Supplemental Material

To evaluate the cluster stability, we utilized a bootstrap resampling technique as outlined by Christian Henning [Hennig, Christian. "Cluster-wise assessment of cluster stability." *Computational Statistics & Data Analysis* 52.1 (2007): 258-271.].

To begin, the complete set of data is clustered. Next a bootstrap sample is drawn (with replacement, matching the size of the original dataset). This bootstrap sample is then clustered using the same value of K. Each cluster in the bootstrap sample is then compared to each of the original K clusters using the Jaccard coefficient (JC), resulting in a measure of overlap of instances in the clusters. The maximum JC is recorded for each of the original K clusters. Clusters with a JC value greater than 0.5 are commonly assumed to be "stable" as the resampled data produced a cluster with a majority of points matching an original clustering. The bootstrap sampling is then repeated several thousand times.

Supplemental Table 1 presents the percentage of bootstrap iterations in which the K original clusters were deemed to be stable (max JC 0.5)

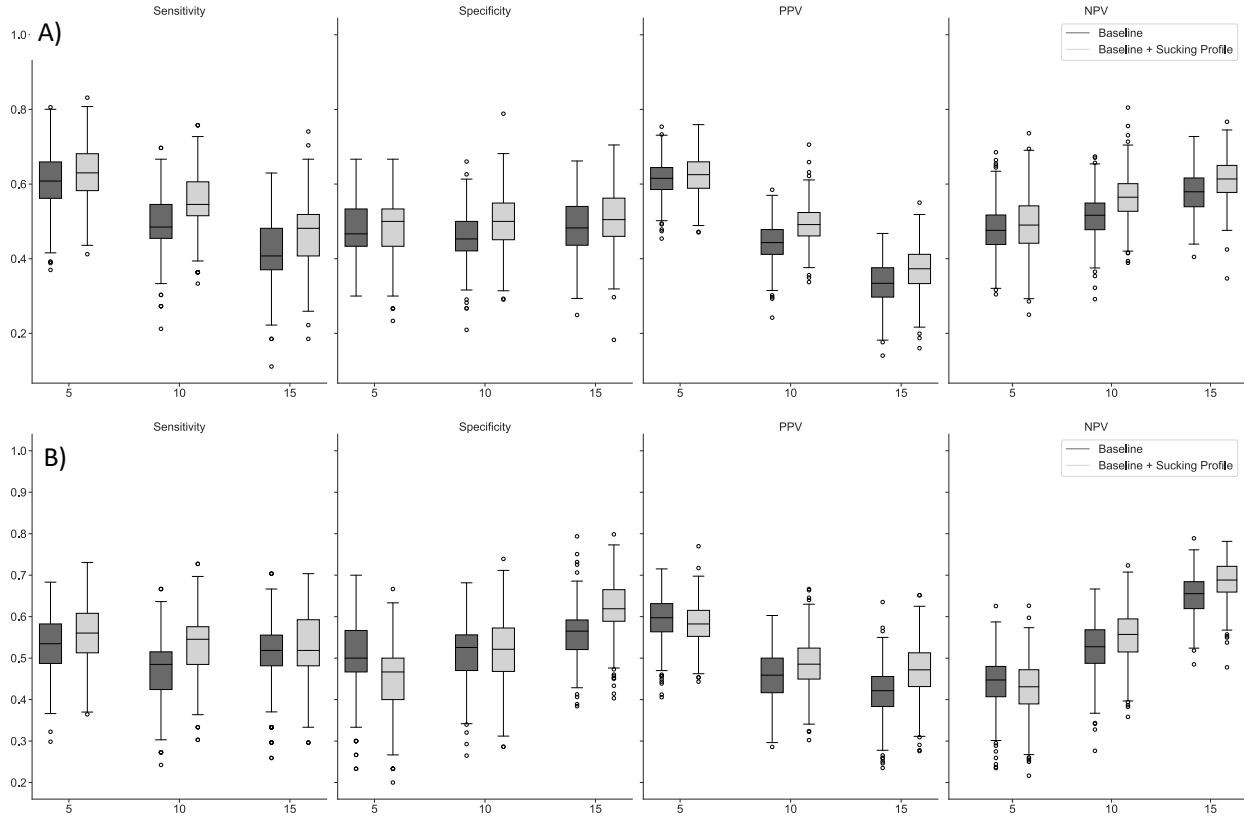
	Cluster ID				
K	**1**	**2**	**3**	**4**	**5**
3	67.8	51.6	71.8	X	X
4	49.5	77.5	39.1	73.0	X
5	43.8	63.9	42.3	52.0	43.4

Supplemental Table S1: Percentage (%) of 1000 bootstrap iterations for which the maximum JC for the original cluster exceeded .5. Cells marked with an "X" are not applicable to the respective K value.

Metric	*Correlation Type*		
	Pearson	**Spearman**	**Kendall**
Peak Amplitude	0.04	0.07	0.05
Max Peak	-0.01	-0.01	-0.01
Effect	0.20	0.23	0.15
Frequency (Hz)	-0.02	-0.04	-0.03
Interval	-0.04	-0.01	-0.01
Smoothness	0.01	0.05	0.03
IP Duration	0.00	0.02	0.01
DP Duration	0.05	0.07	0.05
Peak Amplitude CV	-0.13	-0.09	-0.06
Frequency CV	0.00	0.03	0.02
Smoothness CV	-0.07	-0.10	-0.06
Spectral Centroid	0.00	-0.07	-0.04
Spectral Spread	-0.03	-0.04	-0.03
Entropy	0.04	0.04	0.03

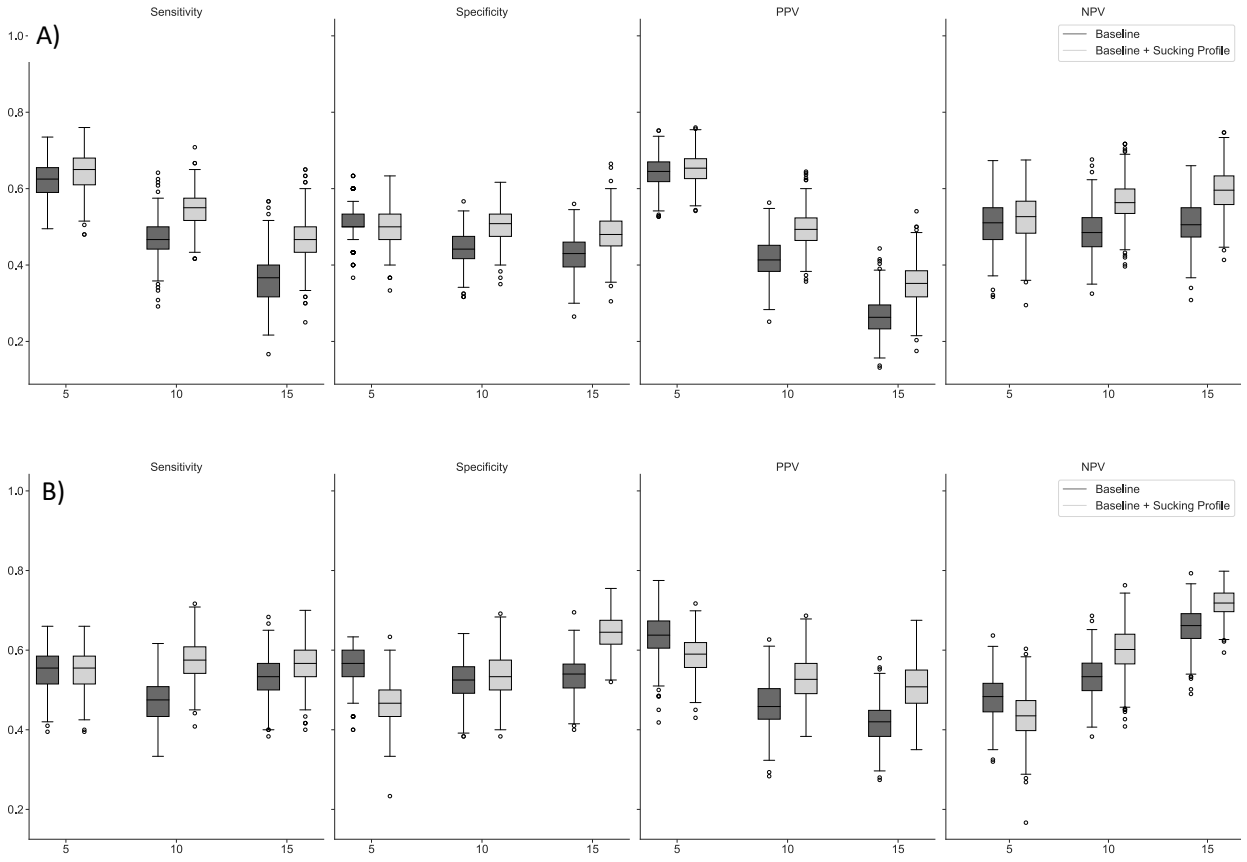
Supplemental Table S2: Correlation between sucking metrics and time since last feed (minutes)

3-Fold Weight Change Results



Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone) vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity/ Specificity/Positive Predictive Value [PPV]/Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.

10-Fold Weight Change Results



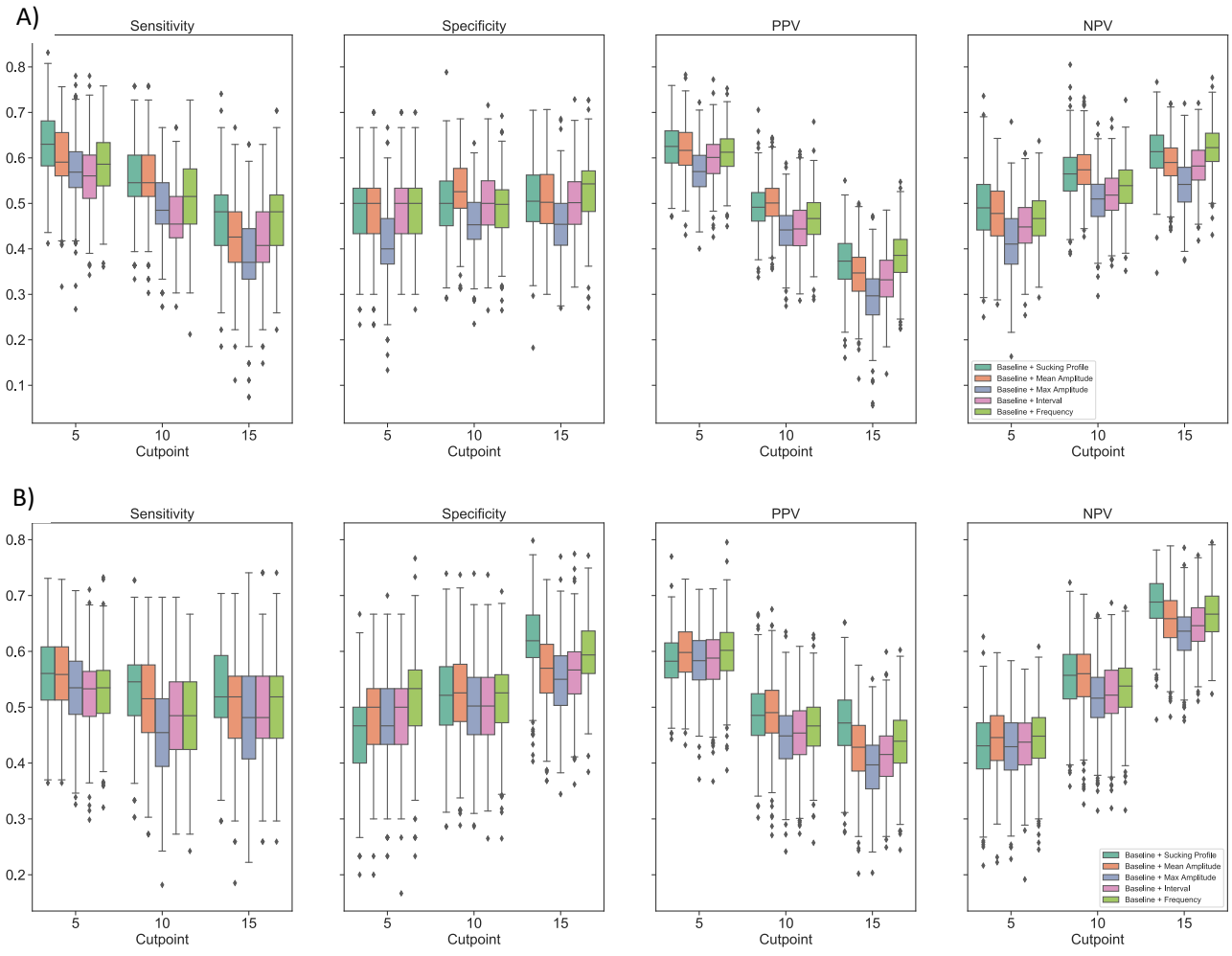
Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone) vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity/ Specificity/Positive Predictive Value [PPV]/Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.

Profile Comparison to Univariate Features

Univariate Sucking Measures

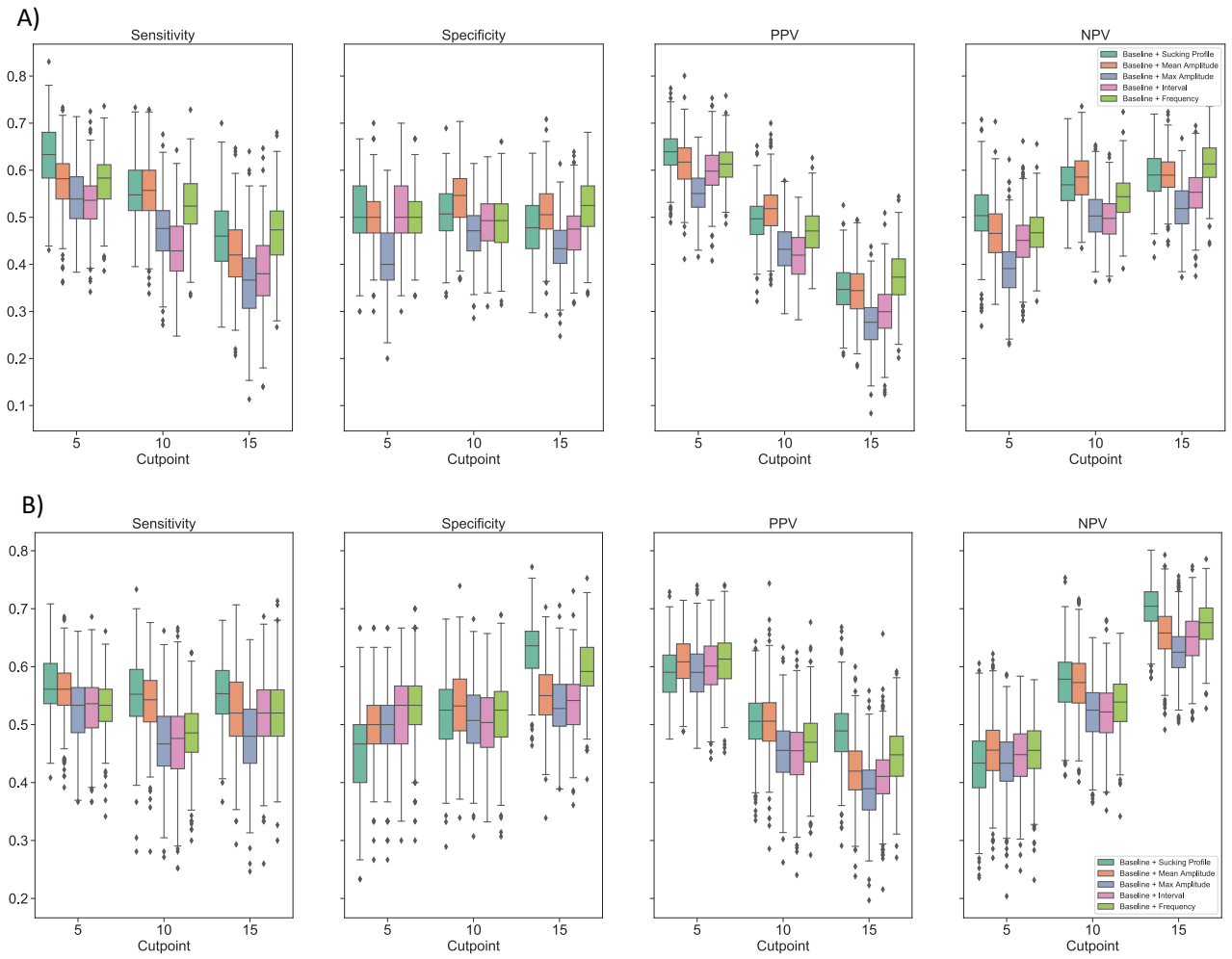
We compared the predictive performance of the sucking profiles (and baseline data) to use of baseline data + a univariate sucking measure (also adjusted for time since last feeding). Models were evaluated using the same repeated k-fold technique as in the main text, with univariate and profile models trained and tested on the same data each iteration. Both KNN and Ridge regression classifiers were used, and results are shown for 3-,5-, and 10-fold configurations.

3-Fold Results



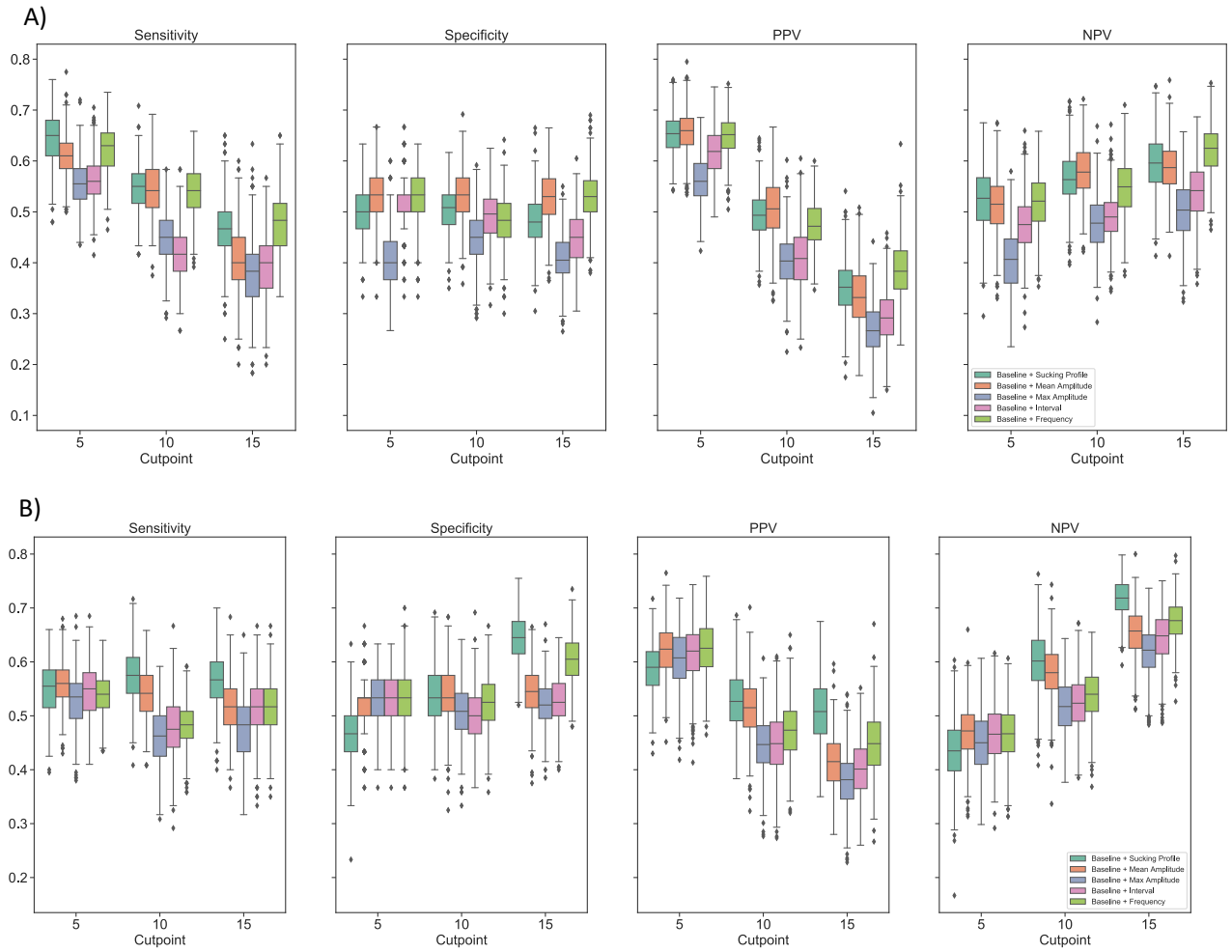
Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI) + one sucking metric [Mean Amplitude, Max Amplitude, Interval, Frequency] vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity / Specificity/Positive Predictive Value [PPV] / Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.

5-Fold Results



Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI) + one sucking metric [Mean Amplitude, Max Amplitude, Interval, Frequency] vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity / Specificity/Positive Predictive Value [PPV] / Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.

10-Fold Results

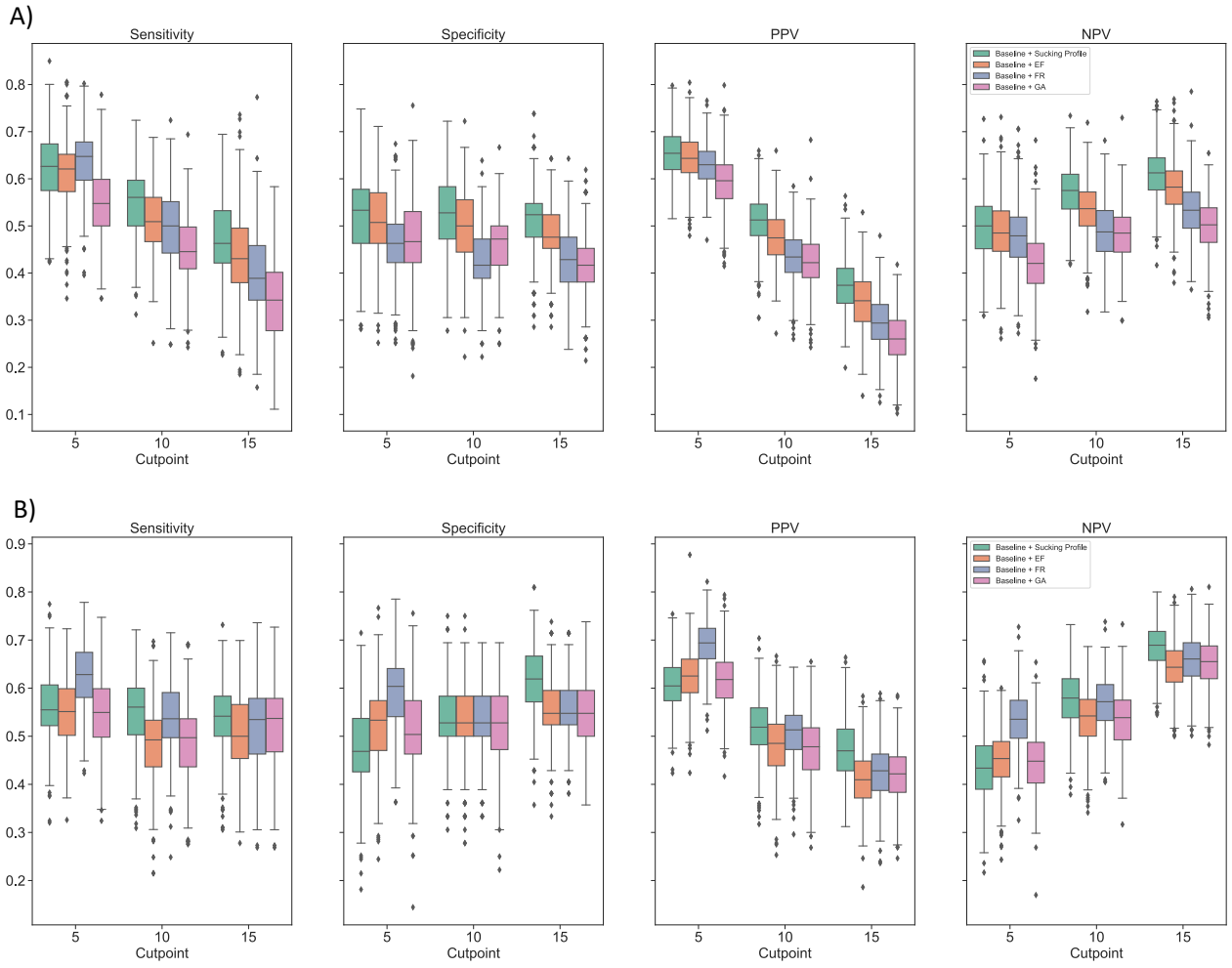


Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI) + one sucking metric [Mean Amplitude, Max Amplitude, Interval, Frequency] vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity / Specificity/Positive Predictive Value [PPV] / Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.

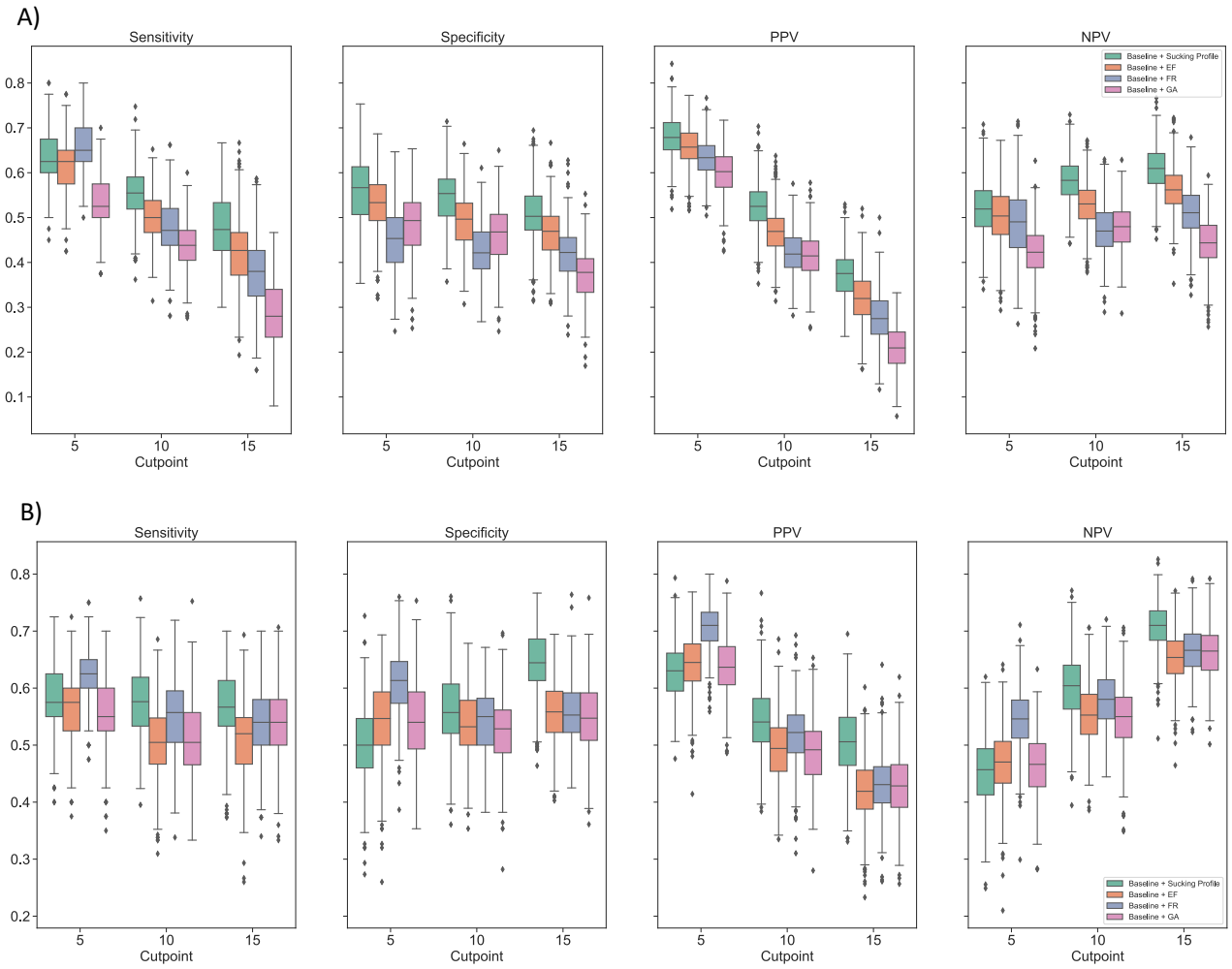
Parent Reported Eating Behaviors (BEBQ) Subscales

In this analysis we compared the predictive performance of the sucking profiles (and baseline data) to the use of baseline data + a subscale of the parent reported BEBQ administered at 4m. Models were evaluated using the same 500-iteration repeated k-fold technique as in the main text, with univariate and profile models trained and tested on the same data each iteration. Both KNN and Ridge regression classifiers were used. Results are again shown for 3-,5-, and 10-fold configurations.

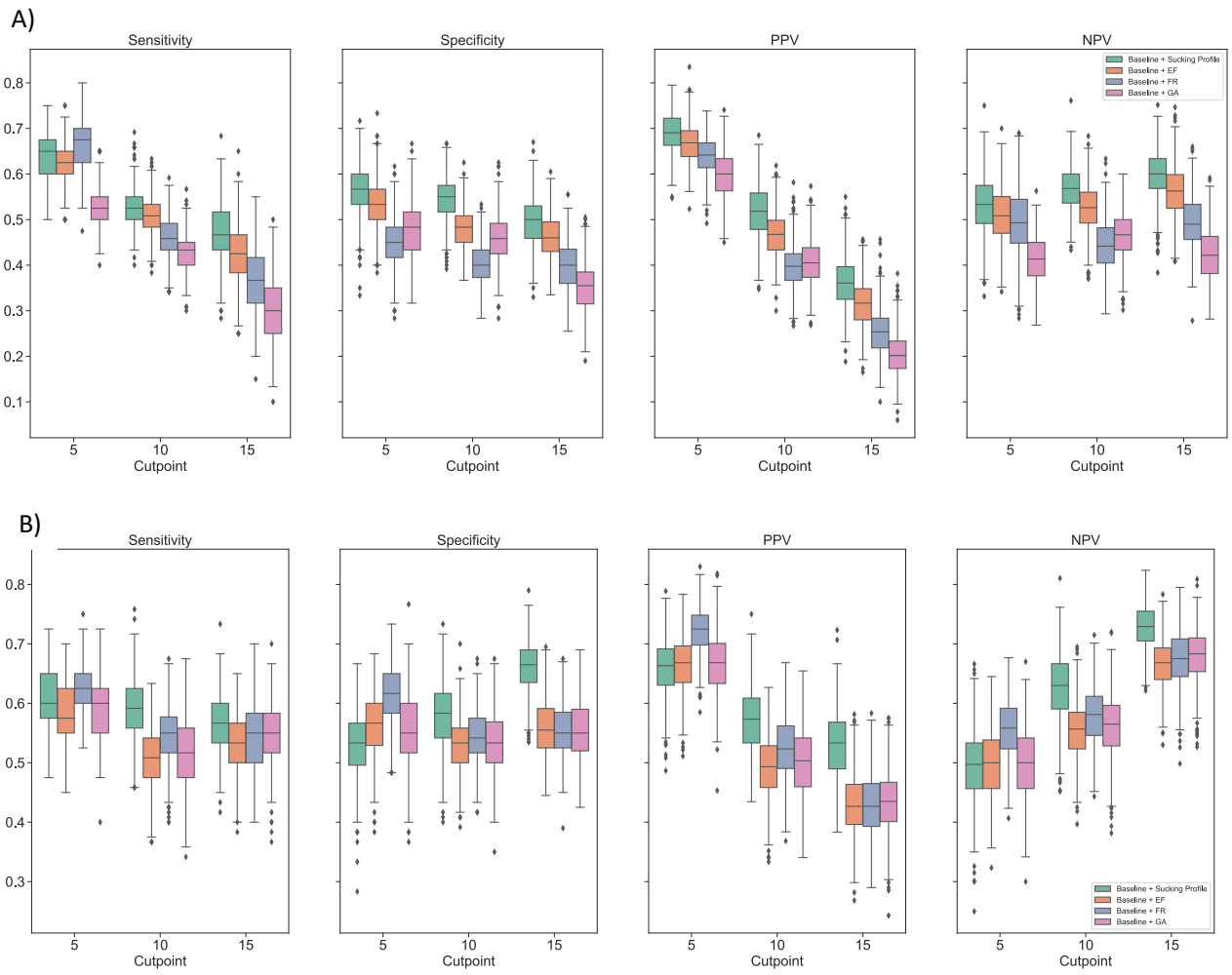
3-Fold Results



Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI) + BEBQ Subscale [EF: Enjoyment of Food, FR: Food Responsiveness, GA: General Appetite] vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity / Specificity/Positive Predictive Value [PPV] / Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.



Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI) + BEBQ Subscale [EF: Enjoyment of Food, FR: Food Responsiveness, GA: General Appetite] vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity / Specificity/Positive Predictive Value [PPV] / Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.



Classification performance of infants by ΔWFA percentiles from 4 to 12 months comparing Baseline predictors (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI) + BEBQ Subscale [EF: Enjoyment of Food, FR: Food Responsiveness, GA: General Appetite] vs. Baseline+Sucking (infant sex, race/ethnicity, birthweight, gestational age, WFA percentile at 4 months and maternal pre-pregnancy BMI alone and sucking profile) predictors. Varying thresholds for high change are found on the X-axis. Each frame of the figure represents a performance metric (Sensitivity / Specificity/Positive Predictive Value [PPV] / Negative Predictive Value [PPV]). Panels A/B represent KNN and Ridge Classifiers respectively. Models adjusted for infant sex, race/ethnicity, birthweight, gestational age, WFA at age 4 months and maternal pre-pregnancy BMI.