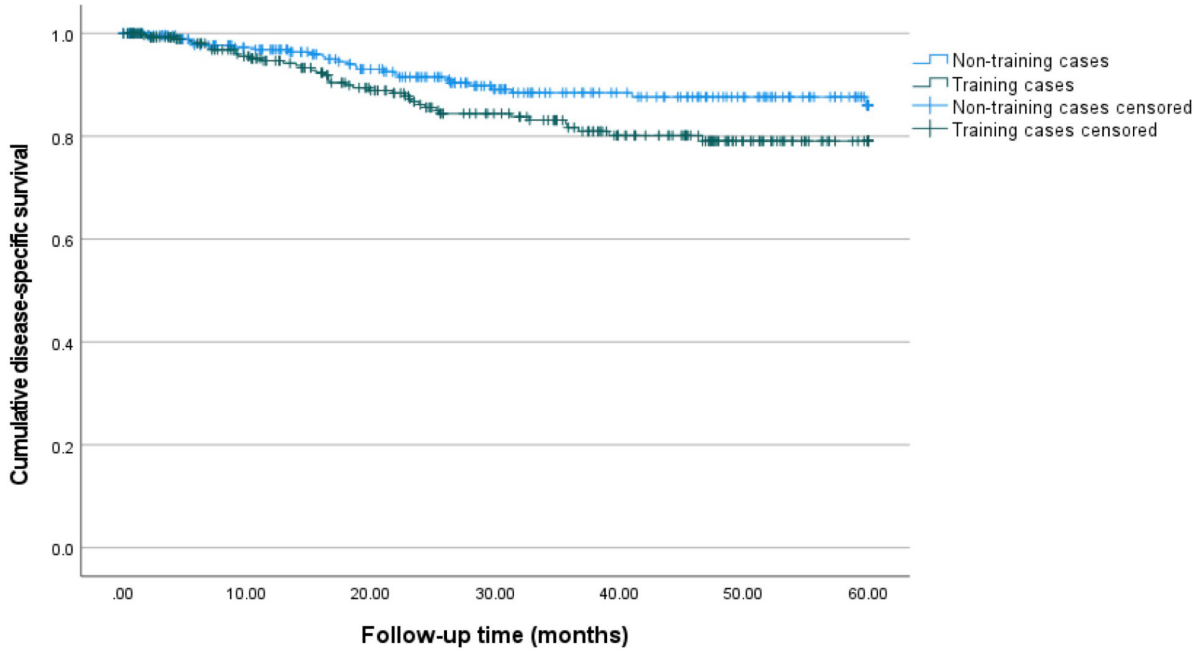


FIGURE 2
5-year disease-specific survival for training and non-training cases. Kaplan–Meier curves for 60 months of follow-up are presented. Training cases are depicted in green and non-training cases are depicted in blue. The estimated 5-year disease-specific survival is 86.1% (95%-CI: 80.0–90.6%) for non-training and 79.1% (95%-CI: 72.6–84.2%) for training cases.

0–1,200 ml, $p = 0.005$) favoring training cases. This did not result in a difference in blood transfusions (as stated above). The rates of conversions (2.4 versus 2.0%, $p = 0.749$) and CCU-admissions (2.5 versus 3.9%, $p = 0.991$) were comparable. Mean operating time was found to be longer in training cases (160 min versus 137 min, $p \leq 0.001$). Lymphoedema rates did not differ between groups (8.4

versus 11.3%, $p = 0.192$). LOS was shorter in training cases (1 day versus 2 days, $p = 0.007$).

Odds and hazard ratios for secondary outcomes are shown in Table 3. Training did not increase the odds ratio for lymphoedema (0.6, $p = 0.110$). Training cases had a lower odds for LOS > 1 day (0.6, $p = 0.004$).

4 Discussion

4.1 Summary of main findings

PBP training had no impact on intra- and post-operative complications, blood transfusions, readmissions < 30 days, return to theater rates and 5-year disease-free and disease-specific survival in RAL surgery for endometrial cancer. Therefore, it can be safely used as a training method for robotic surgery. As expected, operating time was longer in training cases but this did not have a detrimental effect on patient outcome.

4.2 Interpretation of results

We found a significantly shorter median LOS and a lower odds for LOS > 1 day in training cases compared to non-training cases. This might be associated with a gradual increase in the amount of training cases over time (37.2% in 2015 versus 55.3% in 2022) and a simultaneous slight decrease in LOS over time (2 days in 2015 versus 1 day in 2022) due to changed surgical protocols. We found a similar trend in sentinel lymph node procedures. More sentinel lymph node dissections were performed in training cases, which is possibly associated with the gradual increase in training cases over time (37.2% in 2015 versus 55.3% in 2022) accompanied with the simultaneous increase in sentinel lymph node procedures (39.5% in 2015 versus 72.4% in 2022).

In our data median disease-free survival differed between training and non-training cases. However, we also found a difference in tumor grades between training and non-training cases with a lower rate of grade 1 and higher rate of grade 3 tumors in training cases. Grade, stage and age are known predictors for endometrial cancer survival (14). Hence, we corrected for these confounders using multivariate regression analysis. After correction no influence of PBP training on disease-free and disease-specific survival was found.

4.3 Results in the context of published literature

To date no other studies have evaluated the general effect of PBP training in RAL surgery for endometrial cancer on peri-operative and survival outcomes. However, the effect of a learning curve for RAL surgery in endometrial cancer on peri-operative outcomes has been identified by two single-surgeon studies (32, 33). By comparing peri-operative outcomes between cases performed in the early stages of the learning curve and cases performed in later stages of the learning curve, we can roughly compare these results with our training and non-training cases. However, it must be noted that these studies were performed by single surgeons and only assessed a limited number of peri-operative outcomes.

One study (32) observed less estimated blood loss in cases performed early in the learning curve compared to later cases, which was also observed in our cohort. BMI is a possible confounder of EBL, with more blood loss and more blood transfusions in higher BMI groups (34). Since BMI was significantly higher in non-training cases, this is another possible explanation for the significant difference in EBL although no difference in blood transfusion rates was found. Our results on operating time are in line with two other studies (32, 33), that found significant improvements in operating time between cases performed in early stages of the learning curve and later cases.

Obese patients undergoing laparoscopic surgery are more prone to surgical and post-operative complications compared to non-obese patients (11–13, 35). Therefore, previous studies on surgical outcomes in endometrial cancer have performed case-matching based on BMI (36). On the other hand, a recent study by Uwins et al. (34) on surgical outcomes of robotic surgery for endometrial cancer did not perform matching on BMI and found no negative influence of BMI on hospital stay and conversion rate. In our study BMI differed significantly between training and non-training cases and this might have been a confounding factor for intra- and post-operative outcomes. However, additional univariate analysis showed no influence of BMI on intra- and post-operative outcomes (data not shown).

No studies assessing the learning curve of robot-assisted laparoscopic surgery for endometrial cancer evaluated survival outcomes. However, Baeten et al. (17) assessed 5-year disease-free and disease-specific survival for cervical cancer patients undergoing RAL surgery and found worse outcomes for cases in early stages of the learning curve compared to cases in later stages. Comparable results were found by two more studies (19, 20). We did not find such a trend in our cohort, which might be due to several differences with our study. First, whereas the previously mentioned studies (17, 19, 20) analyzed cases between 2007 and 2018 when there was no set training curriculum, we analyzed cases between 2015 and 2022 in which timeframe PBP training was implemented. Secondly, we did not look into individual learning curves as Baeten et al. (17) did but investigated the overall effect of PBP training on survival outcomes, which renders the possibility of underestimation of our survival outcomes (see limitations). Lastly, the effects of training in RAL surgery might differ between cervical and endometrial cancer.

In 2020 a new guideline for the definition of prognostic risk groups in endometrial cancer was formulated by ESGO/ESTRO/ESP (24). These guidelines incorporate clinicopathological with molecular variables, e.g., p53 and POLE mutation status, and effectively predict survival in endometrial cancer patients (29). Since then local protocols have been updated, but regional disparities in adherence to the guidelines still exist. This needs to be overcome to decrease the use of adjuvant therapies to spare morbidity (37, 38). Radiomics, the field in which a large number of quantitative features from radiological images are analyzed using data-characterization algorithms, is another field that potentially has an added value for the prediction of prognosis for endometrial cancer patients (39).

During our study the ESGO/ESTRO/ESP guidelines were published and our protocols were updated and implemented. However, as this implementation took its time we did not perform molecular analysis for all cases and treatment protocols were being adjusted during our study. Therefore, we chose to assess the risk groups according to the new guidelines to increase comparability with similar cohorts, but not correct for them in our main analysis. Additional analysis showed no impact of training on disease-free and disease-specific survival after correction for ESGO risk groups (data not shown).

Due to a limited number of studies in RAL surgery in gynecological oncology, we looked at other fields of robotic surgery to compare our results. A PBP training curriculum for robotic-assisted radical cystectomy by the European Association of Urologists Robotic Urology Section was recently evaluated (40). As in our cohort, operating time was significantly longer in training cases, but otherwise the trainee showed non-inferiority compared to the experienced surgeon in terms of estimated blood loss, positive soft tissue margins, number of resected lymph nodes, overall and high-grade complications, and 90-day readmissions.

Lastly, our results are in line with a meta-analysis including 19 randomized controlled trials comparing peri-operative and survival outcomes between trainees and experts in laparotomic and laparoscopic colorectal surgery. They observed a longer operating time in training cases and found no difference in survival outcomes for oncological surgery between trainees and experts (41).

So, considering all literature described above the results of our study are within expectations.

4.4 Strengths and weaknesses

Our study has several strengths. First, all procedures were performed by one surgical team in a high-volume tertiary cancer center service resulting in a large cohort with highly comparable surgical circumstances. Moreover, all consultant surgeons had extensive experience in robotic surgery (over 5 years) before subspecialty training was provided and consultant surgeons were trained to provide training in a certified manner (27). Secondly, whereas previous studies have evaluated the performance of only one or two trainees, our study includes a cohort of 18 trainees (32, 33, 40). This makes our results robust and generalizable. Moreover, our results reflect a real-world training setting in an experienced training center. This makes our results likely to be applicable to other training centers. Thirdly, our data was collected prospectively which reduces the chance of information bias and results in a limited amount of missing data. One independent researcher completed the database retrospectively thereby further reducing the likelihood of information bias.

The main limitation of our study is that we did not record which part of the surgery was performed by the trainee. Thereby we were unable to define the effect of performance of specific parts of the surgery by a trainee on peri-operative and survival outcomes possibly underestimating the effect of training in individual steps of RAL surgery on our outcomes. On the other hand, our results highlight a real-world training environment and show some expected differences between training and non-training cases (lower BMI and longer operating time) suggesting that our study has the distinguishing capacities needed to pick up major differences between training and non-training cases.

Compared to other robotic cohorts (34, 36, 42) we have a high grade/high stage cohort which is related to the tertiary referral status of our department. This might limit the generalizability of our results. Direct comparison with other robotic cohorts is needed to further evaluate the effect of PBP training on peri-operative and survival outcomes for RAL surgery in all stages of endometrial cancer.

4.5 Implications for practice and future research

Our results show that PBP training can be used safely to teach RAL surgery for endometrial cancer in a high-volume tertiary cancer service with no difference in peri-operative and survival outcomes. We suggest that a PBP training curriculum for RAL surgery should be implemented in gynecological oncology fellowships. We aim to design prospective trials to further investigate the influence of distinct parts of RAL surgery performed by a trainee on peri-operative and survival outcomes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Clinical Research Committee of the Royal Marsden Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because of the use of patient data in a research setting.

Author contributions

AS: Writing – original draft, Writing – review & editing. MN: Writing – original draft, Writing – review & editing. EN: Writing – original draft, Writing – review & editing. OH: Writing – original draft, Writing – review & editing. TI: Writing – original draft, Writing – review & editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank the team of clinical statisticians at the Royal Marsden Hospital who provided input in the statistical design and performance of statistical testing. We would also like to thank Kees Gerestein from the Department of Gynecological Oncology of the University Medical Centre Utrecht for providing input in the set-up of the study.

Conflict of interest

MN was a proctor for robot-assisted surgery in gynecological oncology for Intuitive Surgical. The hospital receives funding from Intuitive for case observations.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Janda M, Gebski V, Davies LC, Forder P, Brand A, Hogg R, et al. Effect of total laparoscopic hysterectomy vs total abdominal hysterectomy on disease-free survival among women with stage I endometrial cancer: A randomized clinical trial. *JAMA*. (2017) 317:1224–33. doi: 10.1001/jama.2017.2068

2. Walker JL, Piedmonte MR, Spirtos NM, Eisenkop SM, Schlaerth JB, Mannel RS, et al. Recurrence and survival after random assignment to laparoscopy versus laparotomy for comprehensive surgical staging of uterine cancer: Gynecologic oncology group LAP2 study. *J Clin Oncol.* (2012) 30:695–700. doi: 10.1200/JCO.2011.38.8645

3. Nobbenhuis MAE, Gul N, Barton-Smith P, O'Sullivan O, Moss E, Ind TEJ, et al. Robotic surgery in gynaecology: Scientific impact paper no. 71. *BJOG.* (2022) 130:e1–8. doi: 10.1111/1471-0528.17242

4. Melamud O, Eichel L, Turbow B, Shanberg A. Laparoscopic vesicovaginal fistula repair with robotic reconstruction. *Urology.* (2005) 65:163–6. doi: 10.1016/J.UROLOGY.2004.09.052

5. Ind T, Laios A, Hacking M, Nobbenhuis MA. comparison of operative outcomes between standard and robotic laparoscopic surgery for endometrial cancer: A systematic review and meta-analysis. *Int J Med Robot.* (2017) 13:e1851. doi: 10.1002/rcs.1851

6. Ind TEJ, Marshall C, Hacking M, Harris M, Bishop L, Barton D, et al. Introducing robotic surgery into an endometrial cancer service–a prospective evaluation of clinical and economic outcomes in a UK institution. *Int J Med Robot.* (2016) 12:137–44. doi: 10.1002/rcs.1651

7. Wright JD, Burke WM, Tergas AI, Hou JY, Huang Y, Hu JC, et al. Comparative effectiveness of minimally invasive hysterectomy for endometrial cancer. *J Clin Oncol.* (2016) 34:1087–96. doi: 10.1200/JCO.2015.65.3212

8. Bell MC, Torgerson J, Seshadri-Kreaden U, Suttle AW, Hunt S. Comparison of outcomes and cost for endometrial cancer staging via traditional laparotomy, standard laparoscopy and robotic techniques. *Gynecol Oncol.* (2008) 111:407–11. doi: 10.1016/j.ygyno.2008.08.022

9. Moss EL, Morgan G, Martin AP, Sarhanis P, Ind T. Surgical trends, outcomes and disparities in minimal invasive surgery for patients with endometrial cancer in England: A retrospective cohort study. *BMJ Open.* (2020) 10:e036222. doi: 10.1136/bmjopen-2019-036222

10. Corrado G, Ciccarone F, Cosentino F, Legge F, Rosati A, Arcieri M, et al. Role of minimally invasive surgery versus open approach in patients with early-stage uterine carcinosarcomas: A retrospective multicentric study. *J Cancer Res Clin Oncol.* (2021) 147:845–52. doi: 10.1007/s00432-020-03372-x

11. Ind TEJ, Marshall C, Hacking M, Chiu S, Harris M, Nobbenhuis M. The effect of obesity on clinical and economic outcomes in robotic endometrial cancer surgery. *Robot Surg.* (2017) 4:33–7. doi: 10.2147/RSRR.S123108

12. Moss EL, Sarhanis P, Ind T, Smith M, Davies Q, Zecca M. Impact of obesity on surgeon ergonomics in robotic and straight-stick laparoscopic surgery. *J Minim Invasive Gynecol.* (2020) 27:1063–9. doi: 10.1016/j.jmig.2019.07.009

13. Kaur MM, Ind T. Laparoscopic and robotic surgery in obese women. In: Mahmood TA editor. *Obesisty and gynecology*. Amsterdam: Elsevier (2020). p. 223–43.

14. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N. Endometrial cancer. *Lancet.* (2022) 399:1412–28. doi: 10.1016/S0140-673600323-3

15. Abel MK, Chan JK, Chow S, Darcy K, Tian C, Kapp DS, et al. Trends and survival outcomes of robotic, laparoscopic, and open surgery for stage II uterine cancer. *Int J Gynecol Cancer.* (2020) 30:1347–55. doi: 10.1136/ijgc-2020-001646

16. Roth K, Kaier K, Stachon P, von Zur Mühlen C, Jungmann P, Grimm J, et al. Evolving trends in the surgical therapy of patients with endometrial cancer in Germany: Analysis of a nationwide registry with special emphasis on perioperative outcomes. *Arch Gynecol Obstet.* (2023) 308:1635–40. doi: 10.1007/s00404-023-07127-0

17. Baeten IGT, Hoogendam JP, Schreuder HWR, Jürgenliemk-Schulz IM, Verheijen RHM, Zweemer RP, et al. The influence of learning curve of robot-assisted laparoscopy on oncological outcomes in early-stage cervical cancer: An observational cohort study. *BJOG Int J Obstet Gynaecol.* (2021) 128:563–71. doi: 10.1111/1471-0528.16399

18. Holub Z, Jabor A, Bartoš P, Hendl J, Urbánek Š. Laparoscopic surgery in women with endometrial cancer: The learning curve. *Eur J Obstet Gynecol Reprod Biol.* (2003) 107:195–200. doi: 10.1016/S0301-211500373-1

19. Ekdahl L, Wallin E, Alfonzo E, Reynisson P, Lönnerfors C, Dahm-Kähler P, et al. Increased institutional surgical experience in robot-assisted radical hysterectomy for early stage cervical cancer reduces recurrence rate: Results from a nationwide study. *J Clin Med.* (2020) 9:3715. doi: 10.3390/jcm9113715

20. Eoh KJ, Lee J-Y, Nam EJ, Kim S, Kim SW, Kim YT. The institutional learning curve is associated with survival outcomes of robotic radical hysterectomy for early-stage cervical cancer-a retrospective study. *BMC Cancer.* (2020) 20:152. doi: 10.1186/s12885-020-6660-7

21. Larcher A, De Naeyer G, Turri F, Dell'Oglio P, Capitanio U, Collins JW, et al. The ERUS curriculum for robot-assisted partial nephrectomy: Structure definition and pilot clinical validation. *Eur Urol.* (2019) 75:1023–31. doi: 10.1016/J.EURURO.2019.02.031

22. Society of European Robotic Gynaecological Surgery. *SERGS curriculum pathway – SERGSSERGS*. (2019). Available online at: https://www.sergs.org/sergs-education/sergs-curriculum-pathway/ (assessed March 10, 2023).

23. British and Irish Robotic Gynaecological Surgeons. *BIARGS TRAINING CURRICULUM*. (2020). Available online at: https://biargs.org.uk/biargs-training-curriculum/ (accessed March 10, 2023).

24. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *Int J Gynecol Cancer.* (2021) 31:12–39. doi: 10.1136/ijgc-2020-002230

25. Rusch P, Ind T, Kimmig R, Maggioni A, Ponce J, Zanagnolo V, et al. Recommendations for a standardised educational program in robot assisted gynaecological surgery: Consensus from the Society of European Robotic Gynaecological Surgery (SERGS). *Facts Views Vis ObGyn.* (2019) 11:29.

