Received: 15 October 2025 | Revised: 2 December 2025 | Accepted: 15 December 2025 | Published online: 29 December 2025

RESEARCH ARTICLE

Smart Wearable Technology
2026, Vol. 2, A3, 20 Pages
DOI: 10.47852/bonviewSWT52027605

BON VIEW PUBLISHING

Assessing Pulse Rate Variability from a Wrist-Worn PPG Device Against ECG-Derived Heart Rate Variability in Ambulatory Settings

Jie Huang[1,†], Yuhong Wu[2,†], Fang Li[2], Dan Zhang[2,*] and Shuping Tan[1,*]

[1] *HuiLongGuan Clinical Medical School, Peking University, China and Beijing HuiLongGuan Hospital, China*

[2] *Department of Psychological and Cognitive Sciences, Tsinghua University, China*

Abstract: This study systematically evaluates the consistency and applicability limits of photoplethysmography (PPG)-derived pulse rate variability (PRV) versus electrocardiogram (ECG)-derived heart rate variability (HRV) in real-world settings. It integrates three methodological dimensions: 24-hour multi-context monitoring, dual-level consistency analysis (inter- and intra-individual), and controlled motion intensity via a 27-level acceleration gradient. Data from 14 healthy participants were collected using synchronized wrist-worn PPG, portable ECG, and triaxial accelerometry. Standardized preprocessing and motion artifact suppression based on acceleration thresholds enabled the extraction of time-domain, frequency-domain, and nonlinear HRV and PRV metrics. Consistency was assessed using Pearson correlation and root mean square error, with false discovery rate-corrected significance testing. Results show strong PPG-ECG agreement during sleep ($r > 0.9$ for HR, MeanNN, Prc80NN) but marked degradation under high motion. Notably, Prc20NN demonstrated exceptional robustness across contexts, retaining significant correlation even during active phases. Stringent motion filtering substantially improved correlations. These findings delineate metric-specific validity boundaries for wearable PRV, distinguishing motion-induced errors from inherent physiological discrepancies, and offer evidence-based recommendations for deploying PRV in context-appropriate applications such as sleep monitoring, passive health tracking, and longitudinal stress assessment.

Keywords: HRV, PRV, PPG, ECG, wearable sensors, ambulatory monitoring, motion artifacts

1. Introduction

Heart rate variability (HRV) has emerged as a critical biomarker for assessing autonomic nervous system function and predicting adverse health outcomes across multiple clinical domains [1, 2]. Traditionally measured through an electrocardiogram (ECG), HRV quantifies the temporal fluctuations between consecutive heartbeats, offering insights into the dynamic balance between sympathetic and parasympathetic nervous system activity. Over the past three decades, mounting evidence has established HRV as a robust predictor of cardiovascular mortality, a physiological marker of mental health disorders, and a measurable target for therapeutic interventions.

However, the clinical utility of ECG-based HRV monitoring faces significant practical constraints in real-world applications. Standard ECG systems require specialized clinical environments, while ambulatory Holter monitors often suffer from user discomfort, limited battery life, and poor compliance during extended monitoring periods. These limitations have catalyzed growing interest in photoplethysmography (PPG), a noninvasive optical sensing technology that has been rapidly integrated into consumer-grade wearable devices such as smartwatches and fitness trackers. PPG offers compelling advantages including unobtrusiveness, continuous monitoring capability, and seamless integration into daily life activities [3].

Despite widespread adoption, the validity of PPG-derived pulse rate variability (PRV) as a surrogate for ECG-derived HRV remains incompletely characterized, particularly in naturalistic, ambulatory settings. While controlled laboratory studies have demonstrated promising agreement under resting conditions, the performance of PPG deteriorates substantially during physical activity due to motion artifacts, skin tone variations, and physiological changes in peripheral vasculature [4–6]. Furthermore, existing comparative studies exhibit considerable heterogeneity in measurement protocols, device types, and analytical approaches, resulting in conflicting conclusions regarding which HRV metrics can be reliably estimated from PPG signals across different contexts [4].

Critical gaps persist in our understanding of context-dependent boundaries for PPG-ECG concordance. First, most validation studies focus on aggregate-level correlations rather than examining intra-individual consistency patterns that are essential for personalized health monitoring. Second, the differential impact of motion intensity on specific HRV metrics—spanning time-domain, frequency-domain, and nonlinear dimensions—has not been systematically quantified through controlled threshold analyses. Third,

***Corresponding authors:** Dan Zhang, Department of Psychological and Cognitive Sciences, Tsinghua University, China. Email:dzhang@tsinghua.edu.cn and Shuping Tan, HuiLongGuan Clinical Medical School, Peking University, China and Beijing HuiLongGuan Hospital, China. Email: shupingtan@126.com
[†]**Co-first authors.**

no consensus exists regarding which HRV features demonstrate cross-context robustness suitable for implementation in real-world wearable applications.

2. Literature Review

2.1. Clinical importance of HRV

HRV shows how the sympathetic and parasympathetic branches of the autonomic nervous system work together in a complex way. HRV is a noninvasive marker of autonomic regulation that provides us with a unique look at how the cardiovascular system adapts to changes in both the body and the mind. The fundamental guidelines set forth by Malik et al. [7] established the basis for standardized HRV metrics and their physiological interpretation, thereby facilitating their clinical implementation.

Over the past three decades, HRV has emerged as a robust biomarker for risk stratification in cardiovascular disease. In a seminal study of the Framingham cohort, Szurhaj et al. [8] established that diminished HRV, specifically lower Standard Deviation of Normal-to-Normal (SDNN) intervals and Root Mean Square of Successive Differences (RMSSD) values, significantly forecasted sudden cardiac death. In a similar study, Herman et al. [9] discovered that diminished HRV in chronic heart failure patients correlated with increased all-cause and cardiac mortality, highlighting its significance as a longitudinal prognostic marker. Liu et al. [10] expanded this comprehension by framing HRV as an indicator of neurovisceral integration, associating autonomic imbalance with increased cardiovascular risk.

Beyond cardiology, HRV also plays an important role in mental health and neuropsychological conditions. Goffi et al. [11] reported consistently lower HRV in patients with major depressive disorder, supporting its relevance as a physiological index of affective dysregulation and increased mortality. Al Jowf et al. [12] further established that HRV serves as a physiological indicator of post-traumatic stress disorder (PTSD), aiding in both diagnosis and therapeutic monitoring. These results have prompted the incorporation of HRV into psych cardiology and psychoneuroimmunology paradigms.

HRV is also responsive to therapeutic modulation. Lehrer et al. [13] demonstrated that HRV biofeedback could significantly improve autonomic flexibility and diminish symptom severity in individuals with hypertension and anxiety. This is in line with what Zucker et al. [14] found when they added HRV tracking to a multisensory virtual reality protocol for PTSD. In their research, wearable sensors recorded variations in HRV throughout treatment sessions, offering objective indicators of psychological response and treatment effectiveness.

Recent clinical studies have identified HRV as a downstream indicator of systemic inflammation and metabolic health. Wang et al. [15] showed that giving probiotics to people on hemodialysis not only helped them sleep better and lowered the levels of inflammatory toxins, but it also raised HRV indices like low-frequency (LF) and low-frequency/high-frequency (LF/HF) ratio significantly, which supports its use in research on the gut–brain–immune axis. In a similar vein, Loh et al. [16] demonstrated that vitamin D deficiency in hypertensive patients was associated with reduced HRV and an increased probability of undergoing coronary revascularization, highlighting the relationship between micronutrient status, autonomic regulation, and cardiovascular intervention.

These findings collectively underscore HRV's multifaceted clinical significance spanning cardiology, psychiatry, immunology, and behavioral medicine. However, translating this clinical potential into real-world applications requires measurement technologies that are accurate, unobtrusive, and scalable—demands increasingly met by consumer-grade wearable devices, yet with validation challenges that remain incompletely characterized.

2.2. HRV monitoring: from lab to real-world settings

The measurement of HRV has experienced a substantial paradigm shift over the past two decades. Early research restricted HRV assessment to controlled laboratory settings, but technological advances and growing recognition of HRV as a dynamic biomarker have driven progressive emphasis on real-world ambulatory contexts.

Pioneering work by Shishavan et al. [17] demonstrated the feasibility of ambulatory HRV monitoring for assessing occupational stress during normal working hours, facilitating HRV's transition from clinical observation to behavioral health monitoring. Kim et al. [18] further validated HRV-derived indices as reliable stress reactivity markers in semi-naturalistic contexts through comprehensive meta-analysis.

As wearable technology matured, researchers increasingly investigated HRV in free-living conditions. Taskasaplidis et al. [19] demonstrated continuous stress monitoring using wrist-worn sensors in academic settings, while Berntson et al. [20] provided a theoretical framework for HRV as a psychophysiological marker in ambulatory research, strengthening methodological validity for out-of-lab measurements. Recent emergence of validated commercial wearables has enabled HRV monitoring in domestic and sleep environments, with studies demonstrating that consumer-grade devices can provide reliable nighttime autonomic profiles comparable to clinical-grade recordings [21, 22]. This evolution from isolated clinical observations to continuous life-integrated monitoring has intensified demands for accurate, scalable measurement approaches—a transition that positions PPG-based wearables as promising solutions, though with context-dependent accuracy boundaries requiring systematic characterization.

2.3. ECG: gold standard for HRV estimation

ECG remains the gold standard for HRV estimation due to its high temporal resolution and signal fidelity in detecting R-peaks within the QRS complex. HRV is calculated as the temporal variation between consecutive R-R intervals (normal-to-normal intervals), providing insights into autonomic nervous system dynamics [7, 23].

ECG enables comprehensive HRV analysis across time-domain metrics (SDNN, RMSSD), frequency-domain parameters (LF, HF, LF/HF ratio reflecting sympathetic-parasympathetic balance) [20], and nonlinear indices. Clinical guidelines from the European Society of Cardiology and North American Society of Pacing and Electrophysiology endorse ECG-based HRV as diagnostic and prognostic tools for cardiovascular and autonomic disorders [7]. HRV has proven utility in risk stratification for heart failure, sudden cardiac death, post-myocardial infarction monitoring, and mental health assessment [24].

However, ECG's practical applicability in continuous real-world monitoring faces significant constraints. Standard 12-lead systems are clinic-bound, while ambulatory Holter monitors and patch-based wearables suffer from user discomfort, movement restriction, and limited battery life [25]. Recent technological innovations have substantially improved ECG wearability: flexible substrate electrodes and dry capacitive sensing eliminate conductive gel requirements [26, 27], and miniaturized single-lead

configurations achieve consumer-grade form factors while maintaining diagnostic quality [28]. Wireless transmission platforms now enable real-time arrhythmia detection at scale. Despite these advances, fundamental limitations persist: all ECG modalities require direct skin contact, electrical sensing remains vulnerable to electrostatic interference [26], and device costs remain substantially higher than optical PPG sensors.

2.4. PPG: a practical wearable alternative

PPG offers a practical alternative to ECG for HRV estimation in wearable applications. Originally developed in the 1930s as a non-invasive method for measuring peripheral blood volume changes [29], PPG has evolved from clinical pulse oximetry to a core biosensing modality in modern wearables. The fundamental principle involves light emission and detection, with returning signals varying with arterial pulsation. Pulse-to-pulse intervals extracted from PPG waveforms enable estimation of HRV metrics [30].

Validation studies demonstrate promising PPG-ECG concordance under controlled conditions. Lu et al. [31] showed the feasibility of accurate time-domain HRV estimation using wristband PPG during rest, while Rienzo et al. [21] validated commercial wrist-worn devices for nocturnal HRV with significant ECG concordance across multiple metrics. Consumer wearables including Empatica E4, Fitbit, and Apple Watch have been evaluated across various activity states (sleep, meditation, cognitive stress), achieving Pearson correlations ≥ 0.85 for RMSSD and SDNN compared to HRV [22, 32–34].

Despite these promising results, PPG faces inherent limitations. Motion artifacts, skin tone variations, ambient light interference, and vasoconstriction degrade signal quality, particularly during movement or physiological stress [22]. Frequency-domain HRV estimation accuracy remains contested due to waveform distortion and pulse transit time (PTT) variations [35]. However, PPG's compelling advantages—compactness, low cost, seamless smartwatch/ring integration, and superior long-term wearability [21]—position it as the pragmatic choice for continuous health monitoring, sleep tracking, and stress analysis in free-living populations, provided its context-dependent accuracy boundaries are systematically characterized.

2.5. Gaps in comparative studies between PPG and ECG

Despite growing literature comparing PRV to HRV estimation, three critical deficiencies persist, limiting clinical translation:

1) Context-dependent validity remains poorly quantified. While multiple studies demonstrate strong PPG-ECG agreement for time-domain metrics (RMSSD, SDNN) during rest or sleep ($r > 0.80$) [34, 36], performance deteriorates substantially during movement [36, 37]. However, no systematic threshold analysis quantifies the motion intensity boundaries separating reliable from invalid estimation. Ishaque et al. [35] highlighted that PPG performs well for mean heart rate but fails to track high-frequency oscillations under dynamic states like exercise, yet the precise activity thresholds defining applicability remain unestablished. This gap prevents practitioners from determining when to trust PRV versus when to withhold clinical interpretation.
2) Metric-specific validation hierarchies are absent. Most comparative studies report aggregate performance or focus on subset metrics, obscuring critical distinctions between HRV features [4, 38]. Time-domain indices may exhibit motion-remediable errors amenable to algorithmic correction, while frequency-domain parameters could suffer fundamental physiological discordance resistant to artifact removal. Mejía-Mejía et al. [37] found PRV systematically overestimates parasympathetic tone due to PTT fluctuations, particularly during hemodynamic changes—an issue unresolvable by noise suppression. Without empirically categorizing which metrics demonstrate correctable versus fundamental errors, algorithm developers risk sophisticated solutions to unsolvable measurement problems.
3) Validation frameworks lack standardization and mechanistic insight. Existing studies differ widely in device types, sampling rates, preprocessing pipelines, and metric selection, undermining reproducibility and meta-analytic synthesis [38]. More critically, most studies emphasize inter-individual correlation (group-level ranking consistency) but omit intra-individual validation (within-person tracking reliability)—the essential criterion for personalized longitudinal monitoring applications. Furthermore, studies report statistical correlation without error decomposition to explain disagreement mechanisms.

We systematically evaluate the consistency between PRV and HRV across 15 established metrics, examining both inter-individual and intra-individual agreement patterns. This study addresses these methodological gaps through three integrated innovations:

1) Comprehensive 24-hour multi-context protocol: Synchronized PPG, ECG, and triaxial accelerometry across sleep, controlled rest, and unconstrained free-living activity in 14 healthy participants, capturing the full spectrum of autonomic states and motion intensities encountered in real-world wearable use.
2) Dual-level consistency framework: Systematic evaluation at both inter-individual (group-level ranking preservation, critical for population screening) and intra-individual (within-person temporal tracking, essential for personalized health monitoring) levels across 15 established HRV metrics spanning time-domain, frequency-domain, and nonlinear dimensions.
3) Systematic motion threshold control: Novel gradient analysis employing 27 thresholds (∞ to 0.1 m/s^2) to quantify the dose-response relationship between motion intensity and measurement concordance, enabling empirical identification of context-specific "reliability boundaries" for each metric.

