Title: What is the future of artificial intelligence in obstetrics? A qualitative study among healthcare professionals

Shortened running title: Future of AI in Obstetrics: A Qualitative Study

Authors: A. M. Fischera, b, \*, A. L. Rietveldb, c, MD, P. W. Teunissend, e, MD, M. Hoogendoorna, P. C. A. M. Bakkerb, c, MD

a Department of Computer Science, Vrije Universiteit, De Boelelaan 1105, Amsterdam, 1081 HV, The Netherlands

b Department of Obstetrics and Gynaecology, Amsterdam UMC Location VUmc, De Boelelaan 1117, Amsterdam, 1081 HV, The Netherlands

c Amsterdam Reproduction and Development, Amsterdam, The Netherlands

d Department of Gynaecology&Obstetrics, Maastricht UMC, P. Debyelaan 25, Maastricht, 6229 HX, The Netherlands

e School of Health Professions Education, Faculty of Health Medicine and Life Sciences, Maastricht University, Universiteitssingel 60, Maastricht, 6229 ER, The Netherlands

\* Corresponding author: De Boelelaan 1117, Amsterdam, 1081 HV, The Netherlands, a.m.fischer@amsterdamumc.nl

# Abstract

## Objective

This work explores the perceptions of obstetric clinicians about Artificial Intelligence (AI) in order to bridge the gap in uptake of AI between research and medical practice. Identifying potential areas where AI can contribute to clinical practice, enables AI research to align with the needs of clinicians and ultimately patients.

## Design

Qualitative interview study.

## Setting

A national study conducted in the Netherlands.

## Sample

## Dutch clinicians working in obstetrics with varying relevant work experience, gender, and age.

## Methods

Thematic analysis of qualitative interview transcripts.

## Results

Thirteen gynaecologists were interviewed about hypothetical scenarios of an implemented AI model. Thematic analysis identified two major themes: perceived usefulness and trust. Usefulness involved AI extending human brain capacity in complex pattern recognition and information processing, reducing contextual influence, and saving time. Trust required validation, explainability and successful personal experience. This result shows two paradoxes: firstly, AI is expected to provide added value by surpassing human capabilities, yet also a need to understand the parameters and their influence on predictions for trust and adoption was expressed. Secondly, participants recognised the value of incorporating numerous parameters into a model, but they also believed that certain contextual factors should only be considered by humans, as it would be undesirable for AI models to utilize that information.

## Conclusions

Obstetricians’ opinions on the potential value of AI highlight the need for clinician-AI researcher collaboration. Trust can be built through conventional means like RCTs and guidelines. Holistic impact metrics, such as changes in workflow, not just clinical outcomes, should guide AI model development. Further research is needed for evaluating evolving AI systems beyond traditional validation methods.

## Funding

A.M. Fischer, MSc, dr. A.L. Rietveld and dr. P.C.A.M. Bakker are funded by a grant of public-private partnerships (PPP) from Amsterdam UMC (grant identifier 2007651). The collaboration project is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. This study has been performed in the context of the COCOON (Combining cord-free uterine electrohysterography and standard clinical measurements for refining the detection of premature birth) study, a cooperation of Stichting VUmc, Stichting VU, Health~Holland and Bloom Technologies NV. All funding bodies played no role in the creation of this paper.

## Keywords

Qualitative research, obstetricians, Artificial Intelligence, interviews, perceptions.

# Funding

AM Fischer, MSc, dr. A.L. Rietveld and dr. P.C.A.M. Bakker are funded by a grant of public-private partnerships (PPP) from Amsterdam UMC (grant identifier 2007651). The collaboration project is co-funded by the PPP Allowance made available by Health~Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships. This study has been performed in the context of the COCOON (Combining cord-free uterine electrohysterography and standard clinical measurements for refining the detection of premature birth) study, a cooperation of Stichting VUmc, Stichting VU, Health~Holland and Bloom Technologies NV. All funding bodies played no role in the creation of this paper.

# Introduction

Artificial intelligence (AI) has rapidly made its way into people's daily lives. The opportunities to apply AI have increased with the collection of big data, and so have expectations. In the medical field, the interest in applying AI has grown considerably, as evidenced by the large number of research publications in the last two decades (1, 2). However, there is a significant gap between the number of scientific papers published and the number of AI applications implemented in clinical practice (3, 4). Up to 2020, only 29 AI-based medical technologies were approved by the U.S. Food and Drug Administration (FDA) (5). Obstetrics ranks within the top 5 clinical conditions featured in the AI literature on PubMed (1); however, as of present, no FDA approvals have been granted for AI applications in this field (5).