26. Ismail A, Wood M, Ind T, Gul N, Moss E. The development of a robotic gynaecological surgery training curriculum and results of a delphi study. *BMC Med Educ.* (2020) 20:66. doi: 10.1186/S12909-020-1979-Y

27. Mackenzie H, Cuming T, Miskovic D, Wyles SM, Langsford L, Anderson J, et al. Design, delivery, and validation of a trainer curriculum for the national laparoscopic colorectal training program in England. *Ann Surg.* (2015) 261:149–56. doi: 10.1097/SLA.0000000000000437

28. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: A new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg.* (2004) 240:205–13. doi: 10.1097/01.sla.0000133083.54934.ae

29. Imboden S, Nastic D, Ghaderi M, Rydberg F, Siegenthaler F, Mueller MD, et al. Implementation of the 2021 molecular ESGO/ESTRO/ESP risk groups in endometrial cancer. *Gynecol Oncol.* (2021) 162:394–400. doi: 10.1016/j.ygyno.2021.05.026

30. Heymans M, Eekhout I. *Applied missing data analysis with SPSS and (R) studio*. Amsterdam: Heymans and Eekhout (2019).

31. Mauger DT, Kauffman GL. Statistical Analysis–specific statistical tests: Indications for use. *Surg Res.* (2001) 24:1201–15. doi: 10.1016/B978-012655330-7/50084-8

32. Lim PC, Kang E, Park DH. Learning curve and surgical outcome for robotic-assisted hysterectomy with lymphadenectomy: Case-matched controlled comparison with laparoscopy and laparotomy for treatment of endometrial cancer. *J Minim Invasive Gynecol.* (2010) 17:739–48.

33. Rajanbabu A, Patel V, Anandita A, Burde K, Appukuttan A. An analysis of operating time over the years for robotic-assisted surgery in gynecology and gynecologic oncology. *J Robot Surg.* (2021) 15:215–9. doi: 10.1007/s11701-020-01094-3

34. Uwins C, Hablase R, Assalaarachchi H, Tailor A, Stewart A, Chatterjee J, et al. Enhanced recovery after uterine corpus cancer surgery: A 10 year retrospective cohort study of robotic surgery in an NHS cancer centre. *Cancers.* (2022) 14:5463. doi: 10.3390/cancers14215463

35. Gunderson CC, Java J, Moore KN, Walker JL. The impact of obesity on surgical staging, complications, and survival with uterine cancer: A gynecologic oncology group LAP2 ancillary data study. *Gynecol Oncol.* (2014) 133:23–7. doi: 10.1016/J.YGYNO.2014.01.041

36. Corrado G, Cutillo G, Pomati G, Mancini E, Sperduti I, Patrizi L, et al. Surgical and oncological outcome of robotic surgery compared to laparoscopic and abdominal surgery in the management of endometrial cancer. *Eur J Surg Oncol.* (2015) 41:1074–81. doi: 10.1016/J.EJSO.2015.04.020

37. Restaino S, Paglietti C, Arcieri M, Biasioli A, Della Martina M, Mariuzzi L, et al. Management of patients diagnosed with endometrial cancer: Comparison of guidelines. *Cancers (Basel).* (2023) 15:1091. doi: 10.3390/cancers1504 1091

38. Di Donato V, Giannini A, Bogani G. Recent advances in endometrial cancer management. *J Clin Med.* (2023) 12:2241. doi: 10.3390/jcm12062241

39. Bogani G, Chiappa V, Lopez S, Salvatore C, Interlenghi M, D'Oria O, et al. Radiomics and molecular classification in endometrial cancer (The ROME Study): A step forward to a simplified precision medicine. *Healthcare.* (2022) 10:2464. doi: 10.3390/healthcare10122464

40. Diamand R, D'Hondt F, Mjaess G, Jabbour T, Dell'Oglio P, Larcher A, et al. Teaching robotic cystectomy: Prospective pilot clinical validation of the ERUS training Curriculum. *BJU Int.* (2023) 132:84–91. doi: 10.1111/BJU.15993

41. Kelly M, Bhangu A, Singh P, Fitzgerald JEF, Tekkis PP. Systematic review and meta-analysis of trainee– versus expert surgeon-performed colorectal resection. *Br J Surg.* (2014) 101:750–9. doi: 10.1002/bjs.9472

42. Mäenpää MM, Nieminen K, Tomás EI, Laurila M, Luukkaala TH, Mäenpää JU. Robotic-assisted vs traditional laparoscopic surgery for endometrial cancer: A randomized controlled trial. *Am J Obstet Gynecol.* (2016) 215:588.e1–588.e7. doi: 10.1016/j.ajog.2016.06.005

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 June 2024
DOI 10.3389/fmed.2024.1387108

OPEN ACCESS

EDITED BY
Florian Recker,
University of Bonn, Germany

REVIEWED BY
Hariharan N. Krishnasamy,
Universiti Utara Malaysia, Malaysia
Zhu Su,
Central China Normal University, China

*CORRESPONDENCE
Robert de Leeuw
✉ r.a.deleeuw@amsterdamumc.nl

RECEIVED 16 February 2024
ACCEPTED 10 May 2024
PUBLISHED 06 June 2024

CITATION
de Leeuw R, Huirne JAF, Rositto C, Mabrouk M, Barri P, Bongers M, Thurkow A, El-Balat A, Vlahos N and Brolmann H (2024) The added value of a face-to-face pan-European course—what makes it worth it?
Front. Med. 11:1387108.
doi: 10.3389/fmed.2024.1387108

The added value of a face-to-face pan-European course—what makes it worth it?

Robert de Leeuw[1,2]*, Judith A. F. Huirne[1,2], Christiano Rositto[3,4], Mohammed Mabrouk[5,6], Pierre Barri[7], Marlies Bongers[7], Andreas Thurkow[8], Ahmed El-Balat[9,10], Nikon Vlahos[11] and Hans Brolmann[1,2]

[1]Department of Obstetrics and Gynaecology, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [2]Amsterdam Reproduction and Development, Amsterdam, Netherlands, [3]Casa di Cura Santa Famiglia, Rome, Italy, [4]Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy, [5]University College of London Hospitals (UCLH), London, United Kingdom, [6]The Cleveland Clinic, London, United Kingdom, [7]Medisch Centrum Maxima, Eindhoven, Netherlands, [8]Onze Lieve Vrouwen Gasthuis, Amsterdam, Netherlands, [9]Spital Uster, Uster, Switzerland, [10]Universitätsspital Zürich, Zurich, Switzerland, [11]Athens Medical School, Aretaieion University Hospital, Athina, Greece

Introduction: Over the past decade, digital education has seen widespread adoption, particularly accentuated during the COVID-19 pandemic. The post-COVID era has further emphasized the advantages of digital education in terms of cost, availability, and sustainability. However, concerns regarding the efficacy of digital education, particularly in skills-based learning and the absence of social interaction, have been raised. This paper will look at the added value of international, face-to-face, skills-based courses.

Method: This study evaluates the potential added value of face-to-face international skills courses using the European "Gynecology Experts Training for Upcoming Professionals" (GET-UP) course. Focus group discussions were conducted with participants and faculty members to explore beliefs, attitudes, and perceptions regarding face-to-face learning. Qualitative analysis was performed using thematic analysis to identify domains of added value.

Results: The GET-UP course, conducted over 4 days with a diverse European faculty and participants, highlighted several added-value domains. Themes including diversity, role models, preparation, live interaction, and community emerged from the analysis, emphasizing the significance of face-to-face interaction in enriching the learning experience beyond attaining learning goals.

Discussion: The study underscores the importance of face-to-face interaction in educational settings, offering insights into diverse teaching methods, role modeling opportunities, enhanced preparation, live interactions, and fostering a sense of community. While digital education continues to evolve with interactive features, this study suggests that the inherent pressure and dynamics of face-to-face learning provide unique benefits that may not be easily replicated in digital environments. Future research should investigate and validate these findings further to inform educational practices effectively.

KEYWORDS

continuous education, European education, intercultural competence, skills education, digital education

1 Introduction

Over the past decade, the education landscape has undergone a profound transformation with the increasing integration of digital technologies (1). This shift has been driven by technological advancements and the need for flexible and accessible learning solutions. Particularly noteworthy is the acceleration of digital education adoption during the COVID-19 pandemic, where widespread school closures necessitated a swift transition to remote learning modalities to ensure the continuity of education (2).

In the wake of the pandemic, we find ourselves in a post-COVID era where digital education has gained further traction, supported by its cost-effectiveness, widespread availability, and environmental sustainability (3). The conveniences afforded by digital platforms, such as anytime, anywhere access to educational resources, have positioned them as indispensable tools in modern medical education. However, amidst the recognition of the benefits of digital education, concerns have been raised regarding its efficacy, particularly in skills-based learning (4).

One of the primary advantages of traditional face-to-face education is the personalized guidance and interaction provided by tutors or instructors, which has been perceived as indispensable, especially in skill acquisition (4). The nuanced feedback, hands-on demonstrations, and individualized support offered in face-to-face settings contribute significantly to the learning experience, fostering deeper understanding and skill mastery. Moreover, the absence of social interaction in most forms of digital education has been identified as a notable drawback, with concerns regarding its impact on student engagement, motivation, and interpersonal skills development (5).

Despite the emphasis on achieving learning goals in educational studies, there has been a noticeable oversight in evaluating the potential additional benefits of educational programs, particularly in fostering social interaction and intercultural competence (6). While learning outcomes remain paramount, the value derived from social interaction, sense of belonging, and cultural exchange in face-to-face settings is increasingly recognized as integral to holistic learning experiences (7).

A potential pitfall of current educational literature is that we forget to evaluate this additional profit, focus on learning goals only, and start providing very effective digital education without realizing what we are missing out on. It is suggested that the added value of face-to-face courses lies in broader goal orientation, sense of belonging, social interaction, and intercultural competence (7). However, despite the recognized benefits of face-to-face education, there remains a gap in understanding its comparative advantages over digitally mediated or locally organized courses.

Understanding the added value of face-to-face education is crucial for several reasons. Firstly, it allows educators and policymakers to make informed decisions about the design and implementation of educational programs, ensuring that they effectively meet learners' diverse needs and preferences (8). Additionally, gaining insights into face-to-face interaction's unique benefits enables optimal educational resources and instructional strategies to maximize learning outcomes. Ultimately, by comprehensively evaluating the added value of face-to-face education, we can enhance the quality and efficacy of educational experiences, fostering holistic development and lifelong learning opportunities for learners.

Therefore, we aim to evaluate the potential added value of a face-to-face international skills course. Once a year, a European course for minimally invasive gynecological surgery (the Gynecology Experts Training for Upcoming Professionals or the GET-UP course) is organized. This course is aimed at novice, experienced minimal invasive gynecological surgeons and/or residents. The GET-UP course is a 4-day course characterized by a productive mix of lectures, short interactive sessions (*SIC*) and hands-on training (HOT). Each morning starts with a plenary session of 30 min, followed by a HOT or *SIC* session. The afternoon starts again with a plenary session, followed by a HOT session for those who had a *SIC* session in the morning and the other way around. The course offers a diverse European faculty, the opportunity to network with peers and upcoming professionals, and cooperation between 25 European medical centers. Participants are invited to participate by local representatives of the GETUP course, local advertisement, or the faculty. The GET-UP board considers geographical spread, and each faculty member can subjoin a maximum of two participants. Finally, 100 participants are invited to participate. The faculty consists of European endoscopic experts with a wide variety of experience, fields of expertise and know-how. The course aims to improve the learner's knowledge, skills, and attitude and enforce collaboration, sharing and endorsing a European community of minimally invasive gynecological surgeons. This study will use the GET-UP course as an example of a face-to-face international skills course to evaluate the potential added value. By examining the nuanced experiences and perceptions of participants and faculty members through qualitative analysis, this study aims to shed light on the multifaceted benefits of face-to-face education and inform educational practices in the evolving digital learning landscape.

2 Methodology

A focus group discussion using thematic analysis is well-suited for exploring the added value of face-to-face education (9). It allows for an in-depth exploration of participants' and faculty members' beliefs, attitudes, and perceptions, capturing diverse perspectives and experiences (10). Thematic analysis is a well-described and accepted way to facilitate the identification of recurring themes and patterns within the qualitative data, providing rich insights into the multifaceted benefits of face-to-face interaction (11). This approach enables researchers to uncover nuanced understandings of the educational process, including the significance of social interaction, role modeling, and community building, thereby comprehensively addressing the research question's complexity (12).

2.1 Study participants

We invited both participants and faculty members to a focus group discussion after the course. Before the course started, all participants and faculty were invited to participate in an evaluation. Only participants who attended the whole course and faculty members who were present during all aspects were included to participate. After the course ended, volunteers approached the author RAL for the focus group (Figure 1).

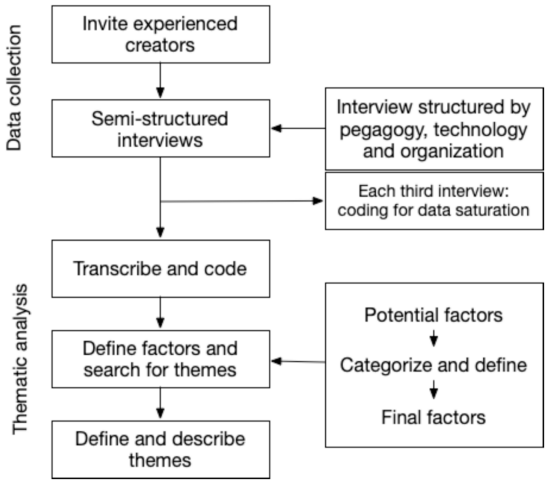


FIGURE 1
Flowchart of methodology.

2.2 Data collection and analysis

Before the start of the interview, we also provided the participants with a questionnaire containing general demographics and base knowledge. To collect the data, we recorded the focus-group discussions after informed consent from the discussion members. The sessions were facilitated by RAL and lasted between 30 and 45 min. The recordings were anonymously transcribed at verbatim. After transcribing, the interview will be analyzed using Max-QD software, and thematic analysis was used to determine the domains of added value.

To perform the data analysis in a structured method, we used the six steps proposed by Braun et al., containing:

1 Familiarizing oneself with the data,
2 Generating initial codes,
3 Searching for themes,
4 Reviewing themes,
5 Defining and naming themes, and
6 Producing the report (13).

2.3 Ethical aspects

All participants were asked to participate before the start of the course and signed an informed consent form. The faculty does not know who refused; therefore, we do not expect a different attitude in the interaction. The Dutch Society for Medical Education gave ethical approval under file number 00833.

3 Results

The GET-UP course took place over 4 days in April 2018. There were 89 participants and 47 faculty members. Of the 59 participants that were included, 83% were in the last year of their residency or just finished, 75% were female, 90% never or rarely had professional contact with a peer from another European country, and 73% never attended another laparoscopy course in their residency (see Figure 2). A total of two interviews were conducted, involving 8 faculty members and 9 participants. Both groups contained participants from Italy, the Netherlands, Great Britain, Denmark, Spain, and Germany and 1 participant came from Romania. 88% of the participants and 25% of the faculty were female.

3.1 Reviewing and defining themes

According to the steps of the thematic analysis, all texts were transcribed by RAL to familiarize oneself with the data. In the second step, the transcription was coded into 62 initial codes. Searching within those codes, five themes emerged; diversity, role models, preparation, live interaction, and community (see Figure 3). Reviewing the codes (step 4) did not reveal any other themes. The next step is defining the themes, which will be explained in more detail.

3.1.1 Theme one: Diversity

The first theme that emerged is "Diversity," which is defined as "a variation in cultural backgrounds, teaching methods, communication skills, and problem-solving methods." Both faculty and participants highlighted the importance of diversity and the specific added value of this domain to digital education. Several aspects were specifically named as beneficial. The course was international in nature, and participants and faculty from different regional and cultural backgrounds were gathered. This showed the participants the different cultural interpretations of health care and problem-solving.

> "*It's so important to experience different ways of approaching the same problem*" (participant 3).

> "*The benefit of mixing with other countries and finding out what's going on outside of your own country is very valuable*" (participant 5).

The faculty also experienced different aspects of diversity as added value. One important example was diversity in teaching methods, which was discussed by both participants and faculty members. In medicine, teaching is done by each level of experience. Therefore, participants usually also teach back at their clinic, and the participants enjoyed learning about diversity in teaching methods.

> "*apart from learning the content of the course, I love to see others teach, so I can adapt my method of teaching*" (participant 2).

3.1.2 Theme two: Role models

The second theme is "Role models," defined as: "a person showing excellence in craft, teaching abilities and personal qualities." The participants described the added value of role models during the course. They got inspiration from a person's excellence in laparoscopic surgery, demonstrated in skills education, videos and presentations. Another aspect was the faculty's teaching skills,

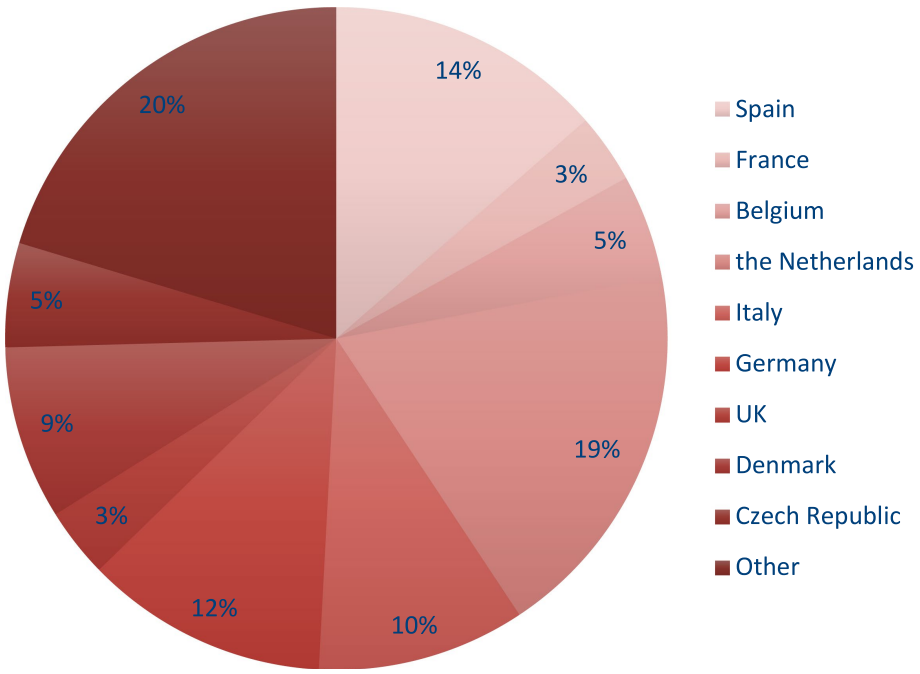


FIGURE 2
Country of residency of participants.