The findings presented herein establish empirical boundaries for PRV estimation, identify metrics with cross-context reliability, and provide application recommendations for integrating wearable-derived autonomic biomarkers into clinical decision-making and consumer health applications. By distinguishing motion-remediable errors from fundamental physiological discordance, this work provides the validation framework necessary for evaluating whether advanced signal processing algorithms—including emerging deep learning approaches—deliver genuine clinical value or merely statistical improvements detached from physiological validity. By revealing the dual-pathway mechanisms underlying PPG-ECG discordance—encompassing both motion artifacts and physiological vascular reactivity—this work advances toward more accurate, scalable, and ecologically valid approaches for continuous cardiovascular health monitoring in free-living populations.

3. Materials and Methods

3.1. Participants and experimental design

This study included 14 healthy adult participants (6 males, 8 females; age range: 21–28 years) with compliant vasculature,

homogeneous skin tones, and no history of cardiovascular disease. All participants were screened from Tsinghua University through structured questionnaires and pre-experimental interviews to confirm they had no history of known cardiovascular diseases and had not taken any medications that could affect autonomic nervous system activity (such as beta-blockers or sedatives) within two weeks prior to the experiment.

This study protocol was approved by the Ethics Committee of Tsinghua University (Approval Number: THU-04-2025-1084), and all participants signed written informed consent forms. To comprehensively assess the consistency of PPG and ECG across different scenarios, the experimental design included the following three tasks:

1) Resting state: Participants sat quietly for 10 minutes in a laboratory with soft lighting and controlled temperature, maintaining limb stillness and avoiding verbal communication, to obtain baseline resting-state PPG, ECG, and three-dimensional acceleration signals;
2) Sleep stages: Participants slept naturally in a home environment, self-reporting their bedtime and wake-up time;
3) Active period: Defined as the daily active period excluding sleep stages during the 24-hour day.

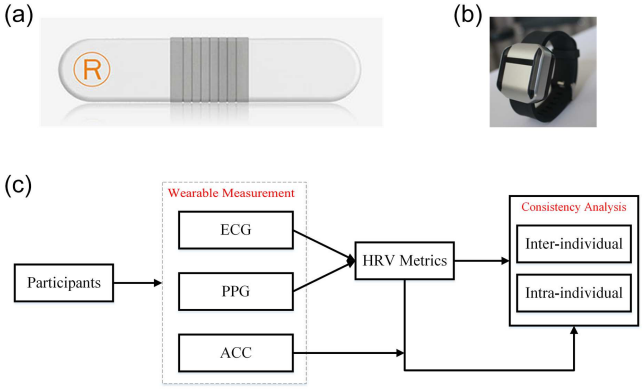
3.2. Data acquisition

A multimodal physiological signal co-acquisition strategy was employed, with hardware time stamps used for calibration to ensure signal temporal alignment. Each participant was instructed to wear a custom-designed wristband and a portable ECG monitor for a one-day multimodal physiological measurement, as shown in Figure 1(a). The specific equipment used is as follows:

Wristband (Psychorus, Huixin, Beijing, China): This custom-designed device, validated in previous daily-context studies by Shui et al. [39], He et al. [40], and Shui et al. [41], continuously collected triaxial acceleration (ACC) signals at 50 Hz (range: ± 2g, sensitivity: 0.000061 g/digit) and PPG signals at 500 Hz. The wristband was worn on the left wrist, 2−3 cm proximal to the radial styloid process, with sufficient tightness to maintain skin contact without obstructing blood flow.

Portable dynamic ECG monitor (ER1, Lepu Medical, Beijing, China): A single-lead dynamic ECG device (250 Hz sampling rate) with electrodes attached to the left anterior chest, suitable for long-term ambulatory monitoring [42, 43].

Figure 1
(a) Portable dynamic electrocardiogram monitor; (b) wristband device; (c) flowchart of data analysis



3.3. Feature extraction

All raw signals underwent feature extraction after standardized preprocessing. The raw triaxial acceleration data is converted to m/s^2 in the International System of Units, the data for each axis are zero-corrected, and the gravity component is estimated by a low-pass filter and subtracted from the signal to obtain the linear acceleration A_x, A_y, A_z of three axes.

The preprocessing of ECG signals aimed at removing noise artifacts and extracting key heartbeat information to guarantee the accuracy of subsequent HRV metrics calculation. The complete ECG processing pipeline consisted of the following steps:

1) Missing data handling: For signal segments containing missing values (NaN), we first removed these values while maintaining index mapping to the original timeline for subsequent peak location restoration.
2) Signal cleaning: Raw ECG signals were processed using NeuroKit2's (version 0.2.0) [44] "ecg_clean()" function, which applied a 0.5–40 Hz band-pass filter to simultaneously remove low-frequency baseline drift, high-frequency EMG interference, and power-line noise. The filtered signal was then demeaned to eliminate the DC offset.
3) R-peak detection: The cleaned signal underwent R-peak extraction using NeuroKit2's "ecg_peaks()" function with the following configuration: ① algorithm: integrated multi-method detector combining Pan-Tompkins, Hamilton, Christov, and other classical algorithms through an ensemble voting mechanism; ② artifact correction: enabled to automatically identify and correct nonphysiological beats; and ③ output: R-peak indices in the cleaned signal timeline.
4) Peak location restoration: Detected R-peak positions were mapped back to the original signal timeline using preserved index information, accounting for any removed NaN values. This comprehensive preprocessing pipeline ensured the continuity and physiological validity of the resulting RR interval time series.

PPG signals in their raw state were often subjected to various interferences such as motion artifacts and environmental noise, necessitating systematic preprocessing to guarantee pulse peak detection accuracy. The complete PPG processing pipeline consisted of the following steps:

1) Missing data imputation: Signal segments containing missing values were first processed using linear interpolation to ensure signal continuity.
2) Band-pass filtering: The BioSPPy (version 2.2.0) [45] signal processing module was employed to apply a 4th-order Butterworth band-pass filter with cutoff frequencies of 0.5–4.0 Hz. This frequency range was specifically selected to encompass typical heart rate ranges (30–200 beats per minute) while excluding motion artifacts and high-frequency measurement noise.
3) Wavelet denoising: The filtered signal underwent additional noise reduction using hybrid wavelet denoising with the following specifications: wavelet basis: db4, decomposition level: 9.
4) Peak detection: Pulse peaks were identified using the Elgendi2013 algorithm (ppg.find_onsets_elgendi2013()), an adaptive dual-threshold detection method specifically designed for PPG signals. This algorithm demonstrated superior robustness against motion artifacts compared to conventional derivative-based methods and has been validated in ambulatory

Table 1
HRV metrics

Metric	Interpretation	Unit
HR	Average Heart Rate	beats per minute (bpm)
MeanNN	Mean NN Interval	milliseconds (ms)
Prc20NN	20th Percentile of NN Intervals	milliseconds (ms)
MedianNN	Median NN Interval	milliseconds (ms)
Prc80NN	80th Percentile of NN Intervals	milliseconds (ms)
SDNN	Standard Deviation of NN Intervals	milliseconds (ms)
RMSSD	Root Mean Square of Successive Differences	milliseconds (ms)
pNN50	Percentage of NN Interval Differences > 50 ms	percentage (%)
pNN20	Percentage of NN Interval Differences > 20 ms	percentage (%)
LF	Low-Frequency Power	milliseconds^2 (ms^2)
HF	High-Frequency Power	milliseconds^2 (ms^2)
TP	Total Power	milliseconds^2 (ms^2)
LFHF	Low-Frequency/High-Frequency Ratio	-
ApEn	Approximate Entropy	-
LZC	Lempel–Ziv Complexity	-

monitoring contexts. The resulting pulse interval (PI) sequence served as the basis for subsequent PRV analysis and cross-modal consistency evaluation with HRV.

In this study, we computed multidimensional characteristics based on the sequence of heartbeat intervals and PI sequences. HRV and PRV metrics were extracted using the open-source toolkit pyHRV [46], which supports HRV calculation based on time-domain, frequency-domain, and nonlinear dimensions, specifically including the feature dimensions in Table 1.

3.4. Segment-based quality control

All signals were processed in segments with a basic unit of 30 seconds, and the following quality control steps were performed:

Integrity check: If the data loss rate for any channel (PPG/ECG/ACC) exceeds 5% in a segment, that segment is discarded.

Motion artifact screening: Calculated the acceleration vector magnitude $A_{mag} = \sqrt{A_x^2 + A_y^2 + A_z^2}$.

Effective segment concatenation: Connected segments that pass the integrity check in chronological order to form continuous analysis segments to enhance the stability of parameter estimation.

3.5. Consistency analysis

3.5.1. Inter-individual consistency analysis

Extract the overall HRV and PRV parameters for each subject during resting, active, and sleep periods. Evaluate the overall consistency from the two signal sources across different subjects by calculating the Pearson correlation coefficients and root mean square error (RMSE) between PPG and ECG for each parameter.

3.5.2. Intra-individual consistency analysis

To further validate stability at the individual level, for each subject, nonoverlapping 5-minute segments were divided according to task phases (sleep stages/active period), and the following operations were performed:

1) Validity criteria: Each 5-minute segment must contain at least 6 valid 30-second sub-segments (validity rate ≥ 60%).
2) Parameter extraction: Calculate PRV from PPG and HRV from ECG for each 5-minute segment.
3) Consistency assessment: Calculate the Pearson correlation coefficient between PRV and HRV parameters for all 5-minute segments of the same subject.
4) Group analysis: Average the correlation coefficients across subjects to obtain the intra-individual consistency metric for that phase.

3.6. Motion strength threshold control

To quantify the impact of acceleration on the PPG signal, a motion intensity threshold control was introduced:

1) Set a series of decreasing thresholds τ (including no control, i.e., $\tau = \infty$): [∞, 10, 8, 6, 4.5, 4.3, 4.1, 3.9, 3.7, 3.5, 3.3, 3.1, 2.9, 2.7, 2.5, 2.3, 2.1, 1.9, 1.7, 1.5, 1.3, 1.1, 0.9, 0.7, 0.5, 0.3, 0.1] m/s^2.
2) Further exclude segments with A_{mag} exceeding τ based on quality control, and assess the trend of consistency as motion intensity decreases.

3.7. Statistical analysis

Consistency analysis employed two complementary statistical metrics to assess PPG-ECG agreement: (1) Pearson correlation coefficient (r) to evaluate the linear association strength between PRV and HRV parameters, indicating whether the two modalities capture the same rank-order patterns, and (2) RMSE to quantify absolute numerical deviations in HRV parameter estimates between the two signals, reflecting measurement precision. Additionally, we have provided Bland–Altman plots for all indicators to visually and statistically demonstrate the limits of agreement and any systematic biases. These Bland–Altman plots are publicly accessible at: https://cloud.tsinghua.edu.cn/d/9629ab21aaba4f03a70d/.

Inter-individual consistency analysis examined agreement across participants by computing correlations for each metric at the

group level. Statistical significance of correlation coefficients was assessed using standard Pearson correlation tests.

Intra-individual consistency analysis evaluated within-person tracking reliability by computing correlations between PPG and ECG within each participant across multiple 5-minute epochs, then performing group-level statistical inference on these individual-specific correlations. A one-sample t-test was employed to determine whether the mean intra-individual correlation coefficient across all participants significantly exceeded the empirical consistency threshold of $r = 0.2$. This threshold represents the minimum correlation required for clinically meaningful longitudinal tracking, below which within-person monitoring is considered unreliable.

The statistical significance threshold was uniformly set at $p < 0.05$ for all analyses. To control for multiple comparisons and reduce false discovery rates (FDRs), corrections were applied using the FDR method with context-specific correction factors: (1) baseline consistency analysis (no motion control, $\tau = \infty$): FDR correction with a factor of 15 (P_{FDR_15}), corresponding to the number of HRV metrics simultaneously examined. This correction addresses the multiple testing problem arising from evaluating 15 distinct features within each experimental context (resting state, activity period, sleep stages); and (2) motion threshold-controlled analysis (τ ranging from ∞ to 0.1 m/s^2): FDR correction with a factor of 27 (P_{FDR_27}), corresponding to the number of acceleration threshold levels examined. This correction strategy reflects the study's focus on identifying stable, threshold-invariant metrics suitable for robust real-world deployment. Since the primary scientific question concerns which HRV features maintain consistency across varying motion intensities (rather than whether motion control improves any single metric at specific thresholds), the correction factor is determined by the number of threshold conditions tested for each metric. Additionally, we calculated the minimum detectable effect (MDE) to make the detection capability boundary of this study clearer. The MDE refers to the smallest true effect that can be reliably detected by statistical testing to specify efficacy.

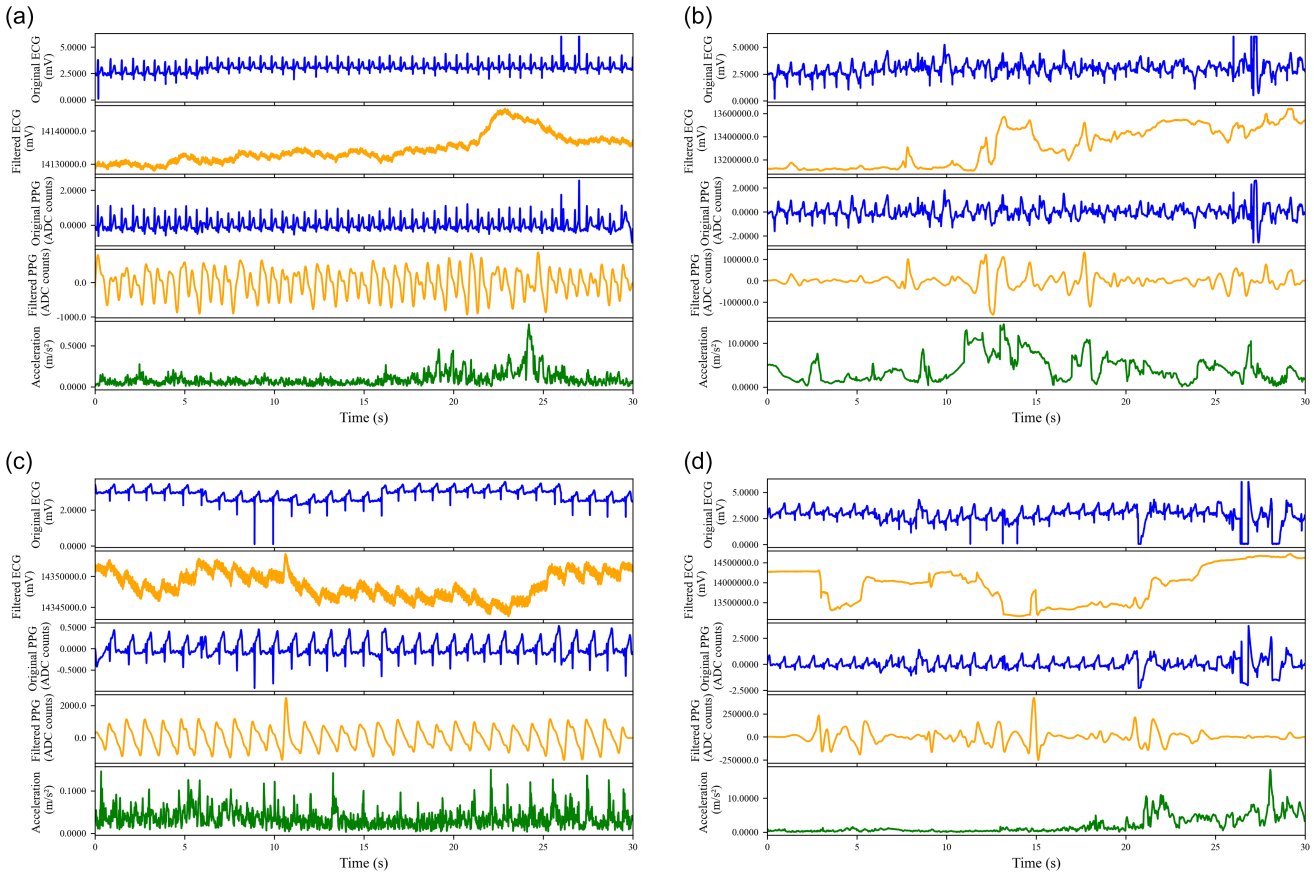
All statistical analyses were performed using Python 3.9 with scipy.stats and statsmodels libraries. Correlation coefficients, p-values, FDR-adjusted p-values, and RMSE values for all comparisons are provided in the supplementary materials.

4. Results

4.1. Data overview

This study collected synchronized data from 14 participants during both active and sleep periods, as shown in Figure 2. Preliminary visual analysis revealed that PPG and ECG signals exhibited varying degrees of motion interference during different tasks. At low exercise intensities, the waveforms of the two signals

Figure 2
(a) Active period with low linear acceleration, (b) active period with high linear acceleration, (c) sleep stages with low linear acceleration, and (d) sleep stages with high linear acceleration



exhibited high synchrony; however, at high exercise intensities, the PPG waveform frequently exhibited pulse wave loss and peak shifts, while the ECG signal remained stable. This phenomenon suggests that exercise interference may affect the quality of the PPG signal, thereby impacting the consistency between the two signals.

4.2. Inter-individual consistency analysis

Figures 3−5 show the scatter plots of various HRV and PRV indicators during resting, active, and sleep states. Figure 6 shows the Pearson correlation coefficients between HRV and PRV parameters in the three states. Table 1 in the S1 of the supplementary materials provides the specific Pearson correlation coefficients, *p*-values, and FDR-corrected *p*-values.

During sleep, multiple parameters exhibited excellent agreement, characterized by high correlations ($r > 0.75$), low RMSE, and strong statistical significance after FDR correction ($P_{FDR\_15} < 0.01$). For instance, HR achieved $r = 0.961$ ($P_{FDR\_15} < 0.001$, 95% CI [0.864, 0.990]) with an RMSE of only 2.17 bpm. MeanNN and MedianNN achieved $r = 0.975$ ($P_{FDR\_15} < 0.001$, 95% CI [0.896, 0.994]) and 0.991 ($P_{FDR\_15} < 0.001$, 95% CI [0.951, 0.998]), respectively, with RMSE below 25 ms. Prc80NN ($r = 0.994$, $P_{FDR\_15} < 0.001$, 95% CI [0.961, 0.999]), SDNN ($r = 0.758$, $P_{FDR\_15} = 0.003$, 95% CI [0.378, 0.920]), and spectral power parameters (LF and TP, $r \approx 0.78$) were also robust. Notably, the LFHF reached statistical significance during sleep ($r = 0.754$, $P_{FDR\_15} = 0.003$, 95% CI [0.370, 0.919]), suggesting potential utility in stable physiological conditions. Complexity measures (ApEn and LZC) showed

Figure 3
Resting state: visualization of PRV and HRV parameters

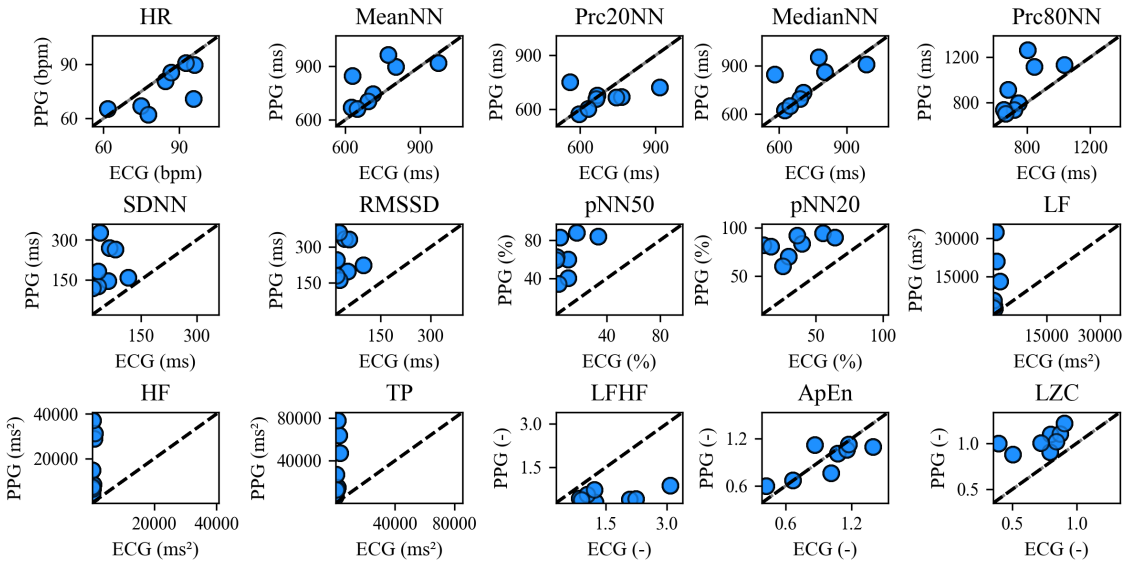


Figure 4
Activity period: visualization of PRV and HRV parameters

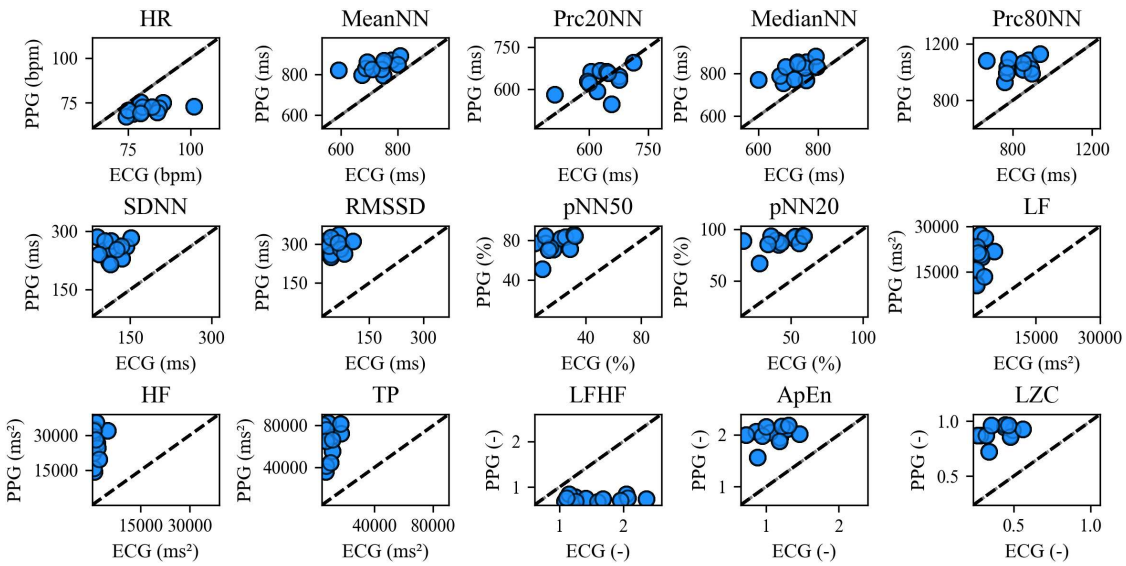


Figure 5
Sleep stages: visualization of PRV and HRV parameters

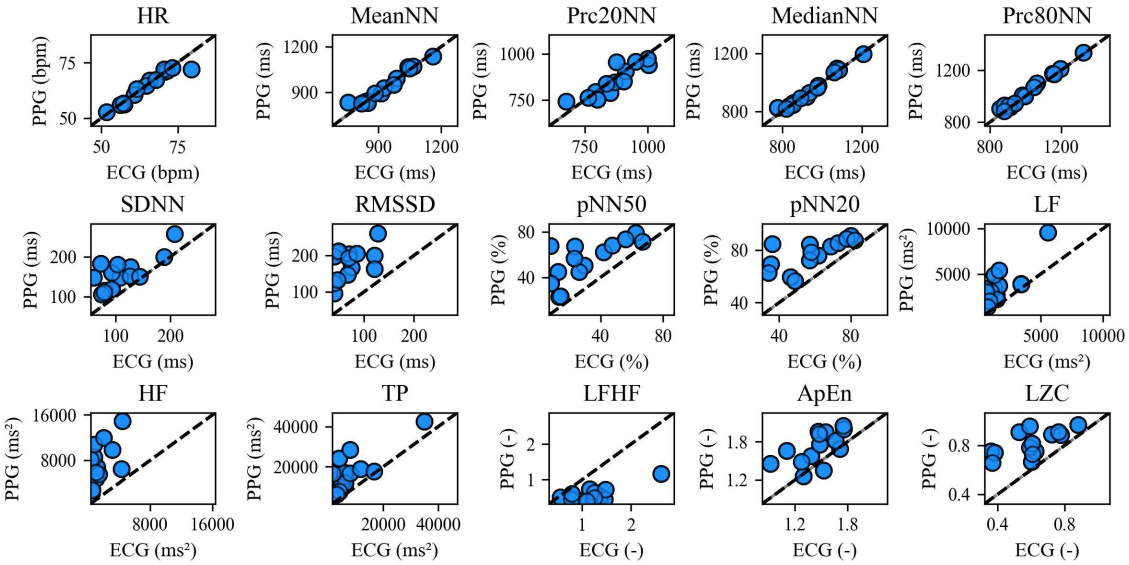
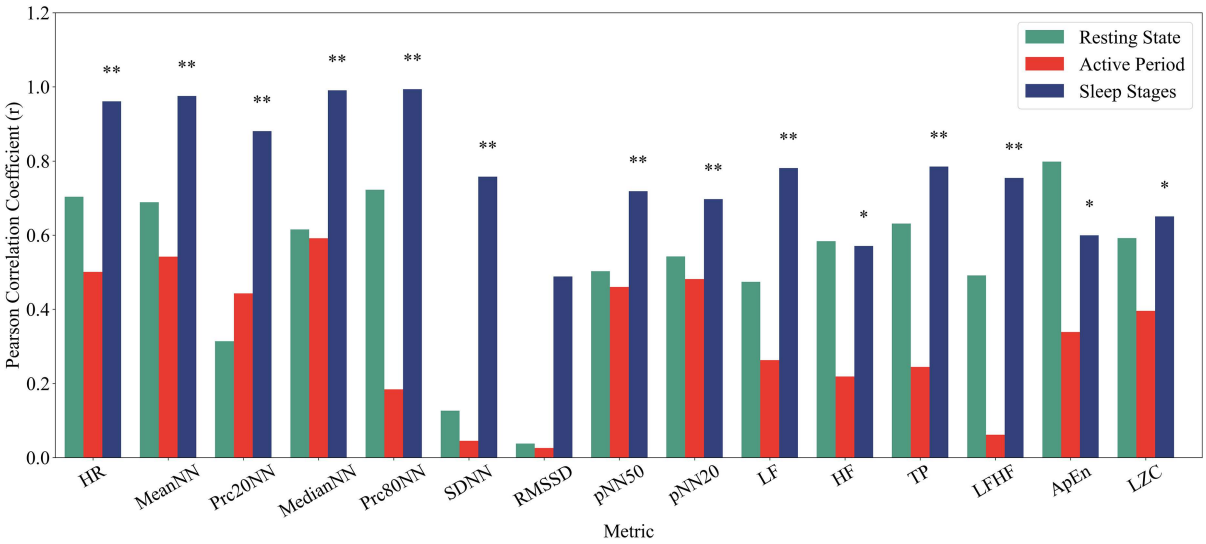


Figure 6
Inter-individual Pearson correlation coefficients between PRV and HRV parameters



Note: **: $P_{FDR\_15} < 0.01$, *: $P_{FDR\_15} < 0.05$

moderate but significant correlations ($r \approx 0.60 - 0.65$), indicating that PPG can partially capture autonomic complexity during sleep.

During the resting state, HR and MeanNN showed moderate-to-good correlations ($r > 0.68$) and relatively low RMSE (HR: 11.07 bpm; MeanNN: 111.6 ms). Although they did not reach statistical significance after FDR correction, they remain practically useful. In contrast, other parameters such as SDNN ($r = 0.126$, $p = 0.766$, 95% CI [−0.423, 0.603]), RMSSD ($r = 0.037$, $p = 0.930$, 95% CI [−0.500, 0.558]), and HF ($r = 0.584$, $p = 0.129$, 95% CI [0.064, 0.852]) were unreliable due to either very low correlation or excessively large RMSE.

During the activity period, all metrics deteriorated substantially: correlations were generally below 0.55, RMSE increased markedly (e.g., SDNN: 139 ms; RMSSD: 240 ms; LF: >18000 ms^2), and none reached statistical significance after FDR correction (minimum $P_{FDR\_15} = 0.372$). Even the best-performing metrics—MeanNN ($r = 0.542$, $p = 0.069$, 95% CI [−0.003, 0.832]) and MedianNN ($r = 0.592$, $p = 0.043$, $P_{FDR\_15} = 0.372$, 95% CI [0.077, 0.856])—exhibited high RMSE (123 and 100 ms, respectively). This indicates that motion artifacts severely disrupt the correspondence between PPG waveforms and true R–R intervals, rendering HRV analysis unreliable in dynamic environments. These findings demonstrate that motion artifacts fundamentally compromise PPG's ability to accurately capture beat-to-beat variability during unconstrained physical activity.

Power analysis indicates that, under the current sample size ($N = 14$), significance level ($\alpha = 0.05$), and target power ($1-\beta = 0.80$), the inter-individual correlation analysis is only sufficiently powered to reliably detect large effect sizes (Pearson's $r \geq 0.688$), which defines the MDE in this analysis. Consistent with this, most metrics that achieved statistical significance also exhibited effect sizes exceeding the MDE, supporting their robustness. However,

during sleep, HF, ApEn, and LZC reached statistical significance despite effect sizes below the MDE ($r < 0.688$). This suggests that their significance may stem from reduced inter-individual variability—due to physiological homogenization during sleep—or favorable sampling fluctuations, thereby carrying a risk of false positives. The actual consistency of these metrics is likely weak and warrants cautious validation in larger cohorts. Conversely, in the resting state, MeanNN, Prc80NN, and ApEn, although not statistically significant, demonstrated effect sizes above the MDE ($r \geq 0.688$), implying that their lack of significance is likely attributable to limited statistical power rather than true null effects. Notably, ApEn showed a strong trend toward individual-specific consistency during rest, reflecting its sensitivity to autonomic complexity in the awake state. In contrast, its statistically significant but low-effect correlation during sleep reveals a lack of individual discriminability.

This contrast underscores a critical principle: the utility of a PRV metric depends not merely on statistical significance but more importantly on its ability to provide robust, individualized characterization within physiologically heterogeneous contexts. The agreement between PRV and HRV metrics is highly state-dependent: sleep represents an optimal condition for signal fidelity; resting state offers only limited reliability—primarily for mean-rate measures; and physical activity largely compromises the validity of PRV estimation. Therefore, when employing wearable devices for PRV assessment, researchers should stratify analyses by physiological state and prioritize metrics that concurrently demonstrate high correlation, low RMSE, and statistical significance within the intended application context.

4.3. Intra-individual consistency analysis

The sleep and active period data of each subject were divided into nonoverlapping 5-minute segments, and HRV indices of PPG and ECG were calculated separately (as shown in Figures 7–11). The intra-individual Pearson correlation coefficients were then calculated based on the corresponding values between these segments. Figures 12 and 13 show the Pearson correlation coefficients and RMSE between PPG and ECG HRV parameters in the two tasks. Table 2 in the S1 of the supplementary materials provides the specific values.

During sleep, most indices demonstrated strong agreement across both linear and absolute dimensions, supported by significant FDR-adjusted p-values. HR ($r = 0.650$, RMSE = 5.06 bpm, $P_{FDR\_15} < 0.001$, 95% CI [0.512, 0.788]) and MeanNN ($r = 0.689$, RMSE = 54.4 ms, $P_{FDR\_15} < 0.001$, 95% CI [0.558, 0.820]) showed the highest consistency. Distribution-related indices such as MedianNN ($r = 0.761$, RMSE = 50.4 ms, $P_{FDR\_15} < 0.001$, 95% CI [0.641, 0.881]), Prc20NN ($r = 0.663$, RMSE = 79.4 ms, $P_{FDR\_15} < 0.001$, 95% CI [0.539, 0.787]), and Prc80NN ($r = 0.536$, RMSE = 80.9 ms, $P_{FDR\_15} = 0.009$, 95% CI [0.323, 0.749]) were also robust ($P_{FDR\_15} < 0.01$). SDNN ($r = 0.404$, RMSE = 80.9 ms, $P_{FDR\_15} = 0.009$, 95% CI [0.310, 0.498]) and ApEn ($r = 0.464$, RMSE = 0.129, $P_{FDR\_15} < 0.001$, 95% CI [0.375, 0.553]) demonstrated moderate but reliable correlations with acceptable RMSE (77–128 ms), suggesting that PPG can effectively estimate multiple HRV dimensions under low-motion conditions.

In contrast, RMSSD, pNN50, pNN20, LF, HF, and TP exhibited poor within-person tracking reliability ($r < 0.25$) with large RMSE values, indicating poor reproducibility within individuals. The LFHF reached marginal significance ($r = 0.289$, RMSE = 1.140, $p = 0.040$, $P_{FDR\_15} = 0.075$, 95% CI [0.195, 0.383]), implying limited interpretability under stable physiological conditions.

During the activity, consistency deteriorated substantially. Only Prc20NN remained significant ($r = 0.516$, RMSE = 94.66ms, $P_{FDR\_15} < 0.001$, 95% CI [0.391, 0.641]). All other parameters degraded to near-random levels, including HR ($r = 0.213$, RMSE = 17 bpm, $p = 0.439$, 95% CI [0.064, 0.362]) and MeanNN ($r = 0.210$, RMSE = 163.3 ms, $p = 0.455$, 95% CI [0.046, 0.374]), accompanied by large RMSE values (HR: 17.3 bpm; SDNN: 192.8 ms;

Figure 7
HR: comparison between ECG and PPG within individuals

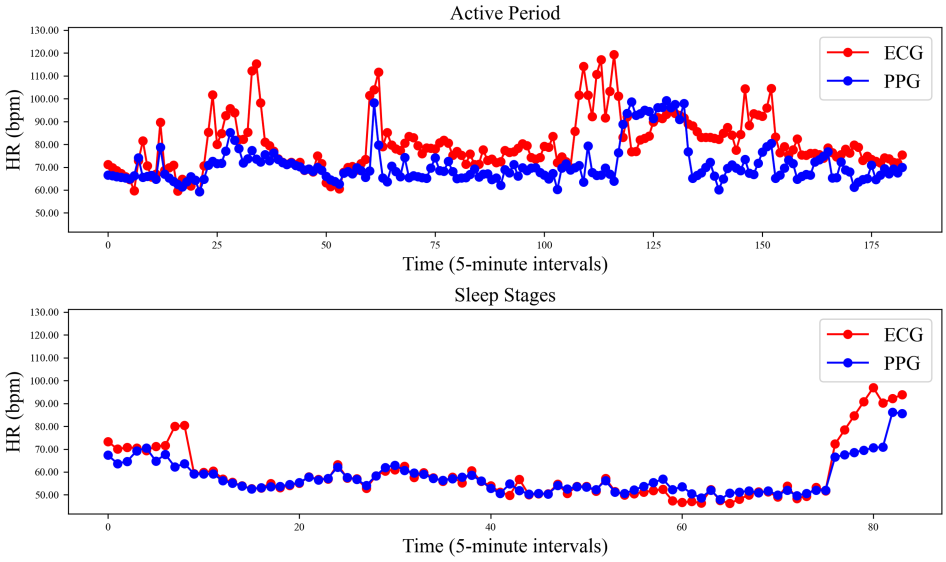


Figure 8
Heartbeat interval distribution: comparison between ECG and PPG within individuals. (a) MeanNN; (b) Prc20NN; (c) MedianNN; (d) Prc80NN

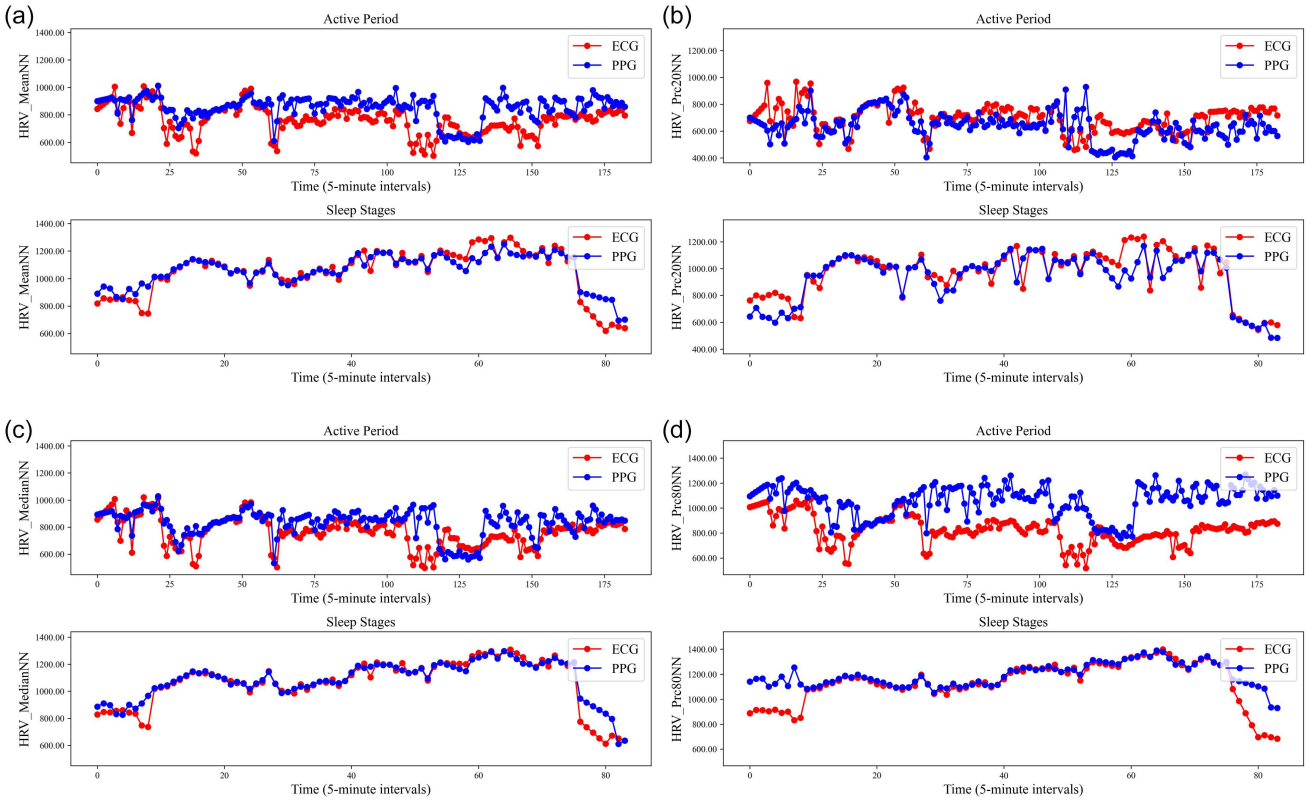


Figure 9
Time-domain metrics: comparison between ECG and PPG within individuals. (a) SDNN; (b) RMSSD; (c) pNN50; (d) pNN20

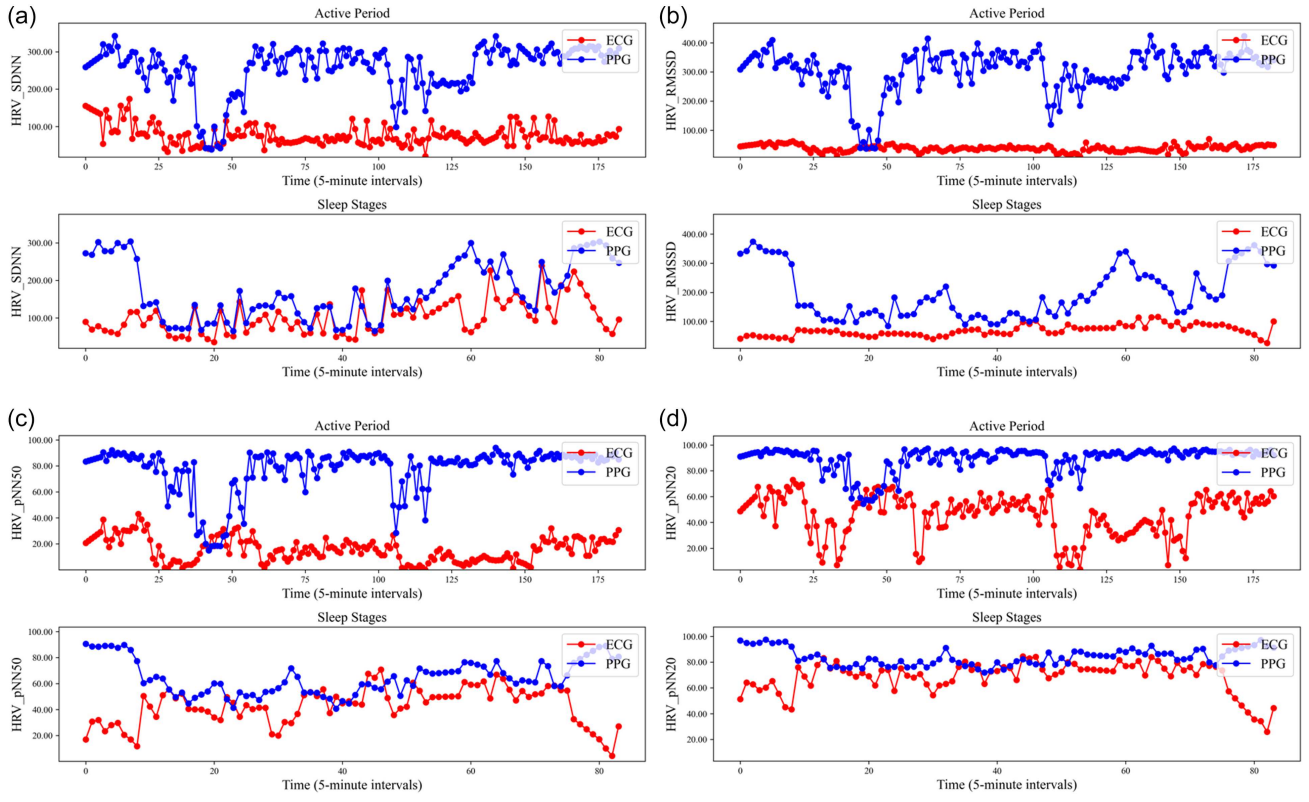


Figure 10
Frequency-domain indicators: comparison of ECG and PPG within individuals. (a) LF; (b) HF; (c) TP; (d) LFHF

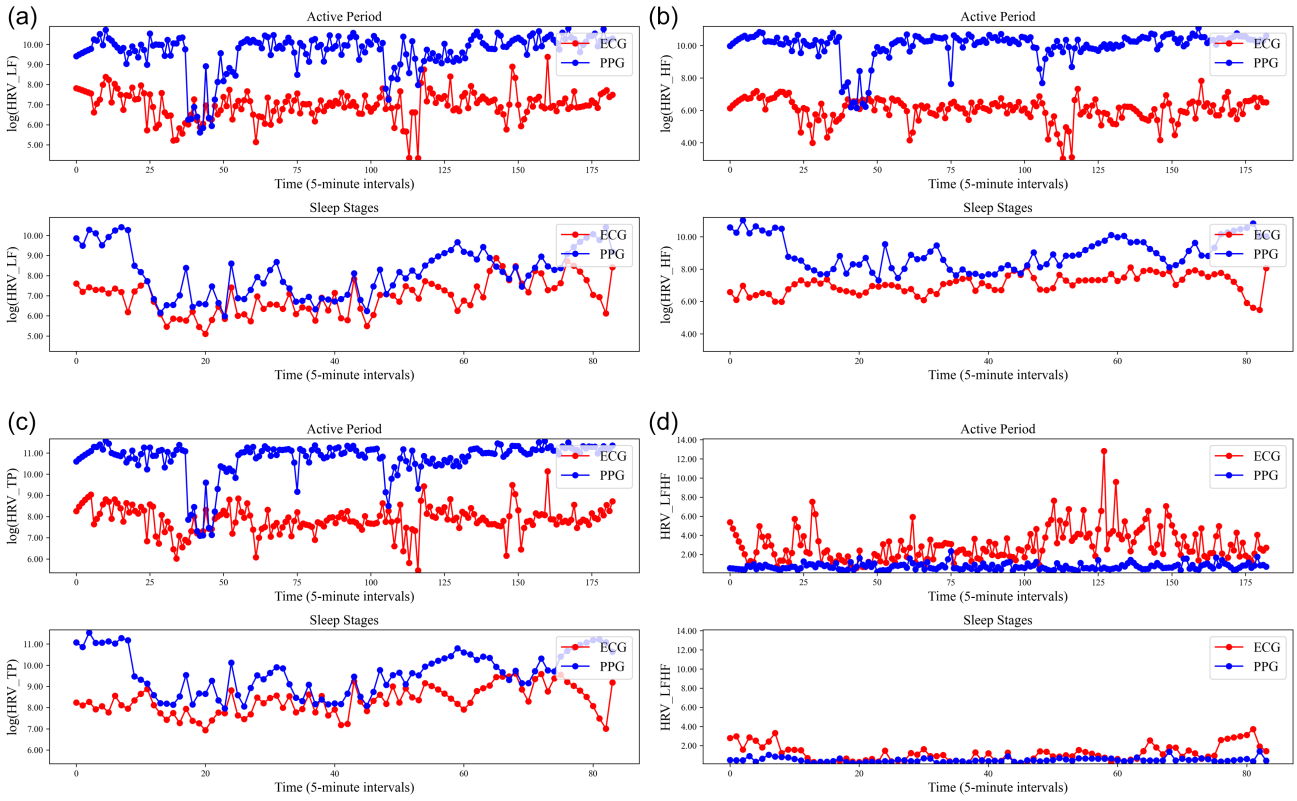
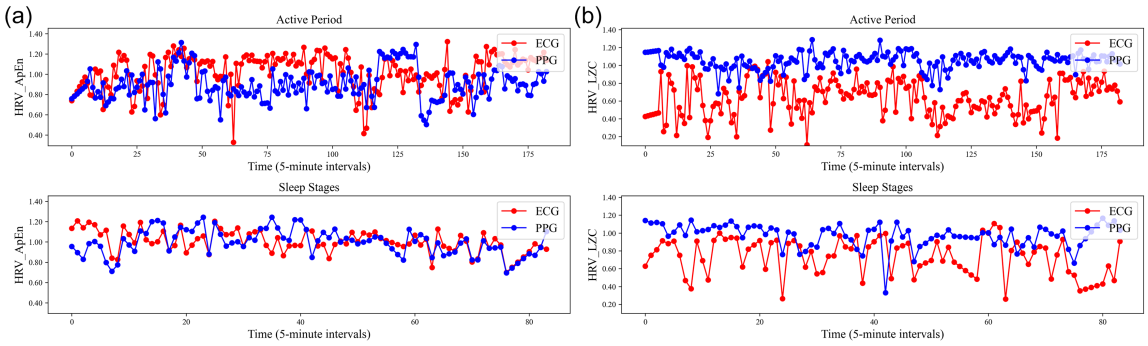


Figure 11
Nonlinear indicators: comparison of ECG and PPG within individuals. (a) ApEn; (b) LZC



LF: >20,000). These findings indicate that motion artifacts during free activity severely impair the capacity of PPG signals to track intra-individual HRV variations.

Power analysis indicates that, with the current sample size (N = 14), significance level (α = 0.05), and target power ($1-\beta$ = 0.80), the one-sample t-test used to evaluate whether within-subject average correlations exceed a threshold of 0.2 is only adequately powered to detect large effect sizes (Cohen's $d \geq 0.702$), which defines the MDE in this analysis. All metrics that achieved statistical significance also exhibited effect sizes exceeding the MDE, reinforcing the practical robustness of their PPG–ECG consistency. In summary, the sleep stage provides the optimal context for intra-individual PRV-HRV consistency, with multiple parameters exhibiting stable and statistically significant agreement. Prc20NN is the only HRV index achieving significant consistency in both activity and sleep, demonstrating high application potential.

4.4. Effect analysis of acceleration threshold control

4.4.1. Inter-individual consistency analysis

By progressively reducing the motion intensity threshold (τ), this study revealed distinct sensitivity patterns of different HRV indicators to motion interference. The integration of Pearson correlation coefficients and RMSE provided a comprehensive assessment of PPG-ECG agreement across measurement contexts. Overall, correlation exhibited an upward trend as the linear acceleration threshold decreased, with the most pronounced improvements observed during the active period and smaller gains during sleep stages. However, error reduction patterns varied substantially across metric types, revealing three distinct response profiles: Pattern 1: Dual-Improvement Metrics (HR and Interval Distribution Parameters); Pattern 2: RMSE-Dominant Improvement Metrics (Time-Domain Variability Indices); and Pattern 3: Correlation-Resistant Metrics

Figure 12
Intra-individual Pearson's correlation coefficient between PPG and ECG HRV parameters variability

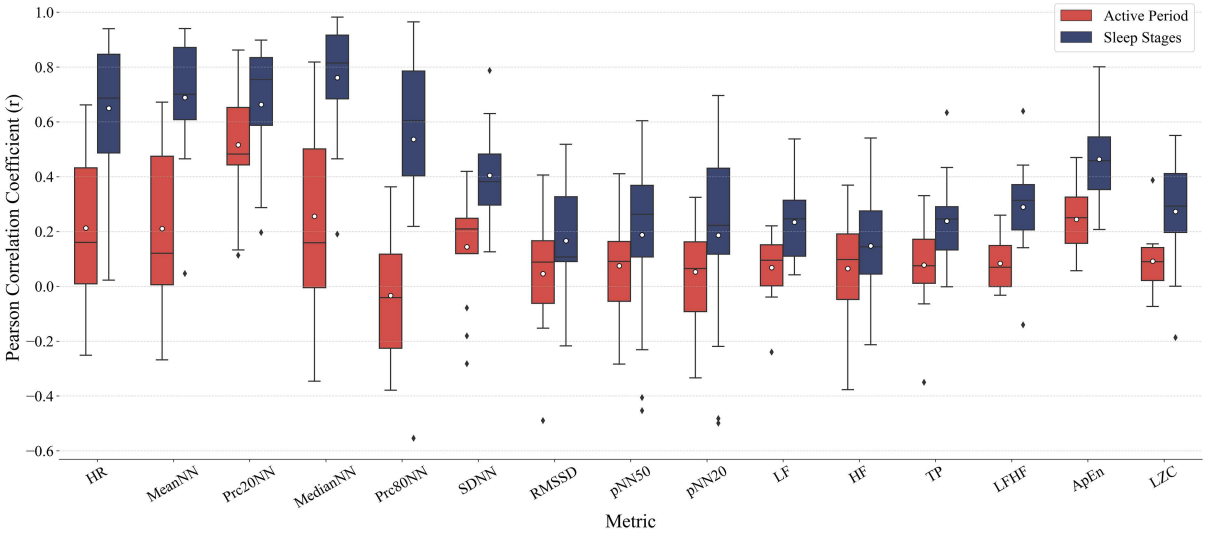
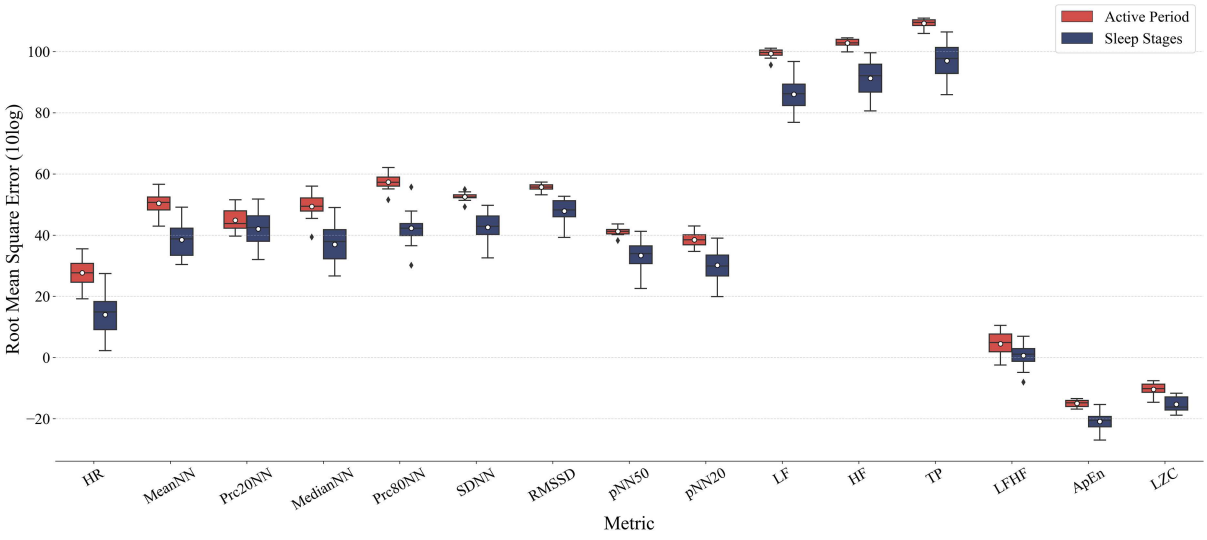


Figure 13
Intra-individual RMSE between PRV and HRV parameters



(Frequency-Domain Parameters). All inter-individual analytical results for consistency and linear acceleration are in "S2, S3, and S4" of the supplementary materials.

HR demonstrated optimal responsiveness to motion control, exhibiting coupled improvements in both correlation and RMSE across all contexts. As illustrated in Figure 14, during the active phase, the correlation increased from 0.501 (p = 0.097, P_{FDR_27} = 0.286) at $\tau = \infty$ to 0.813 m/s^2 (p = 0.001, P_{FDR_27} = 0.020) at τ = 0.1 m/s^2, while RMSE decreased substantially from 13.16 bpm to 5.69 bpm, representing a 57% error reduction. In the resting state, correlation improved steadily from 0.704 (p = 0.051, P_{FDR_27} = 0.053) to 0.898 (p = 0.015, P_{FDR_27} = 0.051), with RMSE decreasing from 11.07 bpm to 7.14 bpm (35% reduction). During sleep stages, HR maintained exceptionally high consistency, with correlation increasing from 0.961 ($p < 0.001$, $P_{FDR\_27} < 0.001$) to 0.993 ($p < 0.001$, $P_{FDR\_27} < 0.001$) and RMSE declining from 2.17 bpm to 1.07 bpm (51% reduction). These parallel improvements in correlation and RMSE validate motion artifacts as the primary error source for HR estimation, establishing PPG as a reliable surrogate for ECG-derived heart rate under controlled motion conditions.

Interval distribution metrics (Prc20NN, MedianNN, Prc80NN) demonstrated context-dependent dual-improvement patterns with notable initial deterioration phases. As shown in Figure 15, Prc80NN during the active phase exhibited a characteristic U-shaped trajectory: correlation decreased from 0.184 (p = 0.567, P_{FDR_27} = 0.638) to a minimum of 0.030 (p = 0.922, P_{FDR_27} = 0.922) before gradually recovering to 0.776 (p = 0.002, P_{FDR_27} = 0.049). Concurrently, RMSE showed parallel dynamics, increasing from 227.32 ms ($\tau = \infty$) to a maximum of 221.60 ms (τ = 10.0 m/s^2) before decreasing to 94.75 ms (τ = 0.1 m/s^2), achieving a 58% error reduction from baseline. During sleep stages, Prc80NN maintained stable high performance with minimal RMSE variation (25.74 ms → 18.13 ms, 30% reduction) and consistently elevated correlation ($r > 0.994$). Resting state showed intermediate behavior with correlation ranging from 0.664 to 0.744 and RMSE fluctuating between 200-215 ms. Prc20NN exhibited superior cross-context

Figure 14
HR: relationship between inter-individual Pearson's correlation coefficient and τ

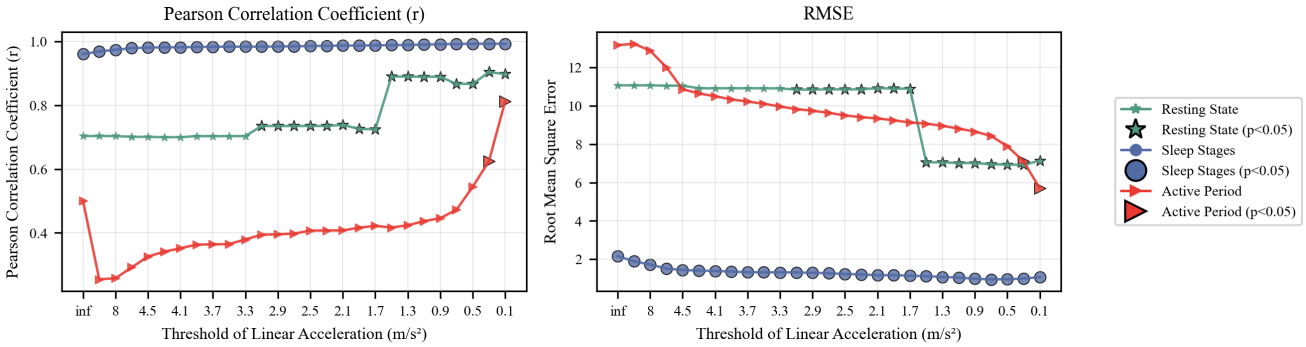


Figure 15
Prc80NN: relationship between inter-individual Pearson's correlation coefficient and τ

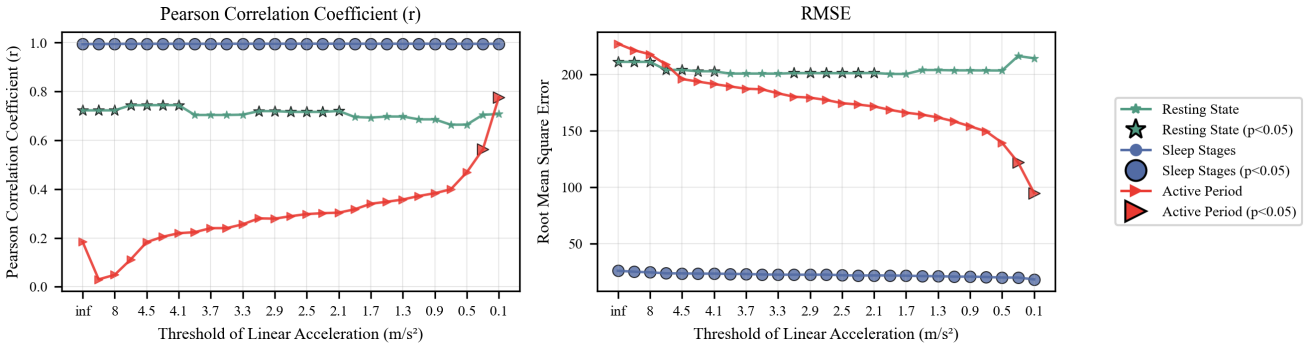
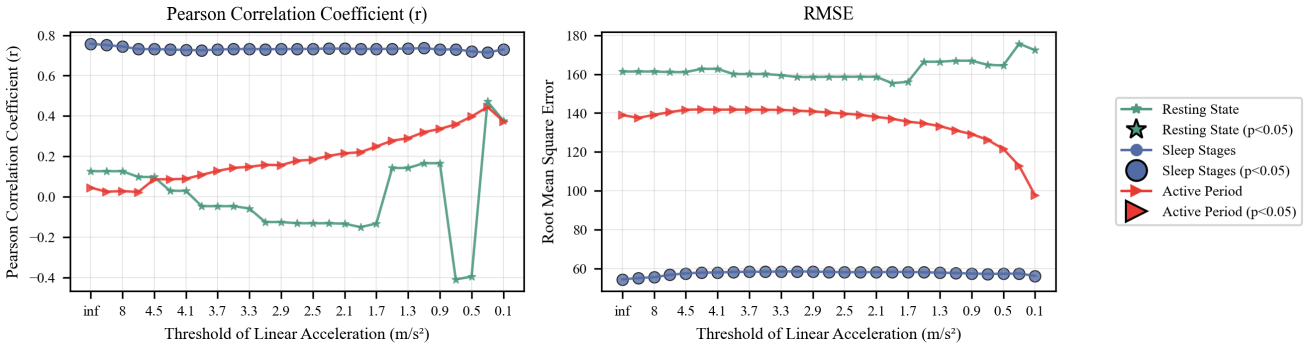


Figure 16
SDNN: relationship between inter-individual Pearson's correlation coefficient and τ



robustness: during activity, correlation increased from 0.443 (p = 0.149, P_{FDR_27} = 0.502) to 0.705 (p = 0.007, P_{FDR_27} = 0.192), with RMSE increasing from 47.74 ms to 83.30 ms; during sleep, correlation reached 0.959 ($p < 0.001$, $P_{FDR} < 0.001$) with RMSE of 41.54 ms.

SDNN and RMSSD exhibited asymmetric improvement patterns characterized by substantial RMSE reduction but minimal correlation enhancement, revealing a fundamental dissociation between absolute error and rank-order consistency. As shown in Figure 16, during the active phase, SDNN correlation increased modestly from 0.045 (p = 0.890, P_{FDR_27} = 0.941) to 0.374 ($p = 0.208$, $P_{FDR\_27} = 0.915$), failing to reach statistical significance despite stringent motion control. Conversely, RMSE decreased substantially from 139.02 ms to 97.65 ms (30% reduction), indicating improved measurement precision without a corresponding enhancement in inter-individual discrimination capability. During sleep stages, SDNN maintained moderate correlation ($r \approx 0.73$) with RMSE fluctuating between 54.38 and 58.49 ms, suggesting that motion artifacts contribute less to SDNN discrepancy during low-activity states. The resting state showed similar patterns with RMSE fluctuating between 155 and 176 ms.

RMSSD showed even more pronounced dissociation: active-phase correlation remained near-zero across all thresholds (0.026 → 0.359, $p > 0.2$), while RMSE decreased from 239.51 to 168.11 ms (30% reduction). Sleep-stage RMSSD exhibited consistently low correlation ($r \approx 0.50$) despite excellent absolute agreement (RMSE: 114.68 → 100.91 ms). This pattern suggests that while PPG can track within-individual SDNN/RMSSD changes under motion control, it fails to preserve the between-individual ranking structure that ECG captures, likely due to PTT variability

and nonlinear PPG-ECG relationships in beat-to-beat interval differences.

Frequency-domain metrics (LF, HF, TP, LFHF) exhibited paradoxical behavior: substantial RMSE reduction coincided with unstable or deteriorating correlations, revealing fundamental PPG-ECG differences beyond motion artifact contamination. During the active phase, LF RMSE decreased from 18278 ms^2 to 8464 ms^2 (54% reduction), yet correlation showed non-monotonic fluctuation from 0.263 (p = 0.409, P_{FDR_27} = 0.442) to 0.204 (p = 0.504, P_{FDR_27} = 0.504), failing to achieve significance at any threshold. HF demonstrated similar patterns with RMSE improving from 25492 ms^2 to 14522 ms^2 (43% reduction), while correlation oscillated between 0.219 and 0.373 without consistent directionality. Sleep stages showed moderately high LF correlation ($r \approx 0.78$) but substantial RMSE (2331 ms^2 → 2022 ms^2), with minimal error reduction despite maintained correlation. LFHF ratio exhibited the most erratic behavior: active-phase correlation fluctuated wildly between -0.113 and 0.355 across thresholds, suggesting near-random agreement.

ApEn and LZC demonstrated moderate sensitivity to motion control with context-dependent patterns. The relatively better performance of complexity metrics compared to frequency-domain indices suggests they may capture complementary autonomic information less susceptible to PTT artifacts, warranting further investigation in larger cohorts.

4.4.2. Intra-individual consistency analysis

Intra-individual consistency analysis examined whether PRV metrics could reliably track within-person fluctuations over time, a critical requirement for personalized health monitoring applications. By segmenting continuous recordings into 5-minute epochs and computing correlations within each participant, this analysis revealed the feasibility of using PPG for longitudinal self-tracking across different physiological states. Intra-individual analysis employed one-sample t-tests to determine whether averaged within-person correlations exceeded the empirical consistency threshold. This threshold represents minimal clinically meaningful tracking capability, where correlations below 0.2 indicate unreliable within-person monitoring. All intra-individual analytical results for consistency and linear acceleration are presented in S5 and S6 of the supplementary materials.

HR demonstrated context-dependent dual-improvement patterns, with both correlation and RMSE showing synchronized enhancement under motion control. As shown in Figure 17, during the active period, baseline HR showed inadequate tracking (r = 0.213, p = 0.439, P_{FDR_27} = 0.455, RMSE = 17.25 bpm), improving dramatically to robust consistency (r = 0.747, p = 0.003, P_{FDR_27} = 0.010, RMSE = 3.61 bpm, 79% error reduction) at τ = 0.1 m/s^2. The correlation achieved significance at $\tau \approx 2.5$ m/s^2 (r = 0.430, p = 0.009, P_{FDR_27} = 0.022), while RMSE crossed clinical acceptability thresholds (< 5 bpm) at $\tau \approx 1.5$ m/s^2(RMSE = 9.68 bpm → 3.61 bpm). This dual convergence validates that stringent motion control ($\tau < 2.5$ m/s^2) is necessary and sufficient for reliable activity-based HR self-tracking. During sleep stages, HR maintained consistently excellent performance: baseline (r = 0.650, $p < 0.001$, $P_{FDR\_27} < 0.001$; RMSE = 5.06 bpm) improved to final (r = 0.806, $p < 0.001$, $P_{FDR\_27} < 0.001$; RMSE = 3.06 bpm, 40% error reduction). The universally significant t-tests across all thresholds combined with clinically acceptable RMSE values (< 6 bpm throughout) establish PPG as a validated tool for overnight HR monitoring without requiring motion filtering algorithms.

Prc20NN demonstrated exceptional baseline robustness combined with significant error reduction potential. As shown in Figure 18, during activity, Prc20NN was the only metric achieving

Figure 17
HR: relationship between intra-individual Pearson's correlation coefficient and τ

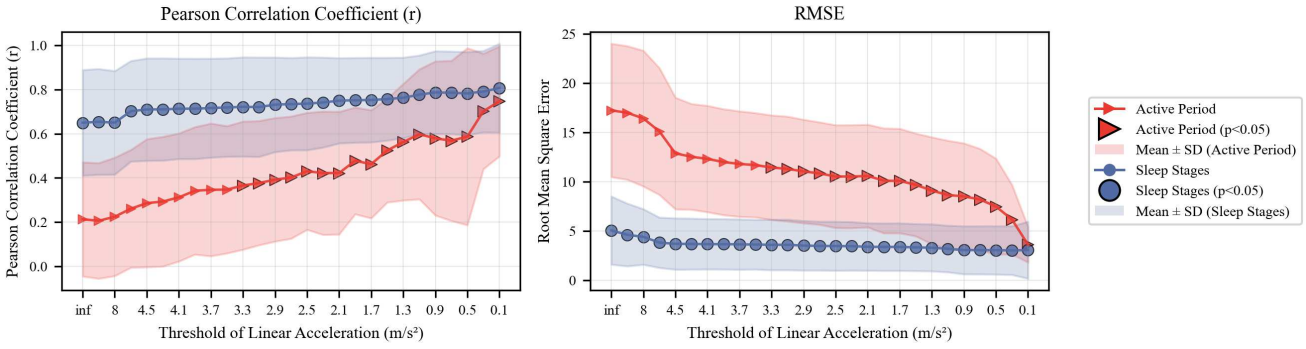
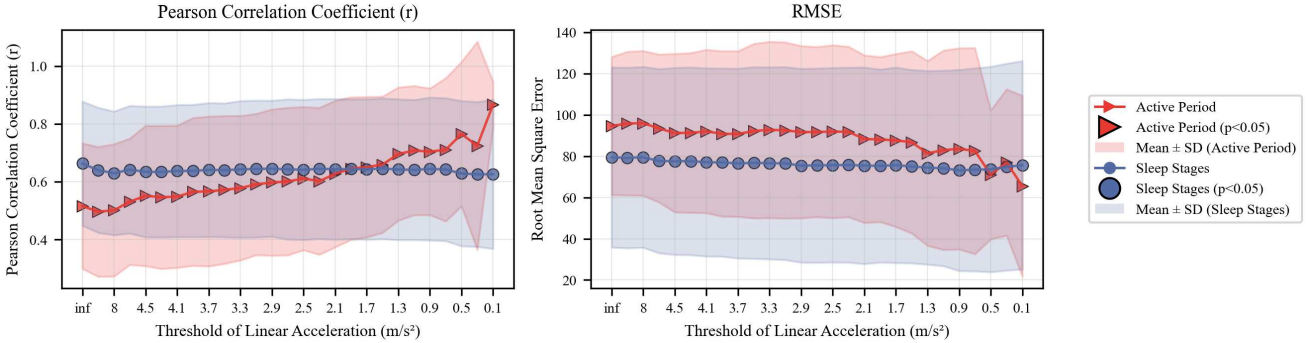


Figure 18
Prc20NN: relationship between intra-individual Pearson's correlation coefficient and τ



significance without motion control ($r = 0.516$, $p < 0.001$, $P_{FDR\_27} < 0.001$), though baseline RMSE was substantial (94.66 ms). Motion control produced dramatic improvements: correlation increased to 0.867 ($p < 0.001$, $P_{FDR\_27} < 0.001$), while RMSE decreased to 65.51 ms (31% reduction) at τ = 0.1 m/s^2. The persistent significance across all 27 thresholds ($p < 0.001$ throughout), combined with monotonic RMSE improvement, validates Prc20NN as the premier metric for unconstrained activity tracking.

Critically, Prc20NN's RMSE trajectory shows two distinct phases: (1) plateau phase ($\tau = \infty$ to τ = 4.5 m/s^2): RMSE fluctuates 91−96 ms with minimal improvement, correlation stable at $r \approx 0.50 - 0.55$; and (2) rapid improvement phase ($\tau < 4.5$ m/s^2): RMSE decreases sharply 91 ms → 65 ms, correlation jumps to $r > 0.65$. This biphasic pattern suggests a critical motion threshold around 4.5 m/s^2, above which artifacts dominate and below which physiological signal prevails. During sleep, Prc20NN maintained stable consistency ($r > 0.62$, all $p < 0.001$) with minimal RMSE variation (79.44 ms → 75.68 ms, 5% reduction).

Prc80NN showed severe activity-phase limitations: despite RMSE improvement (319.49 ms → 70.48 ms, 78% reduction—the largest among all metrics), correlation remained nonsignificant throughout (r = -0.034 → 0.566, all $p > 0.07$, $P_{FDR\_27} > 0.5$). This correlation-RMSE dissociation reveals that while motion control reduces measurement noise for upper-tail intervals (bradycardic beats), it fails to preserve rank-order tracking capability, likely due to differential pulse wave loss during slow heart rates. Sleep-stage Prc80NN performed moderately (r = 0.536 → 0.829, $P_{FDR\_27} < 0.001$; RMSE: 80.92 ms → 39.01 ms, 52% reduction).

MedianNN demonstrated strong dual improvements in both contexts: activity (r = 0.255 → 0.939, $P_{FDR\_27} < 0.001$ at $\tau < 1.9$; RMSE: 150.24 ms → 23.39 ms, 84% reduction) and sleep (r = 0.761 → 0.860, $P_{FDR\_27} < 0.001$; RMSE: 50.42 ms → 32.81 ms, 35% reduction). The exceptional RMSE values during activity (< 25 ms at τ = 0.1 m/s^2, approaching ECG-grade precision) establish MedianNN as a viable alternative to mean-based metrics for within-person tracking, particularly when extreme heart rate values may introduce bias.

Despite substantial RMSE reductions (30–48%), SDNN and RMSSD failed to achieve meaningful correlations, revealing fundamental incompatibility between PRV and HRV for within-person tracking.

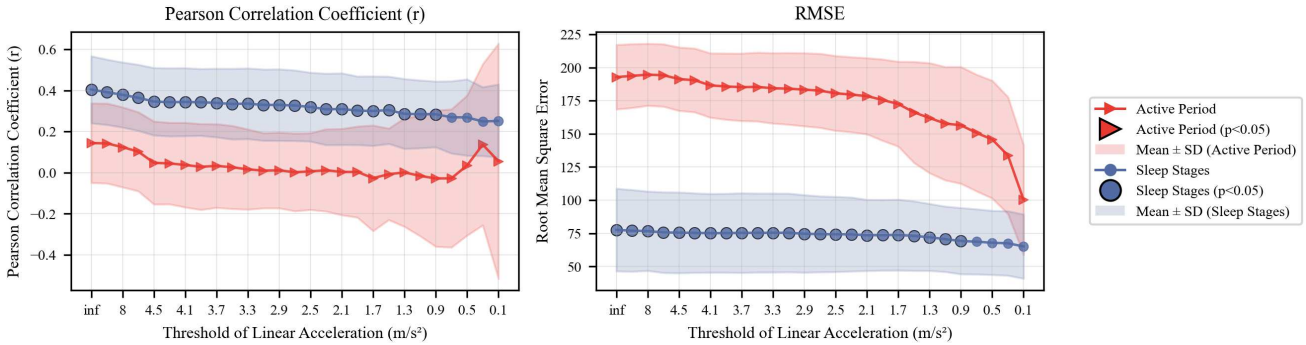
As shown in Figure 19, during activity, SDNN showed near-zero correlations across all thresholds (r = 0.144 → 0.056, all $p > 0.68$), yet RMSE improved markedly from 192.82 ms to 100.30 ms (48% reduction). This pattern indicates that motion control reduces random measurement noise (lowering RMSE) but fails to improve systematic tracking of true within-person SDNN fluctuations (correlation remains near-zero). The dissociation likely reflects individual-specific PTT biases that remain stable across measurements, allowing reduced variance but obscuring true temporal dynamics. Sleep-stage SDNN exhibited even more concerning patterns: baseline significance (r = 0.404, $p < 0.001$, P_{FDR_27} = 0.007) deteriorated progressively under motion control to non-significance (r = 0.252, p = 0.163, P_{FDR_27} = 0.163 at τ = 0.1 m/s^2), despite RMSE remaining stable (77.62 ms → 65.15 ms, 16% reduction). The paradoxical correlation decline suggests stringent thresholds exclude physiologically meaningful low-amplitude movements (e.g., sleep position changes) that genuinely correlate with SDNN variations, providing no net accuracy benefit since baseline motion is already minimal.

RMSSD showed complete failure across all conditions. Activity: correlations ranged 0.046 to -0.264 (all $p > 0.68$), despite dramatic RMSE improvement (263.44 ms → 160.33 ms, 39% reduction). Sleep: correlations consistently near-zero (r = 0.166 → 0.144, all $p > 0.69$), with moderate RMSE reduction (127.86 ms → 105.68 ms, 17% reduction). The universal failure indicates PPG cannot track beat-to-beat variability changes within individuals under any examined condition. The RMSE-correlation dissociation reveals that short-term variability metrics suffer from nonstationary PPG-ECG coupling that varies unpredictably across measurement sessions due to (1) state-dependent pulse wave morphology changes (affecting peak detection differently than ECG R-waves), (2) respiratory modulation of PPG amplitude creating nonequivalent oscillations to ECG heart rate variations, and (3) autonomic effects on peripheral vascular tone that decouple PRV from cardiac electrical variability.

Frequency-domain parameters demonstrated complete failure for intra-individual tracking despite substantial RMSE reductions, confirming fundamental incompatibility with within-person monitoring. Activity: All frequency metrics showed (1) near-zero/negative correlations (LF: 0.068 → 0.088; HF: 0.065 → -0.241; LFHF: 0.084 → 0.234), (2) 100% nonsignificant t-tests across all thresholds ($p > 0.12$), and (3) substantial RMSE improvements (LF: 20903 ms^2 → 7597 ms^2, 64% reduction; HF: 29347 ms^2 → 11915 ms^2, 59% reduction). Sleep-stage results were equally discouraging: (1) persistently low correlations (LF: 0.234 → 0.189; HF: 0.148 → 0.158), (2) all $p > 0.08$, and (3) modest RMSE reductions (LF: 6,19 ms^2 → 3,82 ms^2, 44% reduction; HF: 10,68 ms^2 → 7,47 ms^2, 34% reduction).

The LFHF showed particularly erratic behavior. Activity: RMSE decreased minimally (1.67 → 1.56, 7% reduction—smallest among all metrics), while correlation fluctuated wildly (-0.068 to 0.234) without achieving significance ($p > 0.39$). Sleep: correlation ranged 0.261–0.301 (all $p > 0.03$, all $P_{FDR\_27} > 0.08$), while

Figure 19
SDNN: relationship between intra-individual Pearson's correlation coefficient and τ



RMSE showed paradoxical patterns (1.14 → 0.90, 21% reduction but non-monotonic trajectory).

ApEn showed context-dependent moderate performance with coupled correlation-RMSE improvements during activity but ceiling effects during sleep. Activity: ApEn correlation increased from 0.244 (p = 0.117, P_{FDR_27} = 0.122, marginally nonsignificant) to 0.410 (p = 0.091, P_{FDR_27}= 0.101, marginally nonsignificant at α = 0.05), while RMSE decreased substantially from 0.225 to 0.142 (37% reduction). The failure to achieve formal significance despite approaching the r = 0.2 threshold and substantial error reduction suggests ApEn captures some genuine within-person variability but with insufficient reliability for clinical applications. During sleep, ApEn maintained consistent significance (r = 0.464 → 0.481, all p < 0.001, P_{FDR_27} < 0.001) with excellent RMSE performance (baseline 5.06 → final 3.06, 40% reduction).

LZC showed weak performance in both contexts. Activity: correlations remained nonsignificant throughout despite moderate RMSE reduction (0.361 → 0.253, 30% reduction). Sleep: LZC achieved marginal baseline consistency (r = 0.273, p = 0.101) that improved to borderline significance at stringent thresholds (r = 0.349 at τ = 0.1, p = 0.008, P_{FDR_27} = 0.066), with minimal RMSE improvement (0.223 → 0.174, 22% reduction). The marginal statistical significance, combined with modest error reduction, suggests LZC has limited practical utility for within-person tracking even under optimal conditions.

5. Discussion

This study systematically revealed the consistency boundaries between PRV and HRV parameter estimation through multi-scenario synchronous comparative analysis: PPG can replace ECG for basic time-domain HRV analysis in low-motion interference environments (especially during sleep), but strict motion control and algorithm optimization are required in high-dynamic scenarios or frequency-domain/nonlinear parameter assessments.

During sleep, PPG and ECG exhibit strong correlations in time-domain metrics such as HR, MeanNN, and MedianNN, which align closely with validation results from the Oura ring (nighttime HRV correlation r^2 = 0.98) [47]. Additionally, we computed the MDE based on our sample size (N = 14), significance level (α = 0.05), and target power (1−β = 0.80), confirming that the study was sufficiently powered to detect large effects ($r \geq 0.688$ for correlation analyses). This consistency stems from three physiological foundations: (1) blood flow stability: during sleep, sympathetic tone decreases, weakened peripheral vascular constriction, and improved PPG signal-to-noise ratio [48]; (2) minimization of motion artifacts: the absence of limb movement reduces acceleration interference, avoiding pulse wave loss and peak shifts [48]; and (3) regularity of respiratory rhythm: during slow-wave sleep, respiratory frequency remains stable (0.1–0.3 Hz), reducing modulation of the PPG waveform morphology [49]. Notably, the high robustness of distribution tail metrics such as Prc20NN/Prc80NN during sleep suggests that PPG is better at capturing trends in the interbeat interval distribution than discrepant metrics (e.g., SDNN), which is consistent with the reported sensitivity of PPG to short-term heart rate fluctuations [4].

Compared to the sleep state, consistency significantly decreases during the active state, especially for commonly used HRV time-frequency-domain metrics such as SDNN, RMSSD, and LF, whose PPG estimation results in the active state show almost no correlation with ECG (r < 0.13), indicating that these parameters are highly sensitive to PPG signal quality and susceptible to interference from movement. Especially in intra-individual consistency analysis, basic indicators such as HR and MeanNN even degrade to random levels during activity, further indicating that unprocessed PPG signals struggle to reliably reflect true HRV characteristics under natural conditions. The significant decline in consistency during activity is primarily driven by dual physical-physiological pathways: (1) signal physical level: high linear acceleration causes optical misalignment, leading to PPG waveform distortion, whereas ECG electrical signals exhibit stronger resistance to motion interference [50]; and (2) physiological level: increased sympathetic activation during movement elevates vascular tension, reducing PPG amplitude and increasing peak detection error rates by over fivefold [48]. It is worth noting the universal limitations of frequency-domain and nonlinear indicators: parameters such as LF/HF and ApEn lose their correlation during activity (r < 0.1), consistent with Krolak and Pilecka's conclusion [4] (decreased correlation of frequency-domain features after exercise). The main reason is that these parameters rely on long-term steady-state signals, while PPG motion artifacts disrupt the physiological coupling of heart rate oscillations [49, 51].

By controlling the acceleration threshold, this study quantified the sensitivity of different parameters to exercise interference. As the threshold was gradually reduced, the consistency of key indicators such as HR, MeanNN, and Prc80NN significantly improved, particularly during the active phase, confirming that graded control of exercise intensity is the core strategy for enhancing the accuracy of PRV. This result extends Lin et al.'s finding [50] that PPG can restore HRV estimation capability during the "post-exercise recovery period," while this study further quantifies the exercise intensity threshold. Future research can combine two enhancement approaches: hardware level, using multi-wavelength PPG to suppress motion artifacts [52], and algorithm level, introducing a dynamic weighting model based on exercise thresholds to down-weight high-exercise segments. It is worth noting that some metrics (e.g., LFHF, ApEn, LZC) exhibit low correlation and significant variability across all scenarios, potentially due to their reliance on long-term data segments and signal stability. This suggests that these advanced complexity metrics may be more suitable for high-quality ECG signals rather than dynamic PPG recordings.

Interestingly, Prc20NN, as a distribution metric, exhibits significant consistency during both sleep and active stages, maintaining moderate correlation even under high interference conditions, and remains consistent at both the individual and population levels. Prc20NN's cross-context reliability may stem from intrinsic resistance to pulse PTT variability. This metric measures the 20th percentile of interbeat intervals, representing the "fast heartbeat zone" typically occurring during mild sympathetic activation (light physical activity, emotional arousal). During these tachycardic periods, peripheral vasoconstriction stabilizes PTT at relatively short durations. In contrast, mean-based metrics (MeanNN) aggregate intervals spanning both sympathetic (short PTT) and parasympathetic (long PTT, 250–350ms) dominance, experiencing substantially greater PTT-induced PPG-ECG timing discrepancies. By exclusively sampling from the PTT-stable fast heartbeat zone, Prc20NN minimizes exposure to state-dependent PTT fluctuations that degrade other metrics. Supporting evidence from our data: during activity, Prc20NN maintained significance (r = 0.516, p < 0.001) while MeanNN failed (r = 0.210, p = 0.455), consistent with MeanNN's vulnerability to mixing PTT-stable and PTT-variable intervals. During sleep, both metrics performed well (Prc20NN r = 0.663; MeanNN r = 0.689), suggesting PTT variability rather than absolute PTT magnitude drives the activity-phase divergence. Therefore, Prc20NN can serve as a key indicator for estimating HRV from PPG in dynamic scenarios and warrants further validation and promotion in clinical and home settings.

Our findings provide recommendations for prioritizing future signal processing efforts. Metrics showing dual correlation-RMSE improvements (HR, Prc20NN, MedianNN) represent motion-remediable errors where advanced algorithms—whether deep learning-based artifact removal or adaptive filtering—can expect proportional returns. Conversely, metrics exhibiting RMSE reduction without correlation improvement (SDNN, RMSSD: 30–48% error reduction, $p > 0.1$ for tracking) or complete failure despite substantial RMSE gains (frequency-domain: 40–64% reduction, zero within-person correlation) indicate fundamental physiological discordance resistant to artifact removal alone. For these Tier 2–3 metrics, algorithmic efforts should pivot from noise suppression to physiologically informed correction (e.g., PTT modeling) or alternative approaches leveraging PPG-native features rather than mimicking ECG-derived indices. By establishing which metrics warrant correction versus redesign, our validation framework ensures that future algorithmic innovations target the right problems with appropriate techniques.

6. Study Limitations and Generalizability Considerations

Several limitations warrant careful interpretation of our findings.

1) Sample size and representativeness: This study included 14 healthy young adults (aged 21–28). Although this sample size is sufficient for preliminary exploration, its statistical power in estimating inter-individual correlations is limited. In addition, due to the strict criteria for removing motion artifacts, some subjects may be excluded due to insufficient valid data, further reducing the available samples and limiting the generalizability of the results to a wider population. It is worth noting that this sample represents a best-case scenario for PPG accuracy—young adults exhibit compliant vasculature, minimal arterial stiffening, and homogeneous skin tones that optimize optical signal penetration. Older adults, diverse ethnic groups with varying melanin content, and individuals with cardiovascular diseases (arterial stiffening, autonomic neuropathy) will likely demonstrate worse PPG-ECG concordance. However, this limitation strengthens rather than weakens our core findings: if metrics fail to achieve reliable within-person tracking ($p < 0.05$) in this sample, they will certainly fail in more challenging populations. In intra-individual analysis, statistical power remains adequate despite the small sample size. Each participant contributed multiple 5-minute epochs and each participant's correlation coefficient represents an independent observation derived from dozens of measurements. We adopted conservative statistical practices including FDR correction and explicit confidence interval reporting to guard against false positives. The independent replication of our core finding in Kiran Kumar et al.'s 50-person cohort [53] provides convergent validity evidence, though we acknowledge this cannot substitute for adequate power in our own sample. We therefore frame all inter-individual findings as preliminary estimates requiring larger-sample validation, while presenting intra-individual results with appropriate confidence given adequate statistical power."

For intra-individual analysis—our primary scientific contribution—statistical power remains adequate despite the small sample size. Each participant contributed multiple 5-minute epochs, and each participant's correlation coefficient represents an independent observation derived from dozens of measurements. We adopted conservative statistical practices including FDR correction and explicit confidence interval reporting to guard against false positives. The independent replication of our core finding in Kiran Kumar et al.'s 50-person cohort [53] provides convergent validity evidence, though we acknowledge this cannot substitute for adequate power in our own sample. We therefore frame all inter-individual findings as preliminary estimates requiring larger-sample validation, while presenting intra-individual results with appropriate confidence given adequate statistical power.

2) Device-specific validation: This study employed a single custom-built wristband (Psychorus, Beijing, China) validated in prior daily life studies by Shui et al. [39], He et al. [40] and Shui et al. [41]. Results may not generalize to the diverse ecosystem of consumer wearables (Apple Watch, Fitbit, Garmin) due to differences in optical sensor configurations (wavelength, sampling rate), algorithmic preprocessing (proprietary peak detection and filtering), and wearing placement (wrist vs. finger vs. chest). However, our core contribution—metric-specific validity hierarchies (motion-remediable and fundamentally discordant)—reflects physiological PPG-ECG coupling patterns likely generalizable across devices. For instance, frequency-domain metric failure stems from PTT variability and respiratory-vascular coupling—fundamental physiological phenomena independent of device choice. We recommend that future studies apply our validation framework (dual-criteria assessment with motion threshold analysis) to commercial wearables to determine whether similar Tier structures emerge.
3) Applicability in pathological states is uncertain: Cardiovascular disease patients exhibit increased vascular stiffness, expanding PTT differences between PPG and ECG and thus widening PRV-HRV discrepancies [49, 54]. Autonomic neuropathy (common in diabetes) may alter peripheral vascular reactivity [55], decoupling pulse wave characteristics from cardiac electrical activity in unpredictable ways. Our findings establish PPG validity in healthy populations but cannot be extrapolated to clinical cohorts without dedicated validation. However, the mechanistic insights we provide—distinguishing motion artifacts from physiological vascular reactivity as dual error sources—can guide the design of pathology-specific validation studies.
4) Breathing pattern effects: Controlled breathing significantly impacts HRV metrics, particularly frequency-domain parameters, by synchronizing respiratory sinus arrhythmia [49]. Our study did not monitor respiratory rate or enforce breathing protocols, introducing uncontrolled variance. However, this naturalistic approach enhances ecological validity for real-world wearable applications where breathing cannot be controlled. The persistent failure of frequency-domain metrics despite this "real-world" tolerance suggests their unreliability is not merely an artifact of experimental constraints but reflects fundamental limitations.

These limitations collectively suggest our findings provide necessary but not sufficient conditions for PPG-HRV validity: metrics failing in young healthy adults using controlled devices will certainly fail in broader applications, while metrics succeeding here require further validation across ages, ethnicities, health statuses, and devices before clinical deployment.

Acknowledgment

The authors are grateful to Ms. Xueran Han at the Beijing Huixin Health Technology Co., Ltd, Beijing, China, for the assistance in data collection.

Funding Support

This work is supported by the Grant from The General Office of National Language Commission Research Planning Committee (ZDA145-18), the Graduate Education Innovation Grants of Tsinghua University (202504Z005), and the Education Innovation Grants of Tsinghua University (DX02_20).

Ethical Statement

This study protocol was approved by the Ethics Committee of Tsinghua University (Approval Number: THU-04-2025-1084), and all participants signed written informed consent forms. The authors declare that all potential respondents were fully informed about the survey, and participation was voluntary.

Conflicts of Interest

Dan Zhang is the Editorial Board Member for *Smart Wearable Technology* and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

To protect participant privacy while supporting scientific transparency, anonymized raw physiological signals (ECG, PPG, and accelerometer) and derived HRV/PRV features from five representative participants have been deposited and are publicly accessible at https://cloud.tsinghua.edu.cn/d/9629ab21aaba4f03a70d/.

Author Contribution Statement

Jie Huang: Methodology, Validation, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Yuhong Wu:** Methodology, Validation, Data curation, Writing – original draft, Supervision. **Fang Li:** Validation, Data curation. **Dan Zhang:** Conceptualization, Methodology, Writing – review & editing, Project administration, Funding acquisition. **Shuping Tan:** Conceptualization, Writing – review & editing, Project administration.

References

[1] Agorastos, A., Mansueto, A. C., Hager, T., Pappi, E., Gardikioti, A., & Stiedl, O. (2023). Heart rate variability as a translational dynamic biomarker of altered autonomic function in health and psychiatric Disease. *Biomedicines*, *11*(6), 1591. https://doi.org/10.3390/biomedicines11061591

[2] Jahani, M., Moridani, M. K., & Anisi, M. (2023). Mathematical model presenting to assess variations in heart rate of different age groups. *Bratislava Medical Journal*, *124*(6), 454–465. https://doi.org/10.4149/BLL_2023_070

[3] Vescio, B., Salsone, M., Gambardella, A., & Quattrone, A. (2018). Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short- and long-term recordings. *Sensors*, *18*(3), 844. https://doi.org/10.3390/s18030844

[4] Królak, A., & Pilecka, E. (2022). Analysis and comparison of heart rate variability signals derived from PPG and ECG sensors. In *Biocybernetics and Biomedical Engineering–Current Trends and Challenges: Proceedings of the 22nd Polish Conference on Biocybernetics and Biomedical Engineering*, 9–16. https://doi.org/10.1007/978-3-030-83704-4_2

[5] Kumar, C. K., Manaswini, M., Maruthy, K. N., Kumar, A. V. S., & Kumar, K. M. (2021). Association of Heart rate variability measured by RR interval from ECG and pulse to pulse interval from Photoplethysmography. *Clinical Epidemiology and Global Health*, *10*, 100698. https://doi.org/10.1016/j.cegh.2021.100698

[6] Singh, S., Bennett, M. R., Chen, C., Shin, S., Ghanbari, H., & Nelson, B. W. (2024). Impact of skin pigmentation on pulse oximetry blood oxygenation and wearable pulse rate accuracy: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *26*, e62769. https://doi.org/10.2196/62769

[7] Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

[8] Szurhaj, W., Leclancher, A., Nica, A., Périn, B., Derambure, P., Convers, P., . . . , & Jonckheere, J. D. (2021). Cardiac autonomic dysfunction and risk of sudden unexpected death in epilepsy. *Neurology*, *96*(21), e2619–e2626. https://doi.org/10.1212/WNL.0000000000011998

[9] Herman, R., Vanderheyden, M., Vavrik, B., Beles, M., Palus, T., Nelis, O., . . . , & Bartunek, J. (2022). Utilizing longitudinal data in assessing all-cause mortality in patients hospitalized with heart failure. *ESC Heart Failure*, *9*(5), 3575–3584. https://doi.org/10.1002/ehf2.14011

[10] Liu, S., Cui, Y., & Chen, M.. (2025). Heart rate variability: A multidimensional perspective from physiological marker to brain-heart axis disorders prediction. *Frontiers in Cardiovascular Medicine*, *12*, 1630688. https://doi.org/10.3389/fcvm.2025.1630668

[11] Goffi, F., Maggioni, E., Bianchi, A. M., Brambilla, P., & Delvecchio, G. (2025). Is cardiac autonomic control affected in major depressive disorder? A systematic review of heart rate variability studies. *Translational Psychiatry*, *15*(1), 217. https://doi.org/10.1038/s41398-025-03430-3

[12] Al Jowf, G. I., Ahmed, Z. T., Reijnders, R. A., de Nijs, L., & Eijssen, L. M. T. (2023). To predict, prevent, and manage post-traumatic stress disorder (PTSD): A review of pathophysiology, treatment, and biomarkers. *International Journal of Molecular Sciences*, *24*(6), 5238. https://doi.org/10.3390/ijms24065238

[13] Lehrer, P. M. (2018). Heart rate variability biofeedback and other psychophysiological procedures as important elements in psychotherapy. *International Journal of Psychophysiology*, *131*, 89–95. https://doi.org/10.1016/j.ijpsycho.2017.09.012

[14] Zucker, T. L., Samuelson, K. W., Muench, F., Greenberg, M. A., & Gevirtz, R. N. (2009). The effects of respiratory sinus arrhythmia biofeedback on heart rate variability and post-traumatic stress disorder symptoms: A pilot study. *Applied Psychophysiology and Biofeedback*, *34*(2), 135–143. https://doi.org/10.1007/s10484-009-9085-2

[15] Wang, Y. C., Tsai, I. J., Hsieh, T. H., Wu, C. C., Lu, K. C., & Liao, M. T. (2025). Enhanced sleep quality and reduced indoxyl sulfate levels following probiotic supplementation were linked to gut microbiota modulation in hemodialysis patients. *Journal of the Formosan Medical Association.* Advance online publication. https://doi.org/10.1016/j.jfma.2025.08.036

[16] Loh, H. H., Tay, S. P., Koa, A. J., Yong, M. C., Said, A., Chai, C. S., . . . , & Sukor, N. (2025). Unveiling the benefits of Vitamin D3 with SGLT-2 inhibitors for hypertensive obese obstructive

sleep apnea patients. *Journal of Translational Medicine*, *23*(1), 296. https://doi.org/10.1186/s12967-025-06312-w

[17] Shishavan, H. H., Garza, J., Henning, R., Cherniack, M., Hirabayashi, L., Scott, E., & Kim, I. (2023). Continuous physiological signal measurement over 24-hour periods to assess the impact of work-related stress and workplace violence. *Applied Ergonomics*, *108*, 103937. https://doi.org/10.1016/j.apergo.2022.103937

[18] Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

[19] Taskasaplidis, G., Fotiadis, D. A., & Bamidis, P. D. (2024). Review of stress detection methods using wearable sensors. *2013 IEEE Access*, *12*, 38219–38246. https://doi.org/10.1109/ACCESS.2024.3373010

[20] Berntson, G. G., Thomas Bigger, J., Eckberg, D. L., Grossman, P., Kaufmann, P. G., Malik, M., . . . , & van der Molen, M. W. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology*, *34*(6), 623–648. https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

[21] Rienzo, M. D., Vaini, E., Castiglioni, P., Merati, G., Meriggi, P., Parati, G., . . . , & Faini, A. (2013). Wearable seismocardiography: Towards a beat-by-beat assessment of cardiac mechanics in ambulant subjects. *Autonomic Neuroscience*, *178*(1), 50–59. https://doi.org/10.1016/j.autneu.2013.04.005

[22] Smets, E., Velazquez, E. R., Schiavone, G., Chakroun, I., D'Hondt, E., Raedt, W. D., . . . , & van Hoof, C. (2018). Large-scale wearable data reveal digital phenotypes for daily-life stress detection. *npj Digital Medicine*, *1*(1), 67. https://doi.org/10.1038/s41746-018-0074-9

[23] Grégoire, J. M., Gilon, C., Carlier, S., & Bersini, H. (2023). Autonomic nervous system assessment using heart rate variability. *Acta Cardiologica*, *78*(6), 648–662. https://doi.org/10.1080/00015385.2023.2177371

[24] Polcwiartek, C., O'Gallagher, K., Friedman, D. J., Correll, C. U., Solmi, M., Jensen, S. E., & Nielsen, R. E. (2024). Severe mental illness: cardiovascular risk assessment and management. *European Heart Journal*, *45*(12), 987–997. https://doi.org/10.1093/eurheartj/ehae054

[25] Peltola, M. A. (2012). Role of editing of R–R intervals in the analysis of heart rate variability. *Frontiers in Physiology*, *3*, 148. https://doi.org/10.3389/fphys.2012.00148

[26] Zheng, C., Wang, Z., Li, L., Ma, Y., & Feng, X. (2025). Flexible and wearable bioelectronics for electrocardiography monitoring: A review. *Science China Materials*, *68*(12), 4344–4359. https://doi.org/10.1007/s40843-025-3688-9

[27] Wu, J., Hong, J., Gao, X., Wang, Y., Wang, W., Zhang, H., . . . , & Guo, W. (2025). Recent progress in flexible wearable sensors utilizing conductive hydrogels for sports applications: Characteristics, mechanisms, and modification strategies. *Gels*, *11*(8), 589. https://doi.org/10.3390/gels11080589

[28] Iqbal, S. M. A., Leavitt, M. A., Mahgoub, I., & Asghar, W. (2024). Advances in cardiovascular wearable devices. *Biosensors*, *14*(11), 525. https://doi.org/10.3390/bios14110525

[29] Ebrahimi, Z., & Gosselin, B. (2023). Ultralow-Power Photoplethysmography (PPG) sensors: A methodological review. *IEEE Sensors Journal*, *23*(15), 16467–16480.https://doi.org/10.1109/JSEN.2023.3284818

[30] Lu, S., Zhao, H., Ju, K., Shin, K., Lee, M., Shelly, K., & Chon, K. H. (2008). Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information? *Journal of Clinical Monitoring and Computing*, *22*(1), 23–29. https://doi.org/10.1007/s10877-007-9103-y

[31] Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects. *Journal of Medical Engineering & Technology*, *33*(8), 634–641. https://doi.org/10.3109/03091900903150998

[32] Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., & Gillinov, M. (2017). Accuracy of wrist-worn heart rate monitors. *JAMA Cardiology*, *2*(1), 104–106. https://doi.org/10.1001/jamacardio.2016.3340

[33] Stuyck, H., Costa, L. D., Cleeremans, A., & Bussche, E. V. D. (2022). Validity of the Empatica E4 wristband to estimate resting-state heart rate variability in a lab-based context. *International Journal of Psychophysiology*, *182*, 105–118. https://doi.org/10.1016/j.ijpsycho.2022.10.003

[34] Natarajan, A., Pantelopoulos, A., Emir-Farinas, H., & Natarajan, P. (2020). Heart rate variability with photoplethysmography in 8 million individuals: A cross-sectional study. *The Lancet Digital Health*, *2*(12), e650–e657. https://doi.org/10.1016/S2589-7500(20)30246-6

[35] Ishaque, S., Khan, N., & Krishnan, S. (2021). Trends in heart-rate variability signal analysis. *Frontiers in Digital Health*, *3*, 639444. https://doi.org/10.3389/fdgth.2021.639444

[36] Allado, E., Poussel, M., Moussu, A., Hily, O., Temperelli, M., Cherifi, A., . . . , & Saunier, V. (2022). Accurate and reliable assessment of heart rate in real-life clinical settings using an imaging photoplethysmography. *Journal of Clinical Medicine*, *11*(20), 6101. https://doi.org/10.3390/jcm11206101

[37] Mejía-Mejía, E., May, J. M., Torres, R., & Kyriacou, P. A. (2020). Pulse rate variability in cardiovascular health: A review on its applications and relationship with heart rate variability. *Physiological Measurement*, *41*(7), 07TR01. https://doi.org/10.1088/1361-6579/ab998c

[38] Nelson, B. W., Low, C. A., Jacobson, N., Areán, P., Torous, J., & Allen, N. B. (2020). Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *npj Digital Medicine*, *3*(1), 90. https://doi.org/10.1038/s41746-020-0297-4

[39] Shui, X., Zhang, M., Li, Z., Hu, X., Wang, F., & Zhang, D. (2021). A dataset of daily ambulatory psychological and physiological recording for emotion research. *Scientific Data*, *8*(1), 161. https://doi.org/10.1038/s41597-021-00945-4

[40] He, C., Xu, P., Pei, X., Wang, Q., Yue, Y., & Han, C. (2024). Fatigue at the wheel: A non-visual approach to truck driver fatigue detection by multi-feature fusion. *Accident Analysis & Prevention*, *199*, 107511. https://doi.org/10.1016/j.aap.2024.107511

[41] Shui, X., Xu, H., Tan, S., & Zhang, D. (2025). Depression recognition using daily wearable-derived physiological data. *Sensors*, *25*(2), 567. https://doi.org/10.3390/s25020567

[42] Alkemary, A., Fadlallah, A. S., Helal, I., Alhomrany, M. A., Al-Ghamdi, S., Mousa, D. H., & Alharbi, A. (2024). Post-dilution high-volume online-hemodiafiltration versus high-flux hemodialysis mortality outcome in patients with end-stage kidney disease. *Journal of the American Society of Nephrology*, *35*(10S), 1681–1690. https://doi.org/10.1681/ASN.2024pj0xtjd9

[43] Xu, J., Qiu, B., Zhang, F., & Zhang, J. (2024). Restorative effects of pocket parks on mental fatigue among young adults: A comparative experimental study of three park types. *Forests*, *15*(2), 286. https://doi.org/10.3390/f15020286

[44] Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., . . . , & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

[45] Bota, P., Silva, R., Carreiras, C., Fred, A., & da Silva, H. P. (2024). BioSPPy: A Python toolbox for physiological signal processing. *SoftwareX*, *26*, 101712. https://doi.org/10.1016/j.softx.2024.101712

[46] Caridade Gomes, P. M. (2019). *Development of an open-source Python toolbox for heart rate variability (HRV)*. Master's Thesis, Hamburg University of Applied Sciences.

[47] Kinnunen, H., Rantanen, A., Kenttä, T., & Koskimäki, H. (2020). Feasible assessment of recovery and cardiovascular health: Accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. *Physiological Measurement*, *41*(4), 04NT01. https://doi.org/10.1088/1361-6579/ab840a

[48] Singstad, B. J., Azulay, N., Bjurstedt, A., Bjørndal, S. S., Drageseth, M. F., Engeset, P., . . . , & Martinsen, Ø. G. (2021). Estimation of heart rate variability from finger photoplethysmography during rest, mild exercise and mild mental stress. *Journal of Electrical Bioimpedance*, *12*(1), 89–102. https://doi.org/10.2478/joeb-2021-0012

[49] Jan, H. Y., Chen, M. F., Fu, T. C., Lin, W. C., Tsai, C. L., & Lin, K. P. (2019). Evaluation of coherence between ECG and PPG derived parameters on heart rate variability and respiration in healthy volunteers with/without controlled breathing. *Journal of Medical and Biological Engineering*, *39*(5), 783–795. https://doi.org/10.1007/s40846-019-00468-9

[50] Lin, W. H., Wu, D., Li, C., Zhang, H., & Zhang, Y. T. (2014). Comparison of heart rate variability from PPG with that from ECG. In *The International Conference on Health Informatics*, 213–215. https://doi.org/10.1007/978-3-319-03005-0_54

[51] Castiglioni, P., Meriggi, P., di Rienzo, M., Lombardi, C., Parati, G., & Faini, A. (2022). Heart rate variability from wearable photoplethysmography systems: Implications in sleep studies at high altitude. *Sensors*, *22*(8), 2891. https://doi.org/10.3390/s22082891

[52] Dunn, C. E., Monroe, D. C., Crouzet, C., Hicks, J. W., & Choi, B. (2019). Speckleplethysmographic (SPG) estimation of heart rate variability during an orthostatic challenge. *Scientific Reports*, *9*(1), 14079. https://doi.org/10.1038/s41598-019-50526-0

[53] Kiran Kumar, C., Manaswini, M., Maruthy, K. N., Siva Kumar, A. V., & Mahesh Kumar, K. (2021). Association of Heart rate variability measured by RR interval from ECG and pulse to pulse interval from Photoplethysmography. *Clinical Epidemiology and Global Health*, *10*, 100698. https://doi.org/10.1016/j.cegh.2021.100698

[54] Rinella, S., Massimino, S., Fallica, P. G., Giacobbe, A., Donato, N., Coco, M., . . . , & Conoci, S. (2022). Emotion recognition: Photoplethysmography and electrocardiography in comparison. *Biosensors*, *12*(10), 811. https://doi.org/10.3390/bios12100811

[55] Eleftheriadou, A., Spallone, V., Tahran, A. A., & Alam, U. (2024). Cardiovascular autonomic neuropathy in diabetes: An update with a focus on management. *Diabetologia*, *67*(12), 2611–2625. https://doi.org/10.1007/s00125-024-06242-0

How to Cite: Huang, J., Wu, Y., Li, F., Zhang, D., & Tan, S. (2026). Assessing Pulse Rate Variability from a Wrist-Worn PPG Device Against ECG-Derived Heart Rate Variability in Ambulatory Settings. *Smart Wearable Technology*, *2*, A3. https://doi.org/10.47852/bonviewSWT52027605