Moreover, we observe a large difference between the number of publications on AI applied in obstetrics and gynecology (ObGyn) published in AI/Computer Science core disciplines journals (82%) and ObGyn core disciplines journals, as defined in Web of Science (18%) (6). Generally, ObGyn journals do not report on research of AI-related systems, but prefer to await demonstrated clinical value of AI research (6).

Various computer science journals have reported on the use of AI in pregnancy-related clinical decision making (7-9). In addition, broader oriented research has been conducted on identifying factors within the healthcare domain that play a role in real-world AI adoption (10-15). However, no studies have investigated obstetric clinicians’ opinions and perceptions of *where* and *how* AI might contribute within their field. This is problematic, as insights in individual health professionals expectations and needs are key success factors in the translation of e-health technology to clinical practice (16).

Furthermore, if AI research in ObGyn is developed by and published in the AI community, this could potentially lead to a mismatch between the direction of AI research and the needs of clinicians. Clinicians need to be informed about developments regarding AI applications in obstetrics and should be involved at the earliest possible stage in shaping this future technology.

Understanding what obstetric clinicians consider promising directions for AI application development and what their expectations are within their field is necessary to bridge the gap between research and medical practice. It will help to ensure that AI research is of potential value to their users: clinicians and ultimately patients. Therefore, this work aims to answer the questions: What are the perceptions of obstetric clinicians about AI and where do they see a potential contribution of AI in clinical practice?

# Methods

## Upstream Engagement

Upstream engagement was conducted, since no AI applications have been implemented in obstetrics to date. Upstream engagement means that stakeholders are being involved in a dialogue about emerging areas of science at the earliest possible stage (17). The full value of upstream engagement rests on three pillars: the normative argument, substantive argument and instrumental argument (18). The normative argument states that having a dialogue is valuable in and of itself. It is an important part of the democratic process for making (possibly controversial) decisions. The substantive argument suggests that decisions about emerging technologies should be made with consideration for the ethical and value concerns of affected groups, which will lead to more robust decision making. The instrumental argument proposes that societal dialogue will increase the legitimacy of decisions and generate secondary effects such as greater trust in the policy-making process.

## Participants and recruitment

The study aimed to explore a broad research question, and thus, a diverse group of Dutch clinicians working in obstetrics with varying relevant work experience, gender, and age were approached. A convenience sampling technique was used to recruit gynaecologists who were affiliated with the professional network of clinical research team members. Email invitations were sent out to recruit participants and participation was voluntary. The number of interviews was determined by reaching data sufficiency. This relates to the point when conducting additional interviews is unlikely to lead to the identification of additional dimensions related to the research questions. Data sufficiency is generally expected between 9-24 interviews (19).

## Hypothetical scenarios

Interviews began by exploring clinicians' views on the term AI and what they expected from AI models in healthcare. We based our procedure on that of Tonekaboni et al. (20), who introduced clinicians to a hypothetical scenario of an implemented AI model as the starting point of an interview. We described three types of AI models (decision support for diagnoses, optimizing resources and personalized medicine) that are implemented in clinical practice within the fields of radiology, intensive care unit and internal medicine (21-25). These model types were translated to common scenarios in obstetrics where such AI models could be used in practice. The following scenarios were developed and used to initiate the discussion on the potential value of AI and the necessary preconditions that must be met by an AI model for practical application:

* Diagnoses: an AI model to predict intrapartum fetal asphyxia
* Optimizing workflow/resources: an AI model for hospital discharge of pregnant women with early preterm pre-labor rupture of membranes (PPROM)
* Personalized medicine/treatment: an AI model for personalized treatment for patients with gestational diabetes mellitus (GDM)

The three different categories of AI models vary significantly in output, number of factors considered, and contribution to diagnostic or therapeutic reasoning. We hypothesized that this variation may have an effect on how participants view AI and is therefore important for answering the research questions. Details regarding the scenarios and subsequent questions can be found in the Appendix.