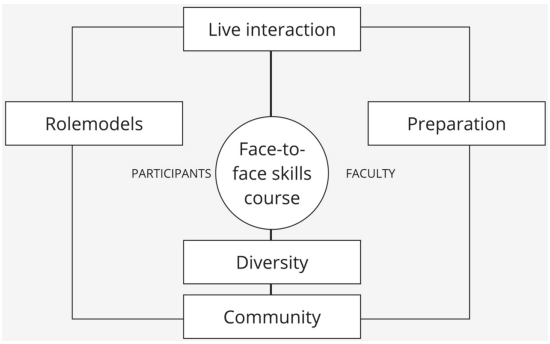


FIGURE 3
The themes showing the added value of face-to-face learning.

creating inspiration to achieve that level. An important value was learning how to teach. Faculty learned from each other and inspired each other to improve their teaching. Finally, the personal qualities demonstrated in how faculty interact with each other and with the participants allowed participants to get a unique insight into their role models. Participants mentioned that they can be role models for each other. Demonstrating different communication skills and interactions with the faculty, depending on the diverse cultural background.

> "*Everybody has a way of working, but we should evaluate this and take the positives from it*" (participant 3).

Finally, participants were inspired by the faculty by showing their vulnerabilities. Saying what they can do, what they cannot do, and where they make mistakes. This showed the participants that their role models are open about their skills and limitations and are willing to get out of their comfort zone. This is illustrated by a quote:

> "*It's very important to get out of your comfort zone, and GET-UP forces you to do so*" (faculty 2).

3.1.3 Theme three: Preparation

The third theme is "Preparation" and is defined as: "the work that has to be done before the course, as part of mandatory lessons, or the creation and/or updating of presentations." Although the participants enjoyed the mandatory preparation, the faculty mainly stressed the added value of their preparation for the course. The faculty experienced a certain amount of peer pressure between faculty members and the feeling of responsibility for the participants to be more prepared than usual.

> "*The dedication and preparation for this course is much higher than to a congress, or even to teaching in the hospital*" (faculty 3).

The possible explanation was that faculty members knew that other faculty members were always present during their presentations and that their sense of community (theme five) pushed them to put in extra effort. They said that when a presentation is due, they usually use an old, existing one. But for this course, they always wanted to do an

update just before the course to be up to date. This was also due to the intimate short communications (two or three faculty members and about 10 participants). Here, the faculty felt more responsible for the quality of their teaching than when they were educating without peers or a larger group. The faculty experienced the preparation as the most important part of their own learning process.

3.1.4 Theme four: Live interaction

The fourth theme is called "Live interaction" and is defined as "the communication (verbal and non-verbal) that faculty and participants experience due to their physical presence at the course." In line with the role models domain, seeing people interact with each other is a very important aspect of a face-to-face course. Both parties said that, apart from learning from short interactions and hands-on training, they learned much from the moments in between. The coffee breaks, the social program in the evening, the pre-and post-course walks and room switching all provided insight into how faculty and participants interacted. The changes in venue, rooms, and assignments (interactive sessions and hands-on training) created an important energy that allowed participants to get more out of the course. This theme overlaps with the role model theme because it addresses communication observation. Yet this was experienced differently because this interaction was about the physical presence of the people themselves. One example is described as:

> "*The highest level of training is training*" (faculty 1).

The faculty members said that physical training and being physically together are the highest levels of training. Another example from the participants was that the act of physically moving around and forcing each other to change locations was of great added value:

> "*The mix between inactive sessions and hands-on training is very important and keeps everybody fresh and active. I have not fallen asleep once*" (participant 1).

3.1.5 Theme five: Community

Finally, the last theme was "Community," which is defined as "the feeling of belonging to a group of peers with comparable interests, passions and calling to improve women's health." This was one of the most important domains described by participant 2 as:

> "*to get other connections from other countries is key for your own development*."

A feeling of unity was created by the yearly recurrent faculty members, who addressed that seeing each other every year created a sense of community between them. Knowing each other's expertise, interests, and partially personal life makes a community stick together. This challenging community stimulates critical thinking and stimulates evidence-based care. As faculty 6 said:

> "*When you are in a critical, international community, you have to think more evidence-based to support your ideas and treatment*."

This sense of community helps faculty members contact each other outside of the course and allows easy contact when faced with difficult cases in the clinic. According to the faculty, this is frequently done, and faculty members enjoy the ability to know European experts for advice and refer patients.

> "*the collective knowledge is huge*" (faculty 2).

Finally, both faculty and participants enjoyed the networking within the community. Networking allowed new research collaborations, easier consultations with experts and the sharing of contact information for future collaborations.

4 Discussion

This study provides five themes that underline the potential additional values of face-to-face learning beyond the initial learning goals. Both faculty and participants provided important insights into aspects rarely evaluated during an educational course. Although diversity in medical education has been frequently addressed in the literature (14), the interpretation of diversity is usually limited to gender and social-cultural aspects. This paper adds unique aspects to the diversity topic, including different teaching styles. Brand et al. write in 2022 that "*Face-to-face meetings facilitate great spontaneity, profound exchange, nuanced communication, personal sharing, and efficient and passionate occurrence of new ideas*," which this study also supports (15). The importance of role models in the education of residents has been described before by van Delft et al. (2018), who show that subconscious behavior by faculty is experienced by participants and is very difficult to mimic in a digital education environment (16). Brand et al. (2022) write that face-to-face meetings allow one to "*gain wisdom from experienced and devoted leaders*," which is a variation of the effect of role models (15). Previous studies have also shown the added value of preparing for education (17). Although most studies address this value for students, this paper also shows that faculty can experience great added value in preparing for a course. Faculty members can experience peer pressure to keep their material up-to-date and evidence-based, which increases the chance that they get this added value from preparing for the course. Cullen et al. demonstrated that face-to-face learning during the COVID-19 crisis was still valuable, especially for skills courses (18) and write that live interaction is viable only during face-to-face learning. Enoch et al. write that "*face-to-face learning will always prevail due to the practical skills that doctors need to acquire. The skills that are central to a doctor's role simply cannot be taught online, and the needs of medical students must be considered in order to produce prepared and competent doctors*" (19).

King et al. compared face-to-face with distance education in 2022 and showed that students found interpersonal interactions an added value of face-to-face learning (20). Michno et al. writes "*students believed that the introduction of online lectures, as a replacement for face-to-face seminars, would have a negative impact on the quality of their medical education*" (21). Another domain from King et al. was the social support network, an important domain in education in general (20). A feeling of community has been described before, and it was found that students were more satisfied. It fostered more meaningful and longitudinal relationships between students and teachers when they experienced a sense of community (22).

The biggest limitation of this study is the focus group members and the sample size. The participants might be biased because they traveled to the course and might want to give socially desirable answers. There is also a selection bias among the participants, where those who did not find benefits from the course did not volunteer for the focus group discussion. The faculty members can be biased toward their opinions as they are part of the face-to-face course. Another limitation is the number of interviews. By analyzing more focus group discussions, other themes might also have appeared. Despite these potential biases and socially desirable answers, this study provides five relevant themes supported by different relevant studies that add value to an international face-to-face skills course. A point can be made that some of these themes might also be possible to address in digital education. More and more digital education is getting interactive; break-out rooms can create a sense of community, and everybody can properly prepare for a digital course. A possible limitation, therefore, lies in the generalizability of the experience in this course, compared to other courses that are provided. Yet this study shows that there is something extra in facing real people. The relative safety of an online environment allows people to hide from their responsibilities, stay hidden inside their comfort zone and prevent exposure to sometimes unwanted but needed social interaction. The underlying pressure of face-to-face education might be something that can and should not always be replaced by a digital variant.

This study is just the beginning of determining the added value of face-to-face learning in a digital world. The themes from this study should be evaluated in different courses with different learning goals. Another step can be determining other domains' presence in those courses. It is possible that other courses, with different content and learning aims, can provide new insights in other domains. Finally, future studies should evaluate a face-to-face course on the five domains from this study and see how each domain is experienced by the whole group to address their possible generalizability.

5 Conclusion

Our study suggests that face-to-face international education may still have a relevant role to play. An in-depth analysis of several domains revealed multiple valuable aspects identified by participants and faculty. While the effort, costs, and sustainability issues related to travel remain concerns, our findings indicate that in an increasingly digital world, there are compelling reasons and benefits for maintaining direct, in-person educational experiences. It is incumbent upon educators to maximize the value of these increasingly vital educational opportunities."

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by NVMO (Nederlands Vereniging Medisch Onderwijs). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

RL: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization. CR: Conceptualization, Formal analysis, Methodology, Project administration, Validation, Data curation, Investigation, Writing – review & editing. JH: Conceptualization, Formal analysis, Methodology, Project administration, Validation, Supervision, Writing – review & editing. MM: Writing – review & editing. PB: Data curation, Methodology, Supervision, Validation, Writing – review & editing. MB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – review & editing. AT: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing – review & editing. AE-B: Conceptualization, Data curation, Investigation, Methodology, Project administration, Validation, Writing – review & editing. NV: Conceptualization, Data curation, Investigation, Methodology, Writing – review & editing. HB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to acknowledge all participants of the course and all faculty members who participated in the interviews.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Lucey CR, Davis JA, Green MM. We have no choice but to transform: the future of medical education after the COVID-19 pandemic. *Acad Med*. (2022) 97:S71–81. doi: 10.1097/ACM.0000000000004526

2. Althubaiti A, Tirksstani JM, Alsehaibany AA, Aljedani RS, Mutairii AM, Alghamdi NA. Digital transformation in medical education: factors that influence readiness. *Health Informatics J*. (2022) 28:14604582221075554. doi: 10.1177/14604582221075554

3. Abovarda A, Vallo Hult H, Master Ostlund C, Palsson P. E-learning as part of residency education. *Stud Health Technol Inform*. (2023) 302:496–7. doi: 10.3233/SHTI230188

4. Van Puyvelde H, Basto M, Chung ASJ, Van Bruwaene S. Making surgery safer in an increasingly digital world: the internet-friend or foe? *World J Urol*. (2020) 38:1391–5. doi: 10.1007/s00345-020-03145-8

5. Hartnup B, Dong L, Eisingerich AB. How an environment of stress and social risk shapes student engagement with social media as potential digital learning platforms: qualitative study. *JMIR Med Educ*. (2018) 4:e10069. doi: 10.2196/10069

6. Venkatesh S, Rao YK, Nagaraja H, Woolley T, Alele FO, Malau-Aduli BS. Factors influencing medical Students' experiences and satisfaction with blended integrated E-learning. *Med Princ Pract*. (2020) 29:396–402. doi: 10.1159/000505210

7. Besser A, Flett GL, Zeigler-Hill V. Adaptability to a sudden transition to online learning during the COVID-19 pandemic: understanding the challenges for students. *Scholarsh Teach Learn Psychol*. (2022) 8:85–105. doi: 10.1037/stl0000198

8. Lee C, Hall K, Anakin M, Pinnock R. Towards a new understanding of uncertainty in medical education. *J Eval Clin Pract*. (2021) 27:1194–204. doi: 10.1111/jep.13503

9. de Leeuw RA, Westerman M, Scheele F. Quality indicators for learner-centered postgraduate medical e-learning. *Int J Med Educ*. (2017) 8:153–62. doi: 10.5116/ijme.58ce.60aa

10. Stalmeijer RE, McNaughton N, Van Mook WN. Using focus groups in medical education research: AMEE guide no. 91. *Med Teach*. (2014) 36:923–39. doi: 10.3109/0142159X.2014.917165

11. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE guide no. 131. *Med Teach*. (2020) 42:846–54. doi: 10.1080/0142159X.2020.1755030

12. Santhosh L, Rojas JC, Lyons PG. Zooming into focus groups: strategies for qualitative research in the era of social distancing. *ATS Sch*. (2021) 2:176–84. doi: 10.34197/ats-scholar.2020-0127PS

13. Braun V, Clarke V. Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. *Health Psychol Rev*. (2023) 17:695–718. doi: 10.1080/17437199.2022.2161594

14. Muntinga ME, Krajenbrink VQ, Peerdeman SM, Croiset G, Verdonk P. Toward diversity-responsive medical education: taking an intersectionality-based approach to a curriculum evaluation. *Adv Health Sci Educ Theory Pract*. (2016) 21:541–59. doi: 10.1007/s10459-015-9650-9

15. Brand JC, Lubowitz JH, Rossi MJ. Is there still a place for face-to-face meetings? *Arthroscopy*. (2022) 38:2771–2. doi: 10.1016/j.arthro.2022.08.007

16. van Delft KWM, de Leeuw RA. How to attract talented juniors to urogynaecology. *Int Urogynecol J*. (2018) 29:323–5. doi: 10.1007/s00192-017-3549-3

17. Mylopoulos M, Brydges R, Woods NN, Manzone J, Schwartz DL. Preparation for future learning: a missing competency in health professions education? *Med Educ*. (2016) 50:115–23. doi: 10.1111/medu.12893

18. Cullen SE, Tiu A, Vaghela KR, Hunter AR. A face-to-face surgical instrumentation course during the COVID-19 pandemic. *Cureus*. (2021) 13:e19266. doi: 10.7759/cureus.19266

19. Enoch TR, Williams RC. Why face-to-face medical education will prevail despite the world's swift acclimatisation to virtual learning. *Postgrad Med J*. (2022) 98:e146–7. doi: 10.1136/postgradmedj-2021-140426

20. Johnson King O, Ryan F, Cunningham S. Postgraduate student perceptions of face-to-face and distance education in orthodontics: a cross-sectional qualitative study. *J Orthod*. (2022) 49:280–7. doi: 10.1177/14653125221077108

21. Michno DA, Tan J, Adelekan A, Konczalik W, Woollard ACS. How can we help? Medical students' views on their role in the COVID-19 pandemic. *J Public Health*. (2021) 43:479–89. doi: 10.1093/pubmed/fdaa271

22. Levine RB, Shochet RB, Cayea D, Ashar BH, Stewart RW, Wright SM. Measuring medical students' sense of community and satisfaction with a structured advising program. *Int J Med Educ*. (2011) 2:125–32. doi: 10.5116/ijme.4ea7.e854

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 09 July 2024
DOI 10.3389/fmed.2024.1363844

OPEN ACCESS

EDITED BY
Sarah M. Cohen,
Hadassah Medical Center, Israel

REVIEWED BY
Pauliana Valéria Machado Galvão,
Universidade de Pernambuco, Brazil
Ali Çetin,
University of Health Sciences, Türkiye

*CORRESPONDENCE
Gezachew Gebeyehu Arega
✉ gizachewg21@gmail.com

RECEIVED 31 December 2023
ACCEPTED 13 June 2024
PUBLISHED 09 July 2024

CITATION
Arega GG, Mitku AA, Mohammed Hussen N, Mamaru Awoke S, Berelie Berehan H and Alem KJ (2024) Spatial variation of short birth intervals and their determinant factors among reproductive women in Ethiopia using a geographically weighted regression model. *Front. Med.* 11:1363844.
doi: 10.3389/fmed.2024.1363844

Spatial variation of short birth intervals and their determinant factors among reproductive women in Ethiopia using a geographically weighted regression model

Gezachew Gebeyehu Arega[1]*, Aweke Abebaw Mitku[2], Nuru Mohammed Hussen[1], Shegaw Mamaru Awoke[3], Haymanot Berelie Berehan[3] and Kasaneh Jigar Alem[4]

[1]Department of Statistics, Samara University, Semera, Ethiopia, [2]Department of Statistics, Bahir Dar University, Bahir Dar, Ethiopia, [3]Department of Statistics, Assosa University, Assosa, Ethiopia, [4]Department of Statistics, Jijiga University, Jijiga, Ethiopia

Background: In low- and middle-income nations, a significant proportion of maternal and infant deaths are caused by a short birth interval (SBI). In Ethiopia, it is the main factor contributing to maternal and infant mortality. Understanding the spatial distribution of SBIs, i.e., birth intervals of less than 33 months, and the factors that influence them is important for categorizing and promoting targeted interventions. This study used a geographically weighted regression model to evaluate the factors associated with SBIs in hot areas of Ethiopia.

Methods: The 2019 Ethiopian Mini Demographic and Health Survey, which is nationally representative, provided the data for this study. The first step in the two-stage cluster design used to collect the data was enumeration areas, and the second stage was households. The survey was conducted between 21 March 2019 and 28 June 2019. A hot spot analysis (local Getis-Ord Gi* statistics) was initially used to investigate spatial variation in SBIs. Geographically weighted regression was used to examine the regional variation in the relationship between SBIs and the factors that cause them.

Result: The study indicated that the overall proportion of SBIs among women in Ethiopia was 43.2%. The values for Global Moran's I (Moran's $I = 0.773$ and $p < 0.001$) showed the presence of significant SBIs clustering in Ethiopian administrative zones in Ethiopia. High-risk areas of the SBIs include Jarar, Doolo, Shabelle, Afder, Liben, Korahe, Nogob, West Harerge, Guji, Sidama, and Assosa zones.

Conclusion: Living in a geographic region with a high proportion of uneducated women, women lacking breastfeeding practices, and followers of Orthodox religions increased the proportion of SBIs. Our full map of hot spots for short birth spacing and the factors that affect them helps in the implementation of precise public health measures for decision-makers.

KEYWORDS

short birth interval (SBI), CSA, EAS, EMDHS, OLS, geographically weighted regression (GWR)

Introduction

The period between the birth of the child being studied (the index child) and the birth that occurred right before it is known as the short birth interval (SBI). Given the implications for fertility, maternal health, and child health, SBIs have drawn more attention in studies on demography and public health (1). The majority of health risks are related to SBIs between pregnancies. There are health risks due to both widely spaced and closely spaced pregnancies (2). Poorly spaced pregnancies have been linked to poor mother and child health outcomes across the world. Every year, 11 million children under the age of 5 are believed to have died naturally, with poor nations accounting for 99% of those deaths (3).

Globally, a birth interval of less than 18 months is associated with a higher risk of maternal, newborn, under-5, and neonatal mortality (3). According to data from 18 developing countries (Africa, Asia, Latin America, and the Middle East) and a global comparison study of 77 countries using Demographic and Health Survey (DHS) data, a birth interval of 3 years or more improves the survival status of mothers, children under 5, and infants (4). Ethiopia is the second most populated country in Africa, with a population of over 100 million and a fertility rate of 4.6 children per woman. Due to sociocultural and religious factors, it has not experienced much progress in reducing births, like many other African nations (5). Due to sociocultural and religious issues, Ethiopia has, like many other African nations, shown little change in the reduction of fertility thus far (6). Early first marriages, the desire for more children, and poor contraceptive use due to religious beliefs all have an impact on fertility. In Ethiopia, SBIs have an impact on neonatal, infant, and childhood mortality rates in addition to fertility (7).

A 2016 Ethiopian DHS found that extending the birth interval by at least 2 years reduces infant mortality by 50% and fertility by 43% (8). Greater benefits come from spacing out successive pregnancies and abortions, as well as by lowering the risk of unsafe abortion-related complications. Additionally, it promotes children's development by improving the previous child's nutritional status (8). Numerous research used various variable sets and statistical techniques, such as survival analysis, logistic regression, and linear regression models, to examine the factors influencing fertility in Ethiopia (9, 10).

The binary logistic regression model and an auto-logistic regression model are used for identifying the relation between SBIs and explanatory variables, as well as to model spatial data, with spatial effects accounting for spatial dependence (9). For instance, regression models that violate the assumptions of independence and normality for errors with a constant variance are handled poorly by all the above classic statistical models and methods (11).

