Multimodal Causal Machine Learning for Fetal Asphyxia Risk Prediction from Imperfect Monitoring Signals

Debjyoti Karmakar

debmelbourne@gmail.com

University of Melbourne

Lochana Mendis

University of Melbourne

Emerson Keenan

University of Melbourne

Marimuthu Palaniswami

University of Melbourne

Enes Makalic

Monash University

Fiona Brownfoot

University of Melbourne

Article

Keywords:

Posted Date: August 13th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7122719/v1

Additional Declarations: Competing interest reported. None of the authors have a direct conflict of interest. FB and EK are directors and shareholders of Kali Healthcare, a company that is commercialising a wearable fetal monitoring device. MP is a shareholder of Kali Healthcare. Rest of the authors have nothing to disclose.

Abstract

Cardiotocography (CTG) is the cornerstone of intrapartum surveillance, yet its predictive value for perinatal asphyxia remains limited. Modern monitors record fetal heart rate (FHR) via Doppler ultrasound or intrapartum fetal electrocardiography (ECG), the latter usable only after cervical dilatation. We analysed 52,757 labours ≥36 weeks' gestation across three hospitals—two in Australia and one in Europe—treating mode-specific FHR dropout (zero-value samples signifying missing valid biosignal) as an informative exposure rather than benign artifact. High dropout (>30% of the CTG recording available for analysis) was evaluated using doubly robust causal machine learning and survival analysis, incorporating inverse-probability weighting, targeted maximum-likelihood estimation (TMLE), and weighted Cox models. In the training cohort (n = 36,792), high ECG-mode dropout tripled the odds of perinatal asphyxia (OR 3.14, 95% CI: 2.20–4.48) and quadrupled the hazard (HR 4.76, 95% CI: 2.98–7.62) compared to ultrasound-mode low dropout. In crude analyses, this corresponded to an absolute risk increase of 5.0% (1 in 20 births). TMLE-adjusted models confirmed a 2.27% absolute risk increase—equivalent to one additional case per 44 labor episodes (95% CI: 28–83). External validation confirmed consistent associations in European (OR 4.02) and Australian (OR 2.22) cohorts. "Imperfect" ECG signals thus offer biologically plausible and trustworthy inputs for standalone or ensemble artificial intelligence models, enabling transparent, mode-aware decision support in real-time fetal monitoring.

Introduction

Perinatal asphyxia, often attributed to intrapartum compromise of gas exchange, remains a significant contributor to neonatal illness and death—though precise incidence and definitions vary globally. Estimates suggest up to one-quarter of neonatal deaths and a large proportion of stillbirths in the third trimester are related to this condition.[1-3] Reported rates of intrapartum-related neonatal encephalopathy vary, with estimates ranging from 1–3 per 1,000 in high-income countries to significantly higher in low-resource settings.[3,4] Despite widespread use of cardiotocography (CTG) for fetal surveillance, outcomes have not improved, while cesarean rates and medicolegal costs have surged.[2,5,6]

Machine learning models for CTG interpretation have shown limited impact, often hindered by small datasets, inadequate adjustment for confounding, and exclusion of signal segments deemed 'noisy' or 'artifact'—segments that may still contain clinically relevant patterns.[7-9] Excessive preprocessing may obscure associations with asphyxia risk. The rarity of asphyxia further limits statistical power, and a lack of causal discovery risks misattributing predictive importance to spurious features.[7,10,11]

Fetal heart rate (FHR) is typically acquired by CTG monitors at ~4 Hz using either Doppler ultrasound or fetal scalp electrode (electrocardiography, ECG). The former captures mechanical motion while the latter directly records cardiac electrical activity, generally offering greater signal fidelity during active labor. Mode selection is subjective by clinician—ECG use generally follows failed ultrasound acquisition during labor. Signal dropout is stored as zeros in backend systems and displayed as trace gaps. Although proprietary algorithms enhance visual readability through autocorrelation and filtering, this may conceal clinically important data loss.[7,12] Dropout differs by acquisition mode: ultrasound is more prone to signal loss due to maternal or fetal factors, while ECG—despite directly capturing cardiac activity—still exhibits dropout, a counterintuitive finding that may

have a relationship with underlying fetal compromise. FDA reports have cited signal quality failures in adverse outcomes.[8]

Across disciplines, the challenge of distinguishing signal loss due to technical limitations versus physiological compromise is shared in many biomedical domains—from ECG and EEG to wearables and implantables.[13] Our findings contribute to this growing body of work, demonstrating that signal dropout can act as a latent biomarker, especially when stratified by acquisition context. This work also speaks to broader trends in AI-driven health monitoring: the transition from pattern recognition to causal, real-time inference from raw and imperfect signals.[11]

As CTG becomes increasingly digitized, signal quality itself may hold prognostic value. We applied causal machine learning to assess whether mode-specific FHR dropout predicts perinatal asphyxia. By treating signal dropout as an informative exposure rather than dismissing it as nuisance noise, we aim to support the development of real-time, mode-aware decision-support tools in fetal monitoring. By uncovering underlying causal drivers using causal machine learning—including those embedded in signal quality—we seek to inform more robust and generalizable prediction models for real-world clinical deployment.

MATERIALS AND METHODS

Study Design and Data Sources

This multicenter cohort study included 52,757 laboring women at ≥36 weeks' gestation across three international datasets (Figure 1). The training cohort consisted of 36,792 births from Mercy Hospital for Women, Melbourne (2010–2021).This is a tertiary hospoital. Inclusion required ≥15 minutes of in-labor CTG data within the final 60 minutes before birth, regardless of labor stage or delivery mode, to ensure inclusion of both vaginal and cesarean births. This aligns with standards in computational fetal monitoring studies.[14,15] Multiple gestations and congenital anomalies were excluded. Active labor was confirmed by midwifery and obstetric annotations.

Data Extraction and Preprocessing

CTG signals were extracted from Philips IntelliSpace™ (Philips, Amsterdam) with technical support from Philips. Traces were linked to clinical data from the Birthing Outcomes System (BOS), hospital morbidity records, and paper charts as needed[7]. All data were stored securely using the University of Melbourne's MediaFlux™ platform, with analysis conducted on the Spartan high-performance computing cluster.

External Validation Datasets

Validation was performed on two cohorts. The first comprised 552 cases from a European dataset -the Czech CTU-UHB database (2010–2012), an open-access European dataset with linked clinical metadata.[16] The second included 15,413 births from Werribee Mercy Hospital (2010–2021),a community hospital in Australia. The data was processed identically to the training dataset. Both required ≥15 minutes of final-hour labor CTG data and excluded multiple gestations and congenital anomalies.

Quality Control

CTG recordings with >90% signal dropout were excluded due to their clinical unreliability.

Exposure Classification

The primary exposure was a high level of missing valid fetal heart rate (FHR) signal, specifically of the *signal dropout* type—defined as the cumulative proportion of zero-valued data points (missing valid biosignal)recorded during active signal acquisition in a live fetus. In Philips systems, these dropouts appear as visual gaps in the CTG trace; however, proprietary signal-processing algorithms (e.g., autocorrelation-based smoothing) may reduce their visibility. Dropout was calculated at a sampling frequency of 4 Hz over the final hour of labor. Based on Delphi consensus from a multidisciplinary team of clinicians, data scientists, and signal-processing experts, high dropout was predefined as >30% signal loss.[7] Maternal–FHR coincidence was not assessed in this analysis.

Participants were categorized into four exposure groups based on monitoring mode (ultrasound vs. ECG) and dropout level (low vs. high):

1. Low dropout in ultrasound mode (reference)
2. High dropout in ultrasound mode
3. Low dropout in ECG mode
4. High dropout in ECG mode

Mode classification was based on the presence and duration of ECG signal. Patients were assigned to the ECG group if ≥10 minutes of ECG was recorded, minimizing misclassification from brief mode switches. Those with <10 minutes were classified as ultrasound. This approach enabled mode-specific dropout analysis while reducing bias from clinician-initiated switching.

Outcome Measures

The primary outcome was perinatal asphyxia, defined using a composite developed by multidisciplinary Delphi consensus and aligned with international standards. Criteria included one or more of: resuscitation at 10 minutes, therapeutic hypothermia, Apgar score ≤6 at 10 minutes or ≤4 at 5 minutes, cord pH <7.05, base excess <−12 mmol/L, NICU admission, or clinically (including seizures) /imaging-confirmed hypoxic–ischaemic encephalopathy (HIE).

Causal Machine Learning Methods

Confounder adjustment was guided by directed acyclic graphs (DAGs) (see Supplement). A doubly robust framework combining propensity score modeling and outcome regression was used to assess the causal effect of mode-specific dropout. All confounders used in causal models are listed in supplementary material and include maternal demographics, obstetric risk factors, labor characteristics, and mode of delivery. Propensity scores were estimated via multinomial logistic regression using maternal, fetal, and intrapartum variables. Inverse probability weights (IPWs), stabilized by clipping between 0.01 and 0.99, were used in weighted logistic regression. Covariate balance was assessed using standardized mean differences. HC3 robust standard errors accounted for heteroskedasticity. Sensitivity analyses were conducted on both complete case and imputed datasets.

Targeted Maximum Likelihood Estimation (TMLE) refined causal estimates for ECG High Dropout vs. Ultrasound Low Dropout. TMLE combined logistic regression and gradient boosting for exposure and outcome modeling. Ridge regularization was applied for logistic regression; the gradient boosting model used a learning rate of 0.1, max depth 3, and 100 estimators. Categorical variables were one-hot encoded; continuous variables (e.g., BMI) were log-transformed and standardized. Additional engineered features, such as labor duration ratios, were included. TMLE used a clever covariate (H) for targeted updates, and influence-curve–based standard errors (via the *zepid* package) for confidence intervals. No additional fetal heart rate features (e.g., decelerations or variability) were modeled directly to isolate the effect of dropout.

Model Validation and Sensitivity Analysis

Robustness was evaluated across dropout thresholds using likelihood ratio tests (LRTs) and E-values for unmeasured confounding. Models were stratified by labor stage and tested on a 20% stratified holdout set from the training cohort. External validation was conducted using both the Czech and Werribee datasets.

Survival Analysis

Time-to-event analysis assessed asphyxia risk from second-stage labor onset to event occurrence. "Survival" denoted remaining free from asphyxia. Exposure was categorized as a four-level composite of monitoring mode and dropout severity. Kaplan–Meier curves were compared using log-rank tests. A weighted Cox proportional hazards model, adjusted using stabilized IPWs derived from a multinomial logistic regression, was used to account for confounding. Model discrimination was assessed using Harrell's concordance index (C-index).

Sample size calculation

Consistent with our prior study on CTG signal quality and perinatal outcomes—conducted on a related but differently grouped cohort—the current analysis builds on a pre-established dataset, including one additional large dataset, for which formal sample size justification had previously been undertaken[7]. Given the exploratory nature of this causal inference study and the rarity of perinatal asphyxia, we analyzed the full cohort (n = 52,757) to maximize statistical power for effect estimation, threshold modeling, and validation across independent populations.

Data Processing and Statistical Analysis

All analyses were performed using Python 3.6(including *lifelines* v0.30.0). and Stata 18.5. CTG data were extracted via Philips IntelliSpace™ and managed on Spartan and MediaFlux™ infrastructures. The detailed statistical analysis plan is enclosed in the supplementary material.

Handling Missing Clinical Data

Missing clinical variables were imputed using chained equations. Sensitivity analyses compared complete-case and imputed models. Maternal BMI had the highest missingness (6.32%, 95% CI: 6.06–6.66%).

Ethical Approval and Reporting Standards

The study was approved by the institutional Human Research Ethics Committee (HREC #R2020-077). It adhered to a pre-specified statistical analysis plan and followed STROBE[17] and TRIPOD+AI guidelines(see checklists in

Supplementary Materials).[18]

RESULTS

Of 36,792 laboring women screened in the training dataset, 32,242 met the inclusion criteria. There were 860 cases of perinatal asphyxia (2.67%, 95% CI: 2.49–2.85%) (Figure1). Table 1 shows baseline sample characteristics. Sample characteristics of the validation datasets are included in supplementary tables 1 and 2. Dropout patterns significantly varied by monitoring modality. Most cases were in the *Ultrasound Low Dropout* group (Group 0, 54.4%, 95% CI: 53.8%–54.9%), which had the lowest asphyxia rate (2.2%, 95% CI: 2.0%–2.4%). In contrast, the *ECG High Dropout group* (Group 3, 1.18%, 95% CI: 1.07%–1.31%) had the highest asphyxia rate (5.2%, 95% CI: 3.4%–7.9%)($p < 0.0001$) (Figure 1).

Maternal and neonatal characteristics differed across groups. *Maternal BMI* was highest in ECG groups, with a median of 25.0 (IQR: 22.0–29.0) compared to 23.0 (IQR: 21.0–27.0) in the US Low Dropout group ($p < 0.0001$). *Gestational age* was similar across groups ($p = 0.09$), but *baby weight* was significantly lower in ECG High Dropout (3280g, IQR: 2970–3610) compared to US Low Dropout (3410g, IQR: 3110–3710, $p < 0.0001$). *Parity* distributions also varied, with nulliparous women being most common in the ECG Low Dropout group (64.7%) versus 47.5% in US Low Dropout ($p < 0.0001$). Figure 2 shows representative modal devices and representative outputs.

Causal Inference Analysis

Logistic regression with inverse probability weighting (IPW) was used to estimate the association between signal dropout and perinatal asphyxia, using *Ultrasound Low Dropout* as the reference group (Figure 3a). The highest risk of asphyxia was observed in the *ECG High Dropout* group (OR = 3.14, 95% CI: 2.20–4.48, $p < 0.001$). The *ECG Low Dropout* group also showed significantly increased risk (OR = 1.51, 95% CI: 1.32–1.71, $p < 0.001$). The *Ultrasound High Dropout* group showed a non-significant risk of asphyxia (OR = 1.25, 95% CI: 0.99–1.58, $p = 0.059$) (table 2, Figure 3a).

E-value analysis assessed robustness to unmeasured confounding. The *ECG High Dropout* group showed the greatest robustness (E-value = 5.73; lower bound = 3.82). *ECG Low Dropout* showed moderate robustness (E-value = 2.39), while *Ultrasound High Dropout*, which was not statistically significant, showed limited robustness (E-value = 1.81). A likelihood ratio test (LRT) confirmed that the inclusion of the 'dropout by mode' classification significantly improved model fitcompared to a model without it (χ^2 (1) = 25.58, $p < 0.0001$) supporting the relevance of mode-specific dropout classification in explaining variation in asphyxia risk.

Causal Machine Learning Analysis

Targeted Maximum Likelihood Estimation (TMLE) refined causal estimates for the two key groups (US Low Dropout as reference and ECG High Dropout with highest asphyxia risk) by integrating machine learning–based propensity score modeling (Gradient Boosting Classifier) and outcome prediction (Gradient Boosting Regressor) to address high-dimensional, nonlinear relationships. TMLE confirmed increased asphyxia risk in ECG High Dropout (adjusted OR = 2.08, 95% CI: 1.21–3.58) relative to US Low Dropout (Table 2). In crude analyses, ECG High Dropout was associated with a 5.0% absolute risk increase—one additional case of asphyxia for every 20 laboring women monitored. After TMLE adjustment, the absolute risk increase was 2.27%, or one extra case per 44 labor episodes (95% CI: 28–83).

We applied TMLE-derived individual risk estimates to assess fetal risk reclassification in the training set. Clinically, cases were considered low risk if labor progressed without operative delivery, and high risk if operative birth (caesarean, forceps, or vacuum) occurred due to suspected fetal distress. TMLE reclassified 9.29% of cases managed as low-risk into a high-risk category based on elevated model-predicted asphyxia risk. Reclassification was significantly more frequent among ECG-monitored cases, suggesting dropout-informed causal modeling may identify high-risk fetuses missed by conventional assessment. TMLE analyses across dropout thresholds showed progressively increasing odds of asphyxia. Dropout <15% was not significantly associated with asphyxia, while increasing dropout thresholds from ≥20% showed a dose-response relationship with rising risk: OR = 1.56 (95% CI: 1.16–2.09, $p < 0.001$) at 20%, increasing to OR = 2.08 (95% CI: 1.21–3.58, $p = 0.01$) at 30% (Figure 3a). SHAP analysis in Supplementary Figure 3 illustrates features driving ECG High Dropout and asphyxia risk.

Survival Analysis (Figure 3b)

Kaplan–Meier estimates stratified by monitoring modality and signal dropout showed the highest asphyxia-free survival in the Ultrasound Low Dropout group, and the steepest decline in the ECG High Dropout group. Differences were statistically significant (log-rank $\chi^2 = 43.6$, $p < 0.0001$).

To adjust for confounding, we applied an inverse probability–weighted Cox proportional hazards model using stabilized weights from a multinomial treatment model. In the fully adjusted model, the hazard of asphyxia was significantly higher in ECG High Dropout versus US Low Dropout (HR = 4.76, 95% CI: 2.98–7.62, $p < 0.0001$). Elevated hazards were also observed for Ultrasound High Dropout (HR = 1.90, 95% CI: 1.39–2.60, $p < 0.0001$) and ECG Low Dropout (HR = 1.42, 95% CI: 1.20–1.69, $p < 0.0001$) (Supplementary Table 3). Harrell's C-index was 0.71, indicating moderate predictive accuracy. Figure 3b displays Kaplan–Meier and IPW-adjusted Cox curves by monitoring modality and dropout level.

External Validation

External validation using the Czech CTU-UHB and Werribee datasets demonstrated similar trends to the model generation cohort(table 3)

Validation on Czech dataset

Causal inference analysis using inverse probability weighting (IPW) demonstrated that the *ECG High Dropout* group was significantly associated with an increased risk of perinatal asphyxia (OR = 4.02, 95% CI: 1.41–11.44, $p = 0.009$) compared to the *Ultrasound Low Dropout* reference group. The ECG *Low Dropout* group showed a paradoxical reduction in asphyxia risk (OR = 0.24, 95% CI: 0.06–0.96, $p = 0.043$), which may reflect differing case mix, small sample size, or differences in mode-switching thresholds between centers. The *Ultrasound High Dropout* group showed a non-significant trend toward increased risk (OR = 1.90, 95% CI: 0.94–3.82, $p = 0.073$). Causal machine learning analysis using targeted maximum likelihood estimation (TMLE) produced consistent findings, estimating an asphyxia odds ratio of 2.83 (95% CI: 2.13–16.36) for the *ECG High Dropout* group compared to the reference *Ultrasound Low Dropout*.

Validation on Werribee dataset

Causal inference analysis using inverse probability weighting (IPW) regression model showed that the *ECG High Dropout* group was significantly associated with increased asphyxia risk (OR = 2.22, 95% CI: 1.61–3.05, $p <$

0.001), followed by the *ECG Low Dropout* group (OR = 1.57, 95% CI: 1.31–1.89, p < 0.001). The *Ultrasound High Dropout* group was not significantly associated with asphyxia (OR = 0.69, 95% CI: 0.33–1.46, p = 0.33), using *Ultrasound Low Dropout* as the reference. Causal machine learning using targeted maximum likelihood estimation (TMLE) showed that the *ECG High Dropout* group showed a significant increase in asphyxia risk (OR = 1.91, 95% CI: 1.11–3.28) compared to the reference *Ultrasound Low Dropout*.

DISCUSSION

Principal Findings

In this multicohort study of over 50,000 births, we show that mode-specific FHR signal dropout—typically overlooked in CTG interpretation—is strongly and causally associated with perinatal asphyxia, particularly in ECG-based monitoring. ECG High Dropout carried the highest risk (IPW OR = 3.14; 95% CI: 2.20–4.48), followed by ECG Low Dropout (OR = 1.51; 95% CI: 1.32–1.71). Ultrasound High Dropout was not significant, reinforcing the hypothesis that dropout in ECG—typically used in advanced labor—may more directly reflect underlying physiological compromise rather than technical factors. TMLE confirmed a dose-response relationship, with dropout ≥30% yielding an OR of 2.08 and an absolute risk increase of one asphyxia case per 44 women. TMLE-based risk profiling reclassified nearly 10% of clinically low-risk cases to high risk. Findings were consistent across external cohorts (Czech: IPW OR = 4.02, TMLE OR = 2.83; Werribee: IPW OR = 2.22, TMLE OR = 1.91), reinforcing the prognostic value of ECG dropout.

Results in the Context of What is Known

Signal dropout in ECG-mode CTG is a strong, clinically meaningful predictor of asphyxia—challenging the assumption that it reflects only technical or user error. Mode selection is subjective, and dropout may get hidden due to interface design. Our findings support making dropout metrics visible. Rather than default mode switching, decisions should be guided by dropout severity: US-mode dropout >30% may warrant switching to ECG, but persistent high dropout in ECG-mode indicates tripled asphyxia risk and should prompt closer surveillance

Clinical Implications

A major strength of this study is the use of causal methods—IPW and TMLE—to estimate the effect of mode-specific dropout. Dropout is non-random and influenced by clinical factors; failure to adjust for this can bias observational analyses. Our approach improved covariate balance and model fit ($\chi^2(1) = 16.2$, p = 0.00006), with E-values (ECG High Dropout: 5.89) supporting robustness to unmeasured confounding.

Unlike traditional machine learning, which may overfit and overlook causal pathways, causal ML methods like TMLE explicitly model exposure–outcome relationships to enhance generalizability.[11,19-21] We used DAGs to select confounders and validated findings across two independent cohorts. Prior models, such as McCoy et al., excluded high-artifact CTGs (>30% dropout), discarding over one-third of cases.[22] While their deep learning model achieved high AUROC (0.85 for pH <7.05), it ignored dropout as a signal. Moreover, pH <7 alone does not reliably stratify neonatal outcomes, as most infants with pH between 6.9 and 7 survive without neurodevelopmental impairment.[23] In contrast, we treat signal dropout as a meaningful exposure with prognostic value and applied a more holistic, clinically accepted composite outcome aligned with established

definitions of neurodevelopmental risk.[5,24,25] While our model focused on signal dropout, it did not incorporate contemporaneous CTG pattern interpretation (e.g., decelerations), which may offer complementary prognostic information in future models.

Research Implications

By integrating mode-specific dropout into causal ML, we demonstrate its utility in identifying high-risk cases missed by conventional approaches. This supports real-time monitoring tools that adapt to dropout severity, aligning with current recommendations for actionable, causally grounded AI in healthcare.[11,19-21] Beyond obstetrics, our findings resonate with challenges in other domains that rely on physiological time series—such as ECG, EEG, and wearable biosensors—where distinguishing noise from latent risk is critical for advancing real-time, interpretable AI applications.

Strengths and Limitations

Our study leverages one of the largest digitized CTG datasets, with validation across multiple cohorts, enhancing generalizability. The integration of causal inference and machine learning improves robustness.[12,19,20,26,27] A limitation is the lack of data on real-time clinician-driven mode-switching decisions. However, consistent validation trends across Czech CTU-UHB and Werribee datasets reinforce the observed associations. Implementation in other settings may face logistical challenges, but prospective validation is essential to assess whether integrating dropout monitoring improves neonatal outcomes. We acknowledge the lack of data on real-time clinician-driven mode-switching or intrauterine resuscitation interventions, which may confound the relationship between dropout and outcome

While our framework enhances risk estimation, clinical deployment would require real-time implementation within a comprehensive decision-support model. Future extensions may adapt approaches such as the AI-ECG sex discordance score, which identifies subtle, clinician-invisible signal shifts—analogous to dropout thresholds in our study.[28]However, developers must also address known demographic disparities in signal-based AI, as shown in recent ECG heart failure models, where performance biases were especially pronounced among young Black women.[29] This study also contributes to emerging efforts across disciplines to better utilize ‘imperfect’ or artifact-prone physiological data streams—turning what was once discarded into features that enhance clinical relevance and model transparency.

Conclusion

Mode-specific FHR dropout—particularly in ECG-mode CTG—is a clinically significant predictor of perinatal asphyxia. Rather than being excluded or ignored, signal quality should be incorporated into predictive modeling. Integrating signal metrics into ensemble models may enhance fetal surveillance and should be prioritized in future validation efforts.

Declarations

Disclosure/Conflict of interest:

None of the authors have a direct conflict of interest. FB and EK are directors and shareholders of Kali Healthcare, a company that is commercialising a wearable fetal monitoring device. MP is a shareholder of Kali Healthcare. Rest of the authors have nothing to disclose.

Funding

FB is supported by a Dame Kate Campbell Fellowship, a University of Melbourne Fellowship, and an National Health and Medical Research Council (NHMRC) Ideas Grant. DK receives postgraduate scholarship support from the NHMRC.Additional funding was provided by the Norman Beischer Medical Research Foundation. LM receives University of Melbourne Graeme Clark Institute and Melbourne Research scholarships. The authors were not precluded from accessing data in the study, and they accept responsibility to submit for publication.

Acknowledgments:

This research was supported by The University of Melbourne's Research Computing Services

and the Petascale Campus Initiative.

Funding:

FB is supported by a Dame Kate Campbell Fellowship, a University of Melbourne Fellowship,

and an National Health and Medical Research Council (NHMRC) Ideas Grant. DK receives

postgraduate scholarship support from the NHMRC.Additional funding was provided by the

Norman Beischer Medical Research Foundation. LM receives University of Melbourne Graeme

Clark Institute and Melbourne Research scholarships. The authors were not precluded from

accessing data in the study, and they accept responsibility to submit for publication.

Author contributions:

DK: Conceptualization, Methodology, Investigation, Data curation, Writing – review & editing

Visualization, Funding acquisition, Writing – original draft

FB: Conceptualization, Methodology, Investigation, Funding acquisition, Project administration, Supervision, Writing – original draft

EM: Methodology, Visualization, Project administration, Supervision, Writing – review & editing

LM: Investigation, Data curation, Writing – review & editing

Writing – review & editing

EK: Supervision, Data curation, Writing – review & editing

Writing – review & editing

MP: Supervision, Writing – review & editing

Competing interests:

None of the authors have a direct conflict of interest. FB and EK are directors and shareholders of Kali Healthcare, a company that is commercialising a wearable fetal monitoring device. MP is a shareholder of Kali Healthcare. Rest of the authors have nothing to disclose.

Data and materials availability:

All individual-level data of the included cohort and modeling codes can be shared upon reasonable request to the corresponding author and completion of data transfer agreement forms.

References

1. Mota-Rojas, D.*, et al.* Pathophysiology of perinatal asphyxia in humans and animal models. *Biomedicines* **10**, 347 (2022).
2. Resolution, N.H.S. Annual Report and Accounts 2019/20. (NHS Resolution, UK, 2020).
3. Schwartz, N. & Young, B.K. Intrapartum fetal monitoring today. (2006).
4. Schiermeier, S.*, et al.* Sensitivity and specificity of intrapartum computerised FIGO criteria for cardiotocography and fetal scalp pH during labour: multicentre, observational study. *Bjog* **115**, 1557-1563 (2008).
5. Brocklehurst, P.*, et al.* Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. *The Lancet* **389**, 1719-1729 (2017).
6. Ayres-de-Campos, D., Spong, C.Y. & Chandraharan, E. FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *International Journal of Gynecology & Obstetrics* **131**, 13-24 (2015).
7. Karmakar, D.*, et al.* Impact of missing electronic fetal monitoring signals on perinatal asphyxia: a multicohort analysis. *NPJ Digit Med* **8**, 233 (2025).
8. Kiely, D.J., Oppenheimer, L.W. & Dornan, J.C. Unrecognized maternal heart rate artefact in cases of perinatal mortality reported to the United States Food and Drug Administration from 2009 to 2019: a critical patient safety issue. *BMC Pregnancy and Childbirth* **19**, 1-10 (2019).
9. Signorini, M.G. & Magenes, G. Advanced signal processing techniques for CTG analysis. in *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016: MEDICON 2016, March 31st-April 2nd 2016, Paphos, Cyprus* 1205-1210 (Springer, 2016).
10. O'Sullivan, M.E.*, et al.* Challenges of developing robust AI for intrapartum fetal heart rate monitoring. *Frontiers in Artificial Intelligence* **4**, 765210 (2021).
11. Prosperi, M.*, et al.* Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence* **2**, 369-375 (2020).
12. Blakely, T., Lynch, J., Simons, K., Bentley, R. & Rose, S. Reflection on modern methods: when worlds collide—prediction, machine learning and causal inference. *International Journal of Epidemiology* **49**, 2058-2064 (2019).

13. Clifford, G.D., Behar, J., Li, Q. & Rezek, I. Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiol Meas* **33**, 1419-1433 (2012).
14. Mendis, L., Palaniswami, M., Brownfoot, F. & Keenan, E. Computerised cardiotocography analysis for the automated detection of fetal compromise during labour: a review. *Bioengineering* **10**, 1007 (2023).
15. Petrozziello, A., Redman, C.W., Papageorghiou, A.T., Jordanov, I. & Georgieva, A. Multimodal convolutional neural networks to detect fetal compromise during labor and delivery. *IEEE Access* **7**, 112026-112036 (2019).
16. Chudáček, V.*, et al.* Open access intrapartum CTG database. *BMC pregnancy and childbirth* **14**, 1-12 (2014).
17. Von Elm, E.*, et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The lancet* **370**, 1453-1457 (2007).
18. Collins, G.S.*, et al.* TRIPOD+ AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *bmj* **385**(2024).
19. Hernán, M.A. & Robins, J.M. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* **183**, 758-764 (2016).
20. Hernán, M.A., Wang, W. & Leaf, D.E. Target trial emulation: a framework for causal inference from observational data. *Jama* **328**, 2446-2447 (2022).
21. Nevo, D., Ogino, S. & Wang, M. Reflection on modern methods: causal inference considerations for heterogeneous disease etiology. *International Journal of Epidemiology* **50**, 1030-1037 (2021).
22. McCoy, J.A.*, et al.* Intrapartum electronic fetal heart rate monitoring to predict acidemia at birth with the use of deep learning. *American Journal of Obstetrics and Gynecology* (2024).
23. Lau, S.L.*, et al.* Neonatal outcome of infants with umbilical cord arterial pH less than 7. *Acta Obstet Gynecol Scand* **102**, 174-180 (2023).
24. Lee, A.C.*, et al.* Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990. *Pediatric research* **74**, 50-72 (2013).
25. Razak, A.*, et al.* Early Neurodevelopmental Assessments for Predicting Long-Term Outcomes in Infants at High Risk of Cerebral Palsy. *JAMA Network Open* **7**, e2413550-e2413550 (2024).
26. Luque-Fernandez, M.A., Schomaker, M., Rachet, B. & Schnitzer, M.E. Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Statistics in medicine* **37**, 2530-2546 (2018).
27. Smith, M.J., Phillips, R.V., Luque-Fernandez, M.A. & Maringe, C. Application of targeted maximum likelihood estimation in public health and epidemiological studies: a systematic review. *Annals of epidemiology* **86**, 34-48. e28 (2023).
28. Sau, A.*, et al.* Artificial intelligence-enhanced electrocardiography for the identification of a sex-related cardiovascular risk continuum: a retrospective cohort study. *The Lancet Digital Health* **7**, e184-e194 (2025).
29. Kaur, D.*, et al.* Race, sex, and age disparities in the performance of ECG deep learning models predicting heart failure. *Circulation: Heart Failure* **17**, e010879 (2024).

Tables

Table 1: Baseline characteristics by mode specific dropout levels

	US mode, low dropout	US mode, high dropout	ECG mode, low dropout	ECG mode, high dropout	P-value
Number of records, % proportion, (95%CI)	17533, (54.38,53.83-54.92)	2363, (7.33,7.05-7.62)	11964, (37.11,36.58-37.64)	382(1.18,1,07-1.31)	0.00
Proportion of CTG signal by US mode	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.04 (0.00-0.35)	0.70 (0.57-0.83)	0.00
Maternal age	32.00 (29.00-35.00)	32.00 (29.00-35.00)	32.00 (29.00-35.00)	32.50 (29.00-36.00)	0.00
Gestational age	39.40 (38.50-40.30)	39.40 (38.50-40.30)	39.50 (38.50-40.40)	39.40 (38.40-40.30)	0.09
Maternal BMI	23.00 (21.00-27.00)	24.00 (21.00-28.00)	25.00 (22.00-29.00)	25.00 (22.00-29.00)	<0.0001
Baby Weight(grams)	3410.00 (3110.00-3710.00)	3380.00 (3070.00-3690.00)	3350.00 (3040.00-3660.00)	3280.00 (2970.00-3610.00)	<0.0001
Parity					
Nulliparous	47.5% (0.468 - 0.483)	35.3% (0.334 - 0.373)	64.7% (0.638 - 0.655)	51.0% (0.460 - 0.560)	<0.0001
Para 1	33.8% (0.331 - 0.345)	46.0% (0.440 - 0.481)	24.3% (0.236 - 0.251)	35.6% (0.310 - 0.405)	
Para 2	12.9% (0.124 - 0.134)	13.2% (0.119 - 0.147)	7.7% (0.073 - 0.082)	9.4% (0.069 - 0.128)	
Para 3 or higher	5.7% (0.054 - 0.061)	5.4% (0.045 - 0.064)	3.2% (0.029 - 0.036)	3.9% (0.024 - 0.064)	
Baby gender					
Female/indeterminate	49.6% (0.489 - 0.503)	50.0% (0.480 - 0.520)	47.6% (0.467 - 0.484)	47.9% (0.429 - 0.529)	0.00
Male	50.4% (0.497 - 0.511)	50.0% (0.480 - 0.520)	52.4% (0.516 - 0.533)	52.1% (0.471 - 0.571)	
Primary Obstetrics Diagnosis					
Low risk pregnancy	44.0% (0.433 - 0.448)	60.0% (0.580 - 0.620)	35.3% (0.344 - 0.361)	48.4% (0.435 - 0.534)	<0.0001

Fetal compromise†	19.9% (0.193 - 0.205)	12.0% (0.107 - 0.133)	25.3% (0.246 - 0.261)	16.0% (0.126 - 0.200)	
Miscellaneous fetal conditions††	3.2% (0.029 - 0.034)	1.7% (0.013 - 0.023)	3.7% (0.033 - 0.040)	1.8% (0.009 - 0.037)	
Maternal medical conditions§	16.9% (0.164 - 0.175)	14.3% (0.129 - 0.157)	19.2% (0.185 - 0.200)	19.1% (0.155 - 0.234)	
Pre labor rupture of membranes	12.1% (0.117 - 0.126)	9.2% (0.081 - 0.104)	13.0% (0.124 - 0.136)	11.5% (0.087 - 0.151)	
Onset of labor					
Induced	49.2% (0.485 - 0.500)	29.2% (0.274 - 0.311)	60.8% (0.600 - 0.617)	43.5% (0.386 - 0.485)	<0.0001
Spontaneous	50.8% (0.500 - 0.515)	70.8% (0.689 - 0.726)	39.2% (0.383 - 0.400)	56.5% (0.515 - 0.614)	
Fetal presentation					
Vertex	97.5% (0.972 - 0.977)	97.4% (0.967 - 0.980)	98.4% (0.982 - 0.986)	97.9% (0.959 - 0.989)	<0.0001
Others	2.5% (0.023 - 0.028)	2.6% (0.020 - 0.033)	1.6% (0.014 - 0.018)	2.1% (0.011 - 0.041)	
Fetal Position					
Non-Vertex	2.5% (0.023 - 0.028)	2.6% (0.020 - 0.033)	1.6% (0.014 - 0.018)	2.1% (0.011 - 0.041)	<0.0001
Vertex, occiput anterior	53.5% (0.528 - 0.543)	46.4% (0.444 - 0.484)	60.6% (0.597 - 0.614)	52.4% (0.473 - 0.573)	
Vertex, not occiput anterior	43.9% (0.432 - 0.447)	51.0% (0.490 - 0.530)	37.8% (0.370 - 0.387)	45.5% (0.406 - 0.506)	
Maternal position					
Dorsal/lithotomy/semi recumbent	90.1% (0.896 - 0.905)	79.6% (0.779 - 0.811)	92.7% (0.922 - 0.931)	88.7% (0.852 - 0.915)	<0.0001
Birth stool/squatting/Standing/Water	1.3% (0.012 - 0.015)	2.5% (0.019 - 0.032)	0.7% (0.006 - 0.009)	0.8% (0.003 - 0.023)	
All fours/kneeling	5.2% (0.049 - 0.056)	12.1% (0.108 -	4.1% (0.038 - 0.045)	6.5% (0.045 - 0.095)	

		0.135)			
Lateral	3.4% (0.031 - 0.036)	5.9% (0.050 - 0.069)	2.5% (0.023 - 0.028)	3.9% (0.024 - 0.064)	
Regional Anesthesia use in labor	42.9% (0.422 - 0.437)	13.2% (0.119 - 0.146)	51.9% (0.510 - 0.528)	24.6% (0.206 - 0.292)	
Duration of labor					
Total	4.97 (2.70-8.35)	3.80 (2.22-6.23)	5.47 (3.20-8.53)	3.62 (2.02-6.40)	<0.0001
First stage	4.17 (2.25-7.25)	3.33 (1.88-5.50)	4.58 (2.58-7.50)	3.12 (1.67-5.66)	
Second stage	0.47 (0.18-1.20)	0.35 (0.18-0.68)	0.50 (0.17-1.22)	0.33 (0.17-0.58)	
Intrapartum passage of meconium	12.9% (0.125 - 0.135)	11.2% (0.100 - 0.125)	14.5% (0.139 - 0.152)	17.3% (0.138 - 0.214)	<0.0001
Sentinel events					
Cord accident	16.5% (0.160 - 0.171)	17.5% (0.160 - 0.191)	15.4% (0.148 - 0.160)	16.5% (0.131 - 0.205)	0.07
Shoulder dystocia/abruption/failed instrumental	1.2% (0.011 - 0.014)	1.4% (0.010 - 0.019)	1.2% (0.010 - 0.014)	1.8% (0.009 - 0.037)	
Mode of birth					
Vaginal	70.4% (0.697 - 0.711)	90.1% (0.888 - 0.912)	52.1% (0.512 - 0.530)	67.3% (0.624 - 0.718)	0.00
Forceps assisted	10.9% (0.105 - 0.114)	2.5% (0.019 - 0.032)	16.5% (0.159 - 0.172)	8.9% (0.064 - 0.122)	
Vacuum assisted	8.4% (0.080 - 0.088)	4.4% (0.036 - 0.053)	13.8% (0.132 - 0.144)	17.5% (0.141 - 0.217)	
CS	10.3% (0.099 - 0.108)	3.0% (0.024 - 0.038)	17.6% (0.169 - 0.182)	6.3% (0.043 - 0.092)	
Incidence of asphyxia* (number of records, % proportion,95%CI)	383, 2.2 (1.98 - 2.41)	57, 2.4% (1.87 - 3.11)	400, 3.3% (3.03 - 3.68)	20, 5.2 (3.41 - 7.94)	<0.0001

Data are median (IQR)(as not normally distributed), n (%,95% Confidence Interval).

Only BMI had missing data as described

* By composite definition; † e.g. Abruption, post-term; § Including hypertensive disease in pregnancy; †† e.g. macrosomia, polyhydramnios, unstable lie

This table summarizes the clinical, obstetric, and neonatal characteristics of training cohort women stratified by fetal heart rate monitoring modality (ultrasound or ECG) and signal dropout level (>30% or ≤30%). Data are presented as medians (IQR) or proportions with 95% confidence intervals. Differences across groups were assessed using appropriate statistical tests.

Table 2: Causal Effects of Mode-Specific Signal Dropout on Perinatal Asphyxia in Training and Validation Cohorts

Dropout Group	Model	Adjusted Odds Ratio (aOR)[a]	95% CI	p-value
Training dataset				
ECG Low Dropout	IPW	1.51	1.32-1.71	<0.001
US High Dropout		1.25	0.99-1.58	0.06
ECG High Dropout		3.14	2.20-4.48	<0.001
ECG High Dropout at threshold 5%	TMLE	1.22	0.84-1.78	0.29
ECG High Dropout at threshold 10%		1.19	0.94-1.51	0.15
ECG High Dropout at threshold 15%		1.21	0.94-1.55	0.14
ECG High Dropout at threshold 20%		1.56	1.16-2.09	0.00
ECG High Dropout at threshold 25%		1.74	1.18-2.56	0.00
ECG High Dropout at threshold 30%		2.08	1.21-3.58	0.01
External Validation Dataset				
Czech dataset				
ECG High Dropout	IPW[b]	4.02	1.41-11.44	0.01
ECG Low Dropout		0.24	0.06-0.96	0.04
US High Dropout		1.90	0.94-3.82	0.07
ECG High Dropout at threshold 30%	TMLE	2.83	2.13-16.36	0.01
Australian dataset 2 (Werribee)				
ECG High Dropout	IPW[b]	2.22	1.61-3.05	<0.001
ECG Low Dropout		1.57	1.31-1.89	<0.001
US High Dropout		0.69	0.33-1.46	0.33
ECG High Dropout at threshold 30%	TMLE	1.91	1.11-3.28	0.02

[a] *All comparisons are made against the Ultrasound Low Dropout group, which serves as the reference category in both IPW and TMLE models.*

[b] *All IPW modelling was on a 30% dropout cutoff*

Adjusted odds ratios (aORs) from inverse probability–weighted logistic regression and dose-dependent targeted maximum likelihood estimation (TMLE) analyses show the association between mode-specific signal dropout and perinatal asphyxia in training and validation cohorts.

Figures

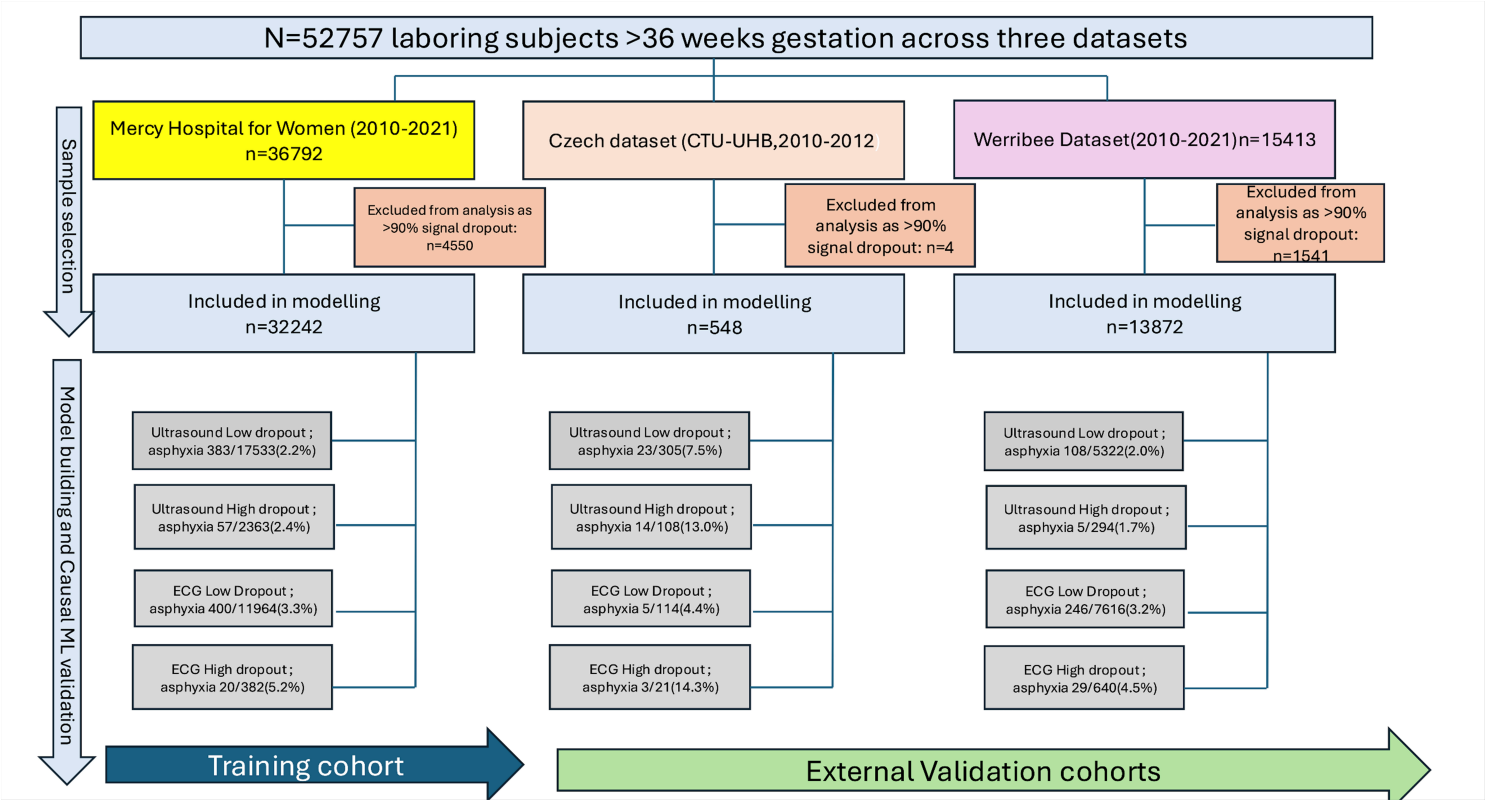

Figure 1

Multicohort sample selection and analysis pipeline.

Flow diagram showing inclusion criteria and sample selection across three cohorts used for model training and validation. The primary training dataset (Australia, n = 36,792), Czech validation set (n = 552), and a second Australian validation cohort (n = 15,413) are shown. All subjects were laboring at ≥36 weeks with interpretable CTG traces ≥15 minutes in the final 60 minutes before birth. Datasets were processed and harmonized using a unified preprocessing pipeline.

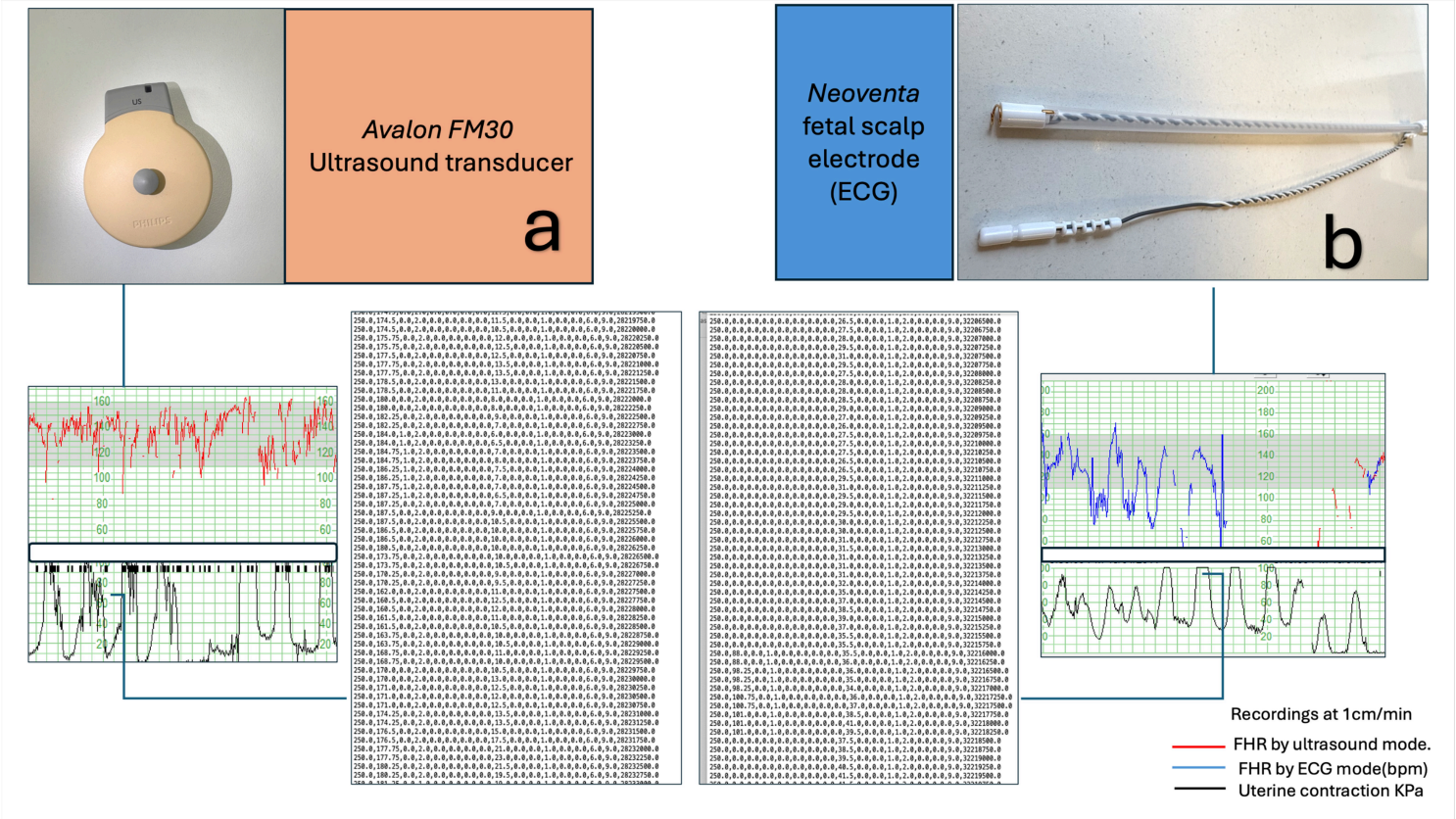

Figure 2

Mode-specific cardiotocography(CTG) traces and dropout patterns.

(a) CTG trace predominantly acquired in ultrasound mode(b) CTG trace predominantly in ECG mode. Representative CTG monitoring devices are shown alongside each trace (device models are illustrative and not necessarily identical to those used in the study). Traces reflect typical patterns observed in each exposure group, with paper speed standardized at 1 cm/min per Australian convention.

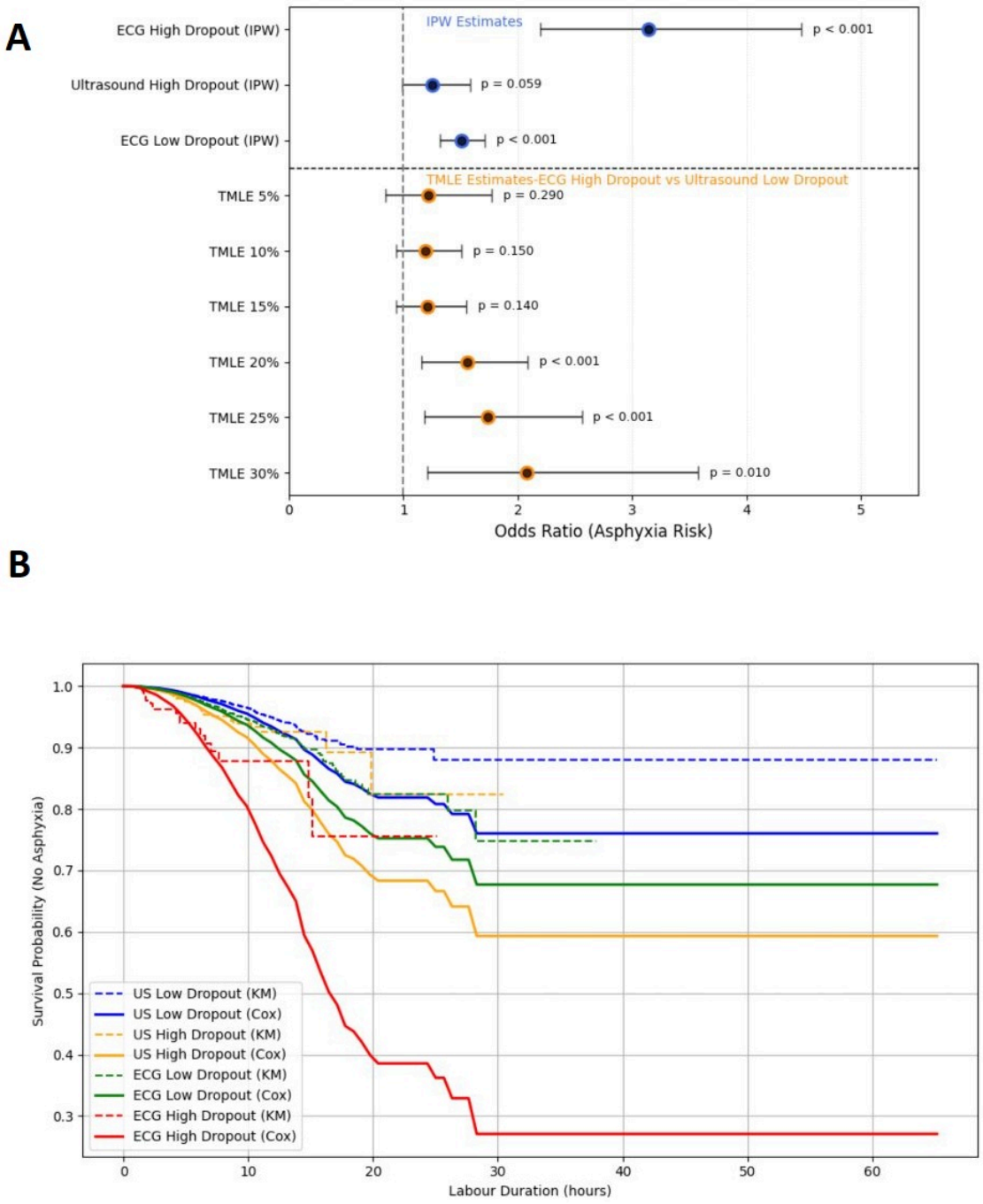

Figure 3

Risk of Perinatal Asphyxia by Monitoring Modality and Signal Dropout: Causal Odds Ratios and Survival Analysis

a) Forest plot summarizing the association between signal dropout and perinatal asphyxia across four mode-dropout exposure groups using causal inference (IPW) and causal machine learning (TMLE) in the primary Australian dataset. Odds ratios (ORs) with 95% confidence intervals are shown. The strongest association was

observed in the ECG High Dropout group, with both IPW and TMLE methods indicating increased risk relative to the reference group (Ultrasound Low Dropout).

b) Kaplan–Meier and IPW-adjusted Cox survival curves showing asphyxia-free survival across four exposure groups defined by monitoring modality and dropout level in the primary Australian dataset. (A) Unadjusted Kaplan–Meier curves demonstrate the steepest decline in the ECG High Dropout group. (B) IPW-adjusted Cox curves confirm persistently elevated hazards in this group, supporting its consistent association with higher asphyxia risk.

Supplementary Files

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

- SupplementaryMaterialmodesubmit.pdf