The Noninvasive Fetal Electrocardiogram During Labor

CME REVIEW ARTICLE

16

The Noninvasive Fetal Electrocardiogram During Labor: A Review of the Literature

**Carlijn Lempersz, MD,*† Lore Noben, MD,*† Bart de Vries, MA,‡
Judith O. E. H. van Laar, MD, PhD,† §
Michelle E. M. H. Westerhuis, MD, PhD,¶ and S. Guid Oei, MD, PhD§ ‖**

*PhD Candidate, Department of Obstetrics and Gynecology, Máxima Medical Centre, Veldhoven; †Department of Electrical Engineering, University of Technology Eindhoven, Eindhoven; ‡Clinical Librarian, Knowledge and Information Center, and §Gynecologist-Perinatologist, Department of Obstetrics and Gynecology, Máxima Medical Centre, Veldhoven; ¶Gynecologist, Department of Obstetrics and Gynecology, Catharina Hospital, Eindhoven; and ‖Professor, Department of Electrical Engineering, University of Technology Eindhoven, Eindhoven, the Netherlands

Importance: The introduction of the cardiotocogram (CTG) during labor has not been found to improve neonatal outcome. The search for a more reliable, less invasive, and patient-friendly technique is ongoing. The noninvasive fetal electrocardiogram (NI-fECG) has been proposed as one such alternative.

Objectives: The aim of this study was to review the literature on the performance of NI-fECG for fetal monitoring during labor. Following the PRISMA guidelines, a systematic search in MEDLINE, EMBASE, and Cochrane Library was performed. Studies involving original research investigating the performance of NI-fECG during labor were included. Animal studies and articles in languages other than English, Dutch, or German were excluded. The QUADAS-2 checklist was used for quality assessment. A descriptive analysis of the results is provided.

Results: Eight articles were included. Pooled analysis of the results of the separate studies was not possible due to heterogeneity. All studies demonstrate that it is possible to apply NI-fECG during labor. Compared with Doppler ultrasound, NI-fECG performs equal or better in most studies.

Conclusions and Relevance: NI-fECG for fetal monitoring is a promising noninvasive and patient-friendly technique that provides accurate information. Future studies should focus on signal quality throughout labor, with the aim to further optimize technical development of NI-fECG.

Target Audience: Obstetricians and gynecologists, family physicians.

Learning objectives: After completing this activity, the learner should be better able to interpret the performance of both noninvasive fetal electrocardiography and Doppler ultrasound for fetal heart rate monitoring during labor; assess the advantages and disadvantages of monitoring fetal heart rate with the current noninvasive fetal electrocardiogram during labor; and summarize the definitions of the different performance measures.

The cardiotocogram (CTG) for fetal heart rate (FHR) and contraction monitoring during labor was introduced in the early 1970s to identify fetuses with hypoxia and to reduce neonatal morbidity and mortality.[1] Unfortunately, neonatal outcome has not improved after the introduction of the CTG.[1]

The 2 most commonly used techniques to acquire FHR for CTG monitoring is via a noninvasive method using Doppler ultrasound (DU) or with a more invasive method, the fetal scalp electrode (FSE). Doppler ultrasound uses a transducer placed on the maternal abdomen and held in place with an elastic band. An advantage of the DU is that it is a noninvasive method that can be used before membranes have been ruptured. Unfortunately, DU is sensitive to signal loss with reported percentages ranging from 5.2% up to 40%.[2–4] This signal loss can partially be due to maternal and fetal movements, a high BMI of the mother, and irregularities of the FHR, that is, decelerations, extrasystolic beats, and other cardiac

arrythmias.[3–6] Furthermore, this method and the means of attaching the DU device to the maternal abdomen can be experienced as uncomfortable.[7] Invasive monitoring via FSE is a more reliable method and is considered the criterion standard for FHR monitoring. However, this method carries an increased risk for complications, such as trauma and infection, and can only be applied after membranes have been ruptured and with sufficient dilation.[8,9]

Overall, the specificity of CTG monitoring is poor.[1] Multiple techniques have been added to increase the detection rate of fetal hypoxia, that is, fetal blood sampling and ST waveform analysis (STAN). However, previous studies have demonstrated that these methods do not significantly decrease neonatal morbidity and mortality.[10–16]

The search for other monitoring techniques that can gather accurate information in a safe and patient-friendly way is still ongoing. Noninvasive fetal electrocardiography (NI-fECG) may be an alternative to conventional monitoring techniques. The NI-fECG retrieves electrophysiologic signals of fetal and maternal heart rate (MHR), as well as the electrohysterogram via electrodes placed on the maternal abdomen. These techniques provide more information than FHR alone as it also provides beat-to-beat information that can be used to assess FHR variability. Furthermore, the NI-fECG provides a complete fetal ECG waveform that could be assessed for morphologic changes possibly indicating fetal hypoxia. In contrast to STAN, the NI-fECG provides a multilead fetal ECG and therefore may overcome the current shortcomings in ST waveform analysis.[17–19]

NI-fECG is not a new technique, as first recordings were made in the early 1900s.[20] However, difficulties in acquiring and processing the electrophysiologic signals limited the development of this technique. Recently, NI-fECG has gained renewed interest due to technical improvements. Over the last years, more research has been performed on NI-fECG as an alternative for intrapartum fetal monitoring.

This article aims to provide a review of the existing literature on the performance of NI-fECG as a method for fetal monitoring during labor.