For the study of spatial data, several studies have limited the use of regression models using geographical factors. Additionally, investigations carried out in Ethiopia were unable to identify the factors associated with SBIs in the mini-EDHS 2019. The current study aimed to investigate the spatial variation of the SBIs and its determinant factors among reproductive women in Ethiopia using a geographically weighted regression (GWR) model.

Abbreviations: CSA, Central Statistical Agency; SBI, Short birth interval; EAS, Enumeration areas; EMDHS, Mini-Ethiopian Demographic and Health Survey; OLS Ordinary least squares; GWR, Geographically weighted regression.

Method

Study area

The study was carried out in Ethiopia, which is located in the Horn of Africa. The country covers 1.1 million km^2 on its high central plateau, which reaches as low as 110 m in the Afar valley and as high as 4,550 m above sea level. There are two administrative cities, Addis Ababa and Dire Dawa, and nine regional states in the country: Tigray, Afar, Amhara, Oromia, Somalia, Benishangul Gumuz, Southern Nations, Nationalities, and People's Region (SNNPR), Gambella, and Harari. The regions are further divided into 74 administrative zones (see Figure 1) (12).

Study design, period, and setting

From 21 March to 28 June 2019, a cross-sectional survey based on the community was conducted. EAs were used as the sample units for the first stage of the stratified, two-stage cluster design used to choose the 2019 mini-EDHS sample, while households served as the sampling units for the second stage. The complete mini-EDHS report includes a detailed description of the sampling process.

Data source and study population

All Ethiopian women in the reproductive age range (15–49) were part of the source population. The study population consisted of women of reproductive age who had given birth within the 5 years before the survey.

Sample size and sampling procedure

The analysis used a weighted sample size of 4,793 people who had given birth within the recent year. To improve the representativeness of the sample data, weighted values were used. Participants were selected through a stratified, two-stage cluster sampling process. The DHS website provided access to each mini-EDHS report, which included a thorough sampling technique.[1]

Data collection tools and procedures

When the request for this study was made and the www.dhsprogram.com website was visited, the DHS Program provided the data. Data from the Ethiopian DHS were gathered using the stratified two-stage sampling method.

Variables

Outcome variable

The study's outcome variable was the SBIs of women. The outcome variable was binary, with 1 indicating an SBI for women and

1 https://dhsprogram.com/

0 indicating not. Finally, for spatial analysis, including spatial regression analysis, the weighted proportion of women's SBIs per EAs, a continuous variable, was used.

Independent variable

The potential explanatory factors used in the current study's exploratory regression were women's education level, wealth index, marital status, religion, survival status of the index child, breastfeeding status, contraceptive utilization, and type of residence.

Data management and analysis

SPSS version 26 and STATA version 16 are the software used to manage the data. We used multiple imputation methods for handling missing data in this study. These methods can account for the complex structure of the data and the relationships between variables while generating multiple imputed datasets to capture uncertainty. Additionally, robust statistical methods were used to handle outliers, ensuring that extreme values do not unduly influence the analysis results. Sample weighting was performed before further analysis, and the spatial analysis was performed using Arc GIS. Before conducting the spatial analysis, each predictor variable was expressed as binary, and after calculating the weighted proportions of the candidate predictor variables and SBIs (the outcome variable) for each enumeration area, ArcGIS version 10.8 was uploaded. Then, global ordinary least squares (OLS) linear regression was performed to model the outcome variable (SBIs) in terms of its relationships to a set of predictor variables. All of the OLS assumptions and diagnoses required by this method were thoroughly assessed and checked. The OLS summary report and diagnostic output tables provided the model summary diagnostics. The corrected Akaike Information Criterion (AIC), the coefficient of determination, the Joint F statistic, the Wald statistic, Koenker's Breusch–Pagan statistic, and the Jarque–Bera statistic are all included in the diagnostics for both. After the OLS assumptions and diagnosis were satisfied, the GWR analysis was then used to model the spatially varying relationships between the SBIs and the candidate predictor variables. Finally, the coefficients of the significantly associated predictor variables were mapped to determine which predictor contributed the greatest influence on the proportion of SBIs across the different geographical locations of Ethiopia.

Spatial autocorrelation

We used Arc GIS 10.8 to identify hot spot zones and perform spatial autocorrelation. The spatial autocorrelation measurement (Global Moran's I) was applied in Ethiopia to determine if SBIs were randomly distributed, clustered, or dispersed. Women's SBIs are spread when the Moran's I value is close to −1, clustered when it is close to +1, and randomly distributed when it is zero (13).

Hot spot analysis

The hot spot analysis was conducted using the proportion of SBIs in each administrative zone of Ethiopia as input. The features with either hot spot or cold spot values are revealed by the Hot Spot Analysis (Getis-Ord Gi* statistic) of the z-scores and significant *p*-values (14). A hot spot refers to the occurrence of a high proportion of SBIs that would be clustered together on the map, whereas a cold spot refers to the occurrence of women who have a low proportion of SBIs that would be clustered together on the map.

Associated factors of the short birth interval of women

Women are nested within a cluster in the EDHS data, and we estimate that these women in the same cluster will be more similar to one another than women across the country. This violates the equal variance across clusters and observation independence requirements of the standard regression model (1, 15). This suggests that the advanced GWR model must be used to account for cluster heterogeneity (16).

Ordinary least squares regression

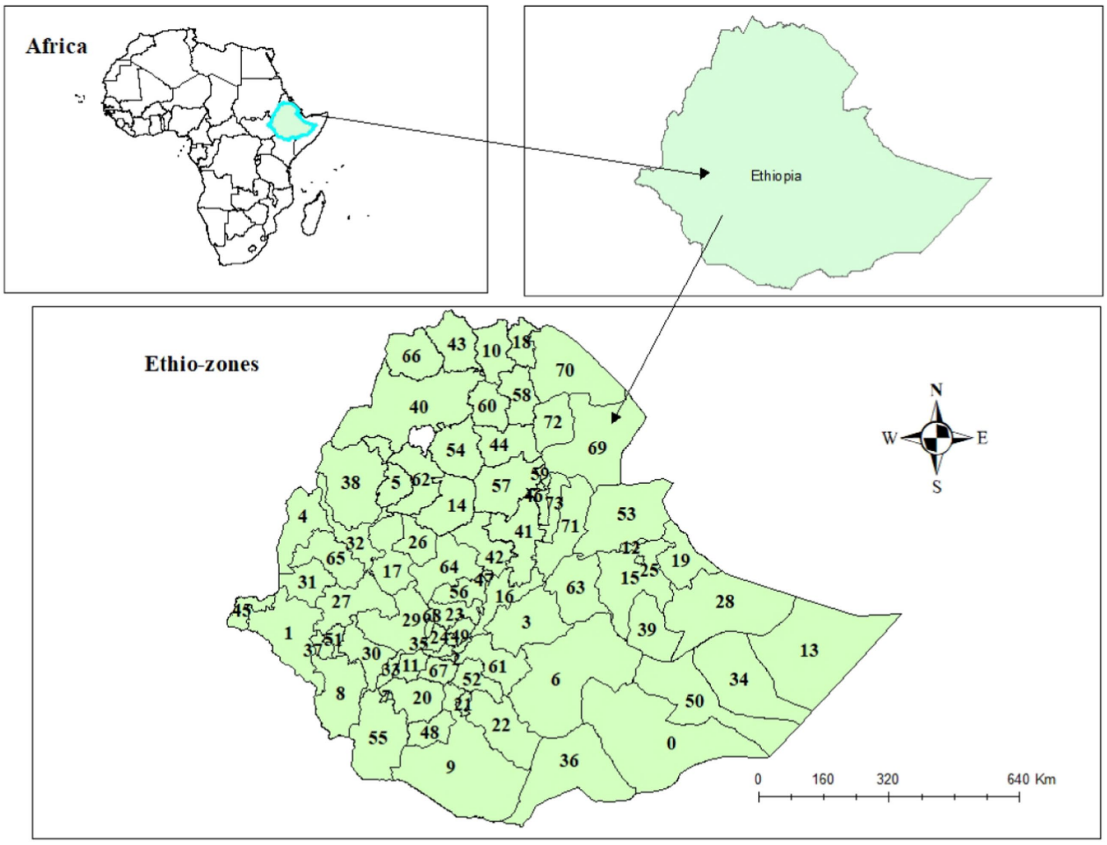
Geographical regression modeling was used to identify indicators of the observed geographical clustering of SBIs when the hot spot zones were identified. It was determined to fit an OLS regression first. The accuracy of the OLS regression results depends on the regression model satisfying all the criteria allowed by this approach. A well-stated OLS model should have a statistically significant coefficient of predictor variables with a positive or negative sign. Furthermore, there should not be any multicollinearity that is not a correlation between the predictor variables. The model should be unbiased (heteroscedasticity or non-stationarity). The residuals should be normally distributed and reveal no spatial patterns. The model should include important predictor variables. The residuals must be free from spatial autocorrelation (17). These assumptions were also checked appropriately. The OLS regression equation (18) is presented as:

$$Y_i = \beta + \sum_{K=1}^{P} \left(\beta_k X_{ik}\right) + \epsilon_i$$

where $i = 1, 2, \ldots n; \beta 0, \beta 1, \beta 2, \ldots \beta p$ are the model parameters, yi is the outcome variable for observation i, X_{ik} are explanatory variables and $_1, \epsilon_2, \ldots \epsilon_n$ are the error term/ residuals with zero mean and homogenous variance $\sigma 2$. Models with high adjusted R2 values are found by exploratory regression, which helps find a model that satisfies the OLS method's basis. Additionally, it finds models that satisfy all of the OLS method's assumptions (19).

Results

Table 1 presents the weighted proportions of SBIs among candidates based on sociodemographic and obstetric variables. The overall proportion of SBIs was 43.2% in this study. Of all respondents



Eth_Zone_2013

FID	ZONENAME	REGIONNAME	FID	ZONENAME	REGIONNAME	FID	ZONENAME	REGIONNAME	FID	ZONENAME	REGIONNAME
0	Afder	Somali	22	Guji	Oromia	44	North Wollo	Amhara	66	Western	Tigray
1	Agnuak	Gambela	23	Gurage	SNNPR	45	Nuer	Gambela	67	Wolayita	SNNPR
2	Alaba	SNNPR	24	Hadiya	SNNPR	46	Oromia	Amhara	68	Yem	SNNPR
3	Arsi	Oromia	25	Hareri	Hareri	47	Region 14	Addis Abab	69	Zone 1	Afar
4	Asosa	Beneshangul Gu	26	Horo Guduru	Oromia	48	Segen Peoples'	SNNPR	70	Zone 2	Afar
5	Awi/Agew	Amhara	27	Ilubabor	Oromia	49	Selti	SNNPR	71	Zone 3	Afar
6	Bale	Oromia	28	Jarar	Somali	50	Shabelle	Somali	72	Zone 4	Afar
7	Basketo	SNNPR	29	Jimma	Oromia	51	Sheka	SNNPR	73	Zone 5	Afar
8	Bench Maji	SNNPR	30	Keffa	SNNPR	52	Sidama	SNNPR			
9	Borena	Oromia	31	Kelem Wellega	Oromia	53	Siti	Somali			
10	Central	Tigray	32	Kemashi	Beneshangul	54	South Gonder	Amhara			
11	Dawro	SNNPR	33	Konta	SNNPR	55	South Omo	SNNPR			
12	Dire Dawa	Dire Dawa	34	Korahe	Somali	56	South West Shewa	Oromia			
13	Doolo	Somali	35	KT	SNNPR	57	South Wollo	Amhara			
14	East Gojam	Amhara	36	Liben	Somali	58	Southern	Tigray			
15	East Harerge	Oromia	37	Majang	Gambela	59	Special Woreda	Amhara			
16	East Shewa	Oromia	38	Metekel	Beneshangul	60	Wag Himra	Amhara			
17	East Wellega	Oromia	39	Nogob	Somali	61	West Arsi	Oromia			
18	Eastern	Tigray	40	North Gonder	Amhara	62	West Gojam	Amhara			
19	Fafan	Somali	41	North Shewa(R3)	Amhara	63	West Harerge	Oromia			
20	Gamo Gofa	SNNPR	42	North Shewa(R4)	Oromia	64	West Shewa	Oromia			
21	Gedio	SNNPR	43	North Western	Tigray	65	West Wellega	Oromia			

FIGURE 1
Area map of Ethiopia with 74 administrative zones.

who have SBIs, 1,569 (75.8%) were observed from rural areas, and only 500 (24.5%) were observed from urban areas. Approximately 600 (29%) SBIs were observed among orthodox religious followers. Similarly, among women who had SBIs, 948 (45.8%), 414 (20%), and 707 (34.2%) were from poor, middle, and rich wealth households, respectively. The majority of women who had SBIs were married

TABLE 1 Association of sociodemographic and obstetric characteristics of women with a short birth interval (SBI) in Ethiopia, MEDHS.

Variable	Category	Weighted frequencies	Short birth interval (SBI)	Non-short birth interval	*p*-value
Residence	Rural	6,024 (67.8%)	1,569 (75.8%)	1,972 (72.4%)	0.007
	Urban	2,861 (32.2)	500 (24.5%)	752 (27.6%)	
Education level	No education	3,589 (40.4%)	1,351 (65.3%)	1,633 (59.9%)	<0.001
	Primary	3,701 (41.7%)	603 (29.1%)	868 (31.9%)	
	Secondary	1,088 (12.2%)	78 (3.8%)	163 (6%)	
	Higher	507 (5.7%)	37 (1.8%)	60 (2.2%)	
Religion	Orthodox	3,685 (41.5%)	600 (29%)	1,192 (43.8%)	<0.001
	Catholic	47 (0.5%)	12 (0.6%)	11 (0.4%)	
	Protestant	2,435 (27.4%)	594 (28.7%)	771 (28.3%)	
	Muslim	2,619 (29.5%)	814 (39.3%)	732 (26.9%)	
	Others	98 (1.1%)	49 (2.4%)	17 (0.6%)	
Wealth index	Poor	3,052 (34.3%)	948 (45.8%)	973 (35.7%)	<0.001
	Middle	1,671 (18.8%)	414 (20%)	559 (20.5%)	
	Rich	4,162 (46.8%)	707 (34.2%)	1,192 (43.8%)	
Survival status of the previous child	No	282 (4.8%)	136 (6.6%)	94 (3.5%)	<0.001
	Yes	5,573 (95.2%)	1,933 (93.4%)	2,629 (96.5%)	
Current contraceptive method	Not using	6,329 (71.2%)	1,373 (66.4%)	1,618 (59.4%)	<0.001
	Using	2,556 (28.8%)	696 (33.6%)	1,105 (40.6%)	
Breastfeeding status	No	1,092 (32.4%)	400 (36.5%)	454 (29.9%)	<0.001
	Yes	2,275 (67.6%)	696 (63.5%)	1,066 (70.1%)	
Marital status	Single	2,580 (29%)	49 (2.4%)	68 (2.5%)	0.003
	Married	5,743 (64.6%)	1,845 (89.2%)	2,447 (89.9%)	
	Widowed	185 (2.1%)	94 (4.5%)	80 (2.9%)	
	Divorced	377 (4.2%)	80 (3.9%)	128 (4.7%)	

women. Likewise, the majority of women who had SBIs also did not use contraceptive methods. In general, the proportion of SBIs decreases as the education level increases.

Spatial autocorrelation of SBIs of women in Ethiopia

The SBIs were spatially clustered in Ethiopian administrative zones with Global Moran's $I = 0.773$ and $p < 0.001$ (Figure 2). The area of study had significant rates of SBIs, as indicated by the clustered patterns on the right sides. With a Z-score of 16.84, it is less than 1% that this clustered pattern is a result of random chance.

Spatial distribution of short birth interval

Figure 3 shows the spatial distribution of the observed proportion of SBIs in Ethiopia across each enumeration area in EMDHS 2019, where high and low proportions of SBIs were represented by red and green colors, respectively.

Hot spot (Getis-Ord Gi*) analysis of short birth interval

Figure 4 displays the hot spot areas of SBIs of reproductive women in Ethiopia. There are significant hot and cold spots in SBIs in Ethiopia, according to the Local Getis-Ord Gi* statistics. A point with a red color indicates significant hot spot areas of SBIs. Accordingly, Jarar, Doolo, Shabelle, Afder, Liben, Korahe, Nogob, West Harerge, Guji, Sidama, and Assosa were the hot spot areas of SBIs. Whereas the cold spot areas were Fafan, East Harerge, Ilubabur, Bench Maji, Keffa, Agnuak, and Nuer zones (Figure 4).

Spatial determinates of short birth interval in the MEDHS data

First of all, the candidate predictor variables were fitted to the OLS regression model. This model satisfied every OLS assumption. The joint Wald statistics showed significance for the whole model ($p < 0.001$), and the predictor variables' coefficient significance was shown by robust probabilities ($p < 0.001$). Additionally,

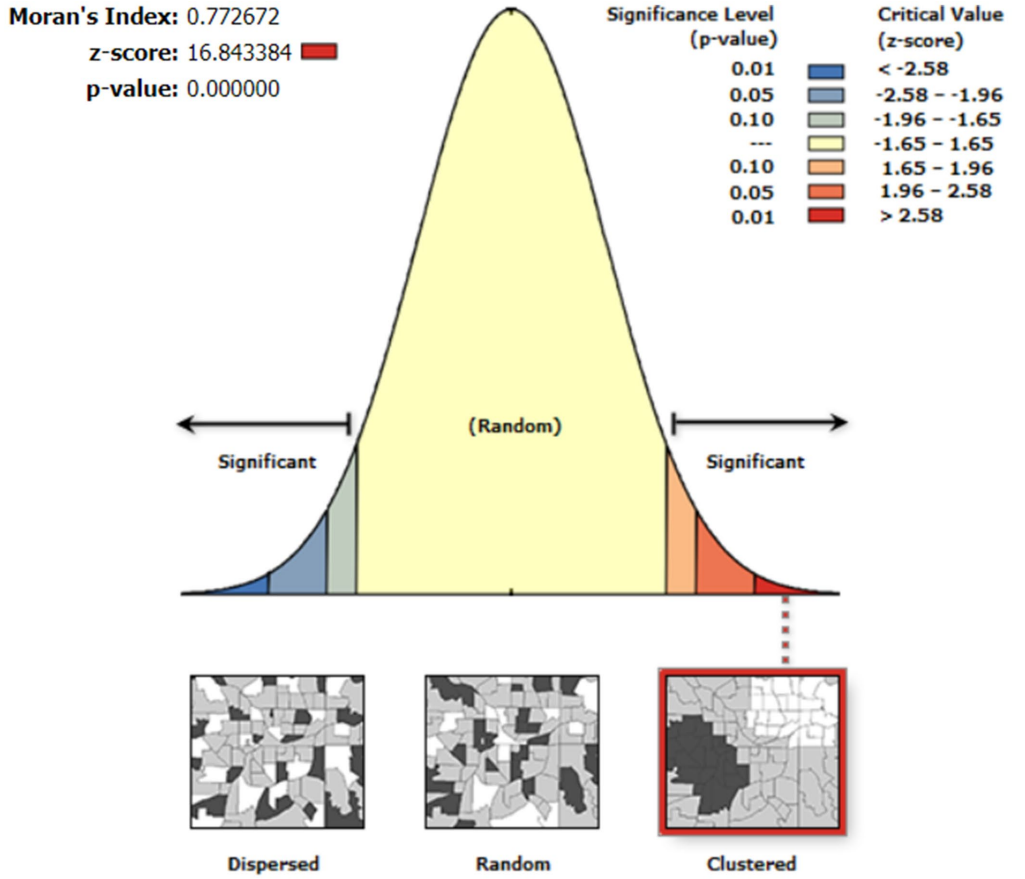


FIGURE 2
Spatial autocorrelation of short birth interval (SBI) in Ethiopia, EMDHS 2019.

multicollinearity was assessed using the coefficient of variance inflation factor, and there were no redundancy issues in the predictor variables (VIF < 7.5). The adjusted R2 = 0.43 indicates that 43% of the variation in SBIs was explained by the model. Therefore, the spatial determinates of hot spot areas of SBIs were illiterate women, women who had not breastfed, and orthodox followers (Table 2).

