

Supplementary Material: Dynamic HRV Assessment Based on Gamma Auditory Stimulation

Supplementary Material S1: Multi-center Standardized Operating Procedures

To minimize inter-center differences and ensure consistency in data collection and assessment, this study established and strictly implemented the following Standard Operating Procedures (SOP):

Initial Training

All participating researchers conducting CRS-R assessments (at least 2 per center) were required to attend a 2-day intensive training session organized by the leading institution (Huashan Hospital). The training content included not only theoretical lectures on the CRS-R scale, but also focused on scoring guidelines for each subscale (auditory, visual, motor, oromotor/verbal, communication, and arousal), standardized operational demonstrations (using pre-recorded standard video cases), and hands-on practice with simulated patients.

Consistency Assessment

Upon completion of training, assessors were required to independently score a standardized video case library containing 10 cases with different levels of consciousness. This case library was carefully designed to cover the full clinical spectrum from VS/UWS to EMCS, and specifically included borderline cases that are easily confused (such as behavioral overlap between MCS- and VS/UWS) to ensure that assessors mastered key distinguishing points. Assessment qualifications were only granted when the Intraclass Correlation Coefficient (ICC; Model: Two-Way Mixed, Type: Absolute Agreement, Single Rater) between their scores and expert consensus scores exceeded 0.85 (see Table 1).

Table 1. Inter-rater reliability assessment results for CRS-R assessors

Center	Assessor ID	ICC Value	95% CI	Qualification Result
Huashan Hospital	Rater-HS-01	0.92	[0.86, 0.96]	Pass
Huashan Hospital	Rater-HS-02	0.94	[0.89, 0.97]	Pass
Zhengzhou First People's Hospital	Rater-ZZ-01	0.89	[0.82, 0.94]	Pass
Zhengzhou First People's Hospital	Rater-ZZ-02	0.91	[0.85, 0.95]	Pass
Shanghai Hebin Rehabilitation Hospital	Rater-HB-01	0.84	[0.76, 0.91]	Re-training required
Shanghai Hebin Rehabilitation Hospital	Rater-HB-02	0.90	[0.84, 0.95]	Pass

Note: ICC = Intraclass Correlation Coefficient. Qualification threshold: ICC > 0.85.

Mid-term Calibration

Every month during the study, all assessors were required to re-take an online video case assessment to prevent assessment standards from drifting over time. If the ICC fell below 0.85, their assessment qualifications were suspended until they passed re-training.

On-site Assessment Supervision

During the initial phase of the study, research coordinators from the leading institution conducted at least one on-site visit to each participating center to observe and ensure that CRS-R assessment procedures strictly followed the SOP.

Supplementary Material S2: Complete Statistical Results for All HRV Indices

Rest Period

Mann-Whitney U Test Results

Non-parametric test, independent of distributional assumptions, robust to outliers. See Table 2.

Wilcoxon Exact Test Results

Non-parametric exact test, prioritizing exact p-value calculations. See Table 3.

Table 2. Mann-Whitney U Test Results: Rest Period

Variable	UWS/V_S_Median	UWS/V_S_IQR	Conscious_Median	Conscious_IQR	U_Statistic	P_Value	Effect_r
HR	75.000	21.000	77.00	12.500	355.5	0.8376	0.0276
SDNN	21.605	20.443	25.77	19.615	329.5	0.5167	0.0874
SDANN	11.470	300.153	17.42	340.635	328.5	0.5022	0.0905
SDNN Index	16.840	12.793	19.45	18.120	356.5	0.8511	0.0253
RMSSD	16.705	7.225	17.35	14.095	336.0	0.5909	0.0725
NN50	2.000	6.250	2.00	15.500	340.0	0.6289	0.0652
pNN	0.505	1.613	0.48	4.470	346.0	0.7058	0.0509
Triangular Index	4.845	2.963	5.31	3.765	291.0	0.1918	0.1760
VLF	135.005	376.340	95.15	426.240	380.0	0.8444	0.0265
LF	50.520	81.700	62.58	94.290	318.0	0.3983	0.1139
HF	42.170	54.950	53.64	112.425	317.5	0.3936	0.1150
LF/HF	1.165	1.448	1.25	1.570	361.0	0.9117	0.0150
Sympathetic%	53.800	31.075	55.60	38.300	351.0	0.7783	0.0380
Parasympathetic%	46.200	31.075	44.40	35.800	375.0	0.9117	0.0150

Note: Values are presented as median (IQR). All comparisons were non-significant ($p > 0.05$).

Table 3. Wilcoxon Exact Test Results: Rest Period

Variable	UWS/V_S_Median	UWS/V_S_Q1	UWS/V_S_Q3	Conscious_Median	Conscious_Q1	Conscious_Q3	U_Statistic	P_Value	P_Type	Effect_r
HR	75	(68.5–89.5)		77	(71.5–84)		355.5	0.8376	Exact	0.0276
SDNN	21.6	(15.18–35.62)		25.77	(19.2–38.81)		329.5	0.5167	Exact	0.0874
SDANN	11.47	(0–300.15)		17.42	(3.98–344.62)		328.5	0.5022	Exact	0.0905
SDNN Index	16.84	(11.11–23.9)		19.45	(10.77–28.88)		356.5	0.8511	Exact	0.0253
RMSSD	16.7	(12.51–19.74)		17.35	(13.04–27.13)		336.0	0.5938	Exact	0.0719
NN50	2	(0–6.25)		2	(0–15.5)		340.0	0.6289	Exact	0.0652
pNN	0.5	(0–1.61)		0.48	(0–4.47)		346.0	0.7058	Exact	0.0509
Triangular Index	4.85	(3.23–6.19)		5.31	(4.22–7.98)		291.0	0.1918	Exact	0.1760
VLF	135	(45.49–421.83)		95.15	(47.71–473.95)		380.0	0.8457	Exact	0.0262
LF	50.52	(24.6–106.31)		62.58	(39.32–133.61)		318.0	0.4014	Exact	0.1131
HF	42.17	(25.93–80.88)		53.64	(26.99–139.42)		317.5	0.3936	Exact	0.1150
LF/HF	1.17	(0.56–2)		1.25	(0.45–2.02)		361.0	0.9117	Exact	0.0150
Sympathetic%	53.8	(35.57–66.65)		55.6	(31.1–69.4)		351.0	0.7801	Exact	0.0376
Parasympathetic%	46.2	(33.35–64.43)		44.4	(33.1–68.9)		375.0	0.9124	Exact	0.0148

Note: Values are presented as median (Q1–Q3). All comparisons were non-significant ($p > 0.05$).

Brunner-Munzel Test Results

Non-parametric test, particularly suitable for situations with unequal variances, providing probabilistic effect size estimates. See Table 4.

Table 4. Brunner-Munzel Test Results: Rest Period

Variable	UWS/V_S_NWS/V_S_Median_IQR	Conscious/Unconscious_Median_IQR	BM_Statistic	P_Value	P_Estimate
HR	32	75 (21)	23	77 (12.5)	0.214 50.07 0.8314 0.517
SDNN	32	21.6 (20.44)	23	25.77 (19.61)	0.661 52.80 0.5112 0.552
SDANN	32	11.47 (300.15)	23	17.42 (340.64)	0.675 48.21 0.5027 0.554
SDNN Index	32	16.84 (12.79)	23	19.45 (18.12)	0.193 47.51 0.8479 0.516
RMSSD	32	16.7 (7.22)	23	17.35 (14.09)	0.521 39.06 0.6050 0.543
NN50	32	2 (6.25)	23	2 (15.5)	0.458 34.04 0.6495 0.538
pNN	32	0.5 (1.61)	23	0.48 (4.47)	0.362 36.48 0.7195 0.530
Triangular Index	32	4.85 (2.96)	23	5.31 (3.77)	1.320 44.81 0.1934 0.605
VLF	32	135 (376.34)	23	95.15 (426.24)	-0.196 40.55 0.8455 0.484
LF	32	50.52 (81.7)	23	62.58 (94.29)	0.841 45.80 0.4047 0.568
HF	32	42.17 (54.95)	23	53.64 (112.43)	0.850 45.48 0.3996 0.569
LF/HF	32	1.17 (1.45)	23	1.25 (1.57)	0.117 46.65 0.9077 0.510
Sympathetic%	32	53.8 (31.08)	23	55.6 (38.3)	0.281 43.50 0.7802 0.523
Parasympathetic%	32	46.2 (31.08)	23	44.4 (35.8)	-0.117 46.76 0.9077 0.490

Note: Values are presented as median (IQR). All comparisons were non-significant ($p > 0.05$). BM_Statistic: Brunner-Munzel test statistic. P_Estimate: probability estimate.

Stimulation Period

Mann-Whitney U Test Results

Non-parametric test, independent of distributional assumptions, robust to outliers. See Table 5.

Table 5. Mann-Whitney U Test Results: Stimulation Period

Variable	UWS/V_S_Median	UWS/V_S_IQR	Conscious_Median	Conscious_IQR	U_Statistic	P_Value	Effect_r
HR	77.000	21.250	76.00	12.500	362.0	0.9252	0.0127
SDNN	20.545	19.450	23.18	26.635	303.0	0.2711	0.1484
SDANN	329.595	365.783	348.72	130.135	318.0	0.3983	0.1139
SDNN Index	12.345	13.460	12.79	22.690	344.0	0.6884	0.0541
RMSSD	15.995	9.888	17.35	14.820	342.0	0.6635	0.0587
NN50	0.500	9.500	4.00	17.000	276.5	0.1082	0.2166
pNN	0.105	2.163	0.94	4.790	280.0	0.1228	0.2081
Triangular Index	4.530	2.748	5.05	3.335	294.0	0.2097	0.1691
VLF	89.930	222.800	52.43	220.025	382.0	0.8178	0.0311
LF	42.775	112.005	32.20	174.445	341.0	0.6511	0.0610
HF	41.965	60.013	90.34	109.360	333.5	0.5618	0.0782
LF/HF	0.730	1.235	1.01	0.735	356.0	0.8444	0.0265
Sympathetic%	42.300	33.025	50.20	23.850	356.0	0.8444	0.0265
Parasympathetic%	57.700	33.025	49.80	23.850	380.0	0.8444	0.0265

Note: Values are presented as median (IQR). All comparisons were non-significant ($p > 0.05$).

Wilcoxon Exact Test Results

Non-parametric exact test, prioritizing exact p-value calculations. See Table 6.

Brunner-Munzel Test Results

Non-parametric test, particularly suitable for situations with unequal variances, providing probabilistic effect size estimates. See Table 7.

Table 6. Wilcoxon Exact Test Results: Stimulation Period

Variable	UWS/V_S_Median_Q1-Q3	Conscious_Median_Q1-Q3	QWQStatistic	P_Value	P_Type	Effect_r
HR	77 (66.75–88)	76 (70.5–83)	362.0	0.9252	Exact	0.0127
SDNN	20.55 (13.39–32.84)	23.18 (15.82–42.46)	303.0	0.2736	Exact	0.1476
SDANN	329.6 (22.84–388.62)	348.72 (280.53–410.67)	318.0	0.4014	Exact	0.1131
SDNN Index	12.34 (7.4–20.86)	12.79 (7.2–29.89)	344.0	0.6884	Exact	0.0541
RMSSD	15.99 (12.54–22.42)	17.35 (11.62–26.44)	342.0	0.6660	Exact	0.0582
NN50	0.5 (0–9.5)	4 (0.5–17.5)	276.5	0.1082	Exact	0.2166
pNN	0.1 (0–2.16)	0.94 (0.12–4.9)	280.0	0.1228	Exact	0.2081
Triangular Index	4.53 (3.1–5.84)	5.05 (3.76–7.1)	294.0	0.2097	Exact	0.1691
VLF	89.93 (37.97–260.77)	52.43 (29.77–249.8)	382.0	0.8193	Exact	0.0308
LF	42.78 (12.62–124.62)	32.2 (18.1–192.55)	341.0	0.6537	Exact	0.0605
HF	41.97 (24.63–84.65)	90.34 (24.88–134.24)	333.5	0.5618	Exact	0.0782
LF/HF	0.73 (0.42–1.65)	1.01 (0.45–1.18)	356.0	0.8444	Exact	0.0265
Sympathetic%	42.3 (29.23–62.25)	50.2 (30.3–54.15)	356.0	0.8457	Exact	0.0262
Parasympathetic%	57.7 (37.75–70.78)	49.8 (45.85–69.7)	380.0	0.8457	Exact	0.0262

Note: Values are presented as median (Q1–Q3). All comparisons were non-significant ($p > 0.05$).

Table 7. Brunner-Munzel Test Results: Stimulation Period

Variable	UWS/V_S_NWS/V_S_Median_IQR	Conscious_N	Unconscious_Median_IQR	BM_Statistic	df	P_Value	P_Estimate
HR	32 77 (21.25)	23	76 (12.5)	0.103	49.49	0.9180	0.508
SDNN	32 20.55 (19.45)	23	23.18 (26.63)	1.115	47.51	0.2704	0.588
SDANN	32 329.6 (365.78)	23	348.72 (130.14)	0.866	52.98	0.3901	0.568
SDNN Index	32 12.34 (13.46)	23	12.79 (22.69)	0.390	38.68	0.6984	0.533
RMSSD	32 15.99 (9.89)	23	17.35 (14.82)	0.416	35.16	0.6803	0.535
NN50	32 0.5 (9.5)	23	4 (17)	1.640	46.05	0.1079	0.624
pNN	32 0.1 (2.16)	23	0.94 (4.79)	1.577	48.22	0.1212	0.620
Triangular Index	32 4.53 (2.75)	23	5.05 (3.34)	1.287	49.02	0.2040	0.601
VLF	32 89.93 (222.8)	23	52.43 (220.02)	-0.230	46.94	0.8190	0.481
LF	32 42.78 (112)	23	32.2 (174.45)	0.457	49.51	0.6494	0.537
HF	32 41.97 (60.01)	23	90.34 (109.36)	0.563	39.47	0.5767	0.547
LF/HF	32 0.73 (1.23)	23	1.01 (0.73)	0.198	48.81	0.8436	0.516
Sympathetic%	32 42.3 (33.03)	23	50.2 (23.85)	0.198	48.81	0.8436	0.516
Parasympathetic%	32 57.7 (33.03)	23	49.8 (23.85)	-0.198	48.81	0.8436	0.484

Note: Values are presented as median (IQR). All comparisons were non-significant ($p > 0.05$). BM_Statistic: Brunner-Munzel test statistic. P_Estimate: probability estimate.

Supplementary Material S3: Test-Retest Reliability (ICC) Analysis of HRV Parameters in Rest and Stimulation States

Data Source

This analysis was based on 6 pDoC patients (3 UWS/Vs, 3 consciousness-preserved) who completed 4 repeated measurements within a single day (8 five-minute segments per patient, totaling 48 records).

Statistical Method

Intraclass Correlation Coefficient (ICC) was calculated using a Two-way mixed effects model for single-measure absolute agreement.

Evaluation Criteria

According to Koo & Li (2016) guidelines, ICC values are interpreted as follows: <0.50 indicates Poor reliability, 0.50–0.75 indicates Moderate reliability, 0.75–0.90 indicates Good reliability, and >0.90 indicates Excellent reliability.

Results Interpretation

Short-term indices reflecting rapid vagal regulation (RMSSD, pNN50, HF) demonstrated extremely high test-retest reliability (ICC > 0.95) in both rest and stimulation states, confirming their stability as core biomarkers for the PAP paradigm. In contrast, indices reflecting overall long-term variability (SDANN, VLF, SDNN) showed lower reliability, which may be attributed to their greater sensitivity to slow physiological fluctuations and environmental factors, making them unsuitable as primary observation indices for short-duration (5-minute) auditory perturbation experiments. Detailed ICC values are presented in Table 8.

Table 8. Test-retest reliability (ICC) of HRV parameters in rest and stimulation states

HRV Parameter	Resting State ICC (95% CI)	Stimulation State ICC (95% CI)	Reliability Rating
<i>Time Domain Indices</i>			
Mean HR	0.87 (0.64–0.98)	0.87 (0.62–0.98)	Good
SDNN	0.76 (0.41–0.96)	0.38 (–0.01–0.84)	Poor
SDANN	0.15 (–0.15–0.71)	0.26 (–0.09–0.78)	Poor
RMSSD	0.98 (0.92–1.00)	0.98 (0.92–1.00)	Excellent
NN50	0.97 (0.91–1.00)	0.87 (0.63–0.98)	Excellent
pNN50	0.97 (0.91–1.00)	0.98 (0.93–1.00)	Excellent
Triangular Index	0.85 (0.58–0.97)	0.80 (0.48–0.96)	Good
<i>Frequency Domain Indices</i>			
VLF	0.44 (0.04–0.87)	0.24 (–0.10–0.77)	Poor
LF	0.58 (0.17–0.91)	0.61 (0.20–0.92)	Moderate
HF	0.96 (0.87–0.99)	0.96 (0.88–0.99)	Excellent
LF/HF Ratio	0.47 (0.06–0.87)	0.70 (0.32–0.94)	Moderate

Note: ICC = Intraclass Correlation Coefficient. Reliability rating based on Koo & Li (2016): Poor <0.50, Moderate 0.50–0.75, Good 0.75–0.90, Excellent >0.90.

Supplementary Material S4: 5-Fold Cross-Validation Feature Selection Statistical Test Report

This supplementary material presents detailed results from the nested feature selection process used in the classification model construction. To prevent data leakage and ensure biomarker robustness, feature selection was independently performed within each fold of the 5-fold cross-validation, with cross-fold consistency used as a key criterion for retaining candidate biomarkers in the final model.

Analysis Specifications

- Analysis Date: 2025-12-24
- Random Seed: 999
- Number of Cross-Validation Folds: 5

Relative Difference Data

Fold 1 Training Set Test Results

Training set samples: UWS/VIS group 26 cases, Consciousness-preserved group 18 cases, Total 44 cases.

Three statistical tests were performed:

- **Mann-Whitney U Test:** Non-parametric test, independent of distributional assumptions, robust to outliers.
- **Wilcoxon Exact Test:** Non-parametric exact test, prioritizing exact p-value calculations.
- **Brunner-Munzel Test:** Non-parametric test, particularly suitable for situations with unequal variances, providing probabilistic effect size estimates.

Fold 2 Training Set Test Results

Training set samples: UWS/VIS group 26 cases, Consciousness-preserved group 18 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set. As an example, Table 9 shows the Brunner-Munzel test results for relative difference data, demonstrating significant differences in NN50 and pNN indices.

Table 9. Brunner-Munzel Test Results for Relative Difference Data: Fold 2 Training Set

Variable	VS_N	VS_Median_IQR	NonVS_N	NonVS_Median_IQR	BM_Statistic	df	P_Value
HR	26	-0.01 (0.02)	18	-0.01 (0.01)	-0.835	42.00	0.4085
SDNN	26	-0.12 (0.27)	18	0.02 (0.45)	1.577	42.00	0.1266
SDANN	26	2.01 (3.19)	18	2.22 (3.73)	0.000	42.00	1.0000
SDNN Index	26	-0.27 (0.45)	18	-0.18 (0.48)	0.643	42.00	0.5239
RMSSD	26	-0.03 (0.16)	18	0.03 (0.22)	1.522	42.00	0.1364
NN50	26	-0.50 (2.75)	18	1.00 (5.50)	3.142	42.00	0.0033**
pNN	26	-0.12 (0.76)	18	0.29 (1.42)	3.327	42.00	0.0019**
Triangular Index	26	-0.09 (0.27)	18	0.03 (0.49)	0.832	42.00	0.4127
VLF	26	-0.17 (1.10)	18	-0.16 (1.96)	0.069	42.00	0.9454
LF	26	-0.20 (0.79)	18	-0.34 (0.86)	-0.830	42.00	0.4117
HF	26	-0.05 (0.28)	18	0.06 (0.73)	0.381	42.00	0.7062
LF/HF	26	-0.06 (0.31)	18	-0.22 (0.54)	-0.640	42.00	0.5271
Sympathetic%	26	-3.80 (14.23)	18	-13.45 (26.62)	-0.729	42.00	0.4719
Parasympathetic%	26	3.80 (14.23)	18	11.60 (25.73)	0.644	42.00	0.5248

Note: Values are presented as median (IQR). ** $p < 0.01$. BM_Statistic: Brunner-Munzel test statistic. df: degrees of freedom.

Fold 3 Training Set Test Results

Training set samples: UWS/VIS group 26 cases, Consciousness-preserved group 18 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Fold 4 Training Set Test Results

Training set samples: UWS/VIS group 25 cases, Consciousness-preserved group 19 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Fold 5 Training Set Test Results

Training set samples: UWS/VIS group 25 cases, Consciousness-preserved group 19 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Feature Consistency Summary for Relative Difference Data

This section displays significance markers for each feature across all 5 folds (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). Candidate biomarkers were retained based on agreement across test methods and cross-fold consistency (e.g., significance in at least 4 out of 5 folds), ensuring generalization stability while reducing chance findings tied to specific splits.

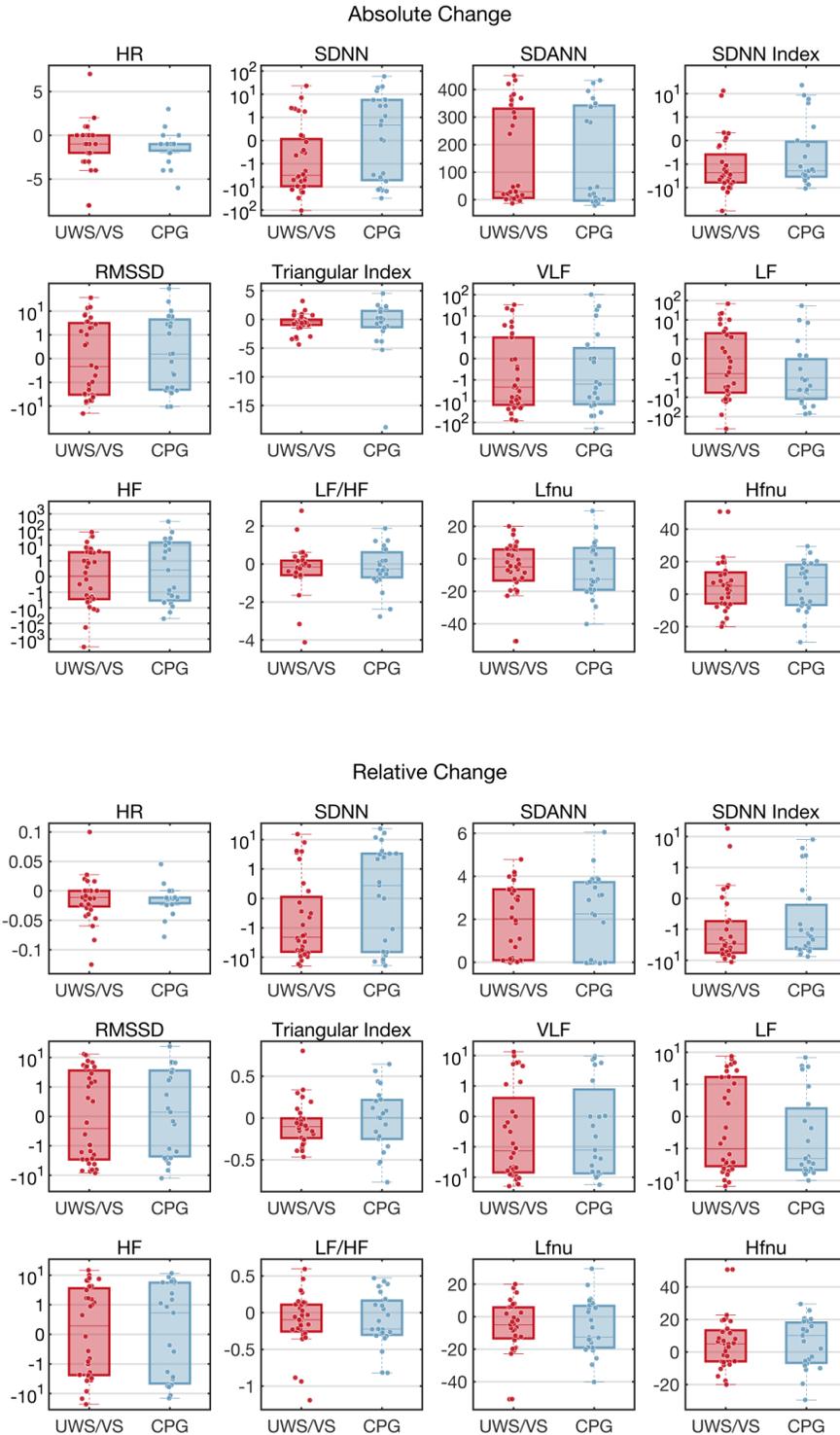


Figure 1. This supplementary figure presents the full spectrum of HRV parameters, complementing the specific biomarkers identified in Figure 4. A comparison between UWS/Vs patients (red, n=32) and CPG patients (blue, n=23) shows broadly overlapping distributions with no systematic significant differences, confirming comparable physiological baselines between groups. Dynamic responses for indices not included in the final model (e.g., SDNN, SDANN, VLF) are also shown in this figure. Unlike the significant vagal modulation observed for NN50 and pNN50 in the main text, these indices generally failed to reach the significance threshold ($p < 0.05$) across the Triple Robustness tests. This highlights the specificity of rapid vagal regulation in distinguishing states of consciousness.

Table 10. Feature consistency summary for relative difference data: Mann-Whitney U Test

Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Consistency
HR	–	–	–	–	–	0/5
SDNN	–	–	–	–	–	0/5
SDANN	–	–	–	–	–	0/5
SDNN Index	–	–	–	–	–	0/5
RMSSD	–	–	–	–	–	0/5
NN50	*	**	–	*	*	4/5
pNN	*	**	–	*	*	4/5
Triangular Index	–	–	–	–	–	0/5
VLF	–	–	–	–	–	0/5
LF	–	–	–	–	–	0/5
HF	–	–	–	–	–	0/5
LF/HF	–	–	–	–	–	0/5
Sympathetic%	–	–	–	–	–	0/5
Parasympathetic%	–	–	–	–	–	0/5

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, – indicates non-significant. Consistency shows the number of folds in which the feature reached statistical significance.

Table 11. Feature consistency summary for relative difference data: Wilcoxon Exact Test

Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Consistency
HR	–	–	–	–	–	0/5
SDNN	–	–	–	–	–	0/5
SDANN	–	–	–	–	–	0/5
SDNN Index	–	–	–	–	–	0/5
RMSSD	–	–	–	–	–	0/5
NN50	*	**	–	*	*	4/5
pNN	*	**	–	*	*	4/5
Triangular Index	–	–	–	–	–	0/5
VLF	–	–	–	–	–	0/5
LF	–	–	–	–	–	0/5
HF	–	–	–	–	–	0/5
LF/HF	–	–	–	–	–	0/5
Sympathetic%	–	–	–	–	–	0/5
Parasympathetic%	–	–	–	–	–	0/5

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, – indicates non-significant. Consistency shows the number of folds in which the feature reached statistical significance.

Table 12. Feature consistency summary for relative difference data: Brunner-Munzel Test

Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Consistency
HR	–	–	–	–	–	0/5
SDNN	–	–	–	–	–	0/5
SDANN	–	–	–	–	–	0/5
SDNN Index	–	–	–	–	–	0/5
RMSSD	–	–	–	–	–	0/5
NN50	*	**	–	*	*	4/5
pNN	*	**	–	*	*	4/5
Triangular Index	–	–	–	–	–	0/5
VLF	–	–	–	–	–	0/5
LF	–	–	–	–	–	0/5
HF	–	–	–	–	–	0/5
LF/HF	–	–	–	–	–	0/5
Sympathetic%	–	–	–	–	–	0/5
Parasympathetic%	–	–	–	–	–	0/5

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, – indicates non-significant. Consistency shows the number of folds in which the feature reached statistical significance.

Absolute Difference Data

Fold 1 Training Set Test Results

Training set samples: UWS/VIS group 26 cases, Consciousness-preserved group 18 cases, Total 44 cases.

Three statistical tests were performed:

- **Mann-Whitney U Test:** Non-parametric test, independent of distributional assumptions, robust to outliers.
- **Wilcoxon Exact Test:** Non-parametric exact test, prioritizing exact p-value calculations.
- **Brunner-Munzel Test:** Non-parametric test, particularly suitable for situations with unequal variances, providing probabilistic effect size estimates.

Fold 2 Training Set Test Results

Training set samples: UWS/VIS group 26 cases, Consciousness-preserved group 18 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Fold 3 Training Set Test Results

Training set samples: UWS/VIS group 26 cases, Consciousness-preserved group 18 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Fold 4 Training Set Test Results

Training set samples: UWS/VIS group 25 cases, Consciousness-preserved group 19 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Fold 5 Training Set Test Results

Training set samples: UWS/VIS group 25 cases, Consciousness-preserved group 19 cases, Total 44 cases.

The same three statistical tests (Mann-Whitney U, Wilcoxon Exact, Brunner-Munzel) were applied to this fold's training set.

Feature Consistency Summary for Absolute Difference Data

This section displays significance markers for each feature across all 5 folds (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). Candidate biomarkers were retained based on agreement across test methods and cross-fold consistency (e.g., significance in at least 4 out of 5 folds), ensuring generalization stability while reducing chance findings tied to specific splits.

Table 13. Feature consistency summary for absolute difference data: Mann-Whitney U Test

Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Consistency
HR	–	–	–	–	–	0/5
SDNN	*	–	–	*	–	2/5
SDANN	–	–	–	–	–	0/5
SDNN Index	–	–	–	–	–	0/5
RMSSD	–	–	–	–	–	0/5
NN50	*	**	–	*	*	4/5
pNN	*	**	–	*	*	4/5
Triangular Index	–	–	–	–	–	0/5
VLFF	–	–	–	–	–	0/5
LF	–	–	–	–	–	0/5
HF	–	–	–	–	–	0/5
LF/HF	–	–	–	–	–	0/5
Sympathetic%	–	–	–	–	–	0/5
Parasympathetic%	–	–	–	–	–	0/5

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, – indicates non-significant. Consistency shows the number of folds in which the feature reached statistical significance.

Table 14. Feature consistency summary for absolute difference data: Wilcoxon Exact Test

Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Consistency
HR	–	–	–	–	–	0/5
SDNN	*	–	–	*	–	2/5
SDANN	–	–	–	–	–	0/5
SDNN Index	–	–	–	–	–	0/5
RMSSD	–	–	–	–	–	0/5
NN50	*	**	–	*	*	4/5
pNN	*	**	–	*	*	4/5
Triangular Index	–	–	–	–	–	0/5
VLF	–	–	–	–	–	0/5
LF	–	–	–	–	–	0/5
HF	–	–	–	–	–	0/5
LF/HF	–	–	–	–	–	0/5
Sympathetic%	–	–	–	–	–	0/5
Parasympathetic%	–	–	–	–	–	0/5

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, – indicates non-significant. Consistency shows the number of folds in which the feature reached statistical significance.

Table 15. Feature consistency summary for absolute difference data: Brunner-Munzel Test

Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Consistency
HR	–	–	–	–	–	0/5
SDNN	–	–	–	*	–	1/5
SDANN	–	–	–	–	–	0/5
SDNN Index	–	–	–	–	–	0/5
RMSSD	–	–	–	–	–	0/5
NN50	*	**	–	*	*	4/5
pNN	*	**	–	*	*	4/5
Triangular Index	–	–	–	–	–	0/5
VLF	–	–	–	–	–	0/5
LF	–	–	–	–	–	0/5
HF	–	–	–	–	–	0/5
LF/HF	–	–	–	–	–	0/5
Sympathetic%	–	–	–	–	–	0/5
Parasympathetic%	–	–	–	–	–	0/5

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, – indicates non-significant. Consistency shows the number of folds in which the feature reached statistical significance.

Summary of Feature Selection Results

As shown in the feature consistency tables above, only NN50 and pNN (both absolute and relative differences) consistently demonstrated statistical significance across 4 out of 5 folds in all three statistical tests. This high consistency (4/5 folds) indicates robust discriminative power that generalizes beyond specific data splits. The fact that these features were selected across multiple folds in all three non-parametric tests provides strong evidence for their validity as biomarkers for distinguishing UWS/VS from consciousness-preserved patients in the PAP framework.

Methodological Notes

The nested feature selection strategy implemented in this analysis represents a rigorous approach to biomarker validation:

1. **Training-Only Feature Selection:** In each cross-validation iteration, feature selection was performed exclusively on the training set, with the test set completely excluded from this process. This prevents data leakage and ensures that classification performance estimates are unbiased.
2. **Triple Robustness Strategy:** For each fold, the same "Triple Robustness Strategy" established in the main analysis was applied, requiring features to pass all three non-parametric tests (Mann-Whitney U, Wilcoxon Exact, and Brunner-Munzel) simultaneously.
3. **Consistency-Threshold Feature Selection:** Features were prioritized for inclusion only if they passed all three non-parametric tests in at least 4 out of 5 training folds. This reduces the risk of chance findings while allowing for limited fold-to-fold variability.
4. **Generalization Stability:** High cross-fold consistency indicates that the identified biomarkers (Δ NN50 and Δ pNN) have discriminative power that generalizes beyond a specific data split.