MATERIALS AND METHODS

This review was registered in Prospero (#CRD42019124807). A systematic search in the electronic databases MEDLINE (1966–present), EMBASE (1974–present), and Cochrane library was performed until the April 24, 2019. The search was conducted following the PRISMA guidelines by 2 independent researchers (L.N. and C.L.) and one trained medical librarian (B.d.V.) from the Máxima Medical Center, Veldhoven, the Netherlands. The following search terms were used: fetus, electrocardiography,

cardiotocography, fetal monitoring, noninvasive, labor, intrapartum (full electronic search is available in Appendix 1). The main outcome measures of interest were accuracy and reliability of the NI-fECG during labor compared with DU and/or FSE.

We only included original research. If there was any overlap between studies, we used the original article. Animal studies and articles in languages other than English, Dutch, or German were excluded.

Articles were initially screened by title and abstract by 2 independent reviewers (L.N. and C.L.). When found appropriate, the full text was evaluated. Furthermore, references of the selected articles were checked for eligible articles. Disagreements were resolved by discussion.

The QUADAS-2 checklist was used as reference for quality assessment of the included studies.[21]

RESULTS

A total of 8 of 658 articles were included in this review after removal of duplicates, title, and abstract screening, reading the full-text articles, and screening reference lists of the included articles. Seven articles describe a prospective study and one article a retrospective study. Figure 1 summarizes the screening and article selection process.

Pooled analysis of the results of the separate studies was not possible due to heterogeneity. Table 1 shows a summary of the quality assessment of the included articles. A summary of the 8 included articles is enclosed in Appendix 2.

Accuracy

Accuracy is defined as the difference in FHR output from the investigational product (NI-fECG or DU) compared with the reference method (FSE) expressed in root mean square error (bpm). This definition of accuracy was reported in 3 included studies, using the FSE as the criterion standard. Euliano et al and Cohen et al reported an overall accuracy of about 5 bpm for NI-fECG. For DU, overall accuracy was reported as 10.9 ± 5.8 bpm by Cohen et al and 14.3 ± 8.2 bpm by Euliano et al.[22,23]

No difference in accuracy of NI-fECG between labor stages was found by Euliano et al, whereas Cohen et al found a slight decrease in accuracy to 7.9 ± 4.2 bpm for the second stage of labor.[22,23]

Ashwal et al[24] reported a higher accuracy of 1.47 ± 0.82 bpm for NI-fECG and 5.39 ± 3.82 bpm for DU, using noncontinuous segments for analysis. Although they used segments from each stage of labor, they only report one accuracy value. Reported accuracy values were higher for NI-fECG compared with DU (see Appendix 2).[22–24]

Reinhard et al chose the correlation coefficient to express accuracy for NI-fECG with DU as reference.
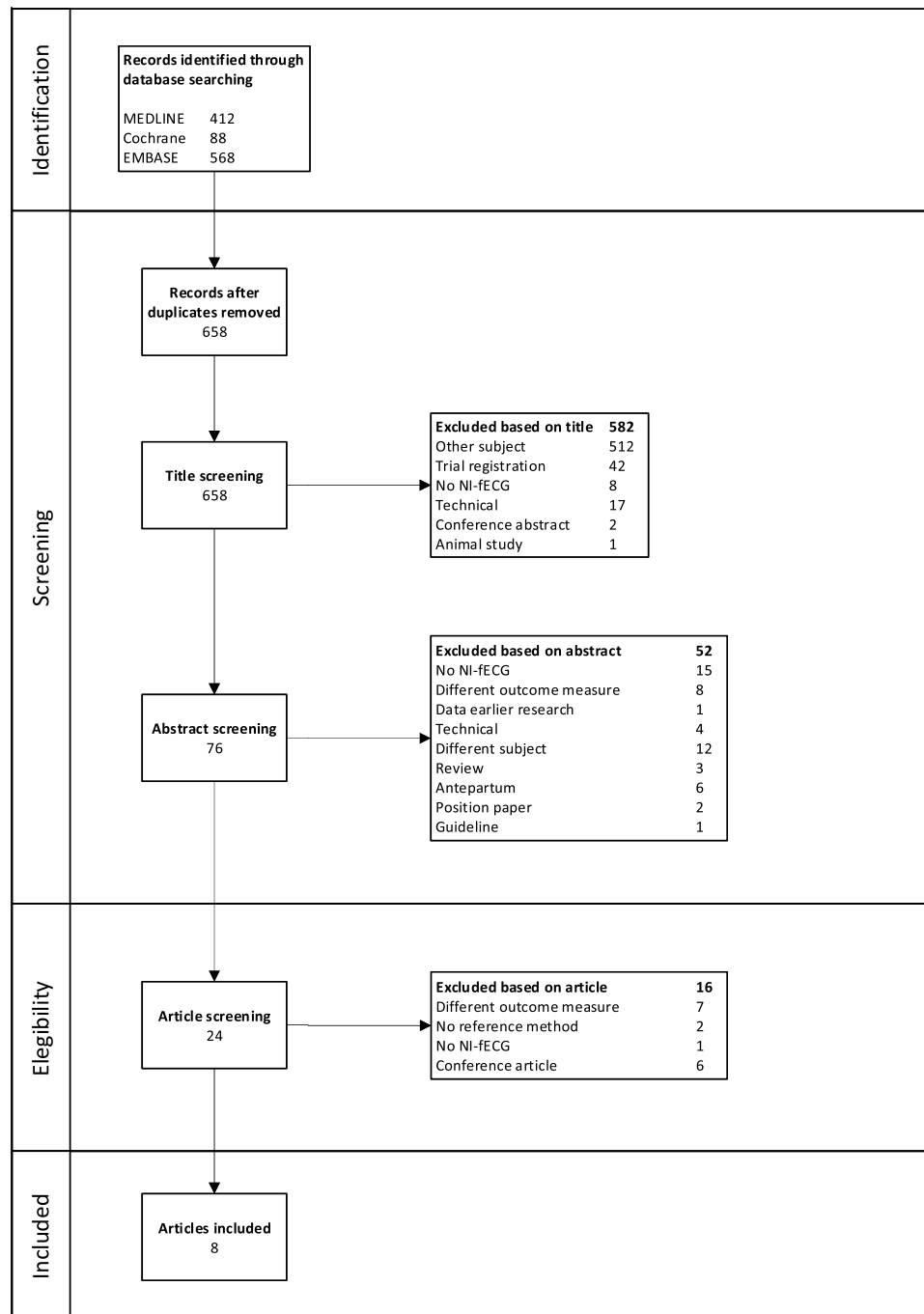
FIG. 1. Flowchart article screening and selection process.

They found a Spearman rank correlation coefficient of 0.94 (range, −0.11 to 0.99) for the first stage of labor and 0.85 (range, −0.73 to 0.99) for the second stage of labor, suggesting a good statistical agreement between both methods.[25]

Frank et al describe 5 cases of laboring women monitored by NI-fECG and FSE. Their definition of accuracy is the absolute difference in the R-R interval.

They reported that 92.6% of the total time of each NI-fECG measurement lay within 1 bpm difference of the FSE measurement.[26]

Reliability

Reliability is defined as the percentage of time that the investigational product (NI-fECG or DU) generates

TABLE 1
Quality Assessment of the 8 Included Articles According to Quadas-2[21]

	Risk of Bias				Applicability Concerns		
Study	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Breuker et al, 1976	?	+	+	+	+	−	+
Frank et al, 1992	−	+	+	+	+	−	+
Stampalija et al, 2012	+	+	−	−	+	−	−
Reinhard et al, 2012	+	+	−	+	+	−	+
Cohen et al, 2012	−	+	+	−	+	+	+
Reinhard et al, 2013	+	+	−	+	+	−	+
Ashwal et al, 2017	+	+	+	+	+	+	+
Euliano et al, 2017	?	+	+	?	+	+	+

an FHR output within 10% of the FHR output of the product used as reference (FSE), expressed as positive percent agreement (PPA). Both Cohen et al and Euliano et al compared NI-fECG with FSE. They found similar results for overall PPA for NI-fECG (81.7% ± 20.5% in Cohen et al and 83.4% ± 15.4% in Euliano et al). For DU, Cohen et al reported an overall PPA of 73.0% ± 24.6% and Euliano et al an overall PPA of 62.4% ± 26.5%, both significantly lower than NI-fECG.

When the first stage of labor was considered separately, Cohen et al found a PPA of 84.9% ± 21.5% for NI-fECG and 74.7% ± 28.2% for DU (<0.001). Euliano et al found a PPA in the first stage of labor of 86.3% ± 14.7% for NI-fECG and 61.3% ± 29.6% for DU (<0.0001). Both Cohen et al and Euliano et al describe a drop in reliability percentages for NI-fECG and DU during the second stage of labor (71.9% ± 20.4% and 77.5% ± 15.1% for NI-fECG in Cohen et al and Euliano et al, respectively, and 61.7% ± 24.8% and 64.8% ± 18.5% for DU). Overall, the reliability of NI-fECG is significantly higher than DU.[22,23] Ashwal et al[24] also used FSE as criterion standard, but found a much higher PPA of 99% ± 1.72% for NI-fECG and 96.6% ± 4.6% for DU. They showed a decrease of 0.5% for NI-fECG and 1.7% for DU during the second stage of labor.

Success Rate

Success rate is defined as the percentage of time that NI-fECG or DU provides any output. Stampalija et al reported an overall success rate of 88.5% ± 16.7% for NI-fECG and 89.4% ± 7.6% for DU (P = 0.77).[27] Cohen et al[22] found an overall success rate of 83.4% ± 20.1% for NI-fECG and 82.5% ± 21.1% (P = 0.38) for DU. Stampalija et al found a success rate of 89.8% ± 16.1% in the first stage of labor for NI-fECG and 89.9% ± 7.9% for DU (P = 0.98). In the second stage of labor, a success rate of 66.5% ± 21.3% for NI-fECG and 83.7% ± 7.4% for

DU (P = 0.001) was found.[27] Cohen et al reported a success rate of 86.4% ± 20.1% for NI-fECG in the first stage of labor and 82.6% ± 24.4% for DU. In the second stage of labor, this was 75.2% ± 19.2% and 77.8% ± 21.1% (P = 0.25), respectively.[22] Reinhard et al also reported on the success rate of NI-fECG and DU. In the first stage of labor, they found a success rate of 97.7% (7.8–100) for NI-fECG and 85.5% (35.1–99.8) for DU. In the second stage of labor, this rate dropped to 85.5% (13.4–100) for NI-fECG, but rose to 92.3% (22.5–99.8) for DU.[25] In 2013, Reinhard et al published another report with results on reliability using the aforementioned definition for success rate. The reliability reported in this article for NI-fECG was 87.1% ± 19.10% for first stage and 70.5% ± 27.90% for second stage of labor.[28]

Signal Loss

Breuker et al reported on quality defined as signal loss; the percentage of time the NI-fECG did not provide an output. The quality of the NI-fECG was assessed by placing recordings in different categories: excellent (<5% signal loss), good (>5% to <10% signal loss), satisfactory (>10% to <20% signal loss), sufficient (>20% to <35% signal loss), deficient (>35% to <50% signal loss), and not interpretable (>50% signal loss). Overall, 17.3% of the cases were classified as excellent, 23.1% as good, 26.6% as satisfactory, 17.9% as sufficient, and 15.0% as deficient. In the first stage of labor, no cases had more than 50% signal loss, whereas in second stage of labor, this was 30%.[29]

Confusion Rate

Stampalija et al, Reinhard et al, and Cohen et al reported on the percentage of time the investigational product confused MHR for FHR.[22,27,28] Stampalija et al and Cohen et al used the term confusion rate (CR) in their article, whereas Reinhard et al used the term MHR/FHR ambiguity. Stampalija et al and Reinhard

et al used the NI-fECG as the reference method for MHR. Cohen et al used pulse oximetry as the reference method for MHR.[22,27,28]

Cohen et al defined CR as the percentage of FHR determinations for which each external device (DU and NI-fECG) calculated an FHR value that was both more than 5% different from that of the FSE and within 5% of the MHR.[22] Stampalija et al defined CR as an FHR within 5 bpm of MHR.[27] Reinhard et al[28] used the same definition but called it MHR/FHR ambiguity.

All 3 studies found a lower CR for the NI-fECG compared with DU. Stampalija et al found a CR in the first stage of labor for DU and NI-fECG of 3.9% ± 4.6% and 1.0% ± 1.9%, respectively. For the second stage, this was 11.3% ± 8.2% and 4.6% ± 5.0%, respectively.[27] Cohen et al found a CR of 9.5% ± 17.8% in the first stage of labor for DU and 11.0% ± 15.4% in second stage of labor, whereas this was 0.3% ± 0.6% and 0.7% ± 0.8% for the NI-fECG, respectively.[22] Reinhard et al showed an ambiguity of DU in the first stage of labor of 1.22% ± 1.9% and for NI-fECG of 0.70% ± 1.2%. For the second stage, of labor ambiguity was 6.20% ± 9.0% for DU and 3.30% ± 4.4% for NI-fECG.[28]

DISCUSSION

The most common method of monitoring fetal well-being during labor is by monitoring the FHR in relation to uterine contractions. Unfortunately, FSE, considered the criterion standard for FHR monitoring, is invasive and carries risks for infection and trauma. Furthermore, FSE can be applied only when sufficient dilation of the cervix is achieved and membranes have ruptured. Doppler ultrasound is a noninvasive method, but shows high percentages of signal loss, especially in obese women, and it is often is experienced as uncomfortable by the patient.[7] NI-fECG is a relatively new method based on electrophysiologic monitoring performed noninvasively using electrodes on the abdomen of the mother. Recent developments in signal processing techniques and improvements of algorithms make it possible to simultaneously monitor FHR, MHR, and uterine contractions with one device in a noninvasive manner. Intrapartum monitoring by NI-fECG may therefore be an alternative for monitoring by FSE and DU. This review evaluates the performance of the NI-fECG technique during the last decade.

Performance Measures of the NI-fECG

The earliest studies describing the use of NI-fECG during labor date back to the 20th century and therefore describe the performance of NI-fECG devices that are outdated.[26,29] However, such studies substantiate the potential added value of NI-fECG during labor, even when development of the technique was in a premature stage. Breuker et al found that only 15% of the recordings were of deficient quality, and in the 5 cases described by Frank et al, 92.6% of the FHR output of the NI-fECG was within 1 bpm of the FHR measured by FSE.[26,29]

Accuracy

All studies found a higher accuracy for the NI-fECG technique compared with DU, when using FSE as reference.[22–24] Both Cohen et al and Euliano et al report an accuracy of about 5 bpm, which is noticeably higher than their reported values for DU. However, there is a risk of selection bias in these studies because they only include women who received FSE for fetal monitoring due to insufficient quality of DU. Therefore, results of the performance measures of the DU may be negatively influenced. The insufficient quality of the registration by DU may be partially explained by the high median BMI of both study populations, because it is known that DU performance worsens with increasing maternal BMI.[6,30]

Ashwal et al found a high accuracy for NI-fECG. Their reported accuracy for DU is also high compared with the literature. Because they used random segments from the total recording, it is likely not to be representative for the total measurement.[24]

Reinhard et al used the correlation coefficient as an outcome measure to reflect the accuracy of their device, using the DU method as reference. They report a good statistical agreement between NI-fECG and DU (Spearman rank correlation coefficient of 0.94 for the first stage of labor and 0.85 for the second stage of labor).[25] A correlation coefficient close to 1 means that there is a high level of agreement between the output of both devices. However, this is an inappropriate method for measuring accuracy, as the correlation coefficient only measures the strength of the linear association between variables.[31] In addition, because Reinhard et al used the DU method, which has poor performance measures compared with FSE, as reference method, this high level of agreement has no clinical importance.

Reliability

A high reliability is an important property for a medical device to be of value in clinical practice. Cohen et al and Euliano et al found similar results for overall reliability (81.7 and 83.4, respectively) for NI-fECG monitoring. Overall reliability for DU reported by Cohen et al and Euliano et al is lower than NI-fECG (73.0 and 62.4). Reliability percentages decrease during the

second stage of labor in both studies, for the NI-fECG as well as the DU technique.[22,23] This decrease in performance is a known disadvantage of the DU method, probably due to maternal movement and increased intra-abdominal pressure during the active pushing phase. Both Euliano et al and Cohen et al found higher reliability percentages for NI-fECG compared with the DU technique, also during the second stage of labor.[22,23] Reliability values reported by Ashwal et al are nearly 100%, for both the NI-fECG and the DU method. As previously described, in this study, random segments from the total measurement were used to analyze reliability.[24] The fact that their reported reliability value for the DU technique is 96.6% whereas other literature shows much lower reliability percentages for DU further supports our explanation that these random segments are not representative for the entire measurement.[22,23]

Success Rate

Only 3 articles reported on the success rate.[22,25,27] Because success rate is defined as the percentage of time the device provides output, it resembles the percentage of signal loss, without providing information on the quality of the registered information. A similar overall success rate for the NI-fECG and DU technique was reported by Cohen et al and Stampalija et al.[22,27]

In all 3 articles, a decrease in success rate of NI-fECG was noticed as labor progressed. This is also a known pitfall of the DU technique.[2] Cohen et al found similar success rates between NI-fECG and DU (83.4% and 82.5%, respectively).[22] Stampalija et al found a significantly higher success rate for DU in the second stage of labor as compared with NI-fECG, which was also described by Reinhard et al.[25,27] They also report a rise in success rate between the first and second stage of labor for DU.[25] These results may demonstrate the limitation of success rate as an outcome measure if other outcome measures are not taken into consideration.

Reinhard et al used a different definition for reliability. They defined reliability as the percentage of available FHR in the recorded period. According to this definition, they found a significant difference in reliability between NI-fECG and DU during the first stage of labor (87.09 vs 85.21) but not during the second stage of labor (70.51% vs 76.46%).[28] Because no reference method was used to compare the FHR output from DU and NI-fECG, interpretation of these results is difficult. In this setting, their definition for reliability better reflects the definition of success rate; the time the investigational device provides an output.

Confusion Rate

From the articles that reported on MHR/FHR confusion, only Cohen et al used a validated method for MHR monitoring, which is pulse oximetry.[22] Reinhard et al and Stampalija et al used the NI-fECG device as a reference method for the performance measures being researched in their study.[27,28] Theoretically, by using NI-fECG, confusion between MHR and FHR is unlikely because electrophysiological signals from the mother are relatively strong compared with those of the fetus. Because NI-fECG is measuring both MHR and FHR by a single device, these signals can be separated very well. All 3 studies demonstrated that confusion of MHR and FHR is significantly lower with NI-fECG as compared with DU.[22,27,28] This is an important characteristic, because confusion of MHR and FHR can lead to unnecessary interventions or failure to intervene where intervention was needed, sometimes leading to a seriously compromised fetus.

General Remarks

Overall, this review demonstrates that there is limited research regarding monitoring by NI-fECG during labor. Studies have small sample sizes and comparing them is difficult due to heterogeneity. Furthermore, 3 studies did not use the FSE as reference.[25,27,28] Therefore, interpretation and clinical validity of their results regarding accuracy, reliability, and CR are difficult.

Despite differences in methodology and type of NI-fECG devices, all included studies in this review demonstrate that it is possible to apply NI-fECG during labor. Compared with the currently used standard method for noninvasive fetal monitoring, which is DU, NI-fECG performs equally well, or better in most studies. Even during the second stage of labor, when a decrease in performance is noticed in most reports, it is shown that NI-fECG still performs equally or better compared with DU.[2] Studies that compare NI-fECG and DU with FSE showed that DU and NI-fECG have comparable success rates. However, compared with DU, accuracy and reliability of NI-fECG is higher and CR is lower.

In 2 studies, the success rate for NI-fECG in the second stage of labor was lower compared with DU.[25,27] These success rates are insufficient according to the FIGO criteria for accepted percentages of signal loss of ≤20%.[32] One of these studies also showed a higher FHR/MHR CR for DU, especially during the second stage of labor.[27] Although DU may have a higher success rate, the output that is generated may not always be as reliable as NI-fECG. Although in fetal monitoring it is generally desirable to have a good signal quality at

all times, it is most important to have a good balance between signal quality and the reliability of the generated output. This review therefore shows NI-fECG to be a more accurate alternative to DU.

In addition to improved test characteristics, patient satisfaction with this type of noninvasive monitoring is also better with NI-fECG, as compared with conventional noninvasive monitoring by DU.[7] Moreover, noninvasive monitoring by NI-fECG also yields several diagnostic opportunities. It may provide information on a preterm fetus, when invasive monitoring is not an option or discouraged due to contraindications. The NI-fECG provides beat-to-beat FHR, which enables the use of spectral analysis.[33] Spectral analysis can monitor the modulation of the autonomic nervous system by evaluating oscillations in beat-to-beat FHR and can differentiate between an asphyxiated and healthy fetus during labor.[34–38] Furthermore, the NI-fECG may provide information on the actual fetal ECG waveform complex, identifying other abnormalities that may indicate fetal distress. This has previously been attempted by combining FSE with STAN.[13–16] Because the NI-fECG uses multiple leads, one of the limitations of STAN, which only uses a single lead scalp electrode, is avoided.[17–19]

To conclude, NI-fECG for FHR monitoring is a promising technique that is noninvasive, is patient-friendly, and provides accurate information. Future studies should focus on evaluating and improving signal quality of the NI-fECG, especially during the second stage of labor. Prospective studies on several diagnostic opportunities of this technique may help implementing NI-fECG in daily clinical practice.

REFERENCES

1. Alfirevic Z, Devane D, Gyte GM, et al. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database Syst Rev*. 2017;2: CD006066.
2. Bakker PC, Colenbrander GJ, Verstraeten AA, et al. The quality of intrapartum fetal heart rate monitoring. *Eur J Obstet Gynecol Reprod Biol*. 2004;116:22–27.
3. Spencer JA, Belcher R, Dawes GS. The influence of signal loss on the comparison between computer analyses of the fetal heart rate in labour using pulsed Doppler ultrasound (with autocorrelation) and simultaneous scalp electrocardiogram. *Eur J Obstet Gynecol Reprod Biol*. 1987;25:29–34.
4. Dawes GS, Visser GH, Goodman JDS, et al. Numerical analysis of the human fetal heart rate: modulation by breathing and movement. *Am J Obstet Gynecol*. 1981;140:535–544.
5. van Geijn HP, Jongsma HW, de Haan J, et al. Analysis of heart rate and beat-to-beat variability: interval difference index. *Am J Obstet Gynecol*. 1980;138:246–252.
6. Cohen WR, Hayes-Gill B. Influence of maternal body mass index on accuracy and reliability of external fetal monitoring techniques. *Acta Obstet Gynecol Scand*. 2014;93:590–595.
7. Thijssen KMJ, Vlemminx MWC, Westerhuis MEMH, et al. Uterine monitoring techniques from patients' and users' perspectives. *AJP Rep*. 2018;8:e184–e191.
8. Kawakita T, Reddy UM, Landy HJ, et al. Neonatal complications associated with use of fetal scalp electrode: a retrospective study. *BJOG*. 2016;123:1797–1803.
9. Leatherman J, Parchman ML, Lawler FH. Infection of fetal scalp electrode monitoring sites. *Am Fam Physician*. 1992;45:579–582.
10. Bretscher J, Saling E. pH values in the human fetus during labor. *Am J Obstet Gynecol*. 1967;97:906–911.
11. van den Berg P, Schmidt S, Gesche J, et al. Fetal distress and the condition of the newborn using cardiotocography and fetal blood analysis during labour. *Br J Obstet Gynaecol*. 1987;94:72–75.
12. Westgate J, Greene K. How well is fetal blood sampling used in clinical practice? *Br J Obstet Gynaecol*. 1994;101:250–251.
13. Westerhuis ME, Visser GH, Moons KG, et al. Cardiotocography plus ST analysis of fetal electrocardiogram compared with cardiotocography only for intrapartum monitoring: a randomized controlled trial. *Obstet Gynecol*. 2010;115:1173–1180.
14. Saccone G, Schuit E, Amer-Wahlin I, et al. Electrocardiogram ST analysis during labor: a systematic review and meta-analysis of randomized controlled trials. *Obstet Gynecol*. 2016;127:127–135.
15. Belfort MA, Saade GR, Thom E, et al. A randomized trial of intrapartum fetal ECG ST-segment analysis. *N Engl J Med*. 2015;373:632–641.
16. Blix E, Brurberg KG, Reierth E, et al. ST waveform analysis versus cardiotocography alone for intrapartum fetal monitoring: a systematic review and meta-analysis of randomized trials. *Acta Obstet Gynecol Scand*. 2016;95:16–27.
17. Vullings R, Verdurmen KMJ, Hulsenboom ADJ, et al. The electrical heart axis and ST events in fetal monitoring: a post-hoc analysis following a multicentre randomised controlled trial. *PLoS One*. 2017;12:e0175823.
18. Hulsenboom ADJ, Vullings R, Van Laar JOEH, et al. How to improve ST-analysis in fetal monitoring: relative versus absolute t/qrs ratio rises. *J Matern Fetal Neonatal Med*. 2014;27(suppl 1):S191–S192.
19. Andriessen P, Zwanenburg A, van Laar JOEH, et al. ST waveform analysis for monitoring hypoxic distress in fetal sheep after prolonged umbilical cord occlusion. *PLoS One*. 2018;13:e0195978.
20. Cremer M. Über die direkte Ableitung der Aktionsströme des menschlichen Herzens vom Oesophagus und über das Elektrokardiogramm des Fötus. *Munchener Medizinische Wochenschrift*. 1906;53:811–813.
21. Whiting P, Rutjes A, Westwood M, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529–536.
22. Cohen WR, Ommani S, Hassan S, et al. Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes. *Acta Obstet Gynecol Scand*. 2012;91:1306–1313.
23. Euliano TY, Darmanjian S, Nguyen MT, et al. Monitoring fetal heart rate during labor: a comparison of three methods. *J Pregnancy*. 2017;2017:8529816.
24. Ashwal E, Shinar S, Aviram A, et al. A novel modality for intrapartum fetal heart rate monitoring. *J Matern Fetal Neonatal Med*. 2019;32:889–895.
25. Reinhard J, Hayes-Gill BR, Schiermeier S, et al. Intrapartum signal quality with external fetal heart rate monitoring: a two way trial of external Doppler CTG ultrasound and the abdominal fetal electrocardiogram. *Arch Gynecol Obstet*. 2012;286:1103–1107.
26. Frank TH, Blaumanis OR, Chen SH, et al. Noninvasive fetal ECG mode fetal heart rate monitoring by adaptive digital filtering. *J Perinat Med*. 1992;20:93–100.
27. Stampalija T, Signaroldi M, Mastroianni C, et al. Fetal and maternal heart rate confusion during intra-partum monitoring: comparison of trans-abdominal fetal electrocardiogram and Doppler telemetry. *J Matern Fetal Neonatal Med*. 2012;25:1517–1520.
28. Reinhard J, Hayes-Gill BR, Schiermeier S, et al. Intrapartum heart rate ambiguity: a comparison of cardiotocogram and abdominal fetal electrocardiogram with maternal electrocardiogram. *Gynecol Obstet Invest*. 2013;75:101–108.

29. Breuker KH, Khalili BD, Bolte A. The abdominal fetal EKG for the analysis of pre and subpartual heart frequency (author's transl [German]). *Arch Gynakol*. 1976;221:211–235.

30. Brocato B, Lewis D, Mulekar M, et al. Obesity's impact on intrapartum electronic fetal monitoring. *J Matern Fetal Neonatal Med*. 2019;32:92–94.

31. Zaki R, Bulgiba A, Ismail R, et al. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PLoS One*. 2012;7:e37908.

32. Rooth G, Huch A, Huch R. FIGO News: guidelines for the use of fetal monitoring. *Int J Gynecol Obstet*. 1987;25:159–167.

33. Peters CHL, van Laar JOEH, Vullings R, et al. Beat-to-beat heart rate detection in multi-lead abdominal fetal ECG recordings. *Med Eng Phys*. 2012;34:333–338.

34. van Laar JOEH, Peters CHL, Houterman S, et al. Normalized spectral power of fetal heart rate variability is associated with fetal scalp blood pH. *Early Hum Dev*. 2011;87:259–263.

35. van Laar JOEH, Porath MM, Peters CHL, et al. Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature. *Acta Obstet Gynecol Scand*. 2008;87:300–306.

36. van Laar JOEH, Peters CHL, Vullings R, et al. Fetal autonomic response to severe acidaemia during labour. *BJOG*. 2010;117:429–437.

37. Verdurmen KMJ, Warmerdam GJJ, Lempersz C, et al. The influence of betamethasone on fetal heart rate variability, obtained by non-invasive fetal electrocardiogram recordings. *Early Hum Dev*. 2018;119:8–14.

38. Warmerdam GJJ, Vullings R, Van Laar JOEH, et al. Detection rate of fetal distress using contraction-dependent fetal heart rate variability analysis. *Physiol Meas*. 2018;39:025008.

APPENDIX 1. SEARCH STRATEGY: MEDLINE (PUBMED, 1966–PRESENT)

1. "Fetus"[Mesh] OR fetus*[tiab] OR foetus*[tiab] OR fetal[tiab] OR foetal*[tiab]
2. "Electrocardiography"[Mesh] OR electrocardiogra*[tiab] OR ECG[tiab]
3. "Cardiotocography"[Mesh] OR cardiotocogra*[tiab] OR CTG[tiab]
4. #2 OR #3
5. #1 AND #4
6. "Fetal Monitoring"[Mesh]
7. (fetal[tiab] OR foetal[tiab] OR fetus*[tiab] OR foetus*[tiab]) AND monitor*[tiab]
8. #6 OR #7
9. #5 OR #8
10. abdominal[tiab] OR non invasive[tiab] OR external[tiab]
11. #9 AND #10
12. "Labor, Obstetric"[Mesh] OR labor[tiab] OR labour[tiab] OR intrapartum[tiab]
13. #11 AND #12

APPENDIX 2. SUMMARY OF THE 8 INCLUDED ARTICLES

Author, Year	Participants	NI-fECG Method	Reference Method	Outcome Measures	Results During Labor	Comments
Breuker, 1976	Inclusion criteria: during pregnancy and labor	Prototype Hewlett & Packard, model number 15174A	FSE	Quality: percentage of signal loss	N = 173 **Results:** Excellent: n = 30 (17.3%) Good: n = 40 (23.1%) Satisfactory: n = 46 (26.6%) Sufficient: n = 31 (17.9%) Deficient: n = 26 (15.0%) Second stage: No clinical evaluation possible: n = 34 (30.3%)	• Used self-defined outcome measure
Frank, 1992	Inclusion criteria: term uncomplicated singleton pregnancy	Prototype Perinatronics Prototype FHR monitor	FSE	Accuracy: differences in R-R interval	N = 5 **Results:** Mean difference R-R interval 0.0056 bpm Accuracy: • Within 1 bpm 92.6% • Within 2 bpm 98.2%	• Different definition of accuracy compared with other studies
Stampalija, 2012	Inclusion criteria: uneventful term singleton pregnancy Exclusion criteria: multiple pregnancies, fetal abnormalities, presence of maternal pathologies	Monica AN24 (Monica Healthcare, Nottingham, United Kingdom)	DU telemetry	(Overall) success rate of FHR monitoring: percentage of the 0.25 s epochs where an FHR was produced Maternal-fetal heart rate confusion (CR): FHR within 5 bpm from MHR	N = 41; n = 21, 1st and 2nd stage; n = 18, 1st stage only **Results:** Success rate: Overall: DU 89.4% ± 7.6%, NI-fECG 88.5% ± 16.7% (P = 0.77) 1st stage: DU 89.9% ± 7.9%, NI-fECG 89.8% ± 16.1% (P = 0.98) 2nd stage: DU 83.7% ± 7.4%, NI-fECG 66.5% ± 21.3% (P < 0.001) Confusion rate: Overall: DU 4.5% ± 4.5%, NI-fECG 1.3% ± 1.9% (P < 0.001) 1st stage: DU 3.9% ± 4.6%, NI-fECG 1.0 ± 1.9 (P < 0.001) 2nd stage: DU 11.3% ± 8.2%, NI-fECG 4.6% ± 5.0% (P = 0.002)	• No criterion standard as reference • 2 cases excluded: no FHR in fECG measurements due to high electrical noise • No criterion standard used for MHR; MHR derived from Monica AN24

APPENDIX 2. *(Continued)*

Author, Year	Participants	NI-fECG Method	Reference Method	Outcome Measures	Results During Labor	Comments
Reinhard, 2012	Inclusion criteria: Singleton pregnancy, admitted to Marien Hospital Witten for delivery	Monica AN24 (Monica Healthcare, Nottingham, United Kingdom)	DU	FHR success rate: Percentage of time an FHR value was reported divided by total time. Percentage of patients with FHR signal loss <20% or <15% Correlation: correlation between FHR success rate and BMI/stage of labor/birth weight/epidural	N = 144; 1st stage n = 138; 2nd stage n = 98 **Results:** Success rate: 1st stage: DU 85.5% (35.1%–99.8%), NI-fECG 97.7% (7.8%–100%) ($P < 0.001$) 2nd stage: DU 92.3% (22.5%–99.8%), NI-fECG 85.5% (13.4%–100%) ($P > 0.05$) Signal loss <20%: 1st stage: DU 89.6%, fECG 81.5% 2nd stage: DU 76.5%, fECG 54.1% Signal loss <15%: 1st stage: DU 80.7%, fECG 77.8% 2nd stage: DU 64.3%, fECG 48% Correlation: 1st stage fECG FHR success rate correlation with: • 2nd stage fECG FHR 0.68 • Birth weight 0.22 • 1st stage CTG FHR 0.34 • 2nd stage CTG 0.15 • EDA 0.10 • BMI 0.00 2nd stage fECG FHR success rate correlation with: • DU 1st stage 0.15 • DU 2nd stage 0.41 • BMI −0.14 • Birth weight 0.19 • EDA −0.03	• No criterion standard as reference • Intermittent DU • For different baseline characteristics, mean and median are given. • Data were excluded if simultaneous DU and fECG <20 min during 1st stage or <5 min during 2nd stage. • Outlier removal: data outside 60–200 bpm, FHR data within 5 bpm from MHR, isolated FHR data points (FHR <10 s in duration with absolute FHR difference from the baseline of >15 bpm)

| Cohen, 2012 | Inclusion criteria: singleton term pregnancy, arrived at the hospital early in or prior to labor
Exclusion criteria: known major fetal anomaly, fetal malpresentation, maternal abdominal skin rash or history of adhesive sensitivity, patients only monitored with the 2 external monitors | Monica AN24 (Monica Healthcare, Nottingham, United Kingdom) | FHR: FSE | MHR: Pulse oximetry | Accuracy: comparison of FHR output from DU and NI-fECG with the average of FHR of the FSE
Reliability: output within 10% of the standard (FSE)
Success rate: Proportion of recording DU and NI-fECG had a nonzero output for FHR.
Confusion rate (CR): Percentage of FHR of DU and NI-fECG that was both more than 5% different from that of the FSE and within 5% of the MHR | $N = 138$; $n = 75$ complete for analysis
Results:
Overall reliability:
Overall: DU 73% ± 24.6%, NI-fECG 81.7% ± 20.5% ($P < 0.01$)
1st stage: DU 74.7% ± 28.2%, NI-fECG 84.9% ± 21.5% ($P < 0.01$)
2nd stage: DU 61.7% ± 24.8%, NI-fECG 71.9% ± 20.4% ($P < 0.01$)
Accuracy:
DU 10.6 bpm, NI-fECG 5.2 bpm ($P < 0.0001$)
1st stage: DU 8.7 bpm ± 5.7, NI-fECG 4.5 ± 2.4 ($P < 0.0001$)
2nd stage: DU 16.1 bpm ± 7.6, NI-fECG 7.9 ± 4.2 ($P < 0.0001$)
Success rate:
DU 82.5% ± 21.1%, NI-fECG 83.4% ± 20.1% ($P = 0.38$)
1st stage: DU 82.6% ± 24.4%, NI-fECG 86.4% ± 21.1% ($P = 0.12$)
2nd stage: DU 77.8% ± 21.1%, NI-fECG 75.2% ± 19.2% ($P = 0.25$)
Confusion rate ($n = 47$):
Overall: DU 8.9% ± 15.2%, NI-fECG 0.4% ± 0.6% ($P = 0.0002$)
1st stage: DU 9.5% ± 17.8%, NI-fECG 0.3% ± 0.6% ($P = 0.0007$)
2nd stage: DU 11.0% ± 15.4%, NI-fECG 0.7% ± 0.8% ($P = 0.0007$) | • FSE was only applied when CTG was abnormal to substitute DU.
• 63 excluded participants
• High mean BMI, 32.6 kg/m^2 |

APPENDIX 2. *(Continued)*

Author, Year	Participants	NI-fECG Method	Reference Method	Outcome Measures	Results During Labor	Comments
Reinhard, 2013	Inclusion criteria: singleton pregnancy, admitted to Marien Hospital Witten for delivery	Monica AN24 (Monica Healthcare, Nottingham, United Kingdom)	DU	Percentage FHR and MHR ambiguity over total recording time: FHR within 5 bpm of MHR. Reliability: percentage of success rate of recording FHR data	N = 144; 1st stage n = 135, 2nd stage n = 98 **Results:** Ambiguity: 1st stage: DU 1.22%, NI-fECG 0.70% ($P < 0.001$) 2nd stage: DU 6.20%, NI-fECG 3.30% ($P < 0.001$) Reliability: 1st stage: DU 85.2%, NI-fECG 87.1% ($P < 0.001$) 2nd stage: DU 76.5%, NI-fECG 70.5% ($P > 0.05$)	• No criterion standard as reference for FHR • No criterion standard for MHR; MHR derived from Monica AN24 • Did not use the right definition for reliability • For multiple characteristics, median and mean are given.
Ashwal, 2017	Inclusion criteria: ≥18 y, singleton term pregnancy, no known fetal anomalies or chromosomal defects, spontaneous rupture of membranes during latent phase of labor. Exclusion criteria: signs suggestive for chorioamnionitis, implanted electronic device, maternal allergy to silver, irritated skin or open abdominal wounds	EUM100Pro (OB tools, Nesher, Israel)	FSE	Correlation: between DU and FSE, NI-fECG, and FSE FHR traces. Accuracy: difference in FHR output of DU and NI-fECG compared with FSE	N = 33 **Results:** Correlation: DU/FSE and NI-fECG/FSE both $r^2 = 0.98$ ($P < 0.001$) Reliability: Overall: DU 96.0%, EUM 98.5% ($P < 0.001$) 1st stage: DU 97.1%, NI-fECG 99.0% 2nd stage: DU 94.9%, NI-fECG 98.5% Accuracy: DU 5.39 bpm, EUM 1.47 bpm	• Only random noncontinuous 30 minutes of the recording time was used for analysis from each phase (latent, 1st stage, and 2nd stage) of labor. • Mean and median are given for correlation.