Considering the fact that OLS analysis identified SBI hot spot area predictors, it makes the assumption that, across the study area, the relationships between each predictor variable and SBIs are stationary. However, this assumption is violated by the significant Koenker (BP) statistic ($p < 0.001$). This problem is better managed by a GWR model or a local model if stationarity is violated. A GWR model was therefore fitted to have a reliable estimator. In this model, the adjusted R2 value obtained from OLS increased from 0.43 in Table 2 to 0.46 using GWR (Table 3). This was further supported and taken into account by the modified AIC, for which the GWR provided a smaller value (AICc = −108.26; Table 3) compared to the OLS model (AICc = −95.303; Table 2). Thus, the GWR model was determined to be better.

OLS diagnostics

Number of observations	305	Akaike's information criterion (AICc)	−95.303
Multiple R-Squared	0.450085	Adjusted R-squared	0.430140
Joint F-Statistic	23.9490890	Prob(> F),(11,293) degrees of freedom:	0.000000*
Joint Wald Statistic	673.47665995	Prob(> chi-squared), (11) degrees of freedom:	0.000000 *
Koenker (BP) Statistic	59.823809	Prob(> chi-squared), (11) degrees of freedom:	0.000000 *
Jarque–Bera Statistic	18.163598	Prob(> chi-squared), (2) degrees of freedom:	0.000103*

Figure 5 shows the geographically weighted standardized residual of SBIs, which is the difference between the predicted and observed proportion of SBIs in the study area. The areas with red points indicate EAs where there is a relatively big difference between the observed and

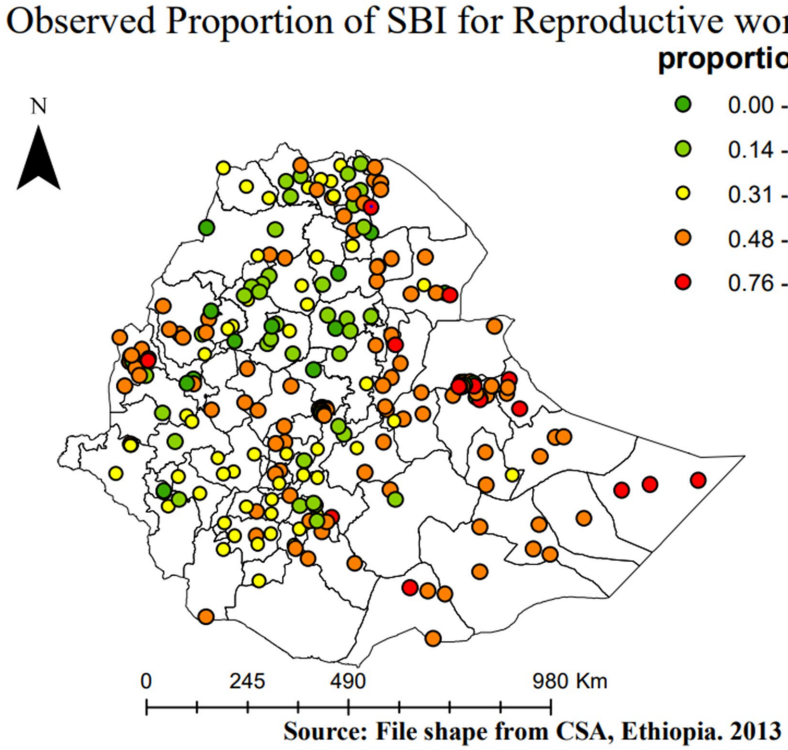


FIGURE 3
Spatial distribution of SBI across Ethiopian administrative zones.

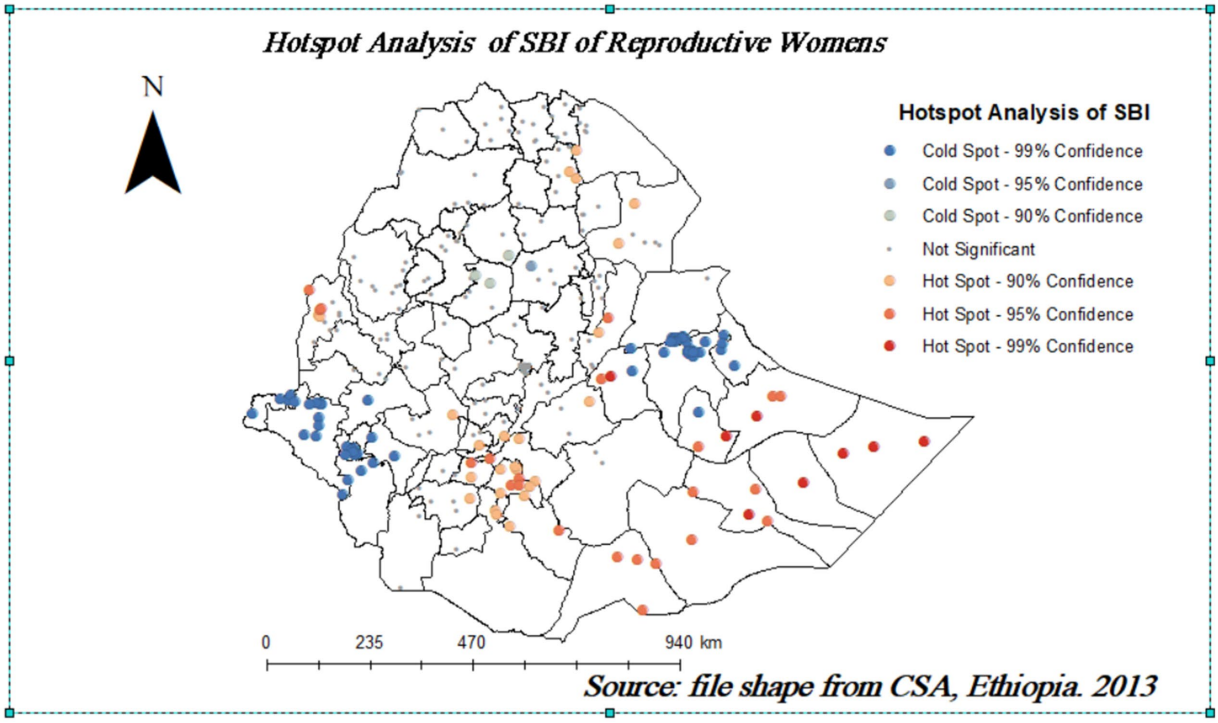


FIGURE 4
Hot spot analysis of SBI in Ethiopia's administrative zones.

TABLE 2 Summary of spatial OLS results of SBI in Ethiopia, mini-Ethiopia demographic, and health survey (EMDHS) 2019.

Variable	Coefficient	Std Error	t-Statistic	Probability	Robus (SE)	Robust_t statistics	Robust_ Probability	VIF
Intercept	0.052520	0.092744	0.566292	0.571635	0.071251	0.737116	0.461634	–
PILLTERATE	0.265947	0.054362	4.892121	0.000002*	0.073742	3.606462	0.000376*	1.993601
PNOTBRE	0.420776	0.050933	8.261432	0.000000*	0.062438	6.739129	0.000000*	1.157890
PORTHO	−0.078091	0.029927	−2.609384	0.009530*	0.026500	−2.946882	0.003474*	1.216877

*Significant at 5% level of significance.

TABLE 3 Geographically weighted regression (GWR) of SBI in Ethiopia, EMDHS 2019.

Predictor variables	Illiterate women, women who do not breastfeed, orthodox follower women
Effective number	25.005
Sigma	0.1976
Akaike information criteria (AICc)	−108.26
Residual square	10.936
Multiple R-square	0.4992
Adjusted R-square	0.4562

predicted (over-predicted) proportions of SBIs. Whereas the green points represent areas where SBIs were under-predicted.

Figures 6–8 provide a list of areas in Ethiopia where the candidate predictor variables had a significant impact on SBIs. For instance, uneducated women had a positive relationship with SBIs, as the intervals of the GWR coefficients of SBIs (Figure 6) are all positive and range from 0 to 1. Thus, as the proportion of uneducated women increases, the proportion of SBIs also increases in a given enumeration area. EAs shaded in red (Figure 6) represent areas where SBIs were highly affected by uneducated women. This is observed in all zones of the Somali region, east Harerge, zone 1, zone 2, zone 3, and zone 4 in the Afar region, northwestern in Tigray, Metekel in the Benishangul-Gumz region, and North Gonder in the Amhara region (Figure 6).

However, the GWR coefficients (Figure 7) indicate the proportion of women in Ethiopia who had no breastfeeding practice, and SBIs have a positive relationship. The red-shaded points represent no breastfeeding practice as a strong positive predictor of SBIs, which ranges from 0 to 1. Thus, as the proportion of women with no breastfeeding practice increases, the proportion of SBIs also increases in a given enumeration area. This circumstance was observed somewhat around all zones of the Somali region, zone 1, zone 2, zone 3, and zone 4 in the Afar region, north Shewa in the Amhara region, central and western Tigray, Guji in the Oromia region, Hadiya in SNNPR, and Asossa (Benishangul-Gumz) region (Figure 7).

The other important spatial predictor of SBIs is the religion (being orthodox) of women, which has a positive relation with the percentage of SBIs (positive GWR regression coefficients). Thus, as the proportion of orthodox women in a given EA increases, the proportion of SBIs increases. Areas shaded in red indicate a high increase in SBI as the proportion of orthodox women increases. This is observed in all zones of Tigray, North and South Gonder, North Wollo, Waghimra, West and East Gojam, North Shewa, Awi (Agew) in the Amhara region, and East Shewa in the Oromia region (Figure 8).

Discussion

Using several spatial analytical techniques, this study explored the spatial clustering and spatial causes of Ethiopian women's SBIs. This study aimed to evaluate the spatial distribution and spatial factors of SBIs in Ethiopia using GWR. The descriptive analysis concluded that over 43.2% of women had SBIs. SBIs were clustered (Moran's $I = 0.773$ and $p < 0.001$) in Ethiopia. This result was in line with the results of several studies from Ethiopia (2, 10, 20) and Bangladesh (21). SBI hotspots were also found in Jarar, Doolo, Shabelle, Afder, Liben, Korahe, Nogob, West Harerge, Guji, Sidama, and Assosa zones of Ethiopia. This claim is consistent with the findings of a study conducted in Ethiopia (22).

This study also found a relatively high prevalence of SBIs in the Afar region and Somali region, where a majority of SBI hot spots are located. These divisional variations in SBI hot spots are due to the division-level variations in sociodemographic and cultural characteristics of women and their partners and their perceptions regarding the desired number of children. Women who had no education, women who did not breastfeed, and women who were orthodox followers were factors in SBI hot spots. The three powerful predictors of SBI in Ethiopia's administrative zones, according to their GWR coefficients. All three strong predictors range from 0 to 1 for women. This finding is in line with the study conducted in Ethiopia (9). In addition to the factors found in this study, the observed geographic variability of SBIs was associated with women who did not breastfeed in the Somali and Afar regions. It has long been recognized that women who breastfeed their children for a longer duration have a longer birth interval. This concept is confirmed by earlier research findings for hot spots for women who did not breastfeed in these regions. This could be caused by many things, such as gaps in media accessibility, knowledge of the best time to begin pregnancy, and the accessibility of medical resources (23). Additionally, variations in socioeconomic, cultural, and demographic aspects among those administrative zones may account for the variation in SBI hot spots. The majority of people in the Afar and Somali regions are pastoralists, who live a lifestyle characterized by seasonal migration (20). Access to health services and information is restricted for residents in certain places. Pastoralist communities are portable, but they also live in highly traditional environments and strongly follow cultural and religious norms (24). For example, research has indicated that women living in pastoralist societies have lower rates of high breastfeeding, which in turn affects the gap between pregnancies (24).

The GWR analysis showed a positive relationship between SBIs and women who had no breastfeeding status this study is consistent with a study by Regasa in Ethiopia (2). The current

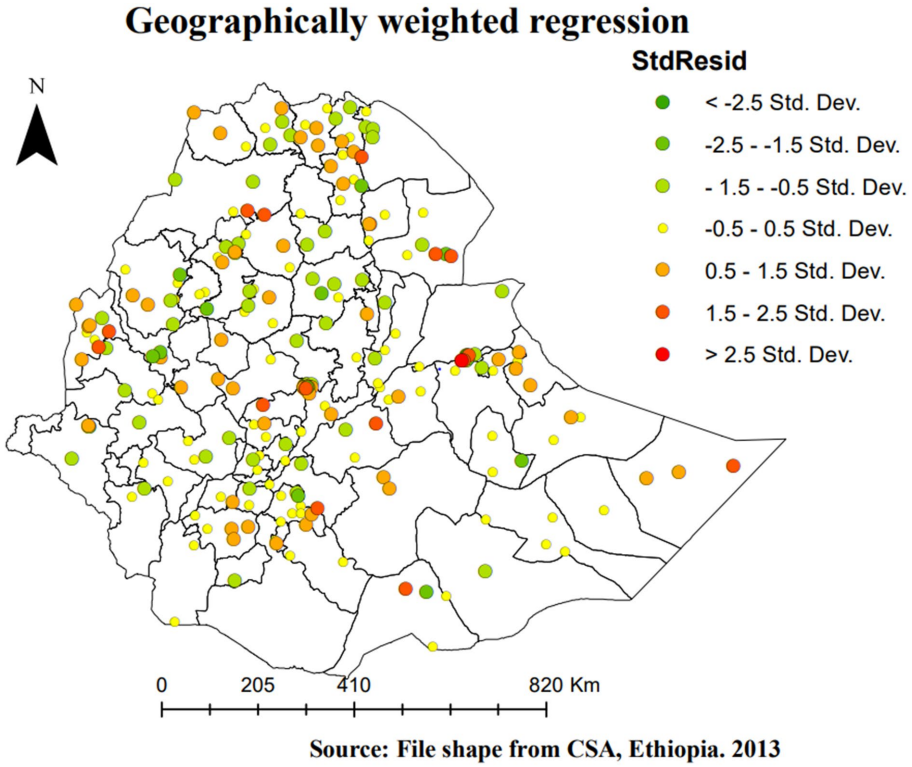


FIGURE 5
Geographically weighted regression StdResid of SBI in Ethiopia, EMDHS 2019.

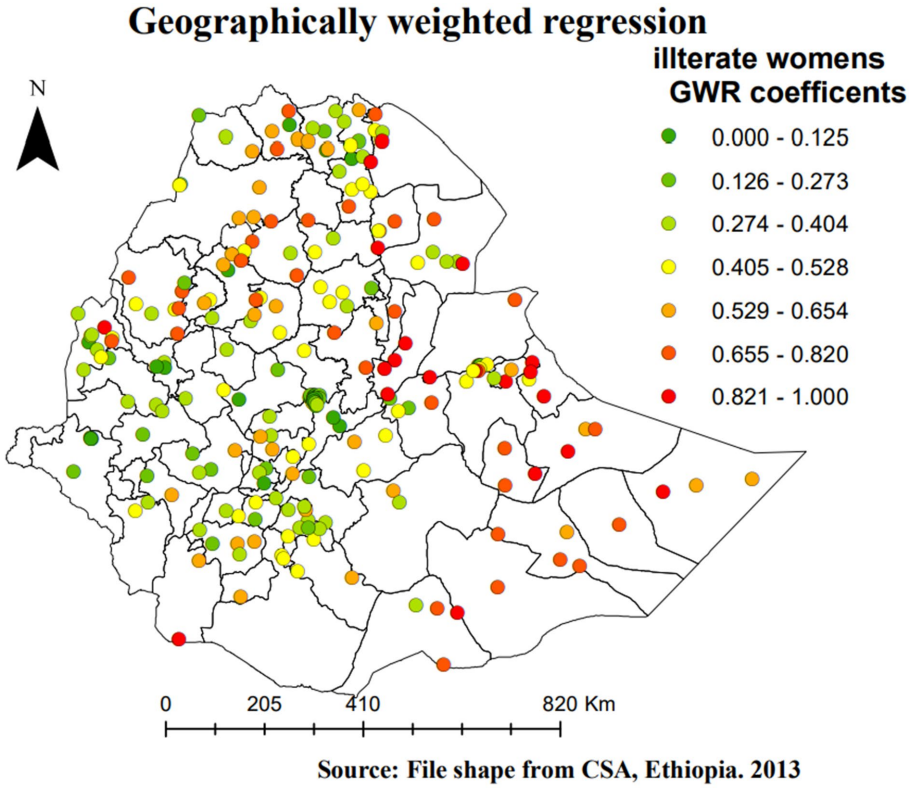


FIGURE 6
Geographically weighted regression coefficients c of illiterate women as a predictor of SBI in Ethiopia, EMDHS 2019.

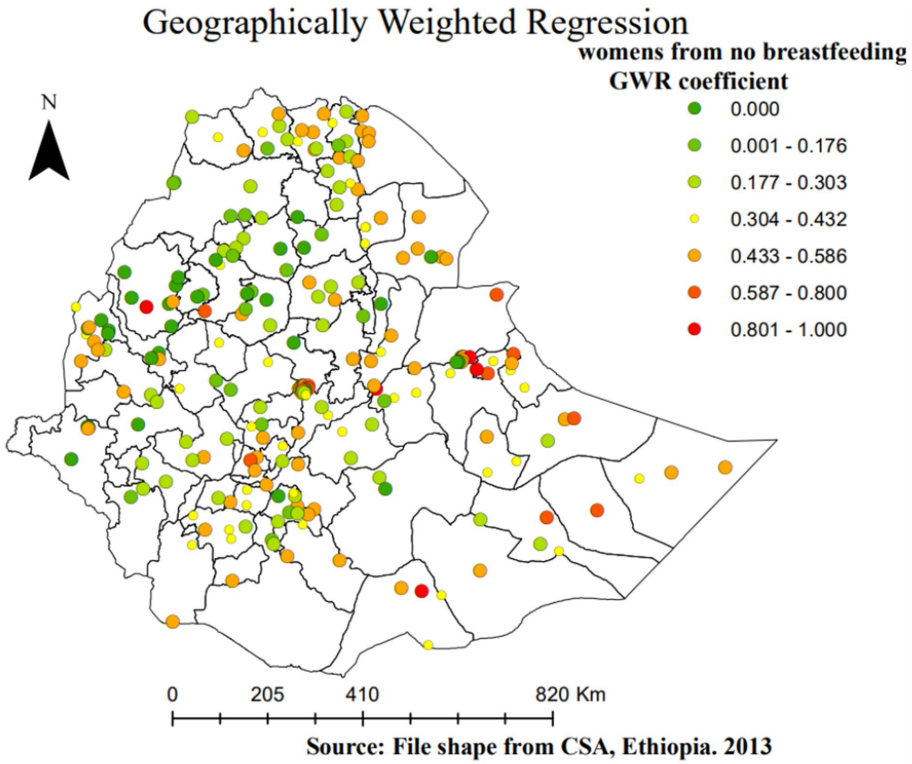


FIGURE 7
Geographically weighted regression coefficients of women who have no breastfeeding practice as a predictor of SBI in Ethiopia, EMDHS 2019.

