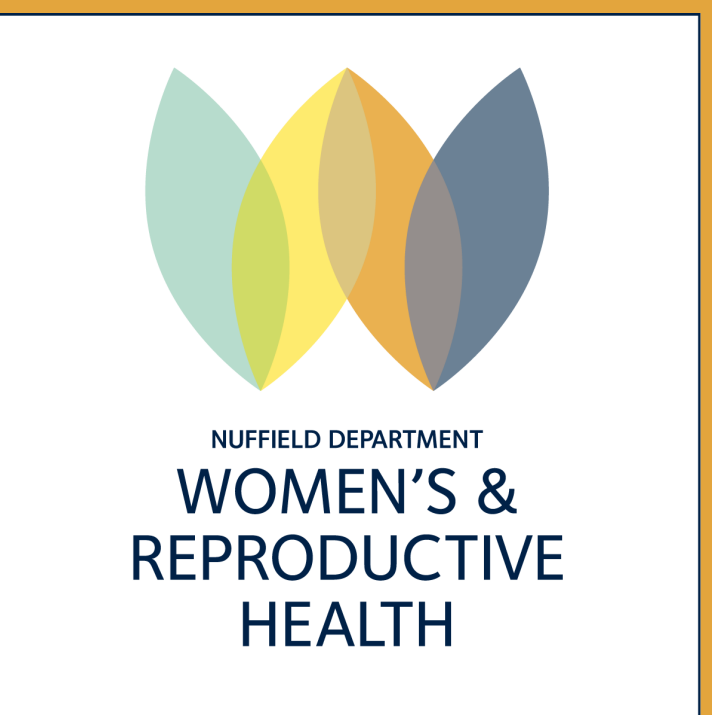




Pattern Evaluation and Recognition in Cardiotocography: Expert Interpretation and Visual Examination (PERCEIVE)

John Tolladay, Beth Albert, William R. Cooke, Manu Vatish, Gabriel Davis Jones

Objective

To assess expert inter-observer variability in antepartum cardiotocography (CTG) interpretation and compare it with the Dawes-Redman (DR) computerised system.

Methods

Dataset: 105 selected 15-minute segments of antepartum CTG traces.

Participants: 5 expert clinicians (3 midwives, 2 obstetricians, >10 years’ experience).

Annotations: Baseline, variability, accelerations and decelerations collected via a structured digital tool.

Majority Voting: Median participant baselines, mode of selected variability, accelerations and decelerations identified by majority have start and end points averaged.

DR Comparison: Same 105 CTG segments analysed using the Dawes-Redman system and compared to majority voted results.

Feature	Inter-observer Agreement	Comparison with Dawes-Redman
Baseline ICC = Intraclass correlation coefficient	Excellent (ICC = 1.0, 91.4% within ± 5 bpm)	Excellent (ICC = 1.0, 90.2% within ± 5 bpm)
Variability	Moderate (Fleiss' κ = 0.39)	Poor (Fleiss' κ = 0.002)
Accelerations PPV = Positive predictive value	Sensitivity: 39.2–87.7% PPV: 63.1–91.3%	Sensitivity: 64.8% PPV: 85.0%
Decelerations PPV = Positive predictive value	Sensitivity: 45.7–97.2% PPV: 54.2–77.8%	Sensitivity: 50.0% PPV: 20.0%

Interpretation & Implications

Clinician Judgement: Baseline interpretation was consistent, but variability and deceleration assessments showed substantial disagreement.

DR System: Accurate for baseline and accelerations; poor sensitivity and PPV for decelerations, and limited agreement on variability.

Clinical Relevance: Discrepancies interpreting decelerations and variability may lead to missed fetal compromise or unnecessary interventions.

Recommendation: Improved training, standardised displays and hybrid AI-clinician approaches may enhance consistency and clinical reliability.

Conclusion

While expert and algorithmic interpretation show high concordance for baseline assessment, variability and decelerations remain inconsistently classified, particularly for patterns associated with fetal compromise. These findings underscore the need for structured training, harmonised display formats, and integrated clinical AI workflows to strengthen antepartum CTG interpretation and decision making.

