



Improving Obstetricians' Accordance in Fetal Monitor Interpretation During Childbirth

Rotem Shapira, Reuven Kedar, Amit Damti and Yael Yaniv
Biomedical Engineering Faculty, Technion, Haifa, Israel; Department of Obstetrics & Gynecology, Carmel Medical Center, Haifa, Israel; Rappaport Faculty of Medicine, Technion, Israel

OBJECTIVE

Despite decades of usage of the fetal monitor and implementation of various guidelines to improve uniformity and accuracy of interpretation, it is evident that the fetal monitor is prone to errors, leads to high false-positive predictive value and to a rise in the rate of cesarean sections, without reducing the rate of cerebral palsy or negative perinatal outcomes. Moreover, past studies found high inter-observer variability in interpretation, emphasizing the inconsistent nature of decision-making in the delivery room.

In this study, we assessed the ability of displaying the fetal heart rate baseline on top of the fetal monitor trace to improve diagnostic performance and agreement between obstetricians of different experience levels in birth management cases.

STUDY DESIGN

This study is a randomized controlled trial done retrospectively using 30 births of women aged 21-38 who gave birth at gestational age range of 35-41 weeks in Carmel Medical Center, Haifa, Israel, between March 2016 and November 2022.

— Baseline Calculation —

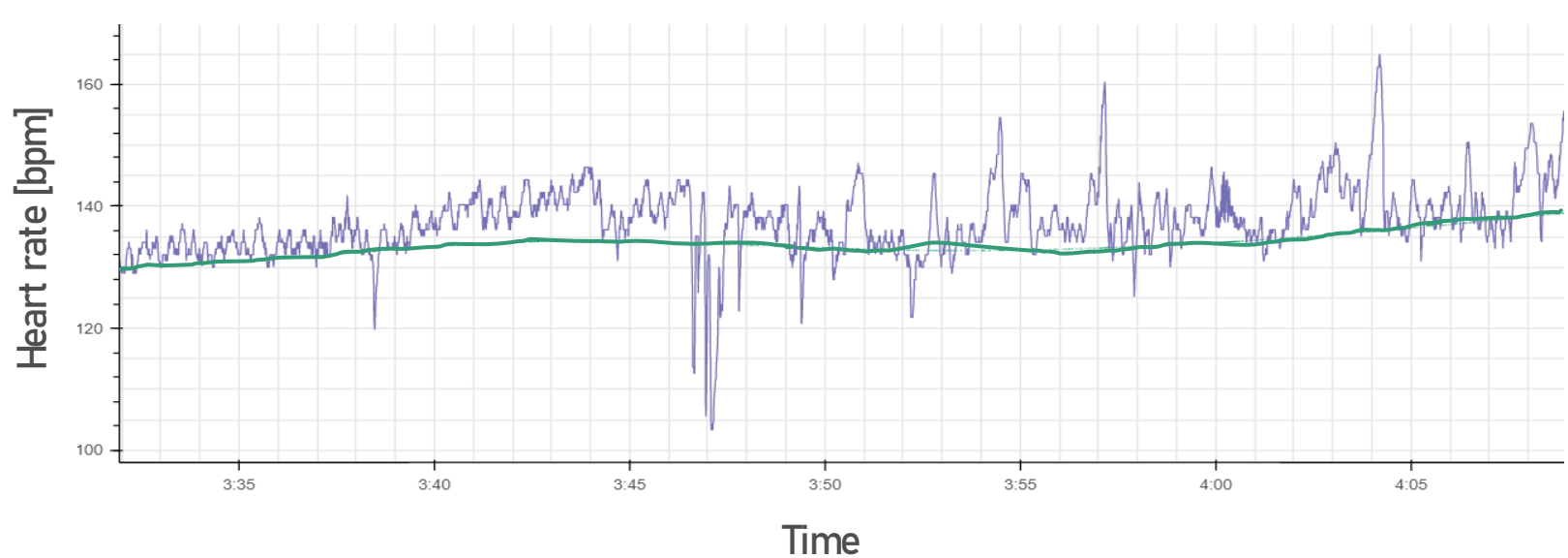
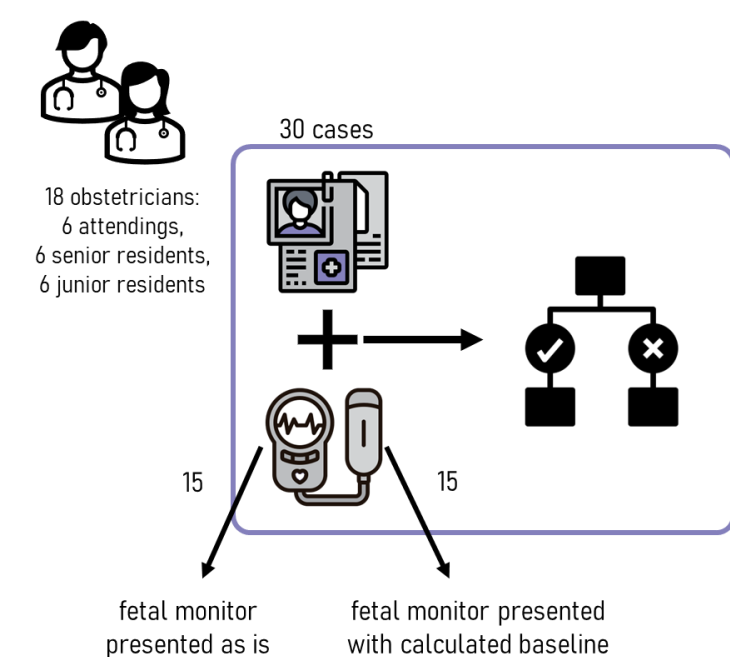


Figure 1: Illustration of the baseline calculation using the auto-regressed double-sided improved asymmetric least squares algorithm. The purple line is the fetal heart rate, and the green line is the calculated baseline, both measured in beats-per-minute.

— Experiment Protocol —



CTG Decision Making

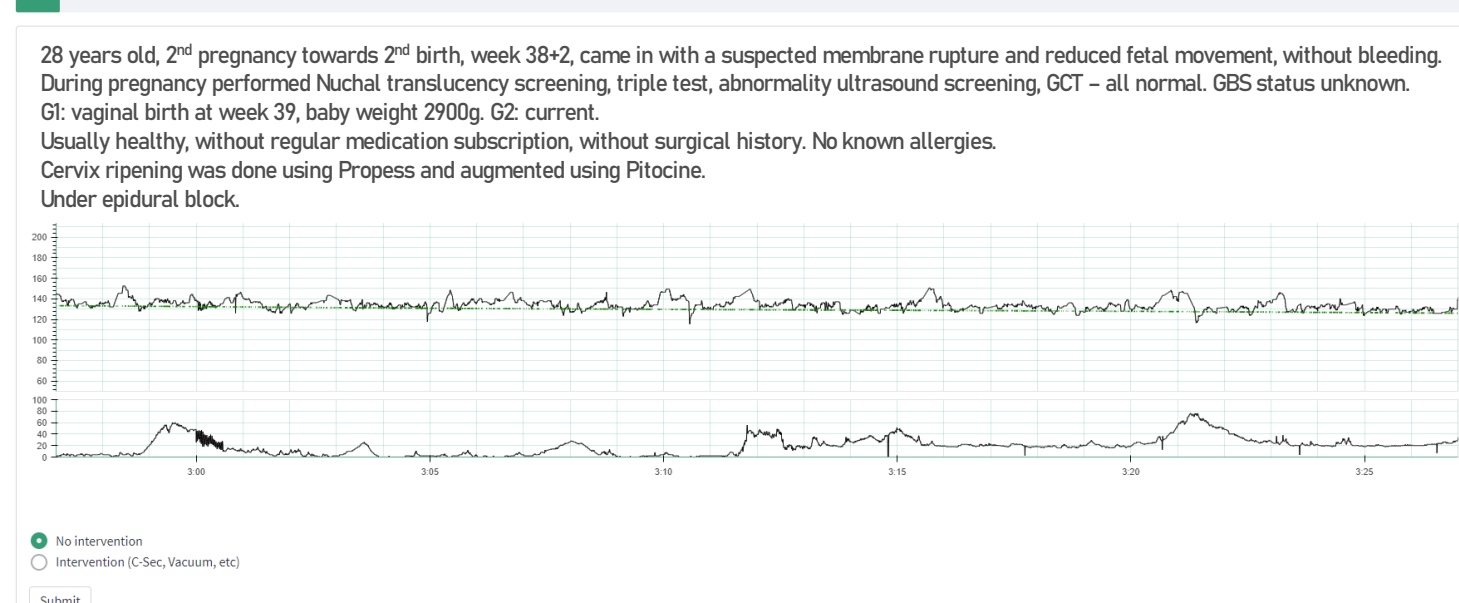
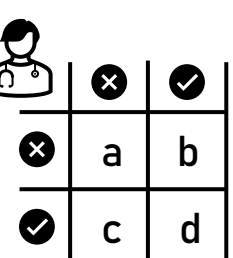


Figure 2: Method section graphical abstract. On the left, illustration of the experiment protocol. 18 obstetricians of different expertise levels were presented with 30 cases via a computerized interface and were asked to decide if an intervention is needed. For each case, a clinical summary of the patient and the CTG trace were given. The control branch included 15 cases in which the CTG traces were presented as is, and the study branch included the other 15 cases in which the calculated baseline was shown on top of the CTG trace. On the right, a screenshot of the user interface participants used to annotate the samples. The "ground truth" of each birth was determined based on the pH of the umbilical artery (<7.10) and the Apgar score of the newborn (<8 after 1, 5, and 10 minutes of birth).

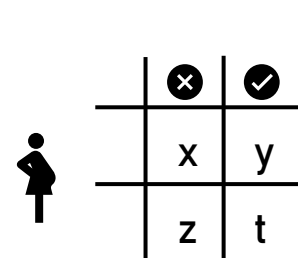
— Statistical Analysis —

Cohen's Kappa



A measure of how reliably two raters measure the same thing.

Fleiss Kappa



$$x+y=z+t=N$$

A measure of how reliably three or more raters measure the same thing.

RESULTS

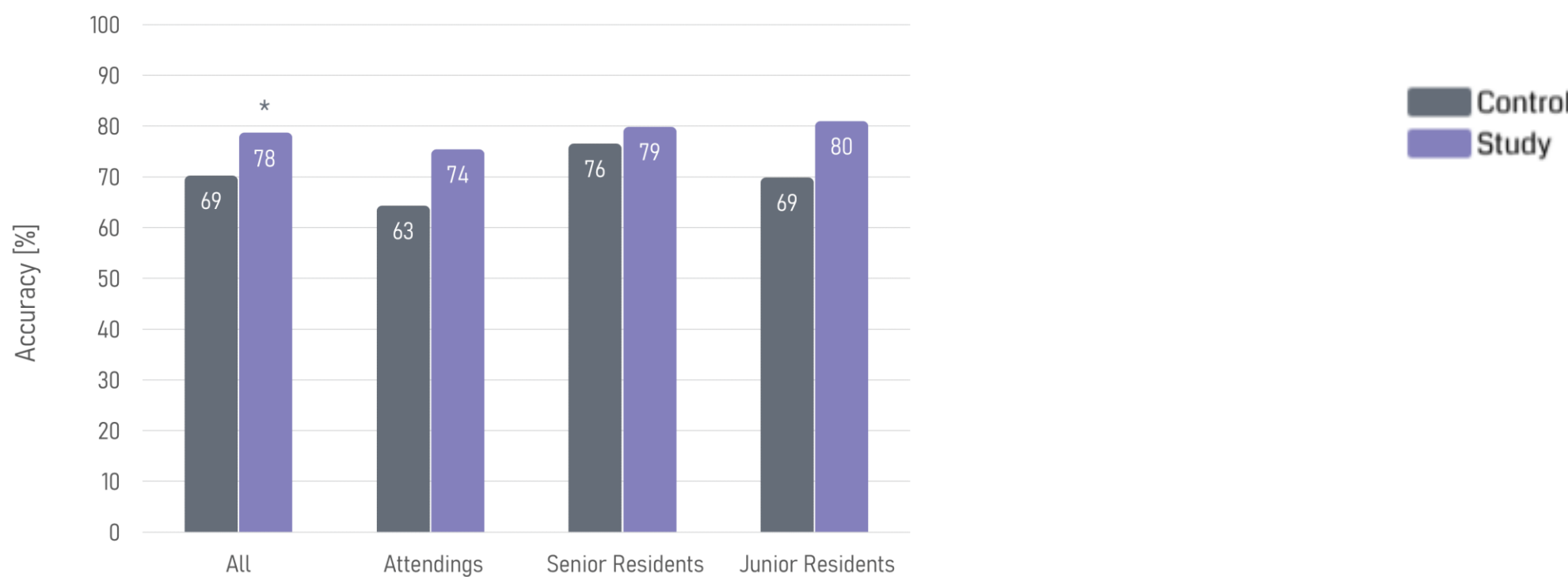


Figure 3: Comparison of accuracy in birth management decision-making between the control and study branches across the different groups of participants. * stands for statistical significance ($p < 0.05$).

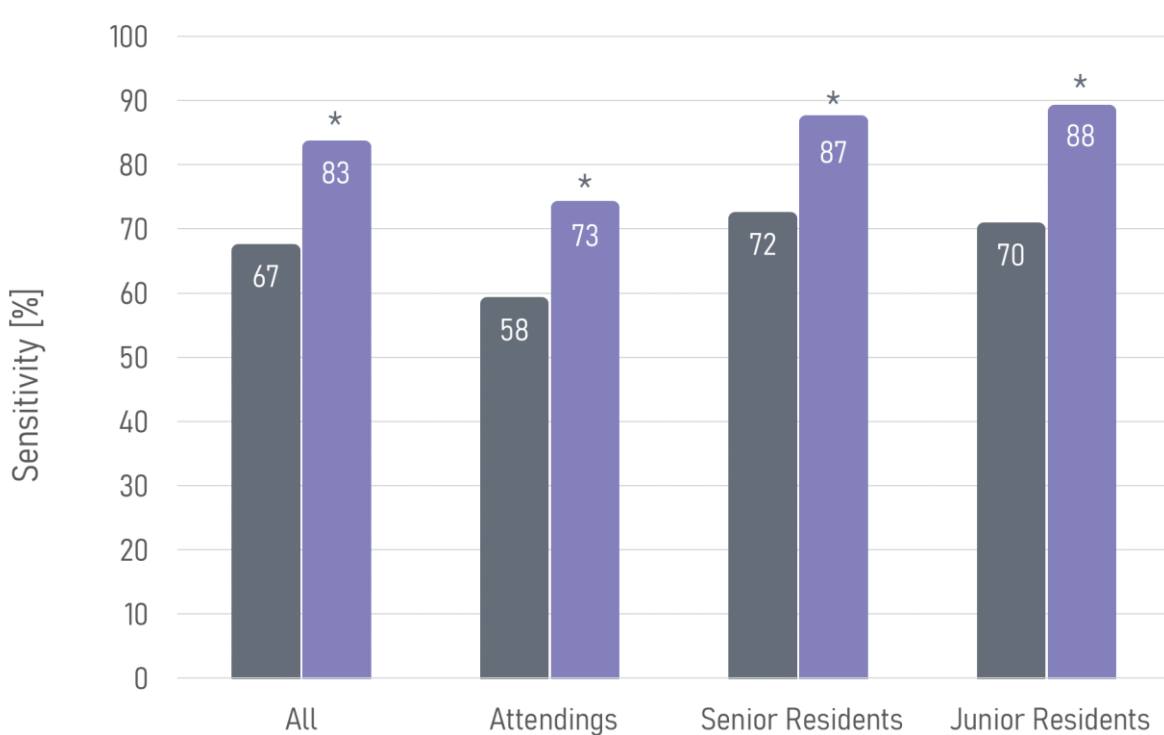


Figure 4: Comparison of sensitivity in birth management decision-making between the control and study branches across the different groups of participants. * stands for statistical significance ($p < 0.05$).

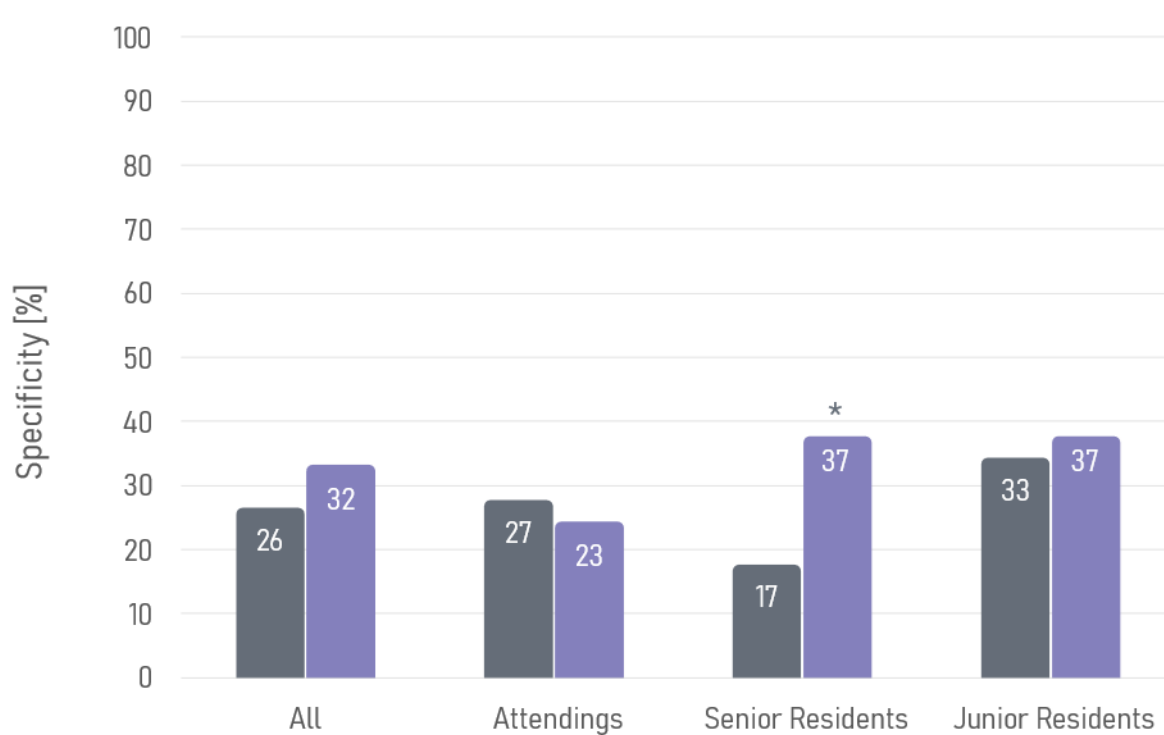


Figure 5: Comparison of specificity in birth management decision-making between the control and study branches across the different groups of participants. * stands for statistical significance ($p < 0.05$).

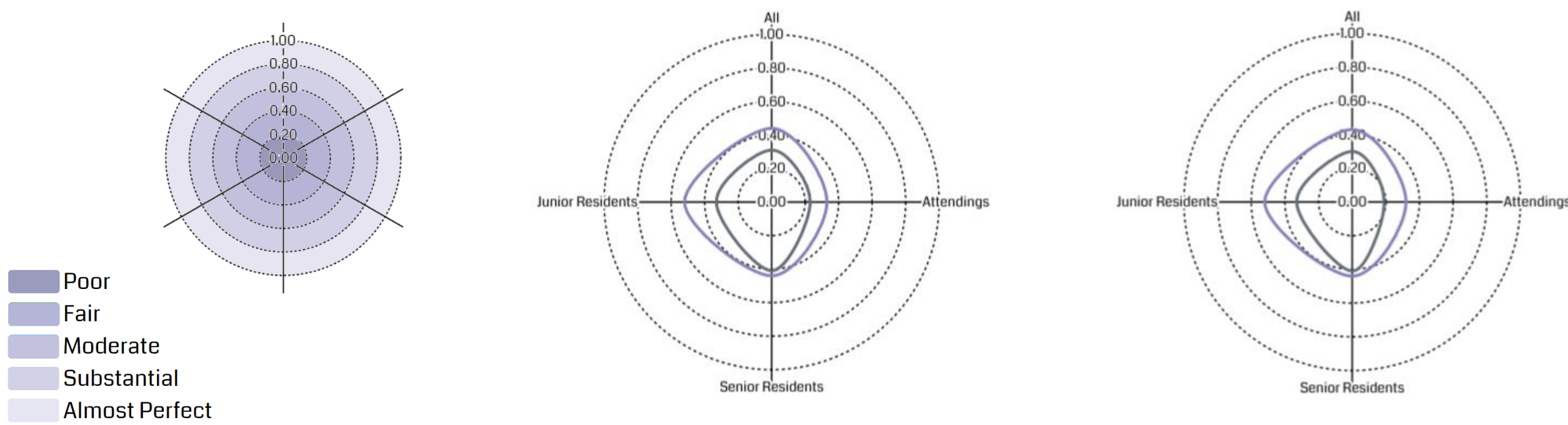
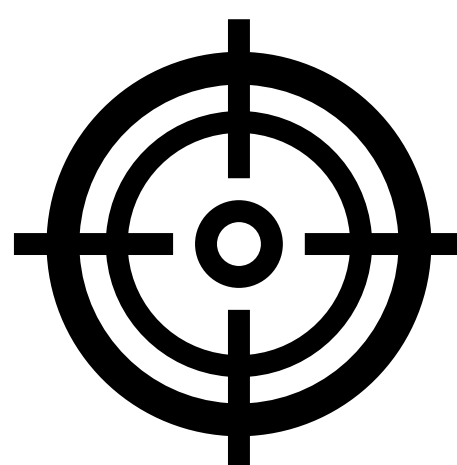


Figure 6: Comparison of inter-observer variability in birth management decision-making between the control and study branches across the different groups of participants. From left to right: kappa ladder legend 0-0.2 is considered poor, 0.2-0.4 fair, 0.4-0.6 moderate, 0.6-0.8 substantial and 0.8-1 almost perfect; Cohen's kappa results in the control branch are 0.31, 0.23, 0.41 and 0.33 for everyone, attendings, senior residents and junior residents respectively, and 0.44, 0.33, 0.44 and 0.52 in the study branch, $p < 0.05$ for everyone, attendings and junior residents; Fleiss' kappa results in the control branch are 0.3, 0.19, 0.41, 0.33 for everyone, attendings, senior residents and junior residents respectively, and 0.43, 0.32, 0.44 and 0.52 in the study branch. The control branch is displayed in grey; the study branch is displayed in purple.

	All	Attendings	Senior Residents	Junior Residents
Control	7.1	7.8	7	6.3
Study	5.1	6.5	4.5	4.3

Table 1: Comparison of the average number of interventions chosen between the control and study branches across the different groups of participants.

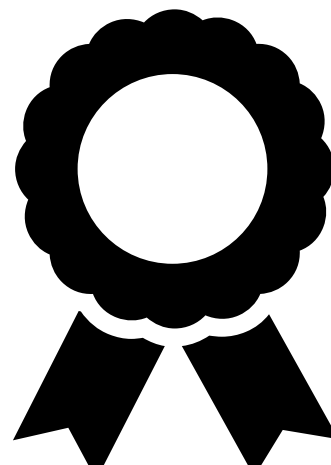
CONCLUSIONS



Improvement of the accuracy, sensitivity and specificity shows better diagnostic precision when the baseline is presented



Both the Cohen's and Fleiss' kappa rose from fair to moderate indicating improved inter-observer agreement in the study branch



Attendings did not outperform the other subgroups



Less interventions in the study branch

MAIN FINDINGS

Our study demonstrated that adding the FHR baseline to the CTG trace may improve accuracy, sensitivity, and specificity of CTG interpretation, along with lower inter-observer variability. Thus, this tool's addition may improve the ability of obstetricians to predict bad outcomes and assist in their decision making accordingly.

ACKNOWLEDGEMENTS

Thank you to the attendings and residents of the Obstetrics & Gynecology department of Carmel Medical Center for their participation in the study.