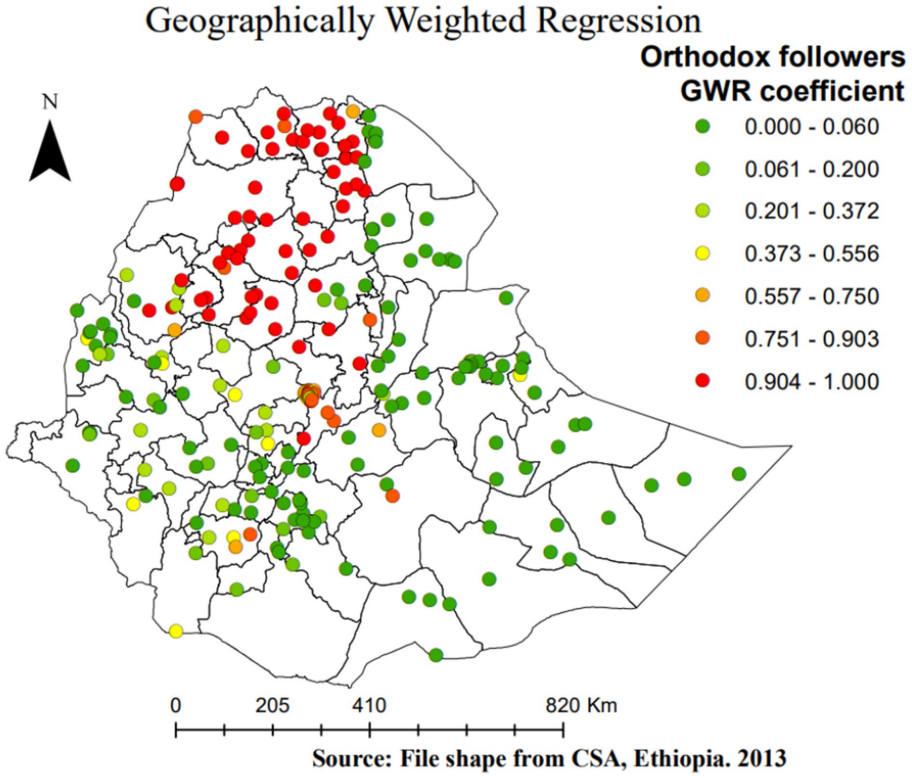


FIGURE 8
Geographically weighted regression coefficients of orthodox follower women as a predictor of SBI in Ethiopia, EMDHS 2019.

study also identified a spatially positive association between orthodox religion followers and SBIs among mothers in Ethiopia. This could be due to the non-use of contraceptive methods among orthodox followers, which is similar to the findings in Ethiopia. In any case, a few Muslims and Orthodox perceive it as a concept that is completely against the principles of Christianity and Islam. Participants who had never utilized contraceptive methods prior to the conception of the last child were more likely to be at risk of having SBIs compared to their counterparts (25, 26). There is some evidence to suggest that religious beliefs and values may influence individuals' attitudes and behaviors regarding contraception and SBIs. For example, some religious teachings discourage the use of contraception, which may contribute to higher SBIs. However, it is important to recognize that there is a great deal of diversity within religious communities, and not all individuals or groups within a particular religion share the same beliefs and practices (27).

Not attending formal education by women had a positive relationship with SBIs. As the proportion of women who had not attended formal education increased, the occurrence of SBIs in the zones of the Somali region, zones of the Afar region, and some zones of the Tigray region increased (28). Additionally, women who did not pursue formal education were positively correlated with hot spots of SBIs in the far east of the Harerge zone in the Somali region. According to the published research, the Somali region has the largest percentage of uneducated women (75.0%), whereas Addis Ababa has the lowest percentage (9.0%) (12). By influencing women's healthcare-seeking behaviors, such as their use of family planning services, education may have an impact on the spacing between pregnancies. Higher-educated women are also more likely to be more aware of health issues, to know more about the available services, to be better able to pay for medical care, and to have more autonomy when it comes to making decisions about their health, including family planning (29, 30).

In this investigation, it was found that different predictors of SBIs (women with no education, women with no breastfeeding status, and orthodox religion-following women) act more/less strongly across administrative zones. For example, women who did not attend formal education and women who did not breastfeed had a strong positive relationship with SBI hot spots in the Somali region, while they had a weak positive relationship with SBIs in the Amhara and Tigray regions. This may be due to differences in the accessibility and availability of various family planning services, as well as differences in socioeconomic status and culture throughout the nation's administrative areas.

Strengths and limitations

This study used a GWR model and the EDHS national datasets to examine the spatial effects of SBIs to identify most hot spot areas, which is considered a strength. The retrospective data collection procedure and the removal of several important variables that affect SBIs from the mini-2019 EDHS datasets resulted in bias in this analysis, which is considered a limitation of this study. In addition, the mini-EDHS data have certain limitations, especially regarding variable selection and biases inherent in self-reported data. While the mini-DHS data cover key demographic and health indicators, the primary limitation is the restricted set of variables collected. It may lack certain variables that could provide a more comprehensive understanding of the population under study. Second, respondents may provide inaccurate information due to memory lapses, social desirability bias, or misunderstanding of the questions. This can lead to inaccuracies and affect the reliability of the data.

Conclusion

Overall, the study revealed geographic variations in Ethiopian women's SBIs, with hot spots of SBIs identified in specific zones. Hot spots with SBIs were found in Jarar, Doolo, Shabelle, Afder, Liben, Korahe, Nogob, West Harerge, Guji, Sidama, and Assosa zones. Factors such as low education levels, lack of breastfeeding practices, and orthodox religious followers were associated with an increased risk of SBIs in these areas. Decision-makers at the national and regional levels are encouraged to prioritize the provision of family planning services in these hot spot communities. Engaging religious leaders is an effective strategy for promoting modern family planning. Community-based programs focusing on women's education and empowerment can also play a significant role in reducing the impact of SBIs. Additionally, future research should consider integrating new data sources, such as socioeconomic status, access to healthcare, cultural norms, and community-level characteristics, to gain a more comprehensive understanding of the determinants of SBIs. Moreover, future studies should consider integrating a mixed-methods approach, combining quantitative and qualitative data, to capture the complex interplay of sociocultural factors influencing SBIs.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the studies involving humans because the data that we used is secondary. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements because the data that we used is secondary.

Author contributions

GA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AM: Methodology, Supervision, Validation, Writing – review & editing. NM: Methodology, Writing – review & editing. SM:

Methodology, Writing – review & editing. HB: Methodology, Writing – review & editing. KA: Methodology, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors are grateful to the DHS Program for their permission to access the DHS dataset for further analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Shifti DM, Chojenta C, Holliday EG, Loxton D. Individual and community level determinants of short birth interval in Ethiopia: a multilevel analysis. *PLoS One*. (2020) 15:e0227798. doi: 10.1371/journal.pone.0227798

2. Regasa Z, Tilahun BC, Mesfin A. Spatial distribution and determinant factors of birth interval among reproductive age group women, based on Edhs 2016, Ethiopia. (2019). [Preprint].

3. Exavery A, Mrema S, Shamte A, Bietsch K, Mosha D, Mbaruku G. Levels and correlates of non-adherence to WHO recommended inter-birth intervals in Rufiji, Tanzania. *BMC Pregnancy Childbirth*. (2012) 12:152. doi: 10.1186/1471-2393-12-152

4. Molitoris J, Barclay K, Kolk M. When and where birth spacing matters for child survival: an international comparison using the DHS. *Demography*. (2019) 56:1349–70. doi: 10.1007/s13524-019-00798-y

5. Seid YM. Innovations in data dissemination at the central statistical Agency of Ethiopia. *IASSIST Q*. (2010) 33:26. doi: 10.29173/iq380

6. Tigabu S, Demelew T, Seid A, Sime B, Manyazewal T. Socioeconomic and religious differentials in contraceptive uptake in western Ethiopia: a mixed-methods phenomenological study. *BMC Womens Health*. (2018) 18:85. doi: 10.1186/s12905-018-0580-6

7. Ayele DG. Determinants of fertility in Ethiopia. *Afr Health Sci*. (2015) 15:546–51. doi: 10.4314/ahs.v15i2.29

8. Kelel HU. Short birth interval and its determinants of reproductive-age women in Ethiopia: Multilevel regression model. Diss. kamuzu university of health sciences. (2020).

9. Shifti DM, Chojenta C, Holliday EG, Loxton D. Application of geographically weighted regression analysis to assess predictors of short birth interval hot spots in Ethiopia. *PLoS ONE*. (2020) 15:e0233790. doi: 10.1371/journal.pone.0233790

10. Efendi A, Ramadhan HW. Parameter estimation of multinomial logistic regression model using least absolute shrinkage and selection operator (LASSO). AIP Conference Proceedings. (2018);2021.

11. Hasinur Rahaman Khan M, Shaw JEH. Multilevel logistic regression analysis applied to binary contraceptive prevalence data. *J Data Sci*. (2021) 9:93–110. doi: 10.6339/JDS.201101_09(1).0008

12. Ethiopia I and Central Stastical A (CSA). Central Statistical Agency (CSA) [Ethiopia] and ICF. Ethiopia demographic and health survey. [internet]. Addis Ababa, Rockvile, Maryland (2016). 134. Available at: www.DHSprogram.com.

13. Dubé J, Legros D. Spatial autocorrelation In: Spatial Econometrics Using Microdata (2014). 59–91.

14. Liyew AM, Kassie A, Teshale AB, Alem AZ, Yeshaw Y, Tesema GA. Exploring spatiotemporal distribution of under-five mortality in Ethiopia: further analysis of Ethiopian demographic and health surveys 2000, 2005, 2011 and 2016. *BMJ Paediatr Open*. (2021) 5:e001047. doi: 10.1136/bmjpo-2021-001047

15. Viktorov VA, Varin AN, Grinval'd VM, Maksimov EP, Fomicheva NN, Zavalishin YK, et al. Current state and prospects for development of domestic equipment for hemodialysis with dialyzate regeneration. *Biomed Eng (NY)*. (2003) 37:16–21. doi: 10.1023/A:1023777427647

16. Tigabu S, Liyew AM, Geremew BM. Modeling spatial determinates of teenage pregnancy in Ethiopia; geographically weighted regression. *BMC Womens Health*. (2021) 21:254. doi: 10.1186/s12905-021-01400-7

17. Warsito B, Yasin H, Ispriyanti D, Hoyyi A. Robust geographically weighted regression of modeling the air polluter standard index (APSI). *J Phys Conf Ser*. (2018) 1025:012096. doi: 10.1088/1742-6596/1025/1/012096

18. Fotheringham AS. Geographically weighted regression white paper. (2009).

19. Ma Z, Xue Y, Hu G. Geographically weighted regression analysis for spatial economics data: a Bayesian recourse. *Int Reg Sci Rev*. (2021) 44:582–604. doi: 10.1177/0160017620959823

20. Ahmed KY, Page A, Arora A, Ogbo FA. Trends and determinants of early initiation of breastfeeding and exclusive breastfeeding in Ethiopia from 2000 to 2016. *Int Breastfeed J*. (2019) 14:1–14. doi: 10.1186/s13006-019-0234-9

21. Shifti DM, Chojenta C, Holliday E, Loxton D. Effects of short birth interval on neonatal, infant and under-five child mortality in Ethiopia: a nationally representative observational study using inverse probability of treatment weighting. *BMJ Open*. (2021) 11:e047892. doi: 10.1136/bmjopen-2020-047892

22. Arega GG, Mitku AA, Fenta HM. Spatio-temporal pattern of short birth interval and associated factors on women in Ethiopia: using Ethiopian demographic and health surveys 2000–2016. *Front Med*. (2023) 10:1131794. doi: 10.3389/fmed.2023.1131794

23. Hailegebreal S, Haile Y, Seboka BT, Enyew EB, Shibiru T, Mekonnen ZA, et al. Modeling spatial determinants of initiation of breastfeeding in Ethiopia: a geographically weighted regression analysis. *PLoS ONE*. (2022) 17:e0273793. doi: 10.1371/journal.pone.0273793

24. Henok A, Takele E. Assessment of barriers to reproductive health service utilization among bench Maji zone pastoralist communities. *Ethiop J Health Sci*. (2017) 27:523–30. doi: 10.4314/ejhs.v27i5.11

25. Kassie SY, Ngusie HS, Demsash AW, Alene TD. Spatial distribution of short birth interval and associated factors among reproductive age women in Ethiopia: spatial and multilevel analysis of 2019 Ethiopian mini demographic and health survey. *BMC Pregnancy Childbirth*. (2023) 23, 23:275. doi: 10.1186/s12884-023-05610-9

26. Gebrehiwot SW, Abera G, Tesfay K, Tilahun W. Short birth interval and associated factors among women of child bearing age in northern Ethiopia, 2016. *BMC Womens Health*. (2019) 19:85. doi: 10.1186/s12905-019-0776-4

27. Srikanthan A, Reid RL. Religious and cultural influences on contraception. *J Obstet Gynaecol Can*. (2008) 30:129–37. doi: 10.1016/S1701-2163(16)32736-0

28. Haque U, Scott LM, Hashizume M, Fisher E, Haque R, Yamamoto T, et al. Modelling malaria treatment practices in Bangladesh using spatial statistics. *Malar J*. (2012) 11:1–11. doi: 10.1186/1475-2875-11-63

29. Raghupathy S. Education and the use of maternal health care in Thailand. *Soc Sci Med*. (1996) 43:459–71. doi: 10.1016/0277-9536(95)00411-4

30. Celik Y, Hotchkiss DR. The socio-economic determinants of maternal health care utilization in Turkey. *Soc Sci Med*. (2000) 50:1797–806. doi: 10.1016/S0277-9536(99)00418-9

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 04 September 2024
DOI 10.3389/fmed.2024.1345316

OPEN ACCESS

EDITED BY
Kathryn Toy Knecht,
Loma Linda University, United States

REVIEWED BY
Daisung Jang,
University of Melbourne, Australia
Sipho Mkhize,
University of KwaZulu-Natal, South Africa

*CORRESPONDENCE
Romana F. Malik
✉ romana_malik@hotmail.com

RECEIVED 14 December 2023
ACCEPTED 06 August 2024
PUBLISHED 04 September 2024

CITATION
Malik RF, Azar P, Taimounti A, Buljac-Samardžić M, Hilders CGJM and Scheele F (2024) How do cultural elements shape speak-up behavior beyond the patient safety context? An interprofessional perspective in an obstetrics and gynecology department.
Front. Med. 11:1345316.
doi: 10.3389/fmed.2024.1345316

How do cultural elements shape speak-up behavior beyond the patient safety context? An interprofessional perspective in an obstetrics and gynecology department

Romana F. Malik[1,2]*, Poyan Azar[3], Achraf Taimounti[4], Martina Buljac-Samardžić[5], Carina G. J. M. Hilders[5,6] and Fedde Scheele[2,7]

[1]Department of Research in Education, OLVG Hospital, Amsterdam, Netherlands, [2]Athena Institute, Faculty of Science, VU Amsterdam, Amsterdam, Netherlands, [3]Department of Human Resources, Bunge, Zaandam, Netherlands, [4]Faculty of Behavioral and Movement Sciences, VU Amsterdam, Amsterdam, Netherlands, [5]Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, Netherlands, [6]Reinier de Graaf Hospital, Delft, Netherlands, [7]Department of Research in Education, Amsterdam University Medical Centre, Amsterdam, Netherlands

Introduction: Interprofessional working and learning thrives with speak-up behavior. Efforts to improve speak-up have mainly focused on isolated techniques and training programs within the patient safety scope, yet sustained improvement requires a cultural shift beyond this scope. This research investigates the influence of culture elements on speak-up behavior in interprofessional teams beyond the patient safety context.

Methods: An exploratory qualitative study design was used in a Dutch hospital's Obstetrics and Gynecology department. A representative sample of stakeholders was purposefully selected, resulting in semi-structured interviews with 13 professionals from different professional backgrounds (nurses, midwifes, managers, medical specialists, and residents). A speak-up pledge was developed by the research team and used to prime participants for discussion. Data analysis involved three-step coding, which led to the development of themes.

Results: This study has identified six primary cultural themes that enhance speak-up behavior. These themes encompass the importance of managing a shared vision, the role of functional hierarchy, the significance of robust interpersonal relationships, the formulation of a strategy delineating when to speak up and when to exercise restraint, the promotion of an open-minded professional mindset, and the integration of cultural practices in the context of interprofessional working and learning.

Conclusion: Six crucial cultural elements have been pinpointed to boost the practice of speaking up behavior in interprofessional working and learning. Remarkably, hierarchy should not be held responsible as the wrongdoer; instead, can be a great facilitator through respect and appreciation. We propose that employing transformational and humble leadership styles can provide guidance on effectively integrating the identified cultural elements into the workplace and provide an IMOI framework for effective interprofessional speak-up beyond patient safety.

KEYWORDS

organizational culture, speak-up, interprofessional education, leadership, hierarchy, postgraduate training, healthcare professionals, communication

Introduction

Healthcare professionals are increasingly highlighting systemic shortcomings within the medical culture (1, 2). *Current efforts to improve communication such as 'speak-up', often focus on techniques and isolated training programs within the patient safety context, whereas sustained improvement requires a culture shift* (1, 3, 4). *Attempts to foster such cultures require medical staff to change their habitual patterns, which in practice is a difficult task, even when speaking up is explicitly encouraged* (5, 6). *In order to create a sustainable set of consistent practices for interprofessional working and learning, it is necessary to understand the underlying cultural elements that influence speak-up beyond the patient safety context.*

In literature, speak-up usually refers to healthcare professionals raising concerns to draw attention to behavior or actions that pose a genuine risk to patient safety (7). The existing speak-up literature already provides valuable insights on the personal factors that can hinder or promote speak-up in the patient safety context (8). When healthcare professionals believe that speaking up will result in meaningful change to patient safety, they are more likely to do so (9). Conversely, they are less likely to speak up when patient safety is not at risk, even if not speaking up would come at the expense of personal (e.g., stress) or organizational interests, such as hindered interprofessional working and learning (6, 8, 10–12). Previous research has identified organizational culture, personality traits, and their interactions as important factors influencing the likelihood of speaking up within the patient safety context (6, 8, 10, 13). For example, a culture that values openness and encourages feedback is more likely to promote speak-up, while personality traits such as introversion may make individuals less likely to speak up in certain situations (14). Although there has been considerable research on speak-up and its benefits, the specific cultural elements important for speak-up beyond the patient safety scope within the context of interprofessional working and learning has received less attention.