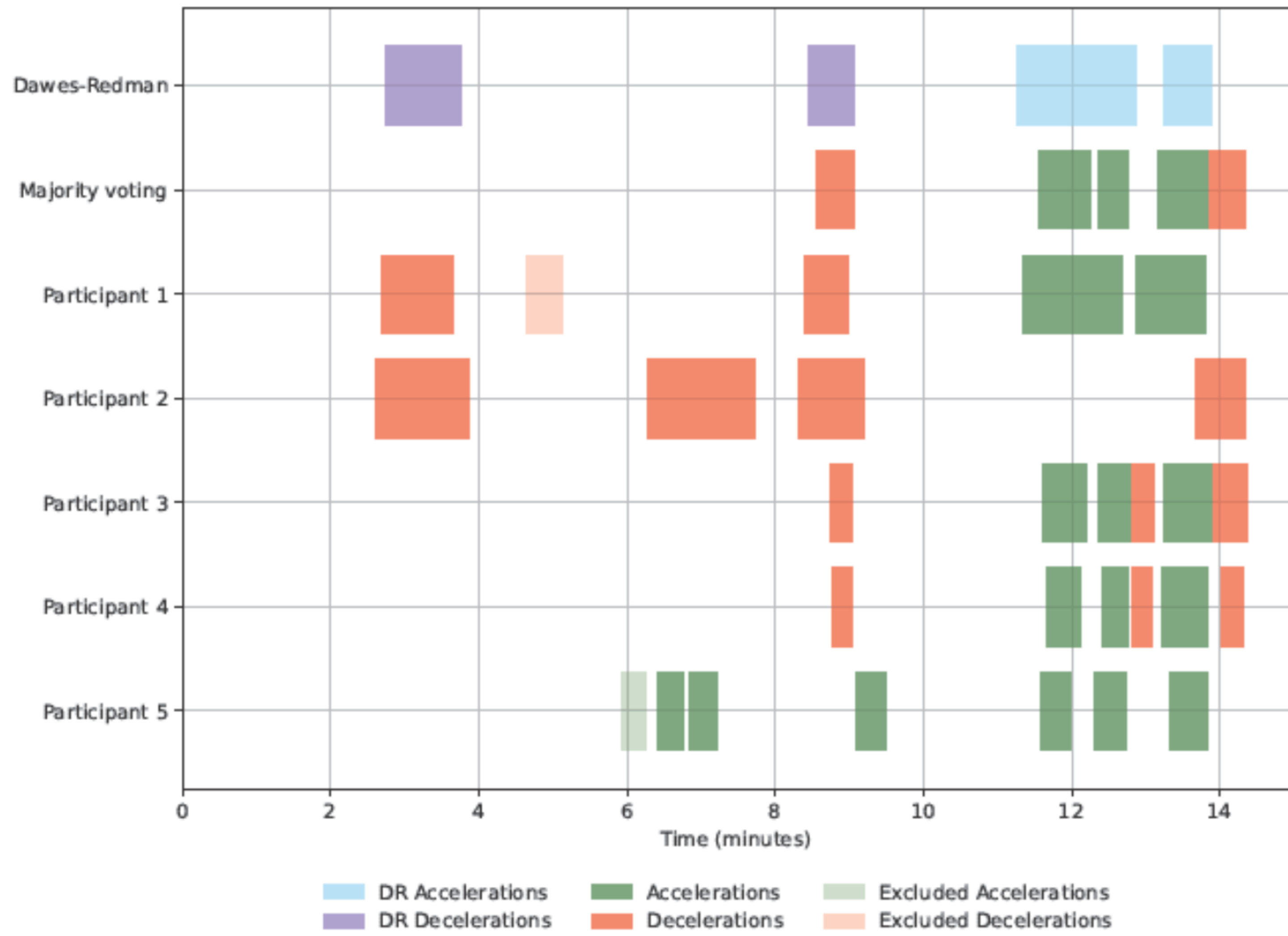


Figure 2 (above): Comparison of pattern annotations by clinicians and the Dawes-Redman system for a single 15-minute fetal heart rate segment. This example illustrates the timing and classification of accelerations and decelerations as identified by five senior clinicians, majority voting, and the Dawes-Redman (DR) system. Green and red boxes represent accelerations and decelerations, respectively, as annotated by individual clinicians. Shaded boxes denote patterns excluded post hoc for failing to meet FIGO criteria (minimum 15 bpm amplitude and 15 second duration). Blue and purple boxes indicate accelerations and decelerations detected by the DR system. Areas of concordance and discordance between DR and clinician annotations are visible, including DR-identified patterns with limited expert agreement and a deceleration marked by the majority of participants but missed by DR. This demonstrates areas of agreement and divergence in pattern recognition across human and automated interpretation.

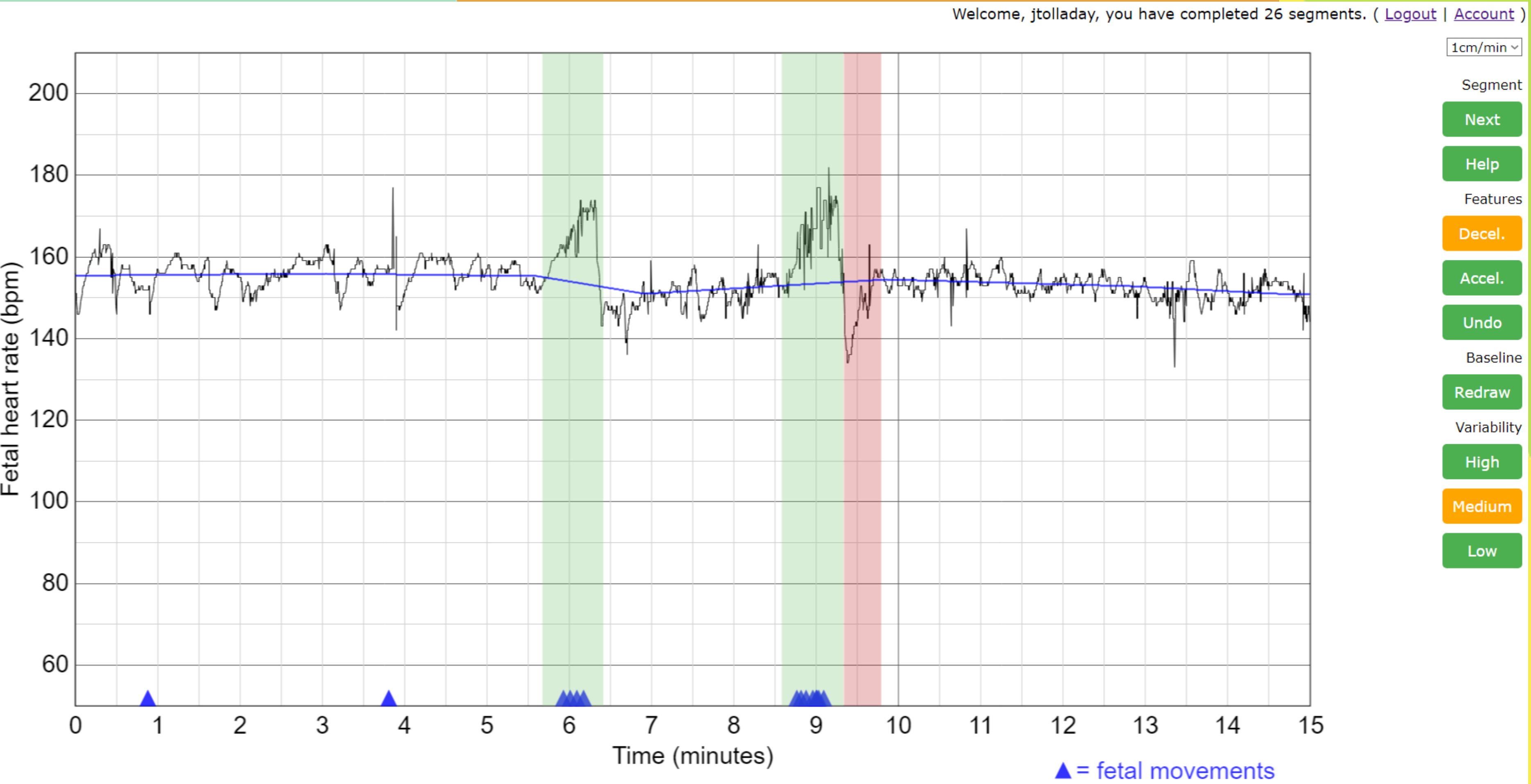


Figure 1: Example FHR segment displayed in the annotation tool used by participants. The FHR (black line) and fetal movements (blue triangles) are displayed to the participant. Accelerations (green shading) and decelerations (red shading) are drawn by dragging non-overlapping boxes over the plot with minimum duration of 15s. The baseline (blue line) is drawn by marking plots along the trace that are then connected. Variability classification is assigned using one of the three buttons on the right hand side: “low”, “medium” or “high”. Basic guidance for annotation provided on clicking “Help” button

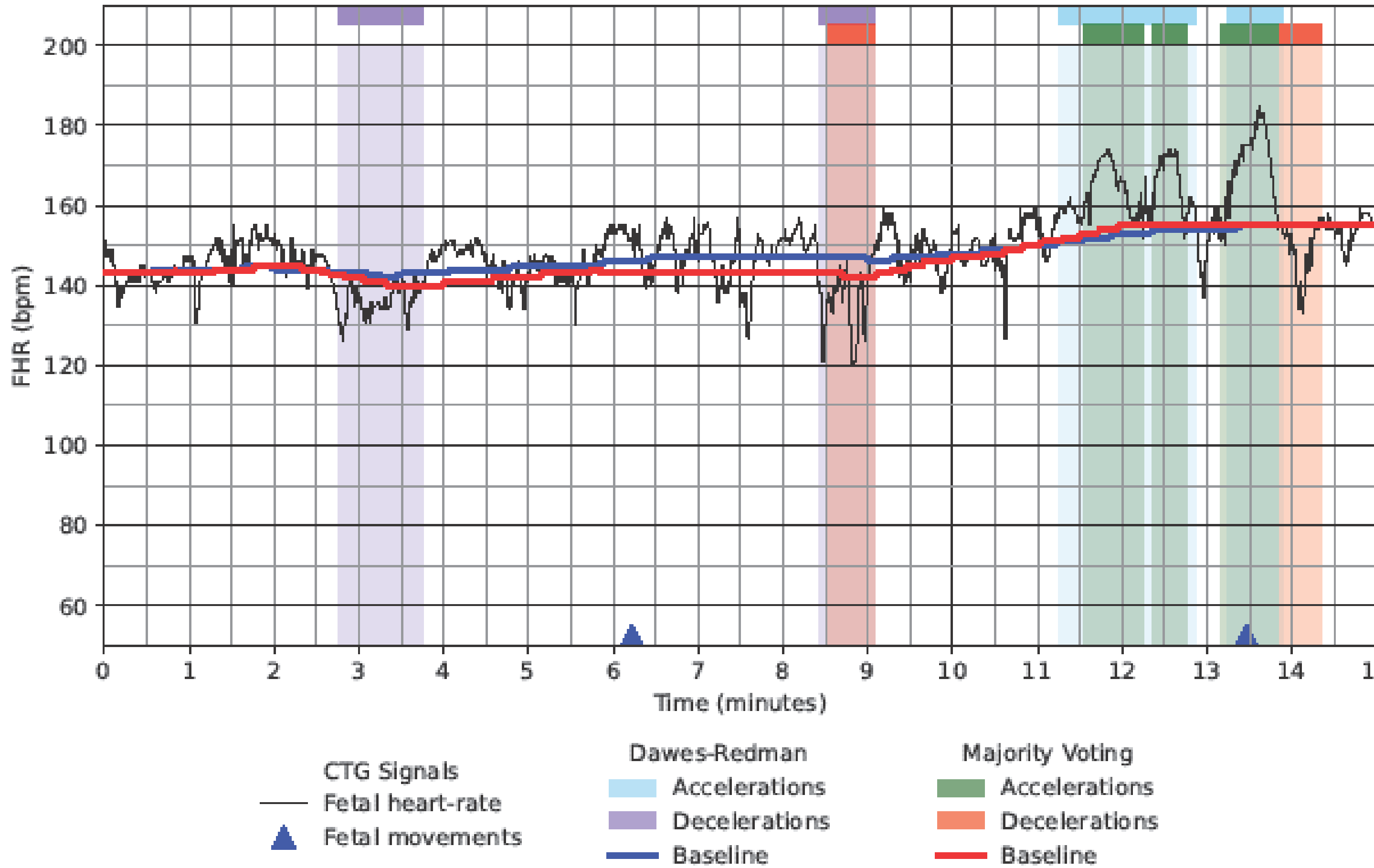


Figure 3 (right): Comparison of baseline assignment and event detection by the Dawes-Redman system and majority-voted clinician annotations for a single FHR segment. A 15-minute FHR trace is shown with superimposed annotations from the Dawes-Redman (DR) algorithm and majority-voted expert clinician interpretations. The black line represents the raw FHR signal. Blue triangles indicate fetal movements. Shaded regions correspond to accelerations (green) and decelerations (red) as identified by majority vote, and to DR-identified accelerations (blue) and decelerations (purple). The baseline determined by each method is plotted in red (majority vote) and blue (DR). This segment was classified as having “medium” variability by both DR and most participants. Differences in event detection are visible, including one deceleration marked by DR but not by clinicians, and a set of accelerations split by clinicians into two separate events, while DR identified a single continuous episode.