Topic Exploration Report [1]

AI assisted electronic fetal monitoring

What is a Topic Exploration Report?

Topic Exploration Reports are not health technology assessments. These reports provide a high-level briefing on new topics submitted to Health Technology Wales and are not based on exhaustive or systematic literature searches. Instead, they rely on a focussed scan of key resources.

What evidence is used in a Topic Exploration Report?

Priority is given to summarising the most relevant or useful evidence, rather than covering all possible evidence. Information reported is typically based on abstracts and study authors' own conclusions, rather than detailed scrutiny of full texts.

What are the aims of a Topic Exploration Report?

Topic Exploration Reports offer an overview of the available evidence on a topic and aim to highlight any uncertainties or gaps in the evidence. These reports outline the quantity and type of evidence found, but no critical appraisal or formal evidence synthesis is conducted.

How should a Topic Exploration Report be used?

Topic Exploration Reports can be used to indicate what evidence may be available for a topic, and do not provide definitive guidance on how a technology should be used. The evidence presented within the reports should be interpreted with caution.

[1] Cyfieithu dogfennau HTW wedi'u cyhoeddi o'r Saesneg i'r Gymraeg
Translation of published technical HTW documents from English into Welsh

Topic exploration report number	TER619
Topic	AI assisted electronic fetal monitoring
Summary of findings	Electronic fetal heart rate monitoring, typically using continuous cardiotocography (CTG), is used during labour in higher risk pregnancies to monitor fetal well-being and to provide early detection of any reduction in fetal oxygen saturation. AI or software assisted fetal monitoring may improve processing, analysis and interpretation of fetal monitoring. We identified two different ways that AI-assisted CTG is deployed: 1) Bedside decision-support software systems which assist in the analysis of CTG at the point-of-care. 2) Central monitoring stations with real-time alerts, which use AI to remotely analyse CTG traces. We identified one systematic review of four randomised controlled trials (RCTs) using four different AI or software assisted systems. Three RCTs, two of which were large UK-based trials, found no evidence for a difference for perinatal outcomes between AI-assisted CTG and conventional CTG. An additional US-based real-world pilot study found no evidence for a difference for perinatal outcomes after the introduction of AI-assisted CTG. One Bulgarian study reported lower rates of adverse perinatal outcomes for AI-assisted CTG, but this may not be generalisable to Wales. A cost-consequence analysis of one of the UK-based RCTs reported no evidence for a difference for maternal health related quality of life or costs between AI-assisted CTG and conventional CTG. We also identified a scoping review of 40 studies that reported on validation outcomes. All studies used previously collected CTG traces. It is uncertain from the scoping review as to how many studies compared AI-assisted or AI-alone CTG interpretation with human interpretation. For AI interpretation, sensitivity of between 63% to 100% and specificity of between 66% and 98% for classifying CTGs was reported. Three further primary studies that compared AI with human interpretation had mixed findings. Overall, there appear to be multiple different AI-algorithms that have been developed to assist in the interpretation of CTG traces, some of which show good diagnostic accuracy when used on previously recorded traces. When compared with conventional CTG in clinical practice most studies did not find evidence for a difference in perinatal outcomes, costs or maternal health related quality of life between AI-assisted and conventional CTG.

Introduction and aims

Fetal heart rate monitoring is used during labour primarily to provide early detection of fetal hypoxia. If not treated, fetal hypoxia can progress to asphyxia leading to brain and heart damage and nervous system failure. In low-risk pregnancies, fetal heart rate is monitored using intermittent auscultation with a Pinard stethoscope or doppler ultrasound at least every fifteen minutes. For higher-risk pregnancies or if there is evidence for increased risk during labour, electronic fetal monitoring (EFM), typically continuous cardiotocography (CTG) monitoring, is used. CTG uses external ultrasound sensors placed on the mother's abdomen and sometimes internal sensors to record fetal heart rate and uterine activity. Both measurements can be displayed on a screen and are traced onto a paper slip which is reviewed for suspicious changes by trained healthcare professionals (NICE 2022).

Despite the existence of clinical guidelines and training in interpretation of EFM outputs, there are concerns about variability between operators and the difficulty of interpreting normal variations in fetal heart rate and uterine activity that occur during labour. Computer algorithms have been developed to assist in interpretation, including algorithms developed using artificial intelligence (AI), such as machine learning, artificial neural networks, deep learning and nature-inspired optimisation techniques. It is suggested that computer or AI assistance may improve the processing, analysis, and interpretation of fetal monitoring. AI developed algorithms are purported to be especially useful for noise suppression, feature detection and fetal state classification. If AI or computer-assistance can improve fetal monitoring, it may lead to more accurate and earlier diagnosis of issues which may, in turn, lead to improved outcomes. HTW researchers identified two CE marked commercial systems, available in the UK. The K2 INFANT-Guardian (K2 Medical Systems) is a bedside monitoring system and Omniview-SisPorto (Duomed) is a central monitoring station for remote CTG.

Health Technology Wales researchers searched for evidence on the clinical and cost-effectiveness of AI assisted electronic fetal monitoring in comparison with conventional electronic fetal monitoring.

Evidence overview

We identified one systematic review and one scoping review, three primary studies published since the reviews and one economic evaluation.

Secondary evidence

Tsipoura et al. (2023) conducted a systematic review of randomised controlled trials that compared perinatal outcomes such as such as caesarean section, operative vaginal delivery, fetal hypoxia, admissions to a neonatal unit and low Apgar score using conventional CTG versus computerised CTG. Searches were conducted up to June 2023. Five studies published between 2016 and 2021, describing four RCTs, were included. A UK-based trial with 7,730 participants (Nunes et al. 2017) and an Italian trial with 28 participants (Saccone et al. 2021) , used central monitoring systems to compare conventional CTG with CTG with computer analysis and real-time alerts. Neither study reported a difference in outcomes between the groups. The other two trials, the UK-based INFANT study (Brocklehurst et al. 2017), with 46,042 participants and a Bulgarian study with 720 participants (Ignatov & Lutomski 2016), used bedside decision-support software systems to assist in the analysis of CTG. The INFANT study reported no difference in perinatal outcomes or developmental assessment at the age of two between the groups. Ignatov & Lutomski (2016) reported lower incidence of fetal hypoxia and metabolic acidosis, lower risk of caesarean delivery and less likelihood of admission to neonatal intensive care in the computer-assisted group. Patients were recruited between March 2008 and March 2011, and it is unclear as to whether this study is generalisable to Wales. The systematic review does not report whether all the computer algorithms used in the

Evidence overview
RCTs were developed using AI, but the INFANT study used the K2 INFANT-Guardian system and Nunes et al. (2017) used Omniview-SisPorto v4.0. Aeberhard et al. (2023) conducted a scoping review of AI-assisted interpretation of CTG . Searches were conducted up until 2020. Forty studies were included, describing at least 20 different algorithms. Five commercially available systems were included, although it is unknown as to how many of the systems are still available, and those that are, may require re-evaluation due to updates to the algorithms. No additional RCTs were identified and most of the included studies were validation studies. It is not completely clear from the review as to how validation was undertaken for most algorithms, or whether there were comparisons with physician interpretation of CTG. For AI interpretation, sensitivity of between 63% to 100% and specificity of between 66% to 98% for classifying CTGs were reported. **Primary evidence** Lowery et al. (2023) conducted a real-world study evaluating an AI-assisted centralised remote electronic fetal monitoring surveillance system in the US. The pilot ran for six months in early 2021. In the six months before implementation there were 2,407 births and during the pilot there were 2,582 births. The study reported no evidence for a difference between groups for neonatal outcomes, although 78.8% of nurses who responded to a survey on safety indicated that they thought that use of the system improved safety. The authors report that there were few adverse events in either group, making it difficult to demonstrate that use of the system improved outcomes. We also identified three additional diagnostic accuracy studies that compared AI analysis of CTG with human analysis. These studies used previously collected CTG traces and were not conducted in real-world settings. Miyata et al. (2025) compared analysis of 50 CTG traces and reported that human diagnosis alone was superior to algorithms developed by machine learning or by deep learning used alone. Analysis by humans with the assistance of deep learning algorithms improved specificity (92% for humans alone vs 98% for AI-assistance) which the authors indicate could reduce false positive rates. Tarvonen et al. (2024) compared vision analysis of CTG recordings and cardiograms from 4,988 singleton term births between expert obstetricians and computer vision analysis, using a model labelled SALKA. Average agreement in recognising increased fetal heart rate variability between experts and SALKA was almost perfect (Cohen's kappa 0.98). Sensitivity for SALKA was 98%, positive predictive rate was 82% and false negative rate was 1%. In the 146 cases with known neonatal acidemia, the agreement was 0.99. Liu et al. (2021) compared an AI model based on deep learning with clinical practice using 3,239 EFM records. There was weak agreement between physicians and AI (Cohen's kappa 0.53). The AI model showed higher sensitivity for predicting fetal compromise (53% vs 13%), but a higher false-positive rate (63% vs 1%) compared with clinical practice. **Economic evaluation** Schroeder et al. (2021) conducted a cost-consequence analysis using individual patient led data from the INFANT study. Outcomes were a composite 'poor neonatal outcome', developmental assessment at two years of age, mean costs per mother and infant dyad from birth to hospital discharge and from hospital discharge to 24 months follow up and maternal health related quality of life at 12 and 24 months. Data were analysed for 46 042 women and 46 614 infants. There was no evidence for a difference between trial arms in any of the clinical outcomes or maternal quality of life. There was no evidence for a difference maternal or infant costs from trial entry to hospital discharge or from hospital discharge to 24-month follow-up. The authors concluded that the decision support software would not lead to additional costs or savings to the NHS.

Areas of uncertainty
• There have been advancements in AI technology since the RCTs were conducted but most newer algorithms appear not to have been tested in clinical settings. • Different AI algorithms report very different sensitivities and specificities • Staff training needs.

Literature search results

Health technology assessments and guidance
NICE. (2022). Fetal monitoring in labour. National Institute for Health and Care Excellence Guideline [NG229]. Available at: https://www.nice.org.uk/guidance/ng229 [Accessed 30 July 2025].
Evidence reviews and economic evaluations
Aeberhard JL, Radan AP, Delgado-Gonzalo R, et al. (2023). Artificial intelligence and machine learning in cardiotocography: A scoping review. Eur J Obstet Gynecol Reprod Biol. 281: 54-62. doi: https://doi.org/10.1016/j.ejogrb.2022.12.008 Tsipoura A, Giaxi P, Sarantaki A, et al. (2023). Conventional Cardiotocography versus Computerized CTG Analysis and Perinatal Outcomes: a Systematic Review. Maedica (Bucur). 18(3): 483-9. doi: https://doi.org/10.26574/maedica.2023.18.3.483
Individual studies
Liu LC, Tsai YH, Chou YC, et al. (2021). Concordance analysis of intrapartum cardiotocography between physicians and artificial intelligence-based technique using modified one-dimensional fully convolutional networks. J Chin Med Assoc. 84(2): 158-64. doi: https://doi.org/10.1097/jcma.0000000000000416 Liu LC, Tsai YH, Chou YC, et al. (2021). Concordance analysis of intrapartum cardiotocography between physicians and artificial intelligence-based technique using modified one-dimensional fully convolutional networks. J Chin Med Assoc. 84(2): 158-64. doi: https://doi.org/10.1097/jcma.0000000000000416 Miyata K, Shibata C, Fukunishi H, et al. (2025). Cardiotocography-Based Experimental Comparison of Artificial Intelligence and Human Judgment in Assessing Fetal Asphyxia During Delivery. Cureus. 17(1): e78282. doi: https://doi.org/10.7759/cureus.78282
Ongoing research
No ongoing trials identified
Individual randomised controlled trials included in the reviews
Ignatov PN, Lutomski JE. (2016). Quantitative cardiotocography to improve fetal assessment during labor: a preliminary randomized controlled trial. Eur J Obstet Gynecol Reprod Biol. 205: 91-7. doi: https://doi.org/10.1016/j.ejogrb.2016.08.023 Nunes I, Ayres-de-Campos D, Ugwumadu A, et al. (2017). Central Fetal Monitoring With and Without Computer Analysis: A Randomized Controlled Trial. Obstetrics & Gynecology. 129(1). doi: https://doi.org/10.1097/AOG.0000000000001799 Saccone G, Tagliaferri S, Grasso A, et al. (2021). Antenatal cardiotocography with and without computer analysis in high-risk pregnancy: a randomized clinical trial. Am J Obstet Gynecol MFM. 3(1): 100284. doi: https://doi.org/10.1016/j.ajogmf.2020.100284 Tarvonen M, Manninen M, Lamminaho P, et al. (2024). Computer Vision for Identification of Increased Fetal Heart Variability in Cardiotocogram. Neonatology. 121(4): 460-7. doi: https://doi.org/10.1159/000538134

Date of search	July 2025
Concepts used	Electronic Fetal Monitoring; Cardiotocography; Artificial Intelligence; AI; Machine learning; computerised

Proposed research question and evidence selection criteria (if selected)

Proposed Research question	**What is the clinical and cost-effectiveness of AI assisted electronic fetal monitoring in comparison with conventional electronic fetal monitoring.**

	Inclusion criteria	Exclusion criteria
Population	Women, at 35 weeks gestation or further	
Intervention	AI-assisted continuous electronic fetal heart rate monitoring	
Comparison/ Comparators	Conventional continuous electronic fetal heart rate monitoring Intermittent auscultation	
Outcome measures	Neonatal outcomes such as birth-related deaths and morbidity, evidence of asphyxia. Apgar score Caesarean section rates Admissions to a neonatal unit Health related QoL Other resource use Economic outcomes	

Proposed speciality	**Obstetrics and gynaecology**