Promoting a culture of open communication within interprofessional healthcare teams is of paramount importance, given that (1) traditional practices and the isolated acquisition of knowledge and skills are no longer effective amidst the rising challenges of multimorbidity and aging patients, and (2) there is a growing emphasis on improving the patient experience, ensuring the well-being of healthcare professionals, and optimizing the overall performance of the healthcare system (15–18). Healthcare providers must work collaboratively across various professions to thrive in this ever-evolving landscape, not only in terms of content (e.g., diverse protocols across various specialties) but also in terms of relational aspects, enabling speak-up, deliberation, and continuous improvement and learning.

Simulation training aimed at promoting speak-up have been effectively used to enhance patient safety (4, 11, 19). Furthermore, various global initiatives have been launched to foster speak-up in hospital departments (6, 20). However, these efforts are typically *ad-hoc* or one-off activities within the patient safety context or overlook the crucial role of organizational culture in the process (20). Implementation research has shown that such activities are seldom effective in the long term and often fail (21). Taking a cultural perspective could potentially result in sustained improvements of these challenges, particularly as numerous healthcare scandals have surfaced as significant drivers for instigating cultural transformation. This is exemplified by the NHS's initiation of the "Freedom to speak up" campaign in 2015, which was prompted by such scandals (22).

Although the available speak-up tools have been proven useful, e.g., for training purposes on the short term in the context of patient safety, healthcare organizations would still benefit from a durable approach to open communication beyond patient safety.

Understanding how cultural factors influence open communication, especially within a broader interprofessional context, can provide valuable insights to bridge existing gaps. This knowledge is crucial for fostering a potential cultural shift that optimizes learning. Therefore, our research question is: What is the impact of cultural elements on the ideal speak-up behavior as perceived by interprofessional teams beyond the patient safety context?

Methods

Study setting and design

The research took place in a general hospital in the Netherlands and focused on the Obstetrics and Gynecology department. This department was selected because interprofessional teamwork is inherent in this field due to the nature and high risks of the job, involving midwives, gynecologists, residents, and nurses etc., and, as a result, well-suited for investigating our research question (17). To gather information, an exploratory qualitative interview study design was used. This approach enabled the researchers to gain a comprehensive understanding of participants' beliefs and thought processes that facilitated or impeded their ability to implement speak-up effectively (23).

Participants and procedure

We used purposeful sampling to select the stakeholders. The participants were selected by two residents who were part of a speak-up workgroup in hopes of generating appropriate and useful data. Two or three stakeholders of each relevant stakeholder group

FIGURE 1
Speak-up pledge.

were selected to form a representative sample of the team in the department.

In order to investigate the impact of cultural elements on speaking up, it was crucial to establish a clear objective and ensure a shared understanding of the ideal concept of speak-up. Therefore, the research team developed a "speak-up pledge" (Figure 1) to prime participants and facilitate the interview discussions. This pledge was developed using speak-up literature and was refined through iterative discussions with the research team. It outlines 'the way we communicate things around here', including a shared vision, clear objectives, preconditions, and general routines for open communication.

From June to August 2020, the research team conducted telephonic semi-structured one-on-one, in-depth interviews with various stakeholders, including nurses, midwives, medical specialists, residents, and operational managers of the department. The operational managers were considered as members of the team for their role as coordinating foremen. All participants were women. After identifying potential participants, fourteen were invited to participate via email, which included attachments providing information about the study, ethical considerations, the speak-up pledge, and the informed consent form. The interviews were scheduled at times convenient for the healthcare professionals, with a clear timeframe established to ensure the interviews could be completed within the

scope of the research. After conducting 13 interviews, data saturation was achieved within the initial cohort, with saturation evaluated by determining the amount of new data generated by each transcript (23). Therefore, the fourteenth interview was not carried out. To prevent misunderstandings, interviewees were provided with a summary of the interview within one week of its completion for approval. Participants were informed about their right to withdraw from the study at any time, and provided written consent indicating their understanding of the study's aims and their voluntary participation, as well as for audio-recording the interview and publishing of data.

Data collection

The interviews were guided by a topic list, following and derived from implementation and organizational culture literature. These concepts combined with concepts on social cognitive and behavioral theory were the units of analysis (Supplementary Material S1). Questions were asked about the current organizational facets to examine whether change is somehow required to successfully act according to the pledge, or a situation wherein healthcare professionals do no longer experience difficulties to speak up. Most questions during the interviews were built around the core question: what do people need from others, from the organization, from and for themselves to be ready for change and to act upon the speak-up pledge. Gaining insight into people's current speak-up behavior, their strengths and weaknesses and what they consider most decisive for a successful implementation was considered necessary information to answer the research question. Other questions were related to a general view of the department, in order to comprehend the contextual factors of speak-up behavior and to make conversation (e.g., built rapport). All interviews lasted for approximately 30 to 60 min.

Data analysis

Each interview was transcribed verbatim, and then analyzed using a three-step coding approach. The first step involved descriptive coding by reading through the data and identifying recurring concepts in at least three interview transcripts. Codes were created based on meaningful text, parts, or statements. Subsequent axial coding led to the creation of theoretical categories. Finally, selective coding was applied to aggregate the theoretical categories, resulting in the development of six themes.

Results

Thirteen interviews were conducted with a range of interprofessional team members including nurses ($n = 3$), gynecologists ($n = 2$), midwives ($n = 3$), residents ($n = 3$) and (operational) managers ($n = 2$). Six themes were identified. The first theme highlights the importance of ensuring a shared vision on speak-up among interprofessional team members. The next three themes focus on three underlying cultural elements: appreciation, cohesion, and open-mindedness, which were identified as crucial competencies for successful speak-up behavior. The last topic explores practical considerations that may contribute to a culture shift by embedding cultural factors alongside conversational techniques in interprofessional working and learning. In the following sections, we will provide further detail on these themes.

Managing vision

Although the speak-up pledge was utilized as a tool during the interviews, it is noteworthy that respondents commonly resonated with the content and acknowledged the inherent cultural significance associated with the pledge. They regarded it as a valuable instrument for delineating vision and expectations surrounding speak-up and fostering deliberate communication. It was also viewed as a useful resource for educating new colleagues on communication systems within the hospital. However, some respondents acknowledged that not everyone may adhere to the pledge. To ensure professionals can effectively speak up, a shared vision appealing to the diverse professional groups and support from supervisors and leadership were identified as crucial factors. In particular, these respondents emphasized the importance of leadership modelling and promoting the pledge to encourage its adoption.

> "And for each professional group or department, there is a necessity to create a sort of an umbrella vision. Because we have nurses, midwives, residents, residents not in training, the gynecologists … everyone has their own thing and that's nice and okay, but how are we going to manage that… I miss that, little attention is paid to it in my experience." (Midwife)

(Dys)functional hierarchy

The functionality of hierarchy in facilitating speak-up stems from the diverse knowledge, experience, and effective decision-making it encompasses. Respondents generally acknowledged that hierarchy can be functional as long as the distinct roles within interprofessional teams are respected. In such cases, hierarchy served as validation for speak-up, contingent upon mutual respect. Respondents emphasized the importance of feeling appreciated and heard during decision-making processes, both within interprofessional teams and among teams within the same profession. However, hierarchy can also become dysfunctional when individual perspectives are devalued or ignored. A significant barrier to speak up was reported by the majority of respondents, who had experienced the detrimental effects of power dynamics and dependency within hierarchical structures.

> "I do not believe that there should be a flat organizational structure. I think that's a bit nonsense actually. I don't think someone who is about to retire who is been in gynecology for 40 years is equal to me. I absolutely don't think so. They have a lot more knowledge and a lot more experience. So I think they should also be able to stand higher than me, for me to look up to. However, I don't think that should stop you from having a discussion with that person." (Resident)

Therefore, leaders need to actively participate in work processes to ensure that they are easily approachable for discussion and need to be mindful of the influence their positions of power have on the work floor, by exhibiting appropriate behavior, vulnerability, and careful language towards their colleagues.

Robust interpersonal relationships

According to respondents, the quality of relationships between professionals affects their willingness to speak up. When professionals work together for an extended period or engage in interprofessional team activities, they gain insight into each other's strengths and weaknesses, which can encourage speaking up.

> "I myself believe that you have to know each other for speaking up. I dare to speak up against all gynecologists. (…) Also just that they know me by name too, that they really see me that they are just, you know, straight up. That honestly makes me feel like I'm treated fairly." (Nurse)

Conversely, lack of familiarity between colleagues can hinder speak-up, particularly for less experienced professionals who may doubt the acceptance of their input.

Young professionals in learning positions are particularly vulnerable due to their reliance on grades and future career prospects. However, respondents also noted that residents tend to develop a more positive attitude towards nurses as they gain experience working with them, regardless of age or experience. Some respondents even suggested that younger professionals could influence the culture positively by speaking up, despite resistance from older colleagues who are less likely to change.

Furthermore, the nature of the relationship and the context in which speak-up occurs are also important factors that influence speak-up behavior, in addition to the duration of the relationship. Informal contact outside of work is considered to be a stimulating factor in increasing cohesion among healthcare professionals. One stakeholder noted that current forms of informal contact are limited to peer-to-peer interactions or within the same profession, and suggested that more interprofessional gatherings should be organized. For example, the residents and gynecologists engage in joint activities like skiing, without involving the nurses. But even the lack of interprofessional lunches, where gynecologists for example could have meals with nurses, is a minor yet noteworthy factor.

> "You know when it's safe, in an honest and safe family, you can share your distress and say 'I think you're doing it wrong' or something like that. But, it really needs to be safe and honest and that's not quite the case. And the logistics, like I said. You should have much more consultations and meeting moments. You should really know each other." (Nurse)

Strategic speak-up strategy

Most respondents believe that speaking up about patient-related matters or medical discussions is less problematic than addressing non-patient-related behavior. However, many respondents choose not to speak up about non-patient-related behavior to maintain a good or at least not worsen an already less satisfactory relationship. Although speaking up is considered professional, it is often taken personally in such contexts, and it can impact the relationship negatively in the long run.

> "… you know, when you're dealing with a person with whom you don't share a nice relationship or merely a good work relationship, then I would really watch my words carefully. And yes, that isn't really safe of course; yet it happens a lot, I would say more than half of the time." (Midwife)

Healthcare professionals strategically choose when to speak up and may choose to avoid certain situations to maintain relationships. This is especially true for situations outside of patient care where feedback may be taken personally and negatively impact the relationship. Additionally, the level of liking or disliking someone influences the ease of giving feedback. Healthcare professionals are more likely to speak up to someone they like as feedback is perceived to be given with good intentions, whereas they may avoid giving feedback to someone they dislike or do not have a good rapport with. This dynamic can be present in both hierarchical and peer-to-peer relationships, as well as relationships between departments.

> "Well, that you feel safe to express things and not afraid of being punished or that you get hit on the head. I do not feel fear with the doctors, but I do have some fellow nurses where I do not dare to say anything because they just have very strong opinions. (…) otherwise you will have to take the bullet and you just don't feel like doing so." (Nurse)

Respondents also noted that not speaking up and being silent on a topic can translate in gossiping. While gossiping or speaking negatively about others behind their backs is often considered a natural human tendency that can provide a sense of relief by releasing pent-up emotions or feelings, it can also lead to the formation of biases and prejudices.

> "We regularly talk critically about each other behind each other's backs. (…) We just like gossiping too much. Listening could sometimes really be better before immediately judging. (.) Not everyone speaks as smooth or fast as others so that should not be the bar. I think everyone should be aware that not everyone is alike; everyone I work with has unique qualities, you cannot say everyone should be like me." (Resident)

Open-mindedness as a professional attitude

The majority of respondents tend to view themselves as approachable for feedback while perceiving others as unapproachable.

> "It depends on how people deal with it, if I give feedback, not necessarily positive feedback, I notice there is tension, because the person who has to receive it must also be able to hear it and be able to do something with it and take it seriously (…) and it should not be that you cut off someone's head or have a judgment.

> So it is really about honesty and the way of saying things and how you deal with it as person to person." (Midwife)

Overall, two important messages were emphasized: the importance of open-mindedness and psychological safety. Open-mindedness was described as the ability to listen without judgment, while psychological safety was described as feeling safe to speak up without fear of punishment or judgment.

When individuals possess open-mindedness, it can foster a sense of psychological safety among them. The ability to feel psychologically safe is determined by multiple factors that relate to open-mindedness, such as the ability of the feedback giver to communicate the message in an honest and skillful manner, the feedback giver's expectations of how the receiver will process the feedback, and how the receiver ultimately responds to it. Open-mindedness allow team members to be transparent and vulnerable by fostering trust, respect, and adaptability. Ultimately, this enables team members to feel comfortable enough to express their thoughts and feelings openly. Reflecting together and sharing thoughts can foster mutual understanding, but respondents suggest that interprofessional reflection is infrequent. Furthermore, treating speak-up as a professional competence instead of a personal message to an individual may make it easier for people to engage in it or identify when others are doing it.

> "I don't care who you are as a person, it shouldn't matter the other way around, we really need each other to work together so it's just really so important not to treat each other like that. You work as a professional (…) you shouldn't let those emotions you have about each other get in the way." (Resident)

Moreover, it may also be exacerbated by certain group dynamics. For instance, one respondent points out that despite the evolving dynamics and increasing female emancipation, men continue to hold a dominant position over women in discussions and the final decision-making process, while another respondent mentions that a group comprised mostly of women may be susceptible to specific female characteristics that affect the way they work together.

> "It's a different group, a different dynamic, maybe because it's mostly women, I don't know but everyone has to have an opinion on everything, it takes a lot of time for them to agree with each other." (Organizational Manager)

From training to practice: cultural practices to embed in interprofessional working and learning

While there is a willingness to change and reflect, concrete action towards encouraging speak-up behavior is still lacking. Although the hospital desires an open culture, employees are not provided with the necessary tools to achieve this.

Respondents report that open communication is only emphasized during acute situation training in the patient context, which occurs once or twice a year, with no other initiatives to stimulate open communication and a lack of tools on how to successfully speak up. In this training context, participants usually speak up without hesitation as it serves the aim of the training and there are no personal stakes involved, even in working environments where speak-up is typically discouraged due to power imbalances. However, outside of this particular context, the level of perceived difficulty and emotions as well as the absence of relevant professional skills on both the part of the person speaking up and the person receiving the message influences the decision to speak up or remain silent.

> "Because you lack those skills to raise your concern properly, so then you often decide to just let it go." (Resident)

To bring about effective open communication, it is important to provide tools that facilitate the development of specific competencies, including conversation techniques on how to speak up, non-judgmental listening, coping with emotions, and interprofessional reflection on practice.

> "There are feedback courses for the trainers and residences, however, interprofessional feedback courses lack (…) For example, a nurse has never learned what feedback they can give to a gynecologist. (.) They are not sufficiently supported and trained in it." (Medical Specialist 2)

The physical context in which healthcare professionals work seems to have a significant impact on their experiences, both formally and informally. One notable contrast is between two locations, where interprofessional teams work together in one room on one location, and nurses and doctors work separately in different rooms on the other. The former setting is generally viewed as advantageous for relationships and communication channels.

> "There should be more mutual understanding here. And I really think that this is something that is reinforced by the way the logistics are arranged, in that the nurses are in one room and the doctors in a separate one. It really feeds the sense of disunity". (Resident)

Discussion

This study explores the impact of cultural factors on speak-up behavior in interprofessional teams beyond patient safety. The study identified six themes that enhance speak-up behavior: the importance of vision management, functional hierarchy, interpersonal relationships, healthcare professionals' speak-up strategy, open-mindedness as a professional attitude, and cultural practices to embed in interprofessional working and learning. These themes can be categorized into four levels; the professional, interprofessional, leader, and organizational levels. Firstly at the professional level an open-minded professional attitude was considered vital for effective communication, next to a corresponding strategy showcasing speak-up on specific subjects while exercising restraint on others. At the interprofessional level, robust relationships between individuals played a crucial role in encouraging speak-up behavior, particularly in situations where power differentials exist. At the leader level the majority of respondents recognized that hierarchy proves effective when professional opinions are sufficiently respected and valued.

Finally at the organizational level, having a shared vision that encourages speak-up is beneficial for alignment, and embedding specific cultural elements into interprofessional working and learning can positively influence speak-up behavior across all levels.

The study's findings are intriguing as it provides new perspectives on existing knowledge regarding team effectiveness in the context of interprofessional speak-up behavior beyond patient safety. Theorists have extensively discussed what makes some teams more effective than others, initially focusing on the outcomes of team performance and viability (24). Eventually, attention shifted to processes that explain why certain inputs affect team effectiveness and viability. However, a single-cycle linear model no longer suffices, as effective teams now operate more than ever within complex, fast-changing, multilevel systems across various times, tasks, and contexts. Ilgen et al. (24) for example, expanded the literature with their widely used Input-Mediator-Output–Input (IMOI) model of effective teams, proposing a broader scope of mediating processes. They also introduced the concept of a cyclical causal feedback loop, indicating that outcomes influence and may serve as initial inputs. Each team experience impacts the next, as both the team and its members grow during these processes. The current landscape of teambuilding is evolving towards greater emphasis on interprofessional collaborative efforts aimed at improving healthcare outcomes and reducing costs. Emerging concepts like shared care, shared leadership, and shared decision-making over the past decade demonstrate this shift (25–27). As a result, interprofessional practices are expanding beyond traditional teams, even though their effects are not conclusively proven (28).

This makes the way professionals interact with each other in the various organizational layers, as well as between different professions, departments, and even different organizations, particularly important, where, according to our findings, hierarchy does not have to be inherently problematic, contrary to previous research (4, 5, 10, 13, 17, 29, 30). The participants in this study acknowledged the functionality of hierarchy when it comes to decision-making for example. Nevertheless, the study emphasizes that the underlying problem lies in the core values and norms surrounding behavior within the hierarchical structure; even with the establishment of a collective vision, there is a challenge in fostering a sense of value for individuals' opinions. These findings align with research by Vatn et al. and Weiss et al. in the healthcare sector (31, 32), but is also consistent with literature in organizational science (33). Bunderson and Reagans found that while power and status can complicate collective learning by disrupting shared goals, risk-taking, and knowledge sharing, the socialized use of power can actually leverage social hierarchy to enhance collective learning (33), making it an essential requirement for leaders in environments where learning is key and power and status differences exist. For leaders to become skilled in the socialized use of power, they must embrace leadership styles that extend beyond traditional theories (34). Considering our results and the literature, hospitals might benefit from concepts such as that of 'Transformational Leadership', 'Humble Leadership', and 'Psychological Safety', as described by Jacobs, Schein and Edmondson, respectively. These concepts are relevant to the study's findings as they address the identified topics of a shared vision, cohesion, appreciation, open-mindedness and leadership approaches that appear necessary to confront the challenges of speak up behavior (35–37).