## Data collection and analysis

We used semi-structured interviews (26) for data collection, as the primary goal of this study is to gather information from key stakeholders who might use AI applications in their field in the near future. All interviews were conducted privately, either on hospital site where the participant worked, or using Microsoft Teams. The interviews took place between November 2022 and February 2023. Mean interview duration was 37 minutes (range 31-57 minutes). A thematic analysis was employed to identify patterns across interviews (27, 28). Data was categorized using open, axial and selective coding. Specifically, all data was dissected into distinct components using open coding, followed by the establishment of associations between codes using axial coding. Finally, selective coding was used to identify a pivotal theme that unified all codes derived from the analysis (29).

Each interview was audio recorded, professionally transcribed, and software system MaxQDA 2020 was utilized to analyse the data. The first five interviews were independently coded by A. F. and A. R., who reviewed each other’s work and discussed any discrepancies until consensus was reached. Questions related to preliminary findings that arose during the coding process were introduced in subsequent interviews. Furthermore, overarching themes and their links between the interviews were identified through discussion among A. F. , A. R. and P. T.

## Research team and reflexivity

The interviews were conducted by Anne Fischer (AF) who attended the formal training Qualitative Research in Healthcare Practice, which covers qualitative research methods, approaches and techniques for data analysis and interview techniques. AF is a PhD candidate with a background in AI and our research team consists of both clinicians working in obstetrics and AI researchers. We considered the combination of both field experiences an advantage that facilitated in-depth interviews for both the clinical and AI sides in this topic.

## Ethical approval

The Medical Ethics Review Committee of Amsterdam UMC, location AMC examined the study protocol (W22\_363#22.431) and judged that an official approval of this study was not required. Informed consent in writing was obtained from every participant.

# Results

Thirteen gynaecologists working in eight different hospitals across the Netherlands were included in the study. Characteristics of the participants are shown in Table 1.

[See separate file for Table 1]

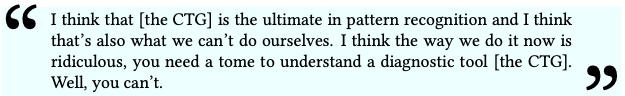
Thematic analysis of the in-depth interviews were organized into two major themes: (1) perceived usefulness of AI models and (2) trust. Both themes are discussed below with illustrative quotes from the interviews.

## Perceived usefulness of AI models

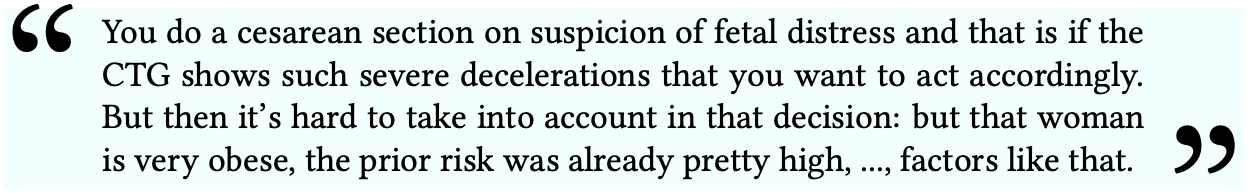
Participants had a positive attitude to at least one of the elucidated scenarios and indicated that such AI models could add value to their current way of working. Since we are not only interested in what makes these specific AI models useful, we will describe overarching factors characterized as useful across models.

### Beyond human brain capacity

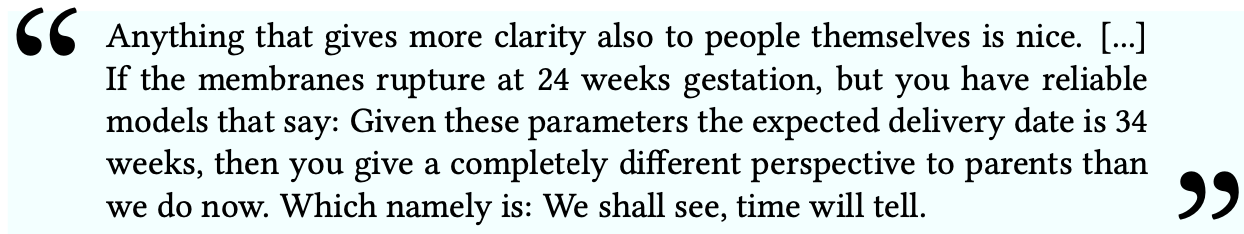
Participants stressed that in order for an AI model to improve their way of working, it must bring something beyond their own human brain capacity or fill a gap in their knowledge. In case of a high-stake and acute scenario, such as predicting perinatal asphyxia, participants recognized that the CTG as a stand-alone tool to monitor foetal well-being has significant limitations. One limitation, and AI as a possible solution to this, was recognised in terms of the complexity related to pattern recognition on the CTG in relation to clinical outcome:



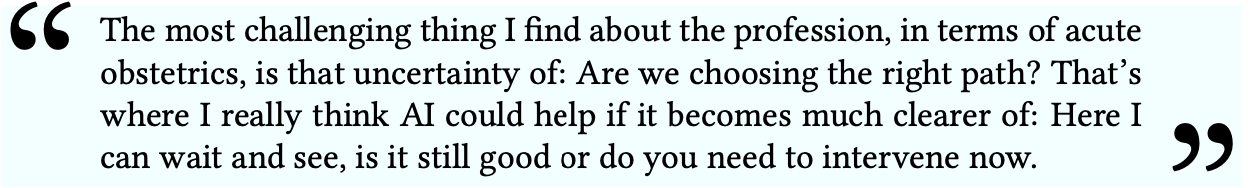
The other described limitation was the multitude of information participants had to take into account during clinical-decision making. An AI model that incorporates multiple factors and facilitates consideration of these factors during clinical decision-making was deemed highly valuable. Although participants indicated that they are trained to consider all pertinent factors during clinical decision-making, this is a challenging task in practice. Moreover, exact knowledge about the relationship between all factors and outcome is unknown:



In situations where participants identified gaps in their knowledge, they expected AI to help to fill these gaps. According to participants, the goal of an AI model should be to enable them to develop more precise treatment plans and communicate more accurately with patients, ultimately enhancing shared decision-making:

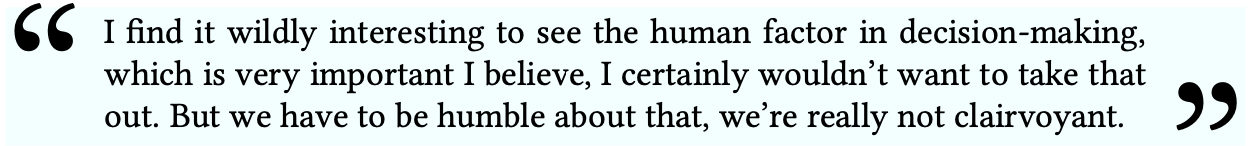


Participants also underlined their expectations of AI adding value especially in what they experience as the so-called ‘grey area’, referring to scenarios where for them it is uncertain if they are choosing the right path during clinical decision making:

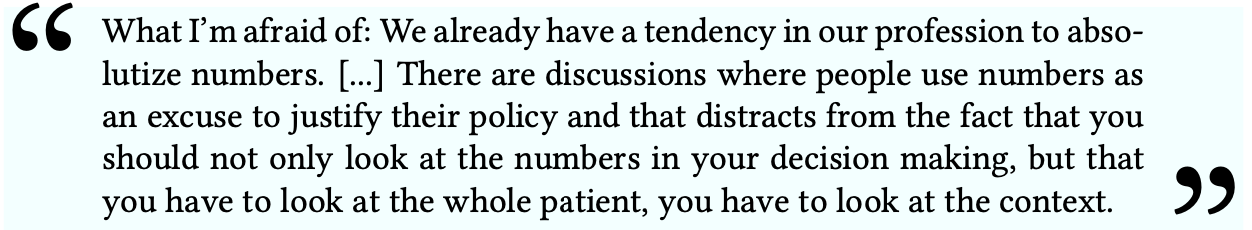


### Contextual factors

Moreover, participants acknowledged the significance of contextual factors, which refer to factors that cannot be (or are difficult to be) captured in data but have a significant influence on clinical decision-making. Diverse perspectives were expressed on the usefulness of an AI model to reduce the influence of subjective factors in their decision-making. Some participants cited that being supported by objective data rather than ‘gut feeling’ could be highly beneficial in preventing or reducing the influence of subjective factors in their decision-making process:

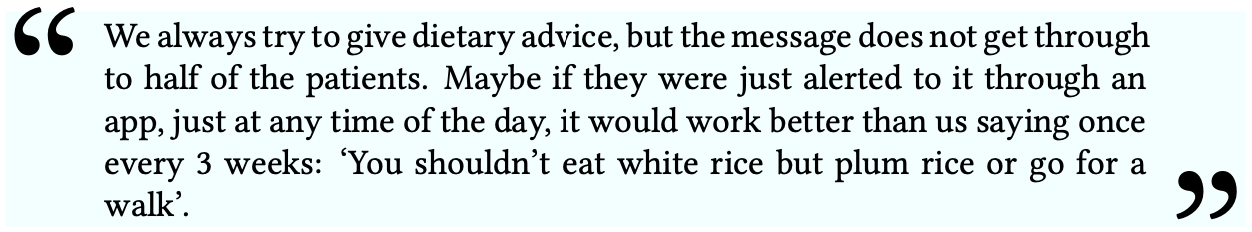


Other participants described their ability to put data into the context of the patient as their most added value as physicians. Participants viewed this as something AI could never master, nor something they perceived as desired for an AI model to do. They advocated for an AI model as a decision support tool, and not as a decisive tool:



### Time constraints

Further advantages were also noted when participants anticipated that an AI model could provide (remote) support to patients that they were unable to offer due to time constraints. They also identified potential benefits when they believed that the current one-size-fits-all strategy could be personalized to better suit individual patients. In this case, the potential benefit of AI lies in its ability to offer a solution that would otherwise be unattainable within the participants' time capacity:

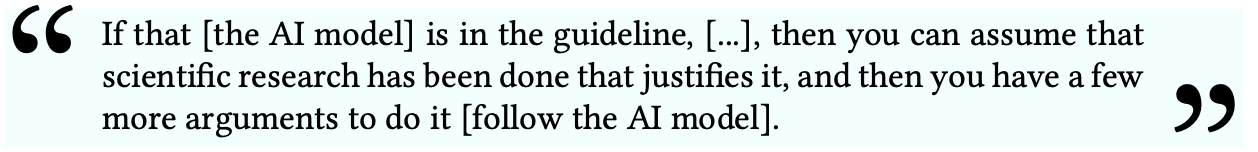


## Trust

The second theme is trust. This theme encompasses three components, namely validation, explainability and successful personal experience.

### Validation

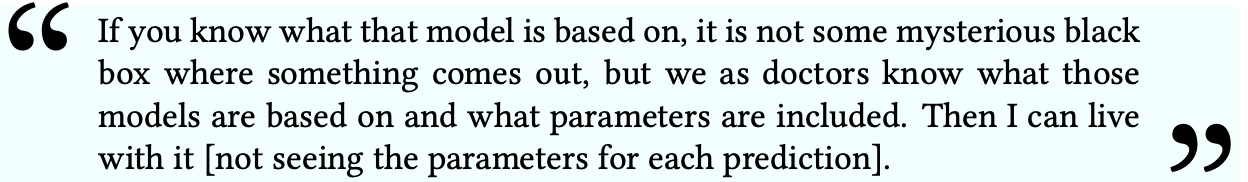
According to the participants, validation of an AI model should consist of two elements: scientific validation and endorsement of the model's integration into the professional standard by incorporating it into guidelines. Most important aspects of scientific validation were considered to be a large cohort (big data) on which an AI model was trained and tested by means of an RCT or similar research. Participants believed that the professional community's endorsement was critical because it was presumed that a model will only be accepted if it has undergone scientific validation:



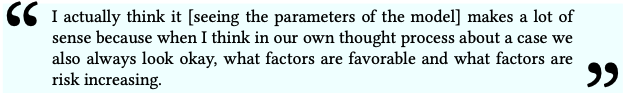
As such, the professionals making the professional guidelines, may act as gatekeepers for introducing new technologies or work practices in the workplace. Furthermore, respondents felt that they were used to working with guidelines, so an addition to the guidelines would best fit their current way of working.

### Explainability

Even though incorporation into the guidelines was considered a trustworthy quality check, participants had different opinions on the type of information deemed important or necessary to gain trust and confidence in working with an AI model. Certain participants held the belief that being informed about the impact of parameters on model predictions is superfluous when using an AI model, provided that the model has undergone scientific validation and the user has a general understanding of its parameters:



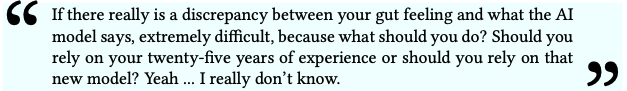
Others felt that seeing how the parameters affected the output of the model will, first, help them understand to some extent how the model works which gives them a degree of confidence during clinical decision-making. Second, they consider this consistent with how they are trained in clinical reasoning:

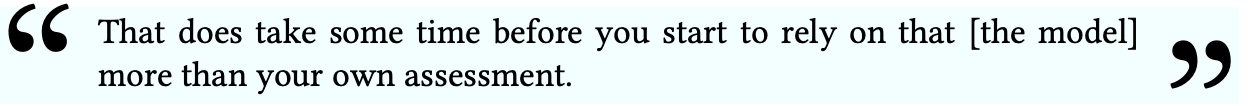


Participants also noted that if the model is based on parameters that are known risk factors related to the clinical outcome, their a priori confidence in the model would be higher and the model's output would be more likely to be considered correct.

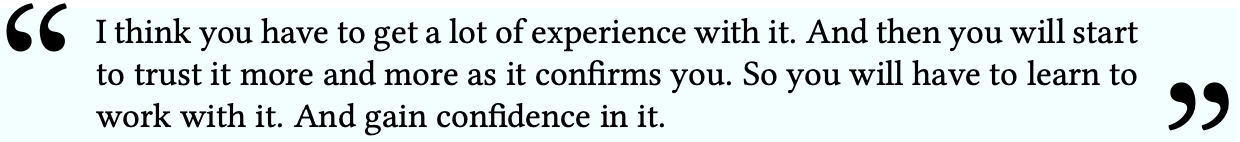
### Successful personal experience

Participants were explicitly asked what steps they would take if the AI model presented them with information that was inconsistent with other information or personal clinical view. Participants emphasized that, especially when they have to work with an AI model in the beginning, other sources of information, such as resources participants have been working with for a longer period or gut feeling, will carry greater weight in their decision-making. Additionally, they found it difficult to foresee how this would work out in practice:





Successful personal experience, i.e., that the AI model correctly predicts a clinical outcome, is considered a critical factor for regular use of the AI model. Participants saw ‘time’ as an important factor to gain experience and build trust in the AI model:



# Discussion

## Main Findings

This study was designed to explore the perceptions of obstetricians about the potential value of AI in clinical practice. Our analysis revealed two major themes: perceived usefulness and trust. We saw two interesting paradoxes between the two major themes, i.e., what the clinicians found most valuable in an AI model and what they considered necessary to trust an AI model.

First, clinicians expect added value from AI models in a matter that is best described as AI being able to grasp more than humans can grasp. At the same time, participants felt that knowing which parameters a model is built on or seeing how parameters have influenced a prediction is a crucial step towards trust and initial uptake of an AI model. The second paradox relates to the fact that clinicians see most value in a model based on as many parameters as possible. While at the same time indicating that some contextual factors can only be taken into account by humans and to some it is even undesirable to let an AI model use this information.

## Interpretation: considerations for future AI applications

The first paradox is relevant for both clinical and technical side. Regarding expectation management for clinicians with respect to the use of AI in complex decision-making, there is a need to consider the limitations of AI in providing absolute certainty. The desire to know what parameters are being used, and more specifically, the desire that these parameters are based on current evidence based medical practice, puts constraints on the technical side of developing an AI model. In recent years, a sub-field within AI, called Deep Learning (DL), has become increasingly popular and has shown to achieve superior model performance, mainly on the medical imaging domain (30). DL differs from more traditional AI models because it develops its own representations from raw data needed for pattern recognition, and unlike other AI models, does not rely on parameters explicitly specified by humans. For more deterministic domains such as medical imaging, less model interpretability might not be a problem and is easier to pinpoint, e.g., location of expected tumour indicated. But for high stake scenarios in obstetrics, where professionals need to be convinced of the action to take, based on the output of the AI model, interpretability might be crucial (31). Moreover, in line with the work of Tonekaboni et al. (20), clinicians find a level of transparency, that allows them to use their domain knowledge to validate model outputs, of great value. This shows that despite the emergence of DL, clinicians may not yet have reached the point of endorsing this technology as this does not rely on parameters that are described in current evidence based medical practice.

The second paradox of clinicians wanting comprehensive AI models with many parameters, yet acknowledging the importance of contextual factors that AI cannot or should not capture, is relevant in light of evidence-based medicine and demonstrating that an AI model adds value. Namely, a traditional evaluation method such as a randomized controlled trial (RCT) may not be sufficient if clinicians base clinical interventions on both an AI model and other contextual factors. In other words, a method to measure effectiveness of an AI model in terms of clinical outcomes is not straightforward and other metrics, such as changes in decision-making or workflow, may make more sense. This would be in line with recent RCTs regarding the usefulness of AI systems in healthcare, where other metrics have been used to provide a more holistic view of its impact (32).

Moreover, there exist a tension between the desire for an AI model that is based on big data and the preference of participants to employ AI technology solely when it is recommended by the guidelines. The process of revising a guideline may require several years and typically necessitates the availability of evidence, preferably from RCTs, which have an average duration of 5.5 years (33). While the strength of AI systems lies in its ability to get an update as more data becomes available and in effect become better at its task. However, a model update can induce variations in the AI system's performance, implying that the efficiency of an AI model may change after an AI model has been implemented in the workplace. An open question remains whether an AI model should undergo a new RCT to again prove its value.

Although new AI-specific extensions on protocols have been incorporated into clinical trial guidelines (34, 35) they explicitly exclude evolving AI systems. To date, no established methods exist for evaluating the quality of such AI models (36). This does not necessarily have to lead to constraints on clinical or technical side, but requires careful coordination with clinicians and AI researchers on what would be the best strategy on incorporating AI models in obstetrics.

## Strengths and limitations

This was the first qualitative study to use the analysis of in-depth interviews to understand obstetric clinicians' perceptions about AI. We managed to distil important themes based on hypothetical scenarios. Drawing conclusions, based on hypothetical scenarios that involve obstetric clinicians interacting with an AI model, may have limitations. What people say they will do in a given scenario is not necessarily the same as what people will actually do. However, since no AI applications have been applied in obstetric clinical settings, it was not possible a priori to observe how people will interact with an AI model.

A second limitation might be selection bias. We applied convenience sampling in our own network leading to only Dutch, research minded participants. Nevertheless, any new technology requires forerunners from the field, and clinicians working in research hospitals may be more likely to encounter such technology.

# Conclusion

Our findings provide insight into what obstetricians see as potential value of AI within their field, opening up a discussion on how clinicians and AI researchers should move forward. Participants emphasised that when faced with new technology in their workflow, the perceived benefit to them was that it improved decision-making, facilitated consideration of multiple factors, or helped them work more efficiently. The establishment of trust in AI models can be achieved through conventional means, including RCTs and incorporation into established guidelines. Moreover, our analysis shows that improving clinical outcome may not be the holy grail when it comes to developing AI models. But rather metrics that provide a holistic view of its impact.

Identifying which AI models are perceived as most useful by clinicians and designed in a way that trust is established among the professional group, should ensure that future AI applications have the greatest chance of successful implementation in clinical practice. Further research endeavours are needed to lay a foundation for evaluating AI systems that are updated over time, for which conventional validation methods such as an RCT may not suffice or be feasible.

# Disclosure of interest

All authors report no conflict of interest.

# Contribution to authorship

Study concept and design was by AF, AR, PT, MH and PB. Acquisition of data was by AF. Analysis and interpretation of data was by AF, AR, PT and PB. AF drafted the manuscript. Critical revision of the manuscript for important intellectual content was by AR, PT, MH and PB.

# Details of ethics approval

Ethical approval was obtained from the medical ethics committee of the Amsterdam UMC hospital, location AMC (reference number: W22\_363 # 22.431, date of approval 11 October 2022).

# Funding

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

# Acknowledgements

We would like to thank all the clinicians who volunteered their valuable time for this interview study.

# References

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology. 2017;2(4).

2. Rong G, Mendez A, Assi EB, Zhao B, Sawan M. Artificial intelligence in healthcare: review and prediction case studies. Engineering. 2020;6(3):291-301.

3. Wolff J, Pauling J, Keck A, Baumbach J. Success factors of artificial intelligence implementation in healthcare. Frontiers in digital health. 2021:51.

4. van de Sande D, Van Genderen ME, Smit JM, Huiskens J, Visser JJ, Veen RE, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health & Care Informatics. 2022;29(1).

5. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ digital medicine. 2020;3(1):118.

6. Dhombres F, Bonnard J, Bailly K, Maurice P, Papageorghiou AT, Jouannic J-M. Contributions of artificial intelligence reported in obstetrics and gynecology journals: systematic review. Journal of Medical Internet Research. 2022;24(4):e35465.

7. Fergus P, Hussain A, Al-Jumeily D, Huang D-S, Bouguila N. Classification of caesarean section and normal vaginal deliveries using foetal heart rate signals and advanced machine learning algorithms. Biomedical engineering online. 2017;16(1):1-26.

8. Idowu IO, Fergus P, Hussain A, Dobbins C, Khalaf M, Eslava RVC, et al., editors. Artificial intelligence for detecting preterm uterine activity in gynecology and obstetric care. 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing; 2015: IEEE.

9. Ahn KH, Lee K-S. Artificial intelligence in obstetrics. Obstetrics & Gynecology Science. 2022;65(2):113-24.

10. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. Journal of medical Internet research. 2020;22(6):e15154.

11. MacDonald S, Steven K, Trzaskowski M. Interpretable AI in Healthcare: Enhancing Fairness, Safety, and Trust. Artificial Intelligence in Medicine: Applications, Limitations and Future Directions: Springer; 2022. p. 241-58.

12. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Annals of Operations Research. 2020;294:567-92.

13. Hu Y, Jacob J, Parker GJ, Hawkes DJ, Hurst JR, Stoyanov D. The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. Nature Machine Intelligence. 2020;2(6):298-300.

14. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Translational Vision Science & Technology. 2020;9(2):45-.

15. Khanijahani A, Iezadi S, Dudley S, Goettler M, Kroetsch P, Wise J. Organizational, professional, and patient characteristics associated with artificial intelligence adoption in healthcare: A systematic review. Health Policy and Technology. 2022:100602.

16. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implementation science. 2016;11(1):1-12.

17. Corner A, Pidgeon N, Parkhill K. Perceptions of geoengineering: public attitudes, stakeholder perspectives, and the challenge of ‘upstream’engagement. Wiley Interdisciplinary Reviews: Climate Change. 2012;3(5):451-66.

18. Fiorino DJ. Citizen participation and environmental risk: A survey of institutional mechanisms. Science, Technology, & Human Values. 1990;15(2):226-43.

19. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qualitative health research. 2017;27(4):591-608.

20. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A, editors. What clinicians want: contextualizing explainable machine learning for clinical end use. Machine learning for healthcare conference; 2019: PMLR.

21. Adhya J, Li C, Eisenmenger L, Cerejo R, Tayal A, Goldberg M, et al. Positive predictive value and stroke workflow outcomes using automated vessel density (RAPID-CTA) in stroke patients: One year experience. The Neuroradiology Journal. 2021;34(5):476-81.

22. Thoral PJ, Fornasa M, de Bruin DP, Tonutti M, Hovenkamp H, Driessen RH, et al. Explainable machine learning on AmsterdamUMCdb for ICU discharge decision support: uniting intensivists and data scientists. Critical care explorations. 2021;3(9).

23. de Vos J, Visser LA, de Beer AA, Fornasa M, Thoral PJ, Elbers PW, et al. The potential cost-effectiveness of a machine learning tool that can prevent untimely intensive care unit discharge. Value in Health. 2022;25(3):359-67.

24. Nimri R, Battelino T, Laffel LM, Slover RH, Schatz D, Weinzimer SA, et al. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. Nature medicine. 2020;26(9):1380-4.

25. Nimri R, Oron T, Muller I, Kraljevic I, Alonso MM, Keskinen P, et al. Adjustment of insulin pump settings in type 1 diabetes management: advisor pro device compared to physicians’ recommendations. Journal of diabetes science and technology. 2022;16(2):364-72.

26. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Family medicine and community health. 2019;7(2).

27. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative research in psychology. 2006;3(2):77-101.

28. Glaser BG, Strauss AL. Discovery of grounded theory: Strategies for qualitative research: Routledge; 2017.

29. Corbin JM, Strauss A. Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative sociology. 1990;13(1):3-21.

30. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. Nature medicine. 2019;25(1):24-9.

31. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics. 2018;19(6):1236-46.

32. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nature medicine. 2022;28(1):31-8.

33. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR mHealth and uHealth. 2016;4(3):e5720.

34. Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ, Ashrafian H, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. The Lancet Digital Health. 2020;2(10):e549-e60.

35. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, Ashrafian H, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. The Lancet Digital Health. 2020;2(10):e537-e48.

36. Lee CS, Lee AY. Clinical applications of continual learning machine learning. The Lancet Digital Health. 2020;2(6):e279-e81.