Transformational leadership is described as a form of leadership that inspires and motivates followers to achieve outcomes beyond expectations, helping followers to grow and develop by responding to their individual needs (38). Jacobs' suggests managers to focus on changing the thought processes that drive behavior by establishing agreement on a set of guiding principles and believes that this will naturally lead to the desired actions from others (35). Humble Leadership emphasizes the development of personal and cooperative relationships, openness, and trust, as opposed to the traditional approach maintaining an appropriate professional distance (37). The focus is on interpersonal and group dynamics, where leaders embrace ambiguity and work to reduce the distance between opposing sides to establish shared commitment based on openness and trust. Simultaneously organizational leaders can enhance their awareness of on-the-ground happenings by cultivating and sustaining effective exchange relationships with frontline employees (39). When properly employed and supported by leaders, this approach is thought to help others feel valued despite the hierarchical structure and improve relationships by actively managing them, encouraging open communication, and preventing dishonest feedback to save face, especially among peers (37). Schein recommends that all teams, regardless of size, perform better when their members feel psychologically safe to share their thoughts and ideas with one another. Psychological safety as to the description of Edmondson refers to an environment where individuals feel secure enough to take interpersonal risks by expressing their concerns, questions, or ideas without fear of negative repercussions (36). This includes being candid and willing to engage in constructive conflict to learn from various viewpoints, while also having the assurance that others will give them the benefit of the doubt when they admit to making a mistake or need help (36).

Whereas the concepts of transformational and humble leadership give emphasis to this study topics of shared vision, cohesion, respect, and openness, the concept of psychological safety relates to the cultural element open-mindedness, which can be seen as a professional attitude that is vital for success. Implementing the aforementioned concepts could facilitate hierarchical structures to become more conducive to effective interprofessional working and learning, yet remains a formidable undertaking in reality. Creating an environment of open communication and open-mindedness, free from prejudices and with a willingness to give others the benefit of the doubt, remains a challenging task. A recent scoping review has identified how team members internalize biases related to dominance and expertise, and how team members adapt to these biases in dynamic ways, resulting in negative effects on interprofessional collaboration (40). In Figure 2, we present an IMOI framework designed to facilitate effective interprofessional speak-up beyond patient safety, emphasizing the need for various mediating processes to enhance interprofessional working and learning, professional well-being, and patient outcomes. The resulting positive outcomes can, in turn, foster a speak-up culture, encouraging further input and engagement.

Recommendations

Knowledge—Traditional interprofessional training often focuses on general attitudes towards working together and the acquisition and demonstration of knowledge and skills (18). To promote effective speak-up behaviors within interprofessional teams, it is

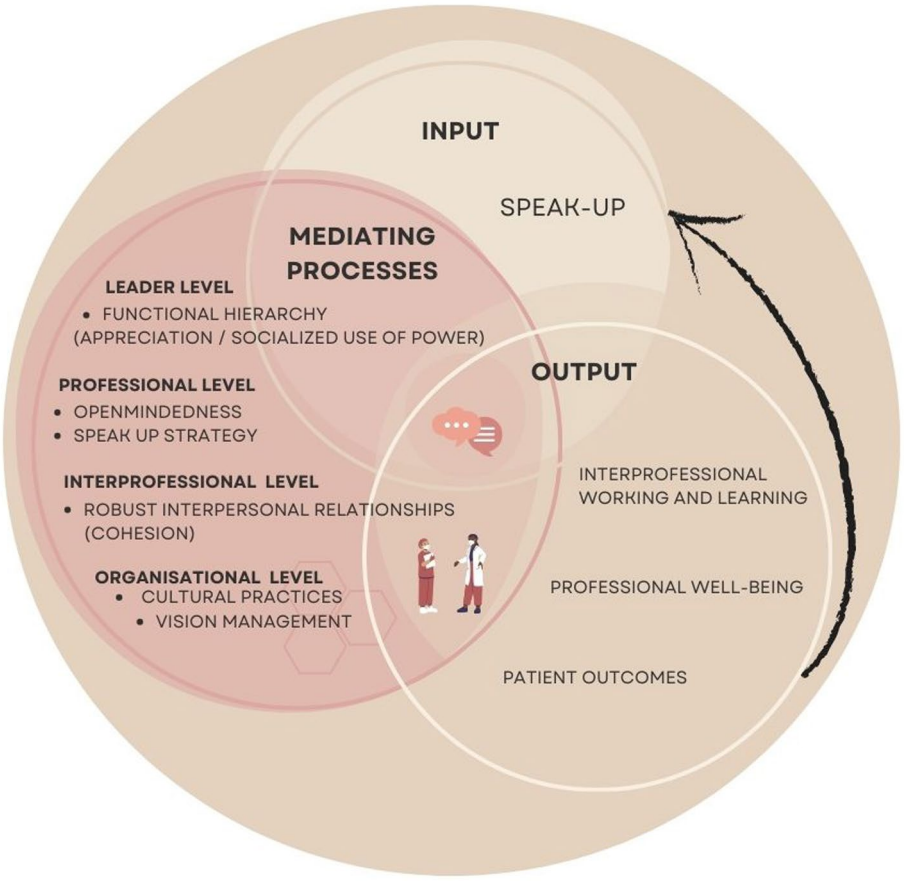


FIGURE 2
IMOI framework for effective interprofessional speak-up beyond patient safety.

important to equip them with technical tools that are tailored to their specific needs. One such tool is the "Humble Inquiry" approach, which emphasizes the use of curious and open-minded questioning (41). This approach requires the questioner to be vulnerable and willing to learn from the responses of others and may be used for speak-up in a subtle way. Ongoing training and education on this topic, in which the six cultural themes found in this study are emphasized, should be key in facilitating interprofessional working and learning. Examining the effect on interprofessional practice is a subject for future research.

Reflective practice—Interprofessional teams must learn to minimize biases, acknowledge and appreciate the valuable insights and perspectives of all members. Interprofessional working and learning that incorporates a shared understanding of roles and responsibilities may contribute to this (42). Other studies have identified that encouraging reflexivity, or reflecting on one's own professional identity and biases, can improve interprofessional working and learning (43, 44), by making people unconsciously suppress their own perspective for that of another (35). Having such a mind-set not only prepares the team for speaking up, but also for receiving feedback and engaging in reflective practices together (45). The socialized use of power is a fundamental prerequisite in this context, particularly because power and status differences can influence the willingness of members to participate in collective learning activities. These differences can impact their perceptions and feelings of psychological safety, their risk assessments, and their inclination to take initiative and independent action (33).

Consolidation in cultural patterns – The incorporation of the six themes uncovered in this study has the potential to be considered as targets for future interventions and bring about a desired shift in culture, promoting speaking-up, and interprofessional working and learning. Appelbaum et al. (18) arrived at a parallel finding indicating that psychological safety and power distance can serve as substantial factors driving cohesion and fostering effective collaboration. However, organizations often fail to address the motivational and supportive factors that impact the target group when preparing for change. As a result, implementation efforts are often unsuccessful in bringing about the desired behavioral change (21). Achieving effective interprofessional working and learning is only possible with the provision of optimal organizational support and resources. This requires individual transformation as well as changes in collective management and leadership practices. Therefore, leaders must possess the ability to envision the organization's role and lead transformational efforts that align with the purpose of embedding the cultural aspects. So first of all it is crucial to establish a shared vision, such as through a speak-up pledge, and ensure that the organization has the ability to effectively receive and respond to feedback (37). To foster effective communication and relationships, leaders can be trained using transformational and humble leadership styles. But most importantly,

they play a crucial role in establishing habitual patterns that drive a cultural shift.

Strengths and limitations

The aim of this research was to provide valuable insights into the cultural influences on speak-up behavior. One of the strengths of this study lies in its comprehensive examination of the potential obstacles associated with implementing speak-up behavior, all the while recognizing the paramount significance of cultural factors in this context.

A limitation of this study is that it was conducted in only one hospital in one country. Although we believe in some transferability of the findings in similar healthcare systems and departments, the complexity of organizational culture means that findings may not be fully transferable to other countries. Therefore, it would be valuable to explore underlying cultural factors in different settings, not just in healthcare but also in other industries. Furthermore, a limitation lies in the selection of respondents, as the two residents who selected the respondents could potentially introduce bias. Respondents' participation and responses may also be influenced by this approach, thus contributing to potential bias. To mitigate this bias, we sought to minimize it by having interviews conducted by an external individual who is not affiliated with the healthcare profession.

Conclusion

This study explores the impact of cultural factors on speak-up behavior for interprofessional working and learning. Six key cultural elements have been identified as enhancers of speak-up behavior within interprofessional working and learning. Rather than assigning blame, hierarchy can serve as a valuable facilitator. Emphasizing managerial vision, in addition to vital traits such as open-mindedness, is imperative. The incorporation of the themes uncovered in this study has the potential to bring about a desired shift in culture, promoting speaking up, and interprofessional working and learning. Transformational and humble leadership approaches offer valuable direction for applying this knowledge in the workplace and strategically utilizing the act of speaking up.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

RM: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Data curation, Supervision. PA: Conceptualization, Writing – original draft. AT: Conceptualization, Writing – original draft, Methodology. MB-S: Conceptualization, Supervision, Writing – original draft. CH: Conceptualization, Writing – original draft, Supervision. FS: Conceptualization, Methodology, Supervision, Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1345316/full#supplementary-material

References

1. Moniz T, Pack R, Lingard L, Watling C. Voices from the front lines: an analysis of Physicians' reflective narratives about flaws with the 'System. *J Med Humanit*. (2021) 42:737–52. doi: 10.1007/s10912-021-09690-6

2. Dixon-Woods M, Aveling EL, Campbell A, Ansari A, Tarrant C, Willars J, et al. What counts as a voiceable concern in decisions about speaking out in hospitals: a qualitative study. *J Health Serv Res Policy*. (2022) 27:88–95. doi: 10.1177/13558196211043800

3. Brennan PA, Davidson M. Improving patient safety: we need to reduce hierarchy and empower junior doctors to speak up. *BMJ*. (2019) 366:l4461. doi: 10.1136/bmj.l4461

4. Garber AB, Posner G, Roebotham T, Bould MD, Taylor T. Facing hierarchy: a qualitative study of residents' experiences in an obstetrical simulation scenario. *Adv Simul*. (2022) 7:34. doi: 10.1186/s41077-022-00232-1

5. Morrow KJ, Gustavson AM, Jones J. Speaking up behaviours (safety voices) of healthcare workers: a metasynthesis of qualitative research studies. *Int J Nurs Stud*. (2016) 64:42–51. doi: 10.1016/j.ijnurstu.2016.09.014

6. Jones A, Blake J, Adams M, Kelly D, Mannion R, Maben J. Interventions promoting employee "speaking-up" within healthcare workplaces: a systematic narrative review of

the international literature. *Health Policy*. (2021) 125:375–84. doi: 10.1016/j.healthpol.2020.12.016

7. Okuyama A, Wagner C, Bijnen B. Speaking up for patient safety by hospital-based health care professionals: a literature review. *BMC Health Serv Res*. (2014) 14:61. doi: 10.1186/1472-6963-14-61

8. Landgren R, Alawadi Z, Douma C, Thomas EJ, Etchegaray J. Barriers of pediatric residents to speaking up about patient safety. *Hosp Pediatr*. (2016) 6:738–43. doi: 10.1542/hpeds.2016-0042

9. Martinez W, Etchegaray JM, Thomas EJ, Hickson GB, Lehmann LS, Schleyer AM, et al. 'Speaking up' about patient safety concerns and unprofessional behaviour among residents: validation of two scales. *BMJ Qual Saf*. (2015) 24:671–80. doi: 10.1136/bmjqs-2015-004253

10. Umoren R, Kim S, Gray MM, Best JA, Robins L. Interprofessional model on speaking up behaviour in healthcare professionals: a qualitative study. *BMJ Leader*. (2022) 6:15–9. doi: 10.1136/leader-2020-000407

11. Pattni N, Arzola C, Malavade A, Varmani S, Krimus L, Friedman Z. Challenging authority and speaking up in the operating room environment: a narrative synthesis. *Br J Anaesth*. (2019) 122:233–44. doi: 10.1016/j.bja.2018.10.056

12. Voogt JJ, Kars MC, van Rensen ELJ, Schneider MME, Noordegraaf M, van der Schaaf MF. Why medical residents do (and Don't) speak up about organizational barriers and opportunities to improve the quality of care. *Acad Med*. (2020) 95:574–81. doi: 10.1097/ACM.0000000000003014

13. Schwappach D, Richard A. Speak up-related climate and its association with healthcare workers' speaking up and withholding voice behaviours: a cross-sectional survey in Switzerland. *BMJ Qual Saf*. (2018) 27:827–35. doi: 10.1136/bmjqs-2017-007388

14. Malik RF, Buljac-Samardžić M, Amajjar I, Hilders C, Scheele F. Open organisational culture: what does it entail? Healthcare stakeholders reaching consensus by means of a Delphi technique. *BMJ Open*. (2021) 11:e045515. doi: 10.1136/bmjopen-2020-045515

15. Donovan AL, Aldrich JM, Gross AK, Barchas DM, Thornton KC, Schell-Chaple HM, et al. Interprofessional care and teamwork in the ICU. *Crit Care Med*. (2018) 46:980–90. doi: 10.1097/CCM.0000000000003067

16. Karam M, Brault I, Van Durme T, Macq J. Comparing interprofessional and interorganizational collaboration in healthcare: a systematic review of the qualitative research. *Int J Nurs Stud*. (2018) 79:70–83. doi: 10.1016/j.ijnurstu.2017.11.002

17. Eddy K, Jordan Z, Stephenson M. Health professionals' experience of teamwork education in acute hospital settings: a systematic review of qualitative literature. *JBI Database System Rev Implement Rep*. (2016) 14:96–137. doi: 10.11124/JBISRIR-2016-1843

18. Appelbaum NP, Lockeman KS, Orr S, Huff TA, Hogan CJ, Queen BA, et al. Perceived influence of power distance, psychological safety, and team cohesion on team effectiveness. *J Interprof Care*. (2020) 34:20–6. doi: 10.1080/13561820.2019.1633290

19. Pian-Smith MC, Simon R, Minehart RD, Podraza M, Rudolph J, Walzer T, et al. Teaching residents the two-challenge rule: a simulation-based approach to improve education and patient safety. *Simul Healthc*. (2009) 4:84–91. doi: 10.1097/SIH.0b013e31818cffd3

20. O'Donovan R, McAuliffe E. A systematic review exploring the content and outcomes of interventions to improve psychological safety, speaking up and voice behaviour. *BMC Health Serv Res*. (2020) 20:101. doi: 10.1186/s12913-020-4931-2

21. Grol R, Wensing M. Implementation of change in healthcare: a complex problem. Hoboken: John Wiley & Sons (2020).

22. NHS. Freedom to speak up United Kingdom (2015). Available at: https://www.england.nhs.uk/ourwork/freedom-to-speak-up/

23. Green J, Thorogood N. Qualitative methods for health research. London: Sage Publication (2018).

24. Ilgen DR, Hollenbeck JR, Johnson M, Jundt D. Teams in organizations: from input-process-output models to IMOI models. *Annu Rev Psychol*. (2005) 56:517–43. doi: 10.1146/annurev.psych.56.091103.070250

25. Aufegger L, Alabi M, Darzi A, Bicknell C. Sharing leadership: current attitudes, barriers and needs of clinical and non-clinical managers in UK's integrated care system. *BMJ Leader*. (2020) 4:128–34. doi: 10.1136/leader-2020-000228

26. Montori VM, Ruissen MM, Hargraves IG, Brito JP, Kunneman M. Shared decision-making as a method of care. *BMJ Evid Based Med*. (2023) 28:213–7. doi: 10.1136/bmjebm-2022-112068

27. Smith SM, Cousins G, Clyne B, Allwright S, O'Dowd T. Shared care across the interface between primary and specialty care in management of long term conditions. *Cochrane Database Syst Rev*. (2017) 2:Cd004910. doi: 10.1002/14651858.CD004910.pub3

28. Reeves S, Pelone F, Harrison R, Goldman J, Zwarenstein M. Interprofessional collaboration to improve professional practice and healthcare outcomes. *Cochrane Database Syst Rev*. (2017) 2018:Cd000072. doi: 10.1002/14651858.CD000072.pub3

29. Siewert B, Swedeen S, Brook OR, Eisenberg RL, Hochman M. Barriers to safety event reporting in an academic radiology department: authority gradients and other human factors. *Radiology*. (2018) 288:693–8. doi: 10.1148/radiol.2018171625

30. Martinez W, Lehmann LS, Thomas EJ, Etchegaray JM, Shelburne JT, Hickson GB, et al. Speaking up about traditional and professionalism-related patient safety threats: a national survey of interns and residents. *BMJ Qual Saf*. (2017) 26:869–80. doi: 10.1136/bmjqs-2016-006284

31. Vatn L, Dahl BM. Interprofessional collaboration between nurses and doctors for treating patients in surgical wards. *J Interprof Care*. (2022) 36:186–94. doi: 10.1080/13561820.2021.1890703

32. Weiss M, Morrison EW, Szyld D. I like what you are saying, but only if i feel safe: psychological safety moderates the relationship between voice and perceived contribution to healthcare team effectiveness. *Front Psychol*. (2023) 14:1129359. doi: 10.3389/fpsyg.2023.1129359

33. Bunderson JS, Reagans RE. Power, status, and learning in organizations. *Organ Sci*. (2011) 22:1182–94. doi: 10.1287/orsc.1100.0590

34. Farrell SE, Bochatay N, Kim S. Embracing or relinquishing sources of power in interprofessional communication: implications for patient-centered speaking up. *J Interprof Care*. (2021) 1-8:1–8. doi: 10.1080/13561820.2021.1975665

35. Jacobs CS. Management rewired: why feedback doesn't work and other surprising lessons from the latest brain science. New York: Portfolio (2010).

36. Edmondson A. The fearless organization: Creating psychological safety in the workplace for learning, innovation, and growth John Wiley & Sons (2018).

37. Schein EH, Schein PA. Humble leadership. Cham: Springer (2021).

38. Ellen BP. Transformational leadership In: A Farazmand, editor. Global encyclopedia of public administration, public policy, and governance. Cham: Springer International Publishing (2016). 1–4.

39. Liu W, Tangirala S, Ramanujam R. The relational antecedents of voice targeted at different leaders. *J Appl Psychol*. (2013) 98:841–51. doi: 10.1037/a0032913

40. Sukhera J, Bertram K, Hendrikx S, Chisolm MS, Perzhinsky J, Kennedy E, et al. Exploring implicit influences on interprofessional collaboration: a scoping review. *J Interprof Care*. (2022) 36:716–24. doi: 10.1080/13561820.2021.1979946

41. Schein EH. Humble inquiry: the gentle art of asking instead of telling. San Francisco: Berrett-Koehler Publishers (2013).

42. Mickan SM, Rodger SA. Effective health care teams: a model of six characteristics developed from shared perceptions. *J Interprof Care*. (2005) 19:358–70. doi: 10.1080/13561820500165142

43. Weiss M, Kolbe M, Grote G, Spahn DR, Grande B. Why didn't you say something? Effects of after-event reviews on voice behaviour and hierarchy beliefs in multi-professional action teams. *Eur J Work Organ Psy*. (2017) 26:66–80. doi: 10.1080/1359432X.2016.1208652

44. Cantaert GR, Pype P, Valcke M, Lauwerier E. Interprofessional identity in health and social care: analysis and synthesis of the assumptions and conceptions in the literature. *Int J Environ Res Public Health*. (2022) 19:14799. doi: 10.3390/ijerph192214799

45. Coelho V, Scott A, Bilgic E, Keuhl A, Sibbald M. Understanding feedback for learners in Interprofessional settings: a scoping review. *Int J Environ Res Public Health*. (2022) 19:10732. doi: 10.3390/ijerph